

University of Louisville

ThinkIR: The University of Louisville's Institutional Repository

Electronic Theses and Dissertations

1-2020

Characterization of adiposity and inflammation genetic pleiotropy underlying cardiovascular risk factors in Hispanics.

Mohammad Yaser (Anwar)
University of Louisville

Follow this and additional works at: <https://ir.library.louisville.edu/etd>



Part of the [Cardiovascular Diseases Commons](#), [Epidemiology Commons](#), [Genetic Phenomena Commons](#), and the [Genetics Commons](#)

Recommended Citation

Yaser (Anwar), Mohammad, "Characterization of adiposity and inflammation genetic pleiotropy underlying cardiovascular risk factors in Hispanics." (2020). *Electronic Theses and Dissertations*. Paper 3530. <https://doi.org/10.18297/etd/3530>

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.

CHARACTERIZATION OF ADIPOSITY AND INFLAMMATION
GENETIC PLEIOTROPY UNDERLYING CARDIOVASCULAR RISK
FACTORS IN HISPANICS

By

Mohammad Yaser (Anwar)
M.D., Herat University, Afghanistan 2008
M.P.H, Johns Hopkins University, MD USA 2014

A dissertation submitted to the Faculty of the
School of Public Health & Information Sciences of University of Louisville
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Public Health Science

Department of Epidemiology
University of Louisville
Louisville, Kentucky

December 2020

CHARACTERIZATION OF ADIPOSITY AND INFLAMMATION
GENETIC PLEIOTROPY UNDERLYING CARDIOVASCULAR RISK
FACTORS IN HISPANICS

By

Mohammad Yaser (Anwar)
M.D., Herat University, Afghanistan 2008
M.P.H, Johns Hopkins University, MD USA 2014

A Dissertation Approved on

December 1, 2020

By the following Dissertation Committee

Dissertation Director
Kira C. Taylor, Ph.D.

Kira North, Ph.D.
2nd Committee Member

Richard Baumgartner, Ph.D.
3rd Committee Member

Bertis Britt Little, Ph.D.
4th Committee Member

Natalie C. DuPré, Sc.D., MS
5th Committee Member

DEDICATION

This dissertation is dedicated to my parents

Mrs. Karimeh Saljoki

&

Dr. Mohammad Anwar Anwar

Words fail to frame how much you sacrificed

to have my aspirations fulfilled

Thank you from bottom of my heart

ACKNOWLEDGEMENT

I would like to take this opportunity to express my gratitude to those who have helped me to pursue this degree and complete this dissertation. First, I would like to thank my advisor Dr. Kira Creswell Taylor whose persistent encouragements and technical insights helped navigate the challenges of the doctoral program, frame my research interests and nurture my research skills since the beginning of the program. I am particularly grateful to have had the chance to work and learn from wonderful Dr. Kari North, a pick of the field scholar in cardiovascular genetic studies whose wisdom and expertise helped me to shape my dissertation, acquire scientific perspectives and continue my development as an upstart professional. I would also like to thank my dissertation committee members Dr. Richard Baumgart, Dr. Bertis Little and Natalie Dupre for their supports and thought evaluations of my work at every step. Special thanks to Dr. Heather Highland and Dr. Mariaelisa Graff and Antoine Baldassari of University of North Carolina, Chapel Hill, for their unwavering technical supports and feedbacks that greatly helped me complete the dissertation objects. I extend my gratitude to all faculty and administrative staff at School of Public Health & Information Sciences for their support along the doctoral program, and the University for providing an opportunity for a doctoral program together with a prestigious fellowship which enabled me to complete my studies.

ABSTRACT

CHARACTERIZATION OF ADIPOSITY AND INFLAMMATION GENETIC PLEIOTROPY UNDERLYING CARDIOVASCULAR RISK FACTORS IN HISPANICS

Mohammad Yaser (Anwar)

11/30/2020

The observed overlap between genetic variants associated with both adiposity and inflammatory markers suggests that changes in both adiposity and inflammation could be partially mediated by common pathways. The pervasive but sparsely characterized “pleiotropic” genetic variants associated with both adiposity and inflammation have been hypothesized to provide insight into the shared biology. This study explored and characterized the genetic pleiotropy underpinning adiposity and inflammation using genetic and phenotypic observations from the Cameron County Hispanic Cohort (CCHC).

A total of 3,313 samples and >9 million single nucleotide polymorphisms (SNPs) were examined in this study. Mixed model genome-wide association studies (GWAS) were performed for 9 phenotypes including C-reactive protein (CRP), Interleukin (IL)-6, IL-8, fibrinogen, body mass index (BMI), waist circumference (WC) in males and females, and waist to hip ratio (WHR) in males and females (separately). GWAS for WHR and WC were meta-analyzed to obtain sex-combined results. Pleiotropy assessment was completed using adaptive Sum of Powered Score (aSPU) test. Three genetic loci with evidence of pleiotropy on chromosome 3, 12 and 18 were fine-mapped to distinguish the set of likely

causal variants. Causal mediation analysis was used to assess whether likely causal variants were independently associated with both inflammation and adiposity.

At least 3 signals, on chromosomes 3, 12, and 12, were identified that suggested the presence of SNPs with strong pleiotropic p-values ($< 5 \times 10^{-6}$). The fine-mapping of these three suspected pleiotropic regions distinguished 22 variants with posterior causality probabilities greater than 50%. The mediation analysis indicated that rs60505812, on chromosome 3, was independently associated with both an inflammatory marker (IL-6) and an adiposity measure (BMI). For the variant rs73093474, on chromosome 12, results indicated both a direct association with CRP and an indirect association (via WHR).

The identification of likely pleiotropic variants indicated that 1) a considerable degree of overlapping genetic pleiotropy exists between adiposity and inflammation, and 2) evidence exists to support both the direct and indirect pleiotropy. The results showed the potential of these genetic variants to provide biological insight, intended to improve the cardiovascular health of the Hispanics, and by extension all populations.

TABLE OF CONTENTS

	PAGE
INTRODUCTION AND BACKGROUND.....	1
1. Epidemiology.....	1
1.1. Obesity trends.....	1
1.2. Public health significance.....	3
2. Pathophysiology.....	4
2.1. Obesity-cardiovascular system association mechanisms.....	5
2.2. Adiposity and inflammation.....	6
2.3. Inflammation as a possible adiposity precursor.....	9
2.4. A case for overlap in the underlying functional mechanism.....	10
3. Genetics of Obesity and Inflammation.....	10
3.1. Genetic overlaps between adiposity and inflammation.....	11
3.2. Study of pleiotropic variants offers novel biologic insight.....	12
4. Gaps in Adiposity-Inflammation Pleiotropy Research.....	13
4.1. Inadequate number of studies.....	13
4.2. Methodological challenges.....	13
4.3. Limited inference.....	14
4.4. Adiposity-inflammation pleiotropy study in Hispanics/latinos.....	16
5. Study Rational.....	16
6. Aims of the Study Project.....	17

7. Tables and Figures.....	19
METHODS	23
1. Study Setting.....	23
1.1. Conception.....	24
1.2. Recruitment process.....	24
2. Data Collection.....	25
2.1. Anthropometric measures.....	25
2.2. Inflammation biomarkers.....	26
2.3. Genetic measurements.....	26
3. Exploratory Analysis of Genetic Measurements and Quality Control.....	27
3.1. Pattern of missingness.....	27
3.2. Sex mismatch.....	28
3.3. Genetic outliers.....	28
3.4. Hardy-Weinberg Equilibrium.....	29
3.5. Duplicated samples and variants.....	29
3.6. Quality control process.....	29
4. Principal Component Analyses.....	30
5. Imputation.....	32
5.1. Pre-imputation quality control.....	32
5.2. Imputation panels.....	32
5.3. Imputation procedure.....	33
5.4. Post-imputation quality control.....	34
6. Phenotypes	34

6.1. Covariates	34
6.2. Inclusion/exclusion criteria.....	35
6.3. Examination of phenotypic outliers.....	35
6.4. The minimalistic approach.....	36
6.5. Hybrid approach.....	37
6.6. Selection of the phenotype set.....	37
7. Statistical Approach.....	38
7.1. Genome-wide association test.....	38
7.2. Pleiotropy assessment.....	39
7.3. Fine-mapping.....	41
7.3.1. PAINTOR.....	41
7.3.2. FINEMAP.....	43
7.4. Causal pathway analysis approach.....	44
7.4.1. Sensitivity analysis.....	46
8. Tables and Figures.....	47
GWAS AND PLEIOTROPY.....	59
1. Background.....	59
2. Methods.....	61
2.1. Study population.....	61
2.2. Covariate selection.....	61
2.3. GWAS test.....	62
2.4. Post-GWAS sensitivity analysis.....	64
2.5. Pleiotropy test.....	65

2.6. In-depth probe of suggestive pleiotropic signals.....	66
2.7. Functional evaluation of pleiotropic signals.....	67
3. Results.....	67
3.1. Descriptive statistics.....	67
3.2. GWAS results.....	68
3.3. Sensitivity analyses.....	68
3.4. Pleiotropy results.....	69
4. Discussion.....	71
5. Tables and Figures.....	75
6. Appendices.....	79
FINE-MAPPING PLEIOTROPIC SIGNALS.....	85
1. Background.....	85
2. Methods.....	87
2.1. Genome-wide pleiotropy summary statistics.....	87
2.2. Statistical framework.....	88
2.3. Defining fine-mapping loci.....	89
2.4. Linkage disequilibrium estimation.....	90
2.5. Functional annotation.....	90
2.6. Fine-mapping process.....	91
2.7. Functional evaluation.....	91
2.8. Independence of signals.....	92
2.9. Visualization of pleiotropic signals.....	93
3. Results.....	93

4. Discussion.....	96
5. Tables and Figures.....	101
6. Appendices.....	107
CAUSAL MEDIATION PATHWAY ANALYSIS.....	109
1. Background.....	109
2. Methods.....	111
2.1. Candidate variant set.....	111
2.2. Phenotypes.....	112
2.3. Covariate selection.....	112
2.4. Statistical analysis.....	113
2.5. Causal mediation pathway analysis.....	114
2.6. Sensitivity analysis.....	116
3. Results.....	117
4. Discussion.....	119
5. Tables and Figures.....	123
6. Appendices.....	127
DISCUSSION.....	129
1. Inflammation-adiposity association.....	129
2. Rule of genetics and ancestry in inflammation-adiposity association.....	130
3. Study object.....	132
4. Cameron County Hispanic Cohort.....	132
5. Genetic signals.....	133
6. Evidence for pleiotropy.....	133

7. Characteristics of likely pleiotropic signals.....	134
8. The functional evaluation of likely pleiotropic variants.....	136
9. Study strengths and limitations.....	138
10. Study summary and conclusion.....	140
REFERENCES.....	142
CURRICULUM VITAE.....	170

LIST OF TABLES

TABLE	PAGE
Table 2.1 Quality control steps completed	47
Table 2.2 The number of variants imputed	48
Table 2.3 List of the phenotypes selected for the study.....	48
Table 2.4 Descriptive Distribution of anthropometry and inflammatory phenotypes.....	49
Table 2.5 Proposed cut-points for exclusion of phenotypic outliers.....	49
Table 2.6 Descriptive distributions of study phenotypes after exclusion of outliers.....	50
Table 3.1 Distributions of inflammatory and anthropometric phenotypes.....	75
Table 3.2 GWAS significant variants in sex-combined meta-analyzed WHR _{BMI-adj} GWAS analysis.....	75
Table 3.3 Functional probe of three loci on chromosome 3, 12 and 18 with suggestion for adiposity-inflammation pleiotropy.....	76
Table 4.1a Characteristics of variants with high posterior causal probability in the target region on chromosome 3 (base pair range: 3,653,147–4,650,039).....	101
Table 4.1b Characteristics of variants with high posterior causal probability in the target region on chromosome 12 (base pair range: 43,981,459–44,980,751).....	102

Table 4.1c Characteristics of variants with high posterior causal probability in the target region on chromosome 18 (base pair range: 61,079,369–62,075,853).....103

Table 5.1 Distributions of the phenotypes and covariates in Cameron County Hispanic Cohort (CCHC) among participants with non-missing values for either C-reactive Protein (CRP) or interleukin (IL)-6.....123

Table 5.2 Univariate regression results with genetic variants as the outcomes and phenotypes as predictors.....124

Table 5.3 Causal mediation analysis for rs60505812 and rs73093474.....125

LIST OF FIGURES

FIGURE	PAGE
Figure 1.1. Obesity increases extracellular matrix levels within the kidney and may cause medullary compression	19
Figure 1.2. Adipocytes expand and replicate to store positive energy balance.....	20
Figure 1.3. Hypertrophied adipocytes upregulate production of pro-inflammatory cytokines, including IL-6.....	20
Figure 1.4. Functional overlap and feedback loop.....	21
Figure 1.5. The conceptual framework for the adiposity-inflammation pleiotropy study.	22
Figure 2.1 Minor Allele Frequency (MAF) distribution of genotyped variants.....	51
Figure 2.2 Proportion of missingness per variant.....	51
Figure 2.3 Per sample missingness vs autosomal heterozygosity.....	52
Figure 2.4 Distribution of F ratio coefficient.....	52
Figure 2.5 Sex vs autosomal heterozygosity.....	53
Figure 2.6 Eigensoft calculated ancestry outliers	54
Figure 2.7 Hardy-Wienberg Equilibrium.....	55
Figure 2.8 Kinship structure comparing identity by descent (IBD) parameters.....	55

Figure 2.9 Projection of the first three PCAir for CCHC over 1000G individuals	56
Figure 2.10 Parallel plot of the first 10 PCs.....	56
Figure 2.11 Screeplot for PCs.....	57
Figure 2.12 Scatter plot of PCAir1 vs PCA1.....	57
Figure 2.13 Distribution of phenotypic markers using observations from visit 1 & 2....	58
Figure 2.14 Log-transformed distributions of age, anthropometry and inflammation phenotypes.....	58
Figure 3.1. Sex-combined meta-analyzed GWAS for WHR _{BMI-adj}	77
Figure 3.2. Pleiotropy Manhattan plot.....	77
Figure 3.3a-c Heatplots of variants with trait-specific GWAS P _{values} of $<5 \times 10^{-5}$ located within the three loci with suggestive evidence for pleiotropy.....	78
Figure 4.1 Regional zoomplot of the suspect pleiotropic locus on chromosome 3.....	104
Figure 4.2. Regional zoomplot of the suspect pleiotropic locus on chromosome 12.....	105
Figure 4.3. Regional zoomplot of the suspect pleiotropic locus on chromosome 18.....	106
Figure 5.1. Sensitivity analysis of observed ACME and ADE.....	126

INTRODUCTION AND BACKGROUND

This chapter presents background information on the epidemiologic, pathophysiologic and genetic aspects of obesity and associated cardiovascular risk factors, particularly in the context of Hispanic/Latino populations. Biologic discussions focus on associations and functional overlaps between adiposity and inflammation that promote adverse cardiovascular outcomes. The review emphasizes gaps in the current literature on genetic factors underpinning biologic pathways affecting adiposity and inflammation in Hispanic/Latinos, and concludes with the scientific aims of the dissertation work.

1. Epidemiology

1.1. Obesity trends

Although a struggle with undernutrition has been a defining feature for most of human history, the industrial revolution resulted in the explosive growth of agricultural productivity over the last two and half centuries, resulting in a relative abundance of food that has significantly contributed to the well-being of societies. Records from industrialized countries show that the average height¹ and weight² of individuals has progressively increased throughout the 19th and the first half of the 20th century.

However, continued uptake of high calorie foods coupled with changes in lifestyles and environmental settings, especially during the latter half of the 20th century, have gradually tipped the balance toward obesity. The populations in more wealthy parts of the world have

begun to gain weight at levels that are disproportionate to the gains made in height³. At the dawn of the new millennium, humans passed a landmark point when adults with excess weight outnumbered those who were underweight⁴. Today, obesity is a major public health challenge⁵.

Even though obesity figures vary between countries, due to differences in socioeconomic characteristics and the speed of epidemic transition [away from infections' dominated landscape toward the rise in metabolic diseases]⁶, recent decades have seen linear upward trend in the prevalence of overweight and obesity almost everywhere⁷. Between 1980 and 2013, the average proportion of overweight increased by one-third, in both men and women, globally⁸. A review of over 2,000 population-based studies published between 1985 and 2017 showed that body mass index (BMI) values, defined as weight (in kilograms) over height squared (in meters), increased from 22.6 to 24.7 in women and from 22.2 to 24.4 in men⁹.

This rising trajectory in obesity appears to include all age groups. Based on a pooled analysis of approximately 2,500 population-based studies, including 128 million individuals aged 5 years and older, the average estimated global prevalence of obesity among children has increased 8-fold, from just under 1% in 1975 to approximately 8% in 2016¹⁰. In the United States, estimates have shown that approximately 69% of adults are classified as overweight (BMI \geq 25), 35% are obese (BMI \geq 30), and 14.5% are moderately to extremely obese (BMI \geq 35)¹¹.

However, average numbers do not reflect the underlying complexities associated with obesity prevalence. At the subnational level, significant disparities in the distribution of overweight and obesity exist among various ethnic and socioeconomic^{12,13}. A higher

proportion of Hispanic Americans appear to suffer from overweight and obesity, with a greater than 12% increase in the prevalence of overweight and a greater than 21% increase in the prevalence of all-type obesity compared with the overall mean for the adult population (>20 years age)¹¹. A review of the National Health and Nutrition Examination Survey (NHANES) biannual reports between 1999-2018 revealed larger annual percentage point increases in obesity among Hispanics compared with non-Hispanics, with severe obesity now occurring in over 10% for this ethnic group¹⁴. It should be noted that increase in obesity is not limited in Hispanic Americans but similarly noted in other countries with majority Latino populations¹⁵, with unequal distribution that disproportionately affects those at lower levels of socioeconomic status¹⁶.

1.2. Public health significance

Increased obesity is an important public health issue due to the associations between obesity and various adverse health outcomes^{17,18}. Central obesity, in which the excessive accumulation of weight occurs in the abdominal region, is a known risk factor for serious diseases¹⁹, and has been associated with insulin resistance²⁰, high blood pressure²¹, dyslipidemia²², and metabolic hormones disorders¹⁸. These metabolic disorders, in turn, increase susceptibility to cardiovascular diseases²³, often with fatal consequences.

Unsurprisingly, given the elevated trends in obesity, the prevalence of cardiovascular risk factors is also higher among ethnic minorities, including Hispanics²⁴. For example, the age-adjusted prevalence ratio of diabetes was >2.5 times higher in Hispanics than in non-Hispanic whites, in one recent survey²⁵; another study suggested that as many as two in three Hispanic adults present with abnormal levels of lipids²⁶.

Yet, despite the elevated prevalence of various metabolic risk factors, Hispanics also appear to have longer life spans and suffer from lower rates of cardiovascular disease-associated mortality^{27,28}. Statistics from NHANES from 1988 to 2006 vividly illustrate the reduced mortality rate in this group compared with other groups (e.g. Hazard ratio 0.42 vs Non-Hispanic Whites among ≥ 50 years old)²⁹. Various factors, including immigration of healthy individuals²⁷, a legume- and fruit-based dietary pattern, supportive social and family structures, the unreliability of the data, and the Salmon bias, which describes the tendency for older, unhealthy individuals to return to their original countries³⁰, have been suggested as possible explanations for this “Hispanic paradox”³¹.

Whether this paradox exists³² and what factors may explain it is an active area of research, rife with arguments and counterarguments. However, the persistence of a lower mortality rate among Hispanics, even in studies that have attempted to control for socioeconomic factors³³, has resulted in a focus on the important but sparsely explored contributions of genetic ancestry to this paradox³⁴.

2. Pathophysiology

Increasing body weight steadily taxes the body’s physiological capacities³⁵, eventually causing major disorders³⁶. Altered pulmonary parameters³⁷, the expansion of cardiac output, increased sympathetic nervous system activities³⁸, increased systolic and diastolic blood pressure³⁹, elevated sodium retention⁴⁰, and the attenuation of immunity⁴¹ are some of the notable consequences of increased body weight. The pathophysiological changes that are triggered by excess weight can affect multiple organs and systems, leading to adverse outcomes in brain⁴², reproductive⁴³, liver⁴⁴, and kidney functions⁴⁵, among others.

The deleterious effects on the heart and vascular system are particularly concerning⁴⁶ because the obese population features a significantly higher mortality rates associated with cardiovascular diseases, with one studying reporting a hazard ratio as high as 2.21 among those with body mass index (BMI) \geq 40, compared to normal weight ($20\leq$ BMI $<$ 25) individuals⁴⁷, even as the BMI is not an ideal predictor of the cardiovascular diseases⁴⁸. The United States National Vital Statistics report in 2017 showed that 23% of all deaths could be attributed to heart diseases, making heart disease the top-ranked cause of mortality, ahead of all types of malignant neoplasia⁴⁹.

2.1. Obesity-cardiovascular system association mechanisms

The precise mechanisms that link obesity with cardiovascular risk factors remain unknown, although many hypotheses have been generated⁵⁰⁻⁵², all of which could partially or substantially contribute to this association. For instance, one proposition is that the obesity-induced decrease in insulin sensitivity promotes increased arterial stiffness⁵³. Insulin-resistance reduces endothelium-dependent vasodilation⁵³, possibly through nitric oxide-dependent processes that act as physiological regulators of vascular tone⁵⁴ and induce smooth muscle proliferation and migration⁵⁴.

Abnormal kidney function and the resulting impacts on elevated blood pressure and hemodynamic imbalance is another area of research focus⁵⁵. Obesity increases tubular reabsorption, shifting pressure natriuresis (sodium excretion) toward higher blood pressure to maintain the water-sodium balance⁵⁶. Elevated reabsorption appears to be mediated through the activation of the renin-angiotensin system, and the buildup of adipose tissue around the kidney and increased extracellular matrix levels within the kidney can cause medullary compression, altering the intrarenal mechanical forces⁵⁷ (Figure 1.1). Obesity is

also associated with tangible renal vasodilation and higher glomerular filtration rates⁵⁸, which can increase sodium reabsorption.

Nonetheless, the focus of the majority of research on obesity and cardiovascular risk factors has been on adipose tissue. Changes in the local distribution⁵⁹ and metabolic characteristics of adipose tissue^{60,61} play decisive roles in the association between obesity and adverse cardiovascular outcomes. The regulation of adipose tissue is a delicate process, and disruptions in adipose regulation can result in detrimental health effects, particularly in visceral and ectopic adipose tissue, which are particularly sensitive to overnutrition⁶². In response to sedentary lifestyle conditions⁶³, genetic predispositions⁶⁴ or exposure to malignant environmental factors⁶⁵, the rapid or sustained accumulation of lipids can disrupt the differentiation of preadipocytes and accelerating the formation of larger, dysfunctional adipocytes^{66,67} (Figure 1.2).

2.2. Adiposity and inflammation

The associations between obesity and metabolic irregularities are often attributed to inflammatory responses induced by adipose tissue^{68,69}. In humans, similar to other mammals, adipose tissues exist in two forms: white adipose tissue and brown adipose tissue. Brown adipose tissue is found primarily in (human) neonates and plays a significant role in the regulation of body temperature through non-shivering thermogenesis during the earlier years of life. White adipose tissue, on the other hand, is thought to be the site of energy storage when a positive imbalance exists between energy intake and expenditure. Adipose tissue also contains preadipocytes, endothelial cells, fibroblasts, leukocytes, and macrophages. Energy, primarily in the form of glucose and fatty acids, is converted to lipids through a well-characterized lipogenesis process⁷⁰.

Adipocytes can expand in size (hypertrophy) and numbers (hyperplasia) to accommodate increased fat storage needs in cases of positive energy imbalance⁷¹ to maintain energy homeostasis. Adipose tissue expansion creates the need for insulin, which is an essential peptide secreted from beta-cells in the pancreas that regulates the absorption of glucose from the bloodstream into the liver, skeletal muscles, and adipocytes.

Sustained anabolic activities, such as growth and differentiation, necessitate a steadier supply of insulin as an adaptive response to excess energy or in response to underlying disease. This intensely pushes the closed-loop of the adipocyte-brain axis, which controls the feedback process⁷², and adipocytes respond with increased hyperplasia and hypertrophic changes.

Enlarged adipocytes feature altered lipolytic profiles, which can be attributed to the enrichment of regulatory proteins, including lipase and perilipin, which are distal in the lipolytic cascade⁷³. These cells have also been shown to be resistant to the lipopenic (blood lipid decreasing) actions of leptin⁷⁴, a hormone that is made in adipose cells (beside enterocytes) and contributes to the regulation of energy balance by inhibiting the hunger sensation⁷⁵.

Ability of enlarged adipocytes to synthesize adiponectin is decreased compared with normal adipocytes, which impairs fat oxidation and glucose regulation and amelioration of inflammatory activities⁷⁶. Simultaneously, highly active hypertrophied adipocytes exhibit distorted cytokine secretion characteristics⁷⁷. They also promote noticeable increase in the circulating levels of several cytokines, including interleukin (IL)-6 and IL-8, noticeably increase^{78,79}.

Adding to the complexity of inflammatory reactions, the physical expansion of cells triggers increased blood flow, resulting in the attraction of macrophages through several signals, including monocyte chemoattractant protein 1 (MCP1)⁸⁰ and increased apoptotic adipocytes⁶⁶. In experimental studies, CD8⁺ T-cells are attracted to and infiltrate obese adipose tissue, which, in turn, facilitates the recruitment and activation of macrophages to this tissue⁸¹. Increased macrophage presence in adipose tissue is associated with increased obesity⁸².

Within pro-inflammatory milieu of expanded adipose tissue, in which increased concentrations of macrophages are admixed with lipolytic and apoptotic adipocytes, cross-talk between these two cell types can promote positive feedback loops, resulting in the amplification of adverse metabolic and inflammatory actions, mediated by an increase in the levels of circulating inflammation markers^{83,50} (Figure 1.3). Moreover, adipocyte-secreted cytokines, including IL-6, are key drivers of C-reactive protein (CRP) production in the liver (through direct liver access via the portal system⁸⁴). As a result, CRP is often used as a significant predictor of cardiovascular diseases⁸⁵.

Enhanced cytokine levels result in the increased production of free fatty acids, which are potent activators of the Toll-like receptor family⁸⁶. Toll-like receptors are pattern recognition receptors that activate a variety of intracellular signaling pathways, including the activation of nuclear factor-kappa B (NF- κ B) and inflammatory cytokine production, which are responsible for activating innate immunity⁸⁷ and triggering the inception of atheromatic processes.

2.3. Inflammation as a possible adiposity precursor

Although most biological research presupposes adiposity to be the trigger for the inflammatory process, some indications have suggested that the adiposity–inflammation relationship could also move in the other direction. For example, compelling evidence suggests that inflammation-induced insulin resistance may dysregulate glucose uptake by adipocytes and promote or exacerbate the accumulation and inflammation of adipose tissue⁸⁸. Insulin resistance-induced hyperinsulinemia can create anabolic pressure⁸⁹, triggering hyperplastic and hypertrophic changes in adipose tissues to alleviate glucose accumulation⁹⁰. If sustained, these cells ultimately reach a tipping point, when further anabolic pressures cannot be accommodated due to constraints on the physical expansion of cells and tissues.

Beside insulin resistance induced adiposity, analysis of inflammatory markers in over 2,000 non-diabetic, middle-aged men found that the highest levels of inflammatory markers at baseline were predictive of weight gain over the 6 years of the study follow-up, regardless of body weight at the start of the study⁹¹.

Specific dietary patterns, particularly those high in sugar and saturated fats, have been reported to cause inflammation in the hypothalamus, resulting in leptin resistance, a precursor event to appetite dysregulation and the altered metabolism of fat and glucose⁹². A similar pattern of leptin resistance was also observed in experimental studies with rats⁹³. Microbiota and bacteria-induced inflammation that precedes obesity is an active area of research, in which structural changes in the permeability of the gut may promote the selective uptake of nutritional components⁹⁴⁻⁹⁶.

2.4. A case for overlap in the underlying functional mechanism

Some of the relationships described above have been suggested to form a circular loop, in which weight gain causes inflammation, resulting in metabolic/structural alterations that, in turn, cause further fat accumulation. Such self-sustained feedback mechanisms have previously been suggested⁹⁷ and merit attention in adiposity–inflammation association studies (Figure 1.4).

Although most of the studies included in this review implicitly assume that these feedback mechanisms move towards a pathophysiological direction, the origin of the trigger remains unclear. The sophisticated and complex natures of both inflammation and adiposity suggest a nuanced underlying mechanism, one that may be induced by certain changes in either adiposity or inflammation and may be mediated by common molecular pathways that result in variations in both domains, indicating a level of interrelatedness between these two domains, linked by some shared biological underpinning. A comprehensive investigative approach remains necessary to discern these relationships by incorporating traditional laboratory-based assessments with genetic insights at both the population and molecular level.

3. Genetics of obesity and inflammation

Although the role played by genetics in the expression of biological traits has long been accepted, scientists have only recently begun to contemplate the exact mechanisms that link heritable elements with the expression of physiological⁹⁸ and other phenotypes⁹⁹. The widespread availability of DNA sequencing and genotyping techniques and the

development of statistical tools that take advantage of ever-increasing computational power has resulted in the advent of GWAS.

In contrast with approaches that test pre-specified regions, the hypothesis-independent GWAS approach scans millions of variants throughout the genome, examining their associations with phenotypic variations. First adopted in early 2000, GWAS-based studies have experienced exponential growth, providing significant insights into our understanding of the genetic underpinnings of diseases^{100,101}.

To date, GWAS Studies have provided strong evidence for associations between genetic variants and adiposity¹⁰²⁻¹⁰⁴ as well as inflammatory markers¹⁰⁵⁻¹⁰⁹, indicating the potential complexity and variability of the underlying biological networks that mediate phenotypic expression, either directly or in tandem with other factors.

3.1. Genetic overlaps between adiposity and inflammation

With the increased availability of GWAS summary results, a number of genomic regions that were originally reported in association with obesity were later found to be similarly associated with inflammatory traits. Evidence for such intersections has also been reported for asthma and obesity¹¹⁰. In an experimental study, high levels of inflammatory markers were observed in mice carrying genes that harbored alleles associated with increased obesity risk¹¹¹. Concurrent inflammatory effects were also reported in studies where the expression levels of adipose tissue genes were regulated^{112,113}. More specifically, single nucleotide variants (SNPs) located in *LRRFIP1*¹⁰⁶, *FTO*^{114,115}, *AdipoQ*¹¹⁶, *CSN1S1*¹¹⁷, *FDFT1*, and *PCCB*¹⁰⁷ were suggested to be associated with both inflammatory markers and obesity traits. The reverse was also observed, with genes including *IL6*, which encodes

an inflammatory cytokine, found to affect energy balance¹¹⁸ and *TNF* interacting with adiposity¹¹⁹.

In addition, the phenotypic synergy between adiposity and inflammatory traits, including the expression of CRP¹²⁰, interleukins¹²⁰, and other cytokines¹²¹, has been consistently observed. This phenotypic synergy provides ample evidence for the existence of overlapping functional pathways that underly both domains and are likely to be affected by genetic variants. Termed pleiotropic variants in the scientific literature¹²², these multi-effect variants could help explain the underlying mechanisms that result in correlated phenotypic expression. The finding that seemingly distinct but co-expressed phenotypes may share genetic pathways highlights the relevance of pleiotropy in understanding these complex traits¹²³.

3.2. Study of pleiotropic variants offers novel biologic insight

Pleiotropic variants are known to be pervasive¹²⁴ and have the potential to do the following. (1) Clarify the molecular functions of phenotype-associated loci and identify common functional pathways between multiple phenotypes¹²⁴. (2) Inform the classification and treatment of patients. Although current treatment strategies combine therapies directed at distinct cardiovascular disease risk factors, an alternative approach, which may potentially be informed by the investigation of pleiotropy, is the identification and targeting of mechanistic “common denominators.” Similarly, drugs that are developed to treat a specific disease phenotype could be repurposed if a common therapeutic target was identified. (3) Prioritize variants for functional follow-up. The vast majority of GWAS-identified variants are not causal but, instead, are correlated with true causals variant through linkage disequilibrium (LD)^{125,126}. However, distinguishing associations that

represent shared effects for a single variant from among multiple independent variants that co-localize is crucial for interpretations of mechanistic models¹²⁷. (4) Increase the statistical power of novel locus discovery¹²⁸. Inconsistencies between estimated and explained heritability¹²⁹ and substantial proportions of variance can be explained when considering all GWAS SNPs^{130,131}. The number of GWAS-identified novel loci continues to increase as sample sizes increase¹²⁸, which suggests that multiple loci remain unidentified. Previous studies by several research groups^{132,133} have demonstrated how harnessing pleiotropy can increase the statistical power of novel locus discovery.

4. Gaps in adiposity–inflammation pleiotropy research

4.1. Inadequate number of studies

Despite their potential for illuminating the origin of complex traits, including inflammation and adiposity, the exploration of pleiotropic variants has been limited, and the underlying mechanisms associated with these SNPs remain largely uncharacterized. The studies referenced in the previous section constitute a substantial proportion of the existing literature on adiposity–inflammation pleiotropy, which highlights the need to perform additional research and in combination with functional studies.

4.2. Methodological challenges

In most instances, whether overlaps in the GWAS-identified loci reflects mediated pleiotropy (i.e., an SNP influences adiposity, which, in turn, affects inflammation)¹²⁷ or represents biological pleiotropy, with distinct SNP effects on each phenotype remain unresolved. Restricting attention to genome-wide significant or even suggestive SNPs will

typically underestimate the extent of pleiotropy, as SNPs with more modest effects will be excluded¹³⁴.

Interrogating pleiotropy by comparing phenotype-specific SNP lists can often fail to distinguish between SNPs that represent biological pleiotropy from SNPs that co-localize due to (1) spurious associations, which may reflect ascertainment bias where SNP associates with co-occurring trait¹³⁵, phenotypic misclassification¹³⁶, allelic structure, or shared controls¹³⁷; or (2) reflect ambiguity in mapping the true underlying causal variant, particularly with non-coding variants that have tissue specific effects¹³⁸.

4.3. Limited inference

Methodological issues and the inadequacy of current pleiotropy research are not the only factors hindering further exploration and functional characterization of multi-trait associated variants. Most existing GWAS and pleiotropic studies have been performed on largely European ancestry populations. The confinement of investigations to specific groups^{139,140} can create a biased view of human variation and hinders the translation of genetic associations into clinical and public health applications that are relevant for all populations¹⁴¹. In addition, these studies fail to leverage the genetic architecture of racial/ethnic minority populations and miss opportunities to pinpoint unique variant-gene combinations.

The conclusions formed from existing studies may warrant limited generalization given the differential distribution of inflammatory markers associated with comparative adiposity configurations¹⁴², patterns of fat accumulation^{143,144}, and the prevalence of cardiometabolic abnormalities in minority groups, such as Hispanics/Latinos¹⁴⁵. These differences have

been persistently observed in various settings; for instance, in the multiethnic Insulin Resistance Atherosclerosis Study (IRAS) cohort, Hispanic participants were found to present significantly higher concentrations of circulatory CRP and plasminogen activator inhibitor-1 (PAI-1) in adiposity adjusted analyses¹⁴⁶. Ethnic disparities were also identified for adipokine levels between participants in the Study of Women's Health Across the Nation (SWAN), even after accounting for adiposity measures¹⁴².

Given the cross-phenotypic effects of genetic loci associated with adiposity¹⁴⁷, the ancestry-specific genetic architecture may underpin the disparities observed in adiposity-associated cardiometabolic risk factors and associations between ethnic groups¹⁴⁸.

Therefore, the investigation of genetic underpinnings is critical for an understanding of the mechanisms that result in disparities in cardiovascular risk factors between Hispanic/Latino populations and other populations. The handful of existing assessments involving minorities have limited inferential utility, largely due to smaller sample sizes, which have unsurprisingly resulted in contrasting conclusions^{149,150}.

Uncovering the genetic underpinnings of ancestry-specific inflammatory pathways that link central adiposity with the metabolic abnormalities that are considered precursors to deleterious cardiovascular outcomes is essential in this population. A genome-wide, systematic, and comprehensive evaluation of pleiotropic genes that overlap between adiposity and inflammation phenotypes in Hispanic/Latino populations is necessary to identify these genetic effects.

4.4. Adiposity–inflammation pleiotropy study in Hispanics/Latinos

The tri-admixed nature of Hispanic/Latinos captures genetic variations from Amerindian, African, and EU populations¹⁵¹, and facilitates the detection of variants that are infrequent/monomorphic in more homogeneous populations, as demonstrated previously^{152,153}.

Hispanics/Latinos also present opportunities for estimating causal probabilities for GWAS inferenced variants (fine-mapping)¹⁵⁴, due to shorter shared haplotypes compared with EU populations¹⁵⁵. Fine-mapping is even more powerful in the context of data from non-European ancestry populations¹⁵⁶, greatly reducing the number of SNPs identified for functional evaluations^{157,158}.

The provision of new genetic and molecular insights into such complex traits in this era of precision medicine is intended for clinical utility and the development of effective preventive strategies designed to improve the cardiovascular health of Hispanics and, by extension, all ethnic groups through the reduction of associated metabolic risk factors.

5. Study Rational

Assuming that adiposity and inflammation are interrelated manifestations of biological mechanisms, the identification and study of the genetic variants that underpin such mechanisms can enable the examination of phenotypic classifications, the repurposing of pharmaceutical interventions, and ultimately, cardiovascular disease prevention. The conceptual framework for the study is graphically illustrated in Figure 1.5.

The selection of a Hispanic population with unique, tri-admixed genetics will highlight functional variants and their associated pathways at the molecular level.

6. Aim of the Study Project

We aim to characterize the adiposity- and inflammation-associated genetic pleiotropy underpinning cardiovascular risk factors in Hispanic Americans by using observations from the Cameron County Hispanic Cohort (CCHC). By utilizing phenotypic and genotypic data, combined with advanced statistical approaches, SNPs that are significantly associated with adiposity and inflammatory traits will be explored, fine-mapped to identify functional variants, and sorted to distinguish biological and pleiotropic variants.

The specific aims of this dissertation are:

1) Identify loci with potential evidence for pleiotropy across inflammation, overall obesity, and central adiposity domains in a Hispanic/Latino population.

By employing measures of overall obesity (BMI), central obesity (waist to hip ratio, waist circumference), and inflammatory traits (CRP, Fibrinogen, IL-6, and IL-8 levels) obtained from blood samples¹⁵⁹, the study intends to perform the following:

a. Estimate genome-wide univariate associations between approximately 10 million SNPs and Aim 1a phenotypes.

b. Apply the adaptive sum of powered score (aSPU) test¹⁶⁰ to the Aim 1a summary statistics to identify potentially pleiotropic loci across phenotypic domains.

2) Leverage allelic associations and variants' functionalities to prioritize SNPs for in-depth statistical and experimental interrogation.

a. Using PAINTOR¹⁶¹ and FINEMAP¹⁶² to conduct fine-mapping on the loci identified in Aim 1b, prioritize SNP candidates for additional statistical and functional evaluation.

3) Distinguish SNPs with evidence for biological pleiotropy.

a. Application of pathway analysis to differentiate Aim-2 prioritized SNPs with evidence of biological pleiotropy (e.g., variants that influence several independent phenotypes) from mediated pleiotropy.

The specific hypotheses are that: (1) genome-wide assessment will highlight loci that potentially affect multiple inflammatory- and adiposity-associated phenotypes (**Aim 1**). (2) At these possibly pleiotropic loci, fine-mapping informed by population structure and functionality will pinpoint functional SNP candidates (**Aim 2**). (3) Some of these candidates will illustrate evidence of biological pleiotropy (**Aim 3**).

By considering obesity and inflammation to be interrelated manifestations of biological mechanisms, rather than in isolation, this proposal will pave the way towards gene and functional variant identification, enabling research examining etiology-based phenotype classification, drug repurposing, and, ultimately, cardiovascular disease prevention.

A Hispanic-Latino population was intentionally selected as a population with a high burden of disease, and that is often underrepresented in genomic studies, which ensures that our proposed study is innovative, particularly as next-generation sequencing, precision medicine, and direct-to-consumer genetic testing become more commonplace, in the context of an increasingly diverse population¹⁶³.

7. Tables and Figures

Figure 1.1. Obesity increases extracellular matrix levels within the kidney and may cause medullary compression, altering the intrarenal mechanical forces. This structural change increases reabsorption of sodium in the Henle Loop and conversely reduce sodium levels in Macula densa. The latter event would trigger renin-angiotensin system and increases glomerular filtration rate and renal blood flow, resulting in elevated extracellular fluid volume, increased hemodynamic load and ultimately hypertension which is a precursor to major cardiovascular adversities (Hall, J.E., 1997) .

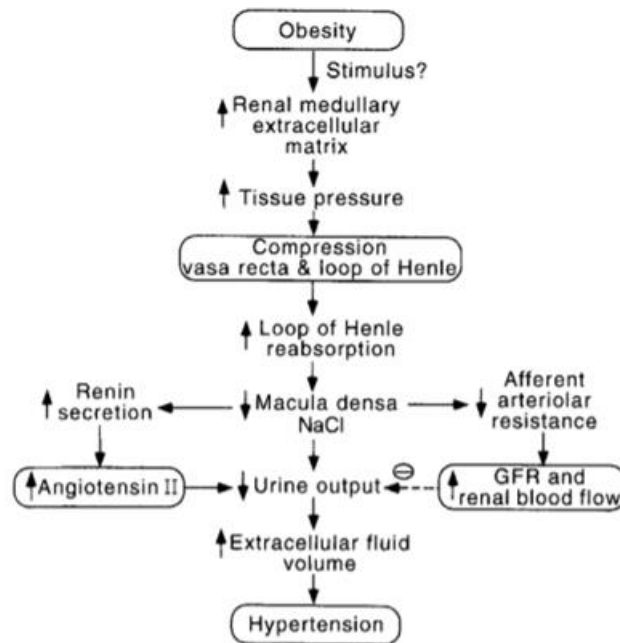


Figure 1.2. Adipocytes expand and replicate to store positive energy balance in the shape of lipids. The rapid or sustained accumulation of lipids can disrupt the differentiation of preadipocytes and accelerating the formation of larger, dysfunctional adipocytes that have altered hormonal and inflammatory profile (Van Gaal et al., 2006).

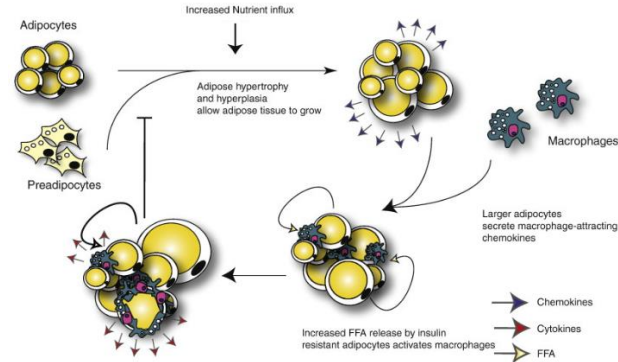


Figure 1.3. Hypertrophied adipocytes upregulate production of proinflammatory cytokines, including IL-6, that are key drivers of C-reactive protein (CRP) production in the liver (through direct liver access via the portal system). Within this inflammatory milieu, increased concentrations of macrophages admixed with lipolytic and apoptotic adipocytes promote positive feedback loops, resulting in the amplification of adverse metabolic and inflammatory actions (Van Gaal et al., 2006).

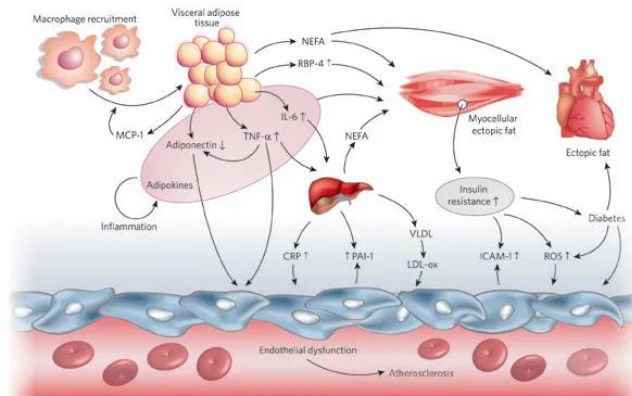


Figure 1.4. Functional overlap and feedback loop has been suggested where weight gain causes inflammation, resulting in metabolic/structural alterations that induce further fat accumulation

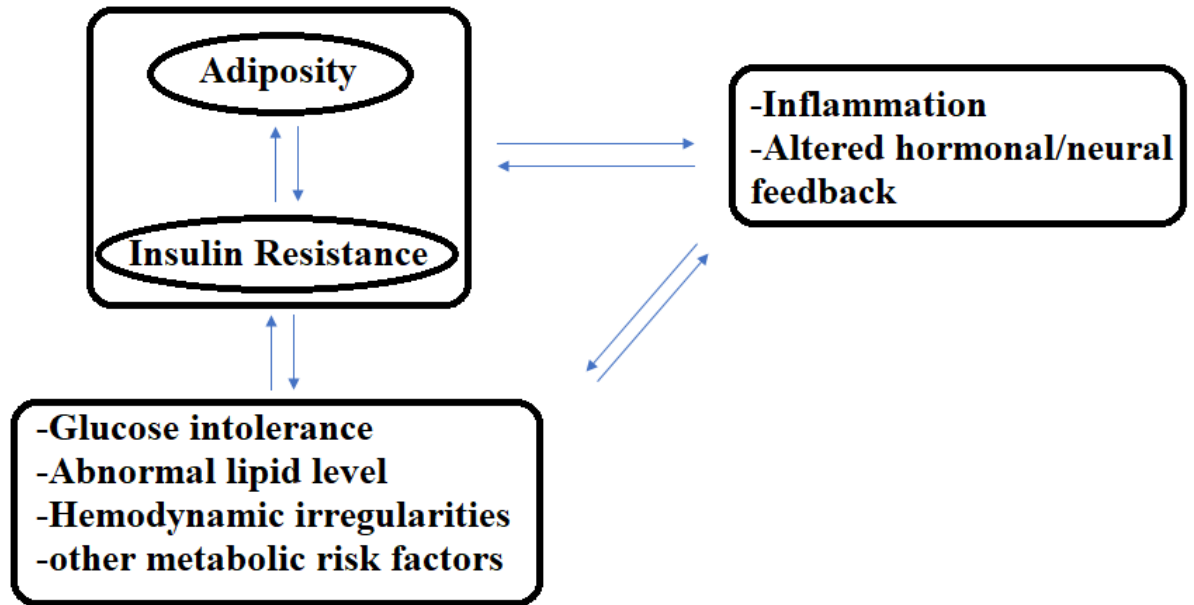
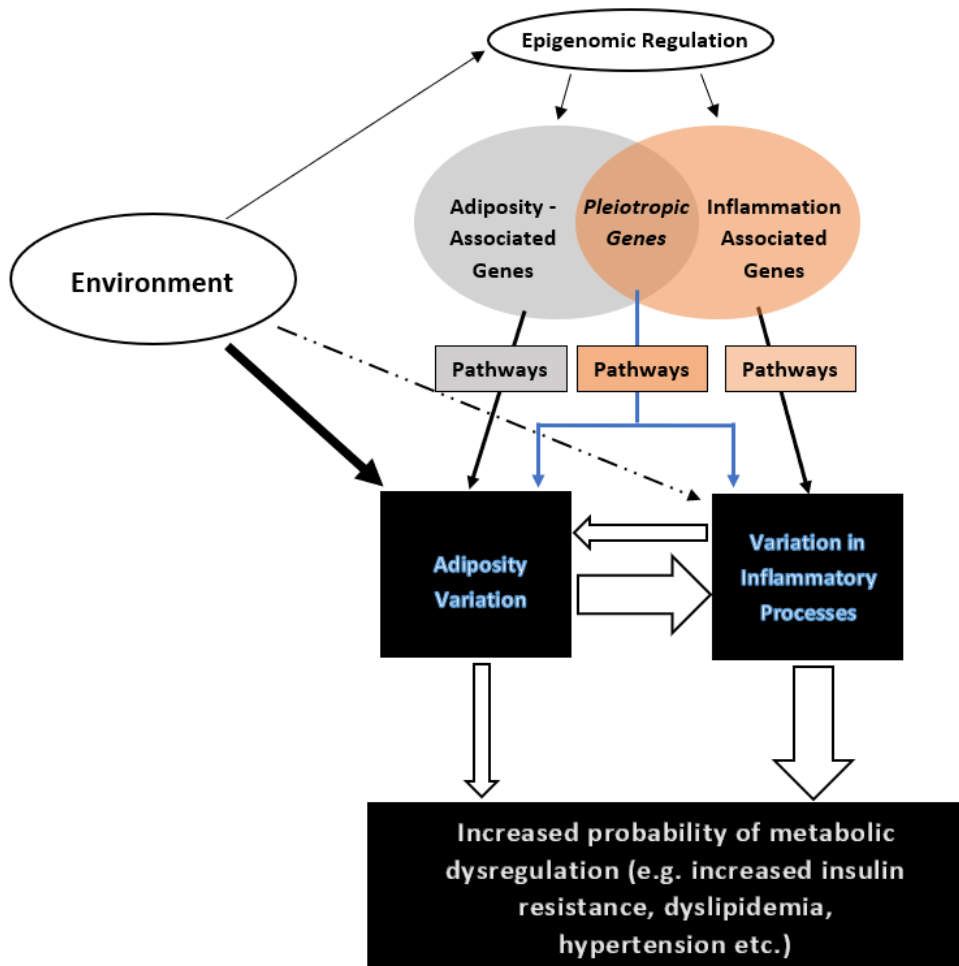


Figure 1.5. The conceptual framework for the adiposity-inflammation pleiotropy study.



METHODS

This chapters present information on study population, quality control procedures, and statistical approaches proposed for the data analysis. The review highlights the study setting and the recruitment process. Discussions entail descriptive analysis of genetic and phenotypic observations, variable selection, genetic data quality assessment and imputation, exclusion/inclusion criteria and outlier examinations. Mathematical summaries of the proposed statistical methods are provided in the latter part of the chapter.

1. Study Setting

Cameron County, in the state of Texas, is the southernmost county in the US. According to a recent Population and Housing Units Estimates report, provided by the US Census Bureau (in 2020) , the total population of the county is 423,163 individuals (census.gov), with the large majority (approximately 88%) self-identifying as Hispanic or Latino.

The county harbors a high percentage of residents who were born in Mexico¹⁶⁴, are economically classified as low-income, and have lower educational attainment than the national average¹⁶⁴. Based on data from 2004-2015 period, only one-third of the individuals in this county were covered by any type of health insurance program¹⁶⁵. The prevalence of obesity (57%), prediabetes (32%), diabetes (28.7%), hypertension (31.0%), and hypercholesterolemia (50.1%) among this population suggests high susceptibility to cardiovascular diseases and chronic inflammatory diseases^{166,167}.

1.1. Conception

The CCHC study started in 2004 and continually developed and expanded ever since. This study is intended to be a “Framingham-like” open enrollment observational cohort consisting of a Hispanic population, providing a setting for the exploration of a variety of health-related topics.

This initiative is primarily funded by the National Institute for Minority Health Disparities (NIMHD). Current research involving this cohort includes the study of diabetes risk, cardiovascular disease, peripheral artery disease, liver disease, cancers (liver/breast), mental health, intervention studies, genomics, immunology, imaging/elastography, clinical trials, and economics.

1.2. Recruitment process

Recruitment is ongoing, and the number of participants is now approximately 5,000 individuals; Medical examinations and interviews are performed at Clinical Research Units (CRUs) in both Brownsville and Laredo, Texas. The first round of recruitment included 2,000 participants and was performed between 2004–2007.

Although most participants have been adults, a small pediatric group (8-17 years old) was also enrolled in the study. Regular follow-up visits, performed every 5–10 years, are planned for participants recruited at baseline to allow for longitudinal data collection, using nested study designs, and the implementation of clinical trials.

Using a 2-stage, stratified sampling design, census tracts were classified into 4 strata according to income level. Targeted sampling was performed on the first (lowest) and third socioeconomic strata. From 42 clusters in the lower socioeconomic stratum, 11 tracts were

selected for random sampling; similarly, 11 tracts were selected from among the 39 clusters in the higher socioeconomic stratum¹⁶⁴. The highest stratum was omitted because it contained predominantly “winter Texans,” who were primarily retired visitors from the northern United States and Canada and were less likely to be Hispanic.

All households within a selected census block were invited to participate in the study. Overall, 71% of approached households elected to participate, with more participants from the lower SES stratum than from the upper stratum (78% vs. 63%); then, one person from each household was randomly recruited for the study. The recruitment process is currently in the second stage and is ongoing. Recruiters are from the local population and are bilingual and bicultural. Each randomly selected household was visited at least 5 times.

The institutional review boards at the University of Texas Health Science Center and the University of Texas at Brownsville reviewed and approved the protocol. They also approved the informed consent forms, which included permission to collect and store de-identified data and specimens for this and other studies.

2. Data Collection

Those who agreed to participate were then invited to visit a CRU for examinations. All questionnaires and explanations were scripted in both English and Spanish, and participants were not asked to fill out any forms. Questionnaires and examinations were conducted by nurses and field workers, who were trained in good clinical practices, in accordance with the National Institutes of Health requirements.

2.1. Anthropometric measures

The weights of the participants were measured with their shoes removed and recorded to the nearest 0.2 kg using a Seca digital scale (Birmingham, UK). Standing height (cm) was measured measured and recorded to the nearest 0.2 cm using a Perspective Enterprises stadiometer (Portage, MI, USA) and weight (kg). BMI was measured by dividing the body weight over the square of the height in meters (m²); waist circumference (visceral adiposity) was measured to the nearest 0.2 cm at the level of the umbilicus, with participants in a standing position and breathing normally¹⁶⁴. All measurements were performed twice, and the mean value was recorded in the dataset.

2.2. Inflammation biomarkers

Participants were asked to fast for at least 10 hours overnight before their visit to the CRU, which was located centrally at Valley Baptist Medical Center. After confirming the duration of their fast, blood samples were obtained for subsequent laboratory measurements and DNA sequencing; participants who had not fasted were rescheduled¹⁶⁴.

Specimens were aliquoted and frozen at -80°C . CRP levels were analyzed at a Clinical Laboratory Improvement Amendment (CLIA)-certified laboratory. IL-6 and IL-8 which are markers of chronic, low-grade inflammation, were measured using a multiplex enzyme-linked immunosorbent assay (ELISA, Milliplex Map, Millipore, CA, USA) and read in a Luminex 200 system (Luminex Corp., Austin TX).

2.3. Genetic measurements

Genetic measurements were completed at the Vanderbilt University Medical Center genotyping core facility, VANTAGE. Genotyping was performed using a MEGA-Ex panel, which included >2 M variants. MEGA also incorporates 400K low-frequency and

rare disruptive variants that were originally derived from a pan-ethnic panel of 40K whole-exome sequenced (WES) individuals and a heavily curated catalog of > 40K clinical variants. Measurements were recorded in binary *PLINK* format ¹⁶⁸.

3. Exploratory Analysis of Genetic Measurements and Quality Control

Out of 2,038,233 total genotyped variants, 1,842,794 single-nucleotide polymorphisms (SNPs) passed initial lab control, whereas the remainder did not perform consistently and were, therefore, dropped from the initial data set.

3.1. Pattern of missingness

Allelic frequency distribution showed that out of 1,842,794 total variants, slightly greater than 50% were classified as rare variants, which are defined as SNPs with minor allele frequencies (MAFs) of < 1% (Figure 2.1). Next, the per variant missingness was assessed (Figure 2.2). Although a sizeable proportion of variants (approximately 24%) had a maximum of 1% missingness, the proportion of variants with missingness equal to or exceeding 5% were negligible (approximately 5%).

Next, per sample missingness were examined. The initial exploration of samples from the first batch (available at the time of assessment) illustrated that approximately 3% of samples were missing 10% of genotyped variants or less, and nearly 4% of samples had a 5% missingness or less.

Finally, the pattern of per sample missingness compared to the distribution of autosomal heterozygosity was examined (Figure 2.3). Using a threshold equal to the median $\pm 3 \times$ the interquartile range, pattern missingness vs. autosomal heterozygosity was assessed, and no

sample was found to be outside of the threshold range. Larger than expected heterozygosity can indicate the possible contamination of DNA samples.

3.2. Sex mismatch

To determine genetic sex, we assessed the inbreeding coefficient for the X-chromosome or the F coefficient ratio¹⁶⁹, which represents the probability that a person receives genetic markers on the X-chromosome from a single ancestor.

Because males have only one X-chromosome (received from the mother), the coefficient is defined to be 1, regardless of inbreeding. For females, the X-chromosome kinship coefficient is either 0 (completely outbred) or close to 0, if any degree of kinship is identified between the parents.

Generally, an F coefficient value < 0.2 is considered to be genetically female, whereas a value > 0.8 is considered to be male. The distribution coefficients generally correlated with the pedigree sex (Figure 2.4), with some mismatches. autosomal allelic heterozygosity compared with X-chromosome (sex) heterozygosity was also assessed (Figure 2.5); results were similar to the F coefficient distribution and similarly included some mismatches between genetic vs. pedigree sex.

3.3. Genetic outliers

CCHC genetic observations were merged with 925 samples from the 1000 Genomes (1000G) project population to assess the alignment of the study participants against reference ancestries using the *smartpca* tool from the *Eigensoft* package. A plot of the first 2 principal components (PCs, Figure 2.6a) suggested a few outliers with different ancestries compared with the rest of the samples; however, these samples did not

consistently separate from the rest of the samples when the third PC was considered (Figure 2.6b and c).

3.4. Hardy-Weinberg Equilibrium

Determining whether genotypic frequencies are in Hardy-Weinberg Equilibrium (HWE) is an important step that must be performed during the quality analysis of genotyped variants. The HWE states that the allelic frequencies of autosomal genes in a population remain unchanged throughout generations in the absence of disruptive forces, such as assortative mating, mutation, and migration. However, a departure from HWE is most likely indicative of potential genotyping errors or population stratification. Yet, given that the cohort under study is a recently admixed population, only extreme deviation from HWE was considered. A negligible number of variants were found outside of the significant thresholds for HWE ($< 10^{-6}$, $< 10^{-20}$, $< 10^{-50}$, Figure 2.7).

3.5. Duplicated samples and variants

After analysis with *Kinship-based Inference for Gwas (KING)*¹⁷⁰, 76 pairs of samples were found to have > 80% genetic concordance with each other. A similar number of pairs were also found during the pedigree reconstruction process using *PRIMUS*¹⁷¹ (Figure 2.8). Additionally, approximately 50K variants were also found to be duplicated.

3.6. Quality control process

Based on empirical observations from prior steps and the template proposed for performing genome-wide association studies (GWAS)¹⁷², the following quality control steps were performed:

- a) removing duplicated variants and indels (insertions/deletions).
- b) removing variants with low MAF (< 1%).
- c) excluding samples with more than 5% missingness and variants with greater than 5% missingness (among all samples).
- d) excluding samples with genetic vs. pedigree mismatch ($F > 0.3$ for females).
- e) removing samples that fall outside of the heterozygosity threshold or HWE significance level ($p\text{-value} < 1^{-10}$), and
- f) removing variants with either allelic or strand mismatch compared against the reference 1000G panel (this step was only completed for dataset used for imputation with 1000G).

In Table 2.1, the rows represent the temporal sequence for the process. All the steps were completed using the *PLINK* package.

4. Principal Component Analyses

In recent years principal component analysis (PCA) has been the most widely adopted approach for both ancestry inference and the correction of population stratification in genomic studies that utilize high-density SNP genotyping data¹⁷³. To calculate PCA, filtered data were used that were previously assessed for quality control standards (as discussed above) and further subsetted these data to include only common variants, with $MAF > 5\%$ (approximately 117,000 SNPs) The selection of common variants can improve ancestry inferencing because these variants are more likely to be observed in reference populations outside of the study.

Using this common variant subset, PCs were calculated using the packages *GENESIS*¹⁷⁴ (for PCAir) and *PRIMUS*¹⁷¹ (PCA). The primary difference between PCAir and PCA is that PCAir adjusts for pairwise kinship. The *GENESIS* method utilizes genomic observations and an efficient algorithm¹⁷⁵ to identify a diverse subset of unrelated samples, which are representative of all ancestries in the data. *GENESIS* uses pairwise relatedness and ancestry divergence measurements on genotyped variants from the autosomal chromosomes and separates the dataset into an ancestry representative subset, u , of unrelated individuals from the total sample, N . PC estimates obtained from the analysis of a subset u would then be used to predict PCs in the set of related individuals, R , based on genetic similarities¹⁷⁵. Samples from 1000G were used as the ancestry references ($N = 2,504$).

Procedurally, PCs were first calculated using a combined set of unrelated CCHC individuals and reference samples. The threshold for relatedness was set to a 3rd-degree relatives, which equates to a first cousin or more distant relative. A slight difference in the number of unrelated individuals was calculated by *PRIMUS* vs. *GENESIS* (2,282 vs. 2,253) due to differences in their respective computational algorithms. Subsequently, PCs were calculated during the first phase and were then projected over the subset of related samples (Figure 2.9).

A sizable majority of CCHC individuals clustered with predominantly Mexican populations, and smaller sets clustered with other ancestry groups (Figures 2.9). The parallel plot of the first 10 PCs also suggested the primacy of Mexican ancestry among the study participants (AMR group in Figure 2.10).

The analysis of explained variance by PC (Figure 2.11) suggested that a negligible proportion of variations in the data is captured by PCs beyond the 3rd PC. No practical difference was observed between the PRIMUS-calculated PCA and PCAir (Figure 2.12); therefore, PRIMUS-derived PCs were selected to ensure consistency with other research studies using the same cohort.

5. Imputation

5.1. Pre-imputation quality control

Using McCarthy group tools (well.ox.ac.uk/~wrayner/tools), genotyped variants underwent further quality control assessments. Steps included strand update (top strand), position check, and the assignment of reference/alternative allele assignments for genotyped PLINK files. The following categories of variants were excluded from the set:

- i) A/T and G/C SNPs if $MAF > 0.4$.
- ii) SNPs with differing alleles (relative to the reference genome), and
- iii) SNPs with > 0.2 allele frequency difference (compared to the reference genome).

5.2. Imputation panels

A cleaned dataset was loaded into the Michigan Imputation Server¹⁷⁶ for imputation using a 1000G Phase 3 panel and a Trans-Omics for Precision Medicine (TOPMed) freeze 8 panel (imputation.biodatacatalyst.nih.gov). The advantages of the 1000G panel include well-characterized reference ancestry groups, along 26 clusters and from all continents and wider adoption in genetic studies, which would facilitate the replication and meta-analysis of different studies using the same set of variants.

TOPMed (nhlbiwgs.org), in contrast, is a consortium of over 30 different study cohorts. Compared with 1000G, the latest freeze 8 reference panel was inferred from the deep sequencing of 186K individuals, which represents a substantial improvement in the number of available variants for imputations and increased diversity of ancestry groups. Currently, this freeze includes 811M SNPs and 66M short indel variants. The availability of larger numbers of SNPs would greatly facilitate genetic studies that involve rare or ancestry specific variants. However, TOPMed is a work in progress and is not yet as widely adopted as 1000G.

5.3. Imputation procedure

With both 1000G and TOPMed servers, imputations were completed with the Minimac4¹⁷⁷ imputation platform. The mean imputation accuracy was set to 10%. For genotype phasing, where each allele is assigned to its proper pair of (paternal or maternal) chromosomes (for haplotypic referencing), the *Eagle* phasing algorithm¹⁷⁸ was utilized. During the TOPMed imputation run, all populations were selected as the parameters for ancestry reference.

The results and the number of variants imputed with TOPMed for each chromosome are presented in Table 2.2. An average of approximately 92% alignment between the reference panel and the genotyped set was achieved. Additionally, the results from the TOPMed panel were converted to a newer genome reference consortium human build 38 (GRChr38)¹⁷⁹.

5.4. Post-imputation quality control

Variants with rare MAF ($< 1\%$) and poor imputation quality ($< 40\%$) or monomorphic (i.e., all individuals having the same genotype at a given site) within the study population were excluded from this analysis. After the completion of the recommended quality control steps, approximately 9 million SNPs remained, which were used for subsequent analyses.

6. Phenotypes

BMI was selected as the measure for overall adiposity, and WHR and WC were selected as reliable indicators of central obesity and visceral adiposity, respectively; all three measures are significantly correlated with inflammatory markers and have been widely adopted in GWAS research on adiposity traits. For inflammation, phenotypes, including CRP, IL6, IL8, and fibrinogen, were included in the assessment (Table 2.3). Exploratory analyses of the target traits showed skewness in the distributions of inflammatory markers, which was suggestive of extreme values at the left tails, although a few high values were also observed for BMI and WHR (Figure 2.13). The data included measurements taken at both visits 1 and 2.

6.1. Covariates

For GWAS analysis, covariates including age (year), sex and BMI were considered. For pathway analysis, age, sex, physical exercise (binary, 150 min of moderate to vigorous activity/week), alcohol consumption (Ounce/week), current smoking status (binary) and homeostatic insulin resistance (HOMA-IR, uIU/mL) were used. Covariates used for pathway analysis were similarly taken from the exams where the primary phenotypes were selected.

6.2. Inclusion/exclusion criteria

The study population was restricted to individuals whose genetic data were available (3,313 out of 4,933). All participants under the age of 15 were excluded from the study because notable differences exist between genes that affect childhood vs. adulthood obesity¹⁸⁰. Prior assessments in larger settings also suggested sexually dimorphic anthropometric variants in regard to central obesity¹⁸¹; therefore, both WHR and WC traits were stratified by sex. Finally, the log-transformed distributions of traits were assessed.

For the IL-6 and IL8 phenotype, the empirical distribution suggested a proportion of observations with the same value (0.64 pg/dL), and this pattern was often observed with measurements taken before January 2010. This value is unlikely to have been the detection threshold for the assay because values smaller than 0.64 were identified. Whether some portion of the values were rounded as a standard reporting practice could not be determined. To minimize bias in the results, the last non-missing observation for each trait in each sample was selected for GWSA analysis (Table 2.4).

6.3. Examination of phenotypic outliers

A considerable number of extreme observations were identified at both tails (Figure 2.13, empirical distributions; Figure 2.14, log-transformed distributions). Although outliers could introduce noise or induce false-positive associations in genetic tests, aggressive filtration can reduce sample sizes and inversely affect power; therefore, a delicate balance is necessary. To address this issue, two comparative methods were utilized:

6.4. The minimalistic approach

Using the minimalistic approach, both biological plausibility and empirical distributions (both raw and log-transformed) were used as criteria for setting outlier cut points. For anthropometric traits, almost all of the observations were in ranges that were biologically plausible, even if clinically abnormal. Therefore, only a few extreme cases that were considerably detached from the rest of the distributions at each tail were excluded.

For CRP, clinical studies have shown that concentration levels above 350 (mg/L) are indicative of underlying severe acute inflammation¹⁸², whereas levels ranging from 10–100 (mg/L) could either indicate active inflammation or be a high normal. Given the empirical distribution and clinical criteria, a few observations above 200 (mg/L) were excluded.

For IL6, the upper range of measurement can be as high as 1000 (pg/dL), especially in individuals with underlying conditions (e.g., infection, cancer, chronic diseases). However, we used both empirical distribution and the reported literature to determine a cut-off point of 100 (pg/dL) as the threshold for outliers, as observations above this level are generally considered to be abnormal and are less likely to be observed in individuals without acute inflammation^{183,184}.

For IL8, levels as extreme as 10,000 (pg/dL) have been reported in severely ill patients¹⁸⁵. Based on the empirical distribution and the highest ranges known for non-acute cases, the threshold was set to 80 (pg/dL)^{186,187}. For all traits, extremely low measurements (observed in the log-transformed distributions, Figure 15) were also excluded. After the filtration of outliers, the remaining measurements were log-transformed and rank-normalized.

6.5. Hybrid approach

Although the minimalistic approach excluded extremes, the dataset continued to include values that were otherwise abnormally high or low; therefore, the possibility of severe or acute underlying diseases being present at the time of measurement could not be excluded. In such scenarios, data is customarily rank-normalized to minimize the likelihood of bias in the results; however, this process can induce additional problems, such as the loss of statistical power¹⁸⁸. Therefore, a more conservative outlier examination was deemed necessary, particularly for inflammatory traits.

First, the distribution of inflammatory markers in similar Hispanic ancestry populations was reviewed in the literature (Table 2.5), which provided inferences for setting standard deviation-based cut-off points for each trait. Subsequently, the data were log-transformed, and the values outside of the proposed cut-off points were then set to missing. Descriptive distributions of the study phenotypes used during the hybrid filtration approach are presented in Table 2.6.

Lastly, outliers were examined to ensure that no specific pattern could be observed.

6.6. Selection of the phenotype set

No discernible difference was observed for either phenotype set after the conclusion of the GWAS tests. Therefore, the results from the hybrid assessment process were selected for subsequent analyses.

7. Statistical Approach

7.1. Genome-wide association test

Generalized linear models are now the primary statistical tools used during association analyses in various scientific fields¹⁸⁹, including biology. Assuming that the subjects are independent, the linear association between the outcome (phenotype) and predictor (SNP variant) can be written as $M1: g(E(Y)) = X\alpha + G\beta$, where Y is a vector for continuous or binary phenotype data, g is a link function, X denotes confounders, and G is the SNP dosage.

However, given the study setting, in which a sizeable proportion of the individuals have kinship relationships with each other, a generalized linear mixed model (GLMM) would be more appropriate to allow for family structure adjustments, beyond the ancestry PCs that account for most distant relatedness and ethnic variations. Hence, a mixed model was implemented in the *GENESIS* package¹⁹⁰ for use during GWAS tests.

GENESIS fits linear mixed models for quantitative phenotypes and penalized-quasi likelihood for binary (or count) traits. In comparison with other packages that use the Likelihood Ratio or Wald tests¹⁹¹⁻¹⁹³, *GENESIS* uses the Lagrange-Multiplier (score)¹⁹⁴ test. A distinguishing feature of score-based tests is that only the null model must be specified to make statistical inference; thus, this type of model is computationally faster than the other two types, for which more complicated models must be fitted.

The package uses a reduced maximum likelihood (REML) algorithm¹⁹⁵, a special form of maximum likelihood estimation, in which the degrees of freedom are used to estimate fixed effects while calculating variance components. Compared with the general maximum

likelihood method, REML operates on linear combinations of observations, chosen such that those combinations are invariant to the values of fixed-effect parameters.

Relatedness due to familial relationships is accounted for by including a random effect term with a covariance matrix proportional to the genetic relationship matrix (GRM) or kinship matrix K . Specifically, the GLMM assumes that the variance matrix for the phenotype Y has the form $V1: Var(Y) = \sigma^2 I + \gamma^2 K$, where I is the identity matrix. In a single SNP analysis, the computing time was greatly reduced by estimating σ^2 and γ^2 from the null model $M0: g(E(Y)) = X\alpha$ and then using these estimates in a score test of $\beta = 0$ in $M1$.

For the family structure effect, the *GENESIS* package uses PC-Relate¹⁹⁶. Briefly, PC-Relate is a PC-based approach for the estimation of commonly used measures of recent genetic relatedness, such as kinship coefficients and identity by descent (IBD), which share probabilities in samples with population structures, admixture, and HWE departures.

PC-Relate is essentially a hypothesis (model)-free approach for inferring recent kinship because it does not require prior estimates of individual ancestry and population-specific allele frequencies, IBD-sharing likelihood models, or the specification of a population genetic model. It uses PCs calculated from genomic data to partition genetic correlations among individuals into two separate groups based on the sharing of recent or more distant common ancestry¹⁹⁶.

7.2. Pleiotropy assessment

Various pleiotropy methods have recently been proposed¹⁹⁷ that take advantage of the increased availability of summary statistics from trait-specific GWAS. Most of these

methods offer superior statistical power compared with multivariate analysis¹⁹⁸. Because the GWAS results in this study were based on score tests, the adaptive Sum of Powered Scores test (aSPU), which itself is essentially a score-based test, refined for multi-trait association testing¹⁹⁹, was used for pleiotropy assessments.

Briefly, suppose $U = (U_1 \dots U_k)^T$ to be the score vector for a set of k trait-SNPs to be tested with the $H_0: \beta = 0$. The score test is: $T_{SCO} = U^T V^{-1} U$, which is equivalent to an asymptomatic Wald test with k degrees of freedom. The popular single-variant test for correlated traits is the univariate minimum P (*UminP*) method, which tests every single variant, one-by-one, and selects the minimum of their p-values²⁰⁰.

Pan et al. showed that the corresponding multi-trait *UminP* for score-based tests is $T_{UminP} = \max_{j=1, \dots, k} U_j^2 / V_{jj}$ ²⁰¹, where U_j is the element of U , and V is the (j,j) diagonal element of the $V = \widehat{Cov}(U)$. However, when the number of traits increases, the test loses power. To address this issue, the sum of score tests was proposed²⁰²: $T_{Sum} = \sum_{j=1}^k U_j$.

This test may also be ineffective when the direction of effects across traits differs. One solution was the sum of squared score (SSU) test²⁰²: $T_{SSU} = U^T U = \sum_{j=1}^k U_j^2$, which was later scaled to a more generalized test that can be adapted for different scenarios²⁰³: $T_{SPU(\gamma)} = \sum_{j=1}^p U_j^\gamma$, where gamma is the power parameter. Note that when the gamma parameter approaches infinity²⁰³, $T_{SPU(\gamma)} \propto (\sum_{j=1}^p |U_j|^\gamma)^{1/\gamma} \rightarrow \max_j |U_j| = T_{SPU(\infty)}$, which is similar to the *UminP* test. Each gamma value may yield a higher power under certain scenarios.

However, because these scenarios cannot be known in advance, an adaptive (aSPU) approach is necessary to maximize the power of a given set of gamma values; hence²⁰⁴: $T_{aSPU} = \min_{\gamma \in \Gamma} P_{SPU(\gamma)}$, where $P_{SP\infty U(\gamma)}$ is the pleiotropic p-value of the test and Γ contains a set of suggested gamma (i.e. 1,2,3... ∞) values, of which at least one value may maximize the detection power.

7.3. Fine-mapping

Although GWAS or pleiotropy analysis can provide crucial insights into the genetic underpinnings of phenotypes, these analyses do not necessarily indicate causality. Due to linkage disequilibrium (LD) between SNPs or coverage quality of genotyping panels (SNP tagging)²⁰⁵, variants with the lowest p-values may not be phenotype-affecting SNPs.

Fine-mapping is a statistical analysis approach that can be used to assign a causal probability to GWAS-inferenced variants. In this study, two complementary methods were utilized for fine-mapping loci with suggestive evidence for multi-trait associations after pleiotropic analysis.

7.3.1. PAINTOR

The PAINTOR (Probabilistic Annotation INTEgratOR) approach takes advantage of both the LD structure and functional annotation. In association analysis for continuous traits, values are marginally regressed on each variant, and the corresponding Z-score ($\hat{\beta}/se(\hat{\beta})$) is obtained.

Assuming that Z_j is the vector of the [pleiotropic] Z-score in the j^{th} locus ($1 \leq j \leq L$) of the size N_j (i.e., total number of SNPs in the locus); Σ_j is the corresponding SNP pairwise

pearson correlation coefficient for the locus j (i.e., LD matrix), derived from a set of unrelated individuals with K annotations ($1 \leq k \leq K$) for each SNP i ; $A_{i,j}$ is a binary vector where $A_{i,j,k} = 1$ if the i^{th} variant at the j^{th} locus is part of the annotation k ; γ_k is the effect size of the k^{th} annotation on the probability that a SNP is causal; λ_i is the vector of non-centrality parameter ; C_j is the vector of causality, where $C_{i,j} = 1$ if SNP i at the locus j is causal; then the prior probability of the Bayes formulation is²⁰⁶:

$$P(Z_j|C_j; \lambda_j) = N(Z_j; \Sigma_j(C_j \circ \lambda_j), \Sigma_j)$$

$$P(C_j; \gamma_k) = \prod_i P(C_{ij}; \gamma_k) \dots (1)$$

$$P(C_{ij}; \gamma_k) = \left(\frac{1}{1 + \exp(\gamma_k A_{ijk})} \right)^{C_{ij}} \left(\frac{1}{1 + \exp(-\gamma_k A_{ijk})} \right)^{1-C_{ij}} \dots (2)$$

PAINTOR computes posterior probabilities of each causal configuration C_j over all possible causal configurations²⁰⁶: $Q_j(|Q_j|) = \sum_{i=0}^S \binom{N_j}{i}$, where S is the potential causal numbers considered for each locus j . With equations above, the posterior probability for each C_j is²⁰⁶:

$$P(C_j|Z_j; \gamma_k, \lambda) = \frac{P(Z_j|C_j; \lambda_j)P(C_j; \gamma_k)}{\sum_{C_j \in Q_j} P(Z_j|C_j; \lambda_j) P(C_j; \gamma_k)} \dots (3)$$

And the posterior probability for each SNP _{i,j} ²⁰⁶:

$$P(C_{ij} = 1|Z_j; \gamma_k, \lambda) = \sum_{C_j \in Q_j; C_{ij}=1} P(C|Z_j; \gamma_k, \lambda) \dots (4)$$

Because posterior probabilities are calculated independently for each locus, the package undertakes further mathematical steps that aggregate results across loci and use these

results to update likelihood parameters, particularly γ_k and λ , which improve posterior probability estimates. The incorporation of functional annotation has been illustrated to significantly improve causal estimates in prior analyses¹⁵⁸.

7.3.2. FINEMAP

Essentially, FINEMAP is a Bayesian approach that incorporates likelihood function, priors, likelihood estimation, and stochastic search algorithm¹⁶². Assuming a linear model $y = X\lambda + \epsilon$, where y is vector of mean-standardized values for a quantitative trait in n individuals and X is a column-standardized SNP genotype matrix of dimension $n \times m$, the maximum likelihood estimate of the causal variants' effects λ depends on X and y through the SNP correlation matrix. Priors for causal variants' effect λ is calculated via $p(\lambda|\gamma) = N(\lambda|0, s_\lambda^2 \sigma^2 \Delta_\gamma)$, where γ is a binary vector, for which $\gamma_l = 1$ if l is causal and 0 otherwise; s_λ^2 is assigned prior variance for causal effects in units of $\sigma^2 = 1$; and Δ_γ is a diagonal matrix.

The probability of each causal configuration γ is assumed as $p(\gamma) = p_k / \binom{m}{k}$ when $\sum_{l=1}^m l = k$. The marginal likelihood of γ is obtained by integrating the causal effects of λ . The posterior probability of any causal configuration is estimated via $p(\gamma|y, X) = \binom{m}{k}^{-1} p_k \times BF(\gamma: NULL)$, where BF is the Bayesian factor and k is the number of causal variant in that specific configuration.

Finally, this approach uses the shotgun stochastic search²⁰⁷ to compare various causal configurations, with each obtained through the process discussed above, to identify those with the highest posterior probability. FINEMAP conducts a pre-defined round of iterations within the space of causal configuration. Each iteration is defined by deleting,

changing, or adding a causal SNP from the prior configuration, which is referred to as manipulative sampling from the preceding iteration.

All evaluated configurations and their posterior probabilities are saved in a list, Γ^* . The posterior probability that SNPs in configuration γ are causal is estimated by normalizing over the Γ^* list: $p(\gamma|y, X) = p^*(\gamma|y, X) / \sum_{\gamma \in \Gamma^*} p^*(\gamma|y, X)$, and for the particular SNP l th to be causal: $p(\gamma_l|y, X) = \sum_{\gamma \in \Gamma^*} 1(\gamma_l = 1) p(\gamma|y, X)$ ¹⁶². The advantages of this method are the minimal input requirements (i.e., Z-scores, minor allele frequencies, and SNP correlations) and the possibility of probing causal configurations with higher numbers of causal SNPs in each locus without incurring a substantial computation penalty.

7.4. Causal pathway analysis approach

Pleiotropy does not distinguish whether an observed association between a genetic variant and a phenotype involves direct (i.e., biologic) or indirect effects due to phenotypic correlations with the primary trait. Mediation refers to a type of conditional analysis, in which the total effect between the predictor and the outcome is appropriated between several factors that contribute to the association and relies on counterfactual principles²⁰⁸. Using this framework, the causal effects can be viewed as the difference between two potential phenotypic outcomes, depending on whether an individual carries the risk alleles or not (at the same time).

Suppose T_i denotes the presence of an effective allele, which would be equal to 1 if present and 0 otherwise; $Y_i(t)$ would denote the potential phenotypic outcome under allelic presence status; then, the causal effect of the variant on the phenotype would be $Y_i(1) - Y_i(0)$ for all values of i . However, because only one state (either 1 or 0) is observable at

any time for each i , the focus is typically shifted toward the estimation of the average causal effect by comparing two groups based on the randomized distribution of the effect allele, defined as $E(Y_i(1) - Y_i(0))$ ²⁰⁹.

Assume that the causal association between a given allele and a phenotypic outcome is totally or partially mediated by another [correlated] factor. Let $M_i(t)$ denote the potential value of a mediator of interest for subject i for a given allele status $T_i = t$. Let $Y_i(t, m)$ denote the potential outcome that would be observed if the allele and mediating variables equal t and m , respectively.

In practice, only one potential outcome is observable, and that outcome, Y_i , is equal to $Y_i(T_i, M_i(T_i))$, where $M_i(T_i)$ represents the observed value of the mediator M_i . The total effect of the allele on the outcome can be written as $\tau_i = Y_i(1, M_i(1)) - Y_i(0, M_i(0))$. The total effect can then be decomposed into the two components: the causal mediation effect²¹⁰ $\delta_i = Y_i(t, M_i(1)) - Y_i(t, M_i(0))$, for the effect allele status (0,1), and direct effect $\zeta_i = Y_i(0, M_i(t)) - Y_i(0, M_i(t))$, for subject i and each allele status (0,1).

For both decomposed elements, including δ_i and ζ_i , because only one effect allele status is observable, the average causal mediation effect (ACME) and the average direct (ADE) could be estimated, as $E\delta_i$ and $E\zeta_i$, respectively. Now the total effect can be re-written as: $\bar{\tau} = E(Y_i(1, M_i(1)) - Y_i(0, M_i(0))) = \frac{1}{2} \sum_{t=0}^1 \{\delta(t) + \zeta(t)\}v$, and under the assumption of no interaction, ACME and ADE sum up to the average total effect²¹¹.

7.4.1. Sensitivity analysis

Causal pathway sensitivity analysis proposed for this study relies on the sequential ignorability²¹² assumption. The test implies that after adjustment for covariates, no unmeasured variable confounds any of: 1) variant-mediator association (i.e. ignorability of effect allele state)²¹²: $Y_i(t, m), M_i(t) \perp\!\!\!\perp T_i | X_i$ for all $t, m = 0, 1$, where X_i is covariate(s), 2) allele-phenotype outcome association: $Y_i(t, m) \perp\!\!\!\perp M_i(t) | (T_i = t, X_i)$ for all $t, m = 0, 1$, and 3) mediator-outcome relationships. The validity of this mediation method largely rely on assumption 1, regardless of whether hierarchical models are fitted to the data or not, which is an important consideration.

8. Tables and Figures

Table 2.1 Quality control steps completed with genotyped variants prior to principal component analysis.

Filter	Samples	Variants
Baseline dataset	3,739	1,849,381
Remove duplicate variants & indels		1,796,564
Remove variants with >10% missingness		1,771,718
Remove samples >3% missingness	3,584	
Remove one pair of duplicate samples (the pair with higher missingness)	3,508	
Remove variants maf<1%		921,083
Remove variants with >5% missingness		917,256
Remove samples with >5% missingness	3,506	
Remove samples outside heterozygosity filter (f coefficient >0.3)	3,505	
Remove samples with pedigree vs genetic sex mismatch	3,387	
Remove variants violating hwe (p-value<1e-10)		914,301
Remove variants with allele/strand mismatch (vs. 1000 genome panel variants)		843,850
Remove control samples	3,313	
Pre principal component analysis sample	3,313	843,850

Table 2.2 The number of variants imputed for each chromosome using TOPMed panel. The total number of variants were 60,941,072.

Chromosome	Imputed Variants	Chromosome	Imputed Variants
1	4,889,931	7	3,575,429
2	5,277,930	8	3,426,745
3	4,370,489	9	2,690,807
4	4,310,328	10	3,008,743
5	3,980,182	11	454,157
6	3,782,515	12	2,931,624
13	2,183,250	19	1,393,198
14	1,948,976	20	1,379,845
15	1,781,331	21	810,711
16	1,984,703	22	865,521
17	1,774,596	X	2,388,886
18	1,731,175	Total	60,941,072

Table 2.3 List of the phenotypes selected for the study.

Selected phenotype	Reason for inclusion
Body Mass index	Associated genes also linked to inflammation ¹¹⁴
WAIST Circumference	A reliable indicator of visceral adiposity, and significantly correlated inflammation and cardiovascular risk factors ²¹³
WAIST to HiP Ratio	A reliable indicator of visceral adiposity, and significantly correlated inflammation and cardiovascular risk factors ²¹⁴
C-reactive Protein	Plays a leading role between adiposity-CVD ^{215,216}
Interleukin-6	Associated genes significantly linked to variation in adiposity in Europeans ²¹⁷
Interleukin-8	Inflammatory marker significantly raised in hyperlipidemia and cardiovascular diseases context ²¹⁸
Fibrinogen	Significant correlation with adiposity markers ²¹⁹

Table 2.4 Descriptive Distribution of anthropometry and inflammatory phenotypes.

Phenotype (unit)	Min.	1st Qu.	Median	Mean	SD	3rd Qu.	Max.	N
<i>BMI (kg/m²)</i>	13.29	26.31	29.88	30.91	6.96	34.31	70.48	3180
<i>WHR-Female(wc/hip*100)</i>	36.39	86.5	91.3	91.19	7.93	96.03	184.79	2087
<i>WHR-Male(wc/hip*100)</i>	66.31	91.6	95.98	95.93	7.59	100.21	215.79	1093
<i>Waist-female (cm)</i>	37	90	99.5	101.03	16.56	110	186	2088
<i>Waist-male (cm)</i>	68	94	102.5	104.16	15.2	112	168	1096
<i>CRP (mg/L)</i>	0.08	2	3.99	7.97	14.3	8.1	242.5	1937
<i>IL-6 (pg/dl)</i>	0.01	1.4	2.68	5.6	20.97	4.91	521.38	1362
<i>IL-8 (pg/dl)</i>	0.0	3.42	4.69	6.53	7.13	6.81	100.01	1499
<i>Fibrinogen (mg/L)</i>	0.05	0.51	0.8	0.95	0.74	1.18	11.89	2544
<i>AGE(Year)</i>	15.00	33.00	45.00	49.92	17.22	59.00	94.00	3180

Table 2.5 Proposed cut-points for exclusion of phenotypic outliers under hybrid approach.

Marker	Unit	Median(IQR)	Mean(SD)	Populations	Proposed Cut-points
CRP	g/dl	2.7 (1.2–6.0)		Hispanics ²²⁰	± 2.4 SD
IL6	pg/dl	1.6 (0.8–2.8)		Hispanics ²²⁰	
IL6	pg/dl	0.5(0.5-10)		Hispanics ²²¹	± 2.8 SD
IL6	pg/dl		0.9(0.8)	Hispanics ²²¹	
IL8	pg/dl		3.7(2.3)	Hispanics ²²²	
IL8	pg/dl	4.4(3.2–5.9)		Hispanics ²²³	± 2.6 SD
Fibrinogen	g/L		2.8(0.75)	Mixed ²²⁴	
Fibrinogen	g/L		3.6(3.4)	Mixed ²²⁵	± 2.8 SD
WHR-F	WC/H				± 3.0 SD
WHR-M	WC/H				± 3.0 SD
WC-M	cm				± 3.0 SD
WC-F	cm				± 3.0 SD

Table 2.6 Descriptive distributions of study phenotypes after exclusion of outliers under hybrid method.

Phenotype (unit)	Min.	1st Qu.	Median	Mean	SD	3rd Qu.	Max.	N
<i>BMI (kg/m²)</i>	16.16	26.30	29.86	30.76	6.53	34.25	57.21	3159
<i>WHR-Female(wc/hip*100)</i>	69.89	86.54	91.30	91.17	7.08	96.00	116.39	2077
<i>WHR-Male(wc/hip*100)</i>	77.87	91.70	96.01	95.94	6.40	100.22	113.58	1084
<i>Waist-female (cm)</i>	62.00	90.00	99.50	100.77	15.52	109.50	158.00	2072
<i>Waist-male (cm)</i>	68.30	94.00	102.50	103.95	14.62	112.00	157.00	1089
<i>CRP (mg/L)</i>	0.35	2.00	3.90	6.23	6.93	7.73	50.03	1880
<i>IL-6 (pg/dl)</i>	0.16	1.44	2.68	4.32	5.74	4.87	47.38	1336
<i>IL-8 (pg/dl)</i>	1.10	3.42	4.67	5.67	3.58	6.65	25.56	1448
<i>Fibrinogen (mg/L)</i>	0.15	0.51	0.80	0.92	0.56	1.17	4.06	2506
<i>AGE(Year)</i>	15.00	33.00	45.00	49.92	17.22	59.00	94.00	3159

Figure 2.1 Minor Allele Frequency (MAF) distribution of genotyped variants.

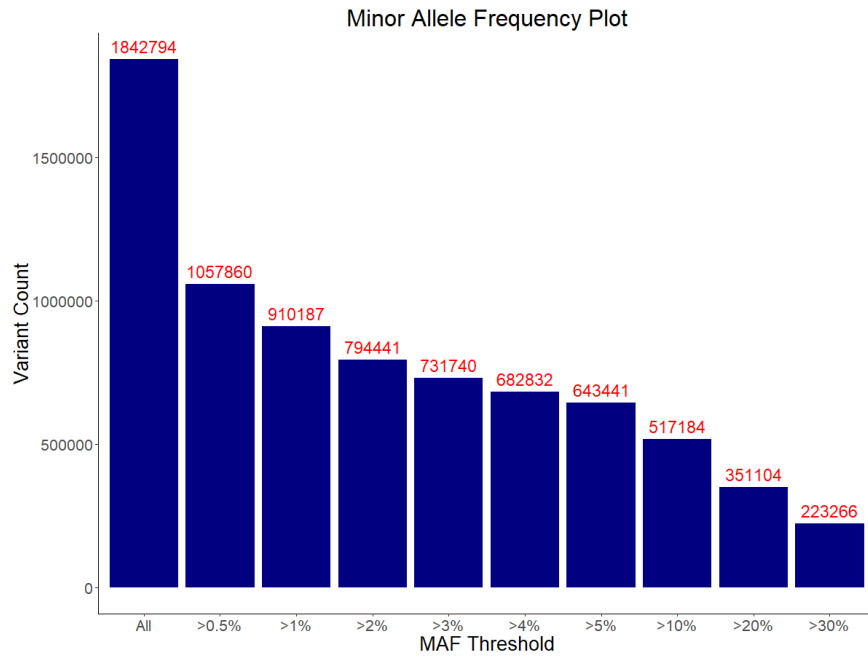


Figure 2.2 Proportion of missingness per variant.

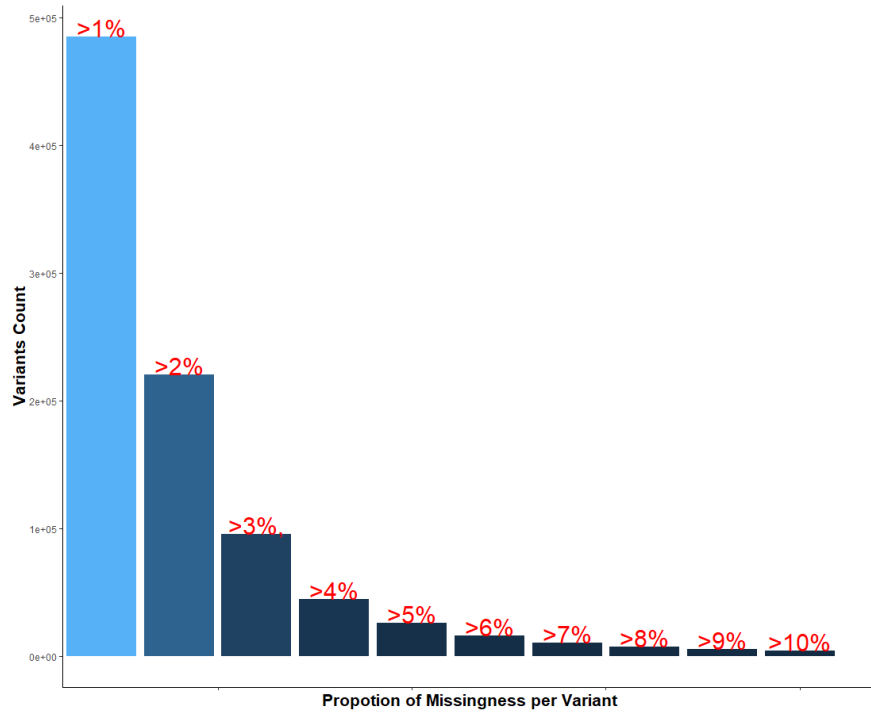


Figure 2.3 Per sample missingness vs autosomal heterozygosity. No sample was observed outside proposed threshold (median $\pm 3 \times$ interquartile range for autosomal heterozygosity) which would have been indicated with black colored data points.

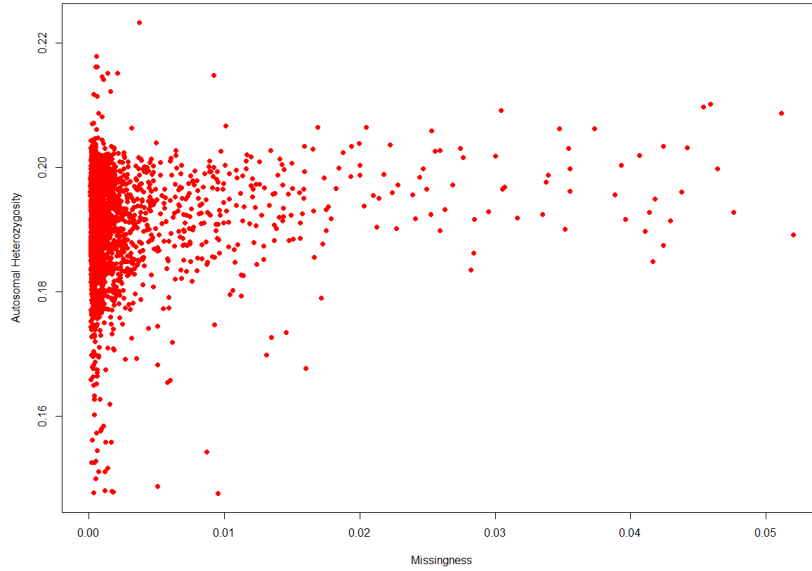


Figure 2.4 Distribution of F ratio coefficient. Each data point is an individual. Green color represents female pedigree sex, and shape indicates genetic sex where round shape suggests ambiguous genetic sex, trigon is male and square is genetic female sex. A number of mismatches between genetic vs pedigree sex observed.

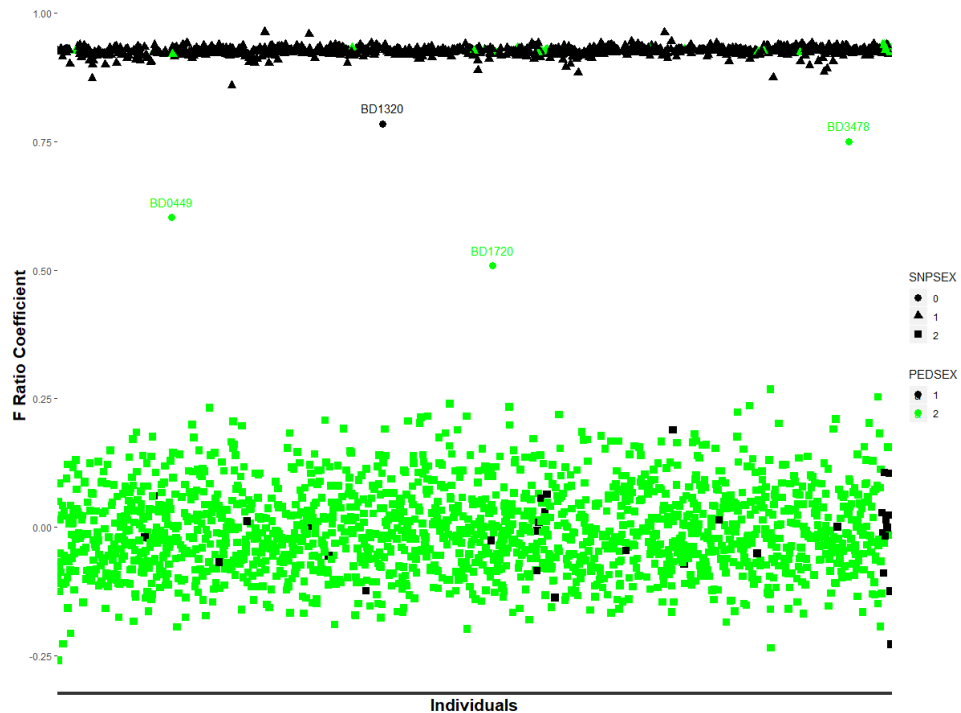


Figure 2.5 Sex vs autosomal heterozygosity. Each data point is an individual; color represents genetic sex where blue is pedigree recorded female and red is male. Results are consistent with F ratio coefficient distribution (Figure 2.4).

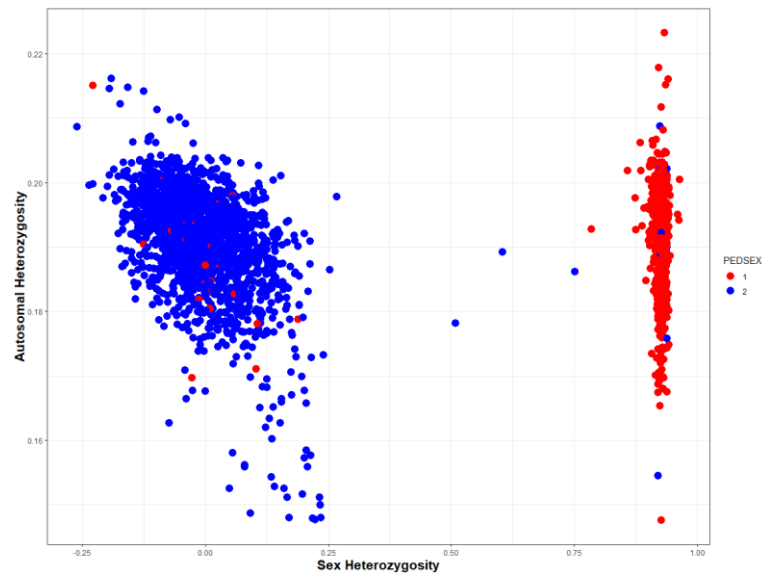


Figure 2.6 Eigensoft calculated ancestry outliers. Note that the pattern of ancestry outliers is not consistent when different PCs are compared.

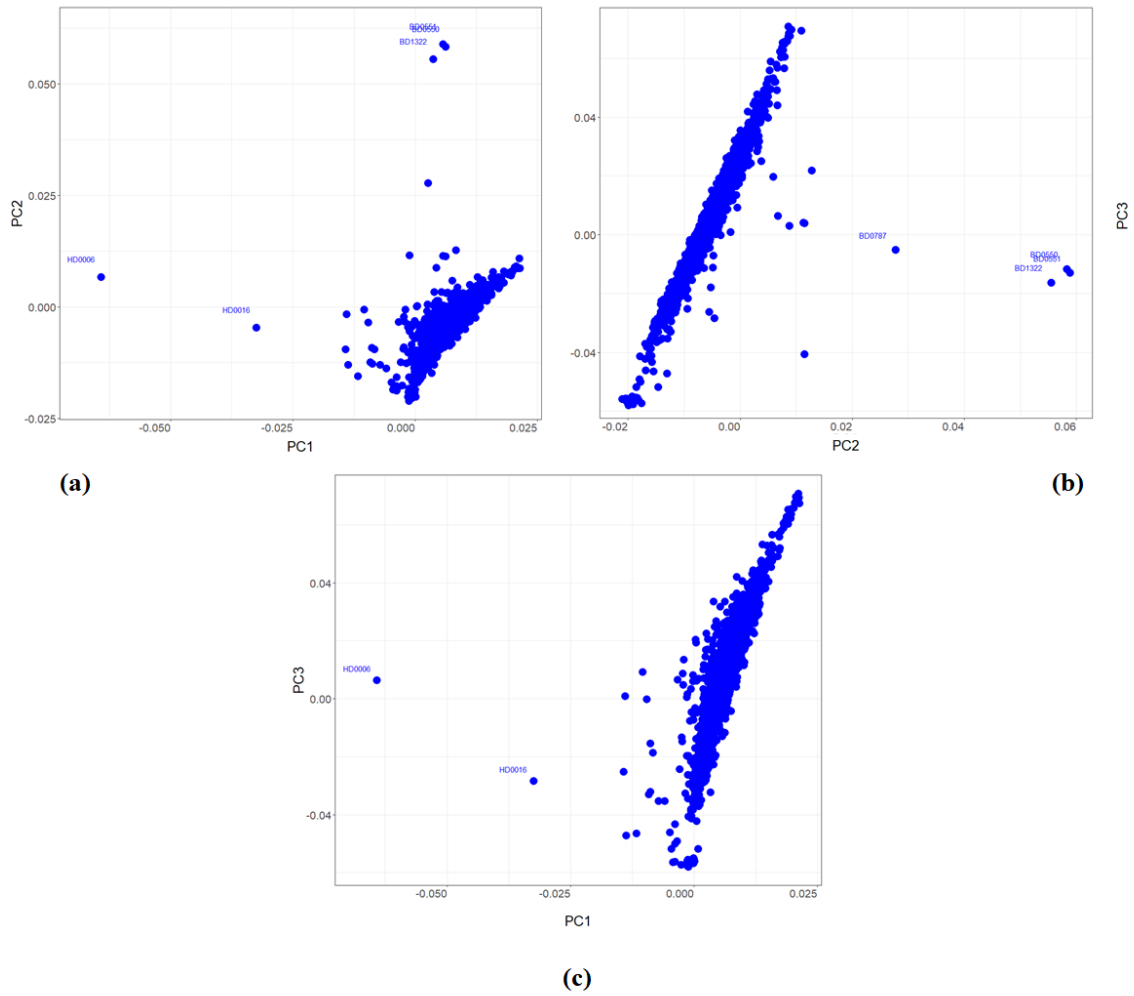


Figure 2.7 Hardy-Wienberg Equilibrium. A negligible number of variants were found outside of the significant thresholds for HWE ($< 10^{-6}$, $< 10^{-20}$, $< 10^{-50}$).

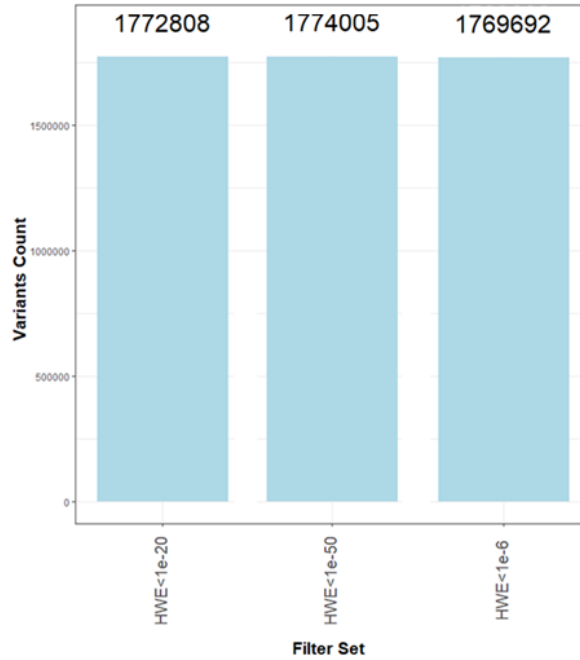


Figure 2.8 Kinship structure comparing identity by descent (IBD) parameters; IBD1 of 1 or close indicates child-parent kinship because they share 50% of their DNA markers. IBD0 is defined ~ 1 for pair of unrelated individuals if no recent shared ancestry. IBD2 of 1 suggest $\sim 100\%$ ancestry genetic concordance which can only be observed in monozygotic twins or duplicated samples. Approximately 76 pairs had $>80\%$ genetic concordance which suggest either duplicity or monozygotic twins (left down corner).

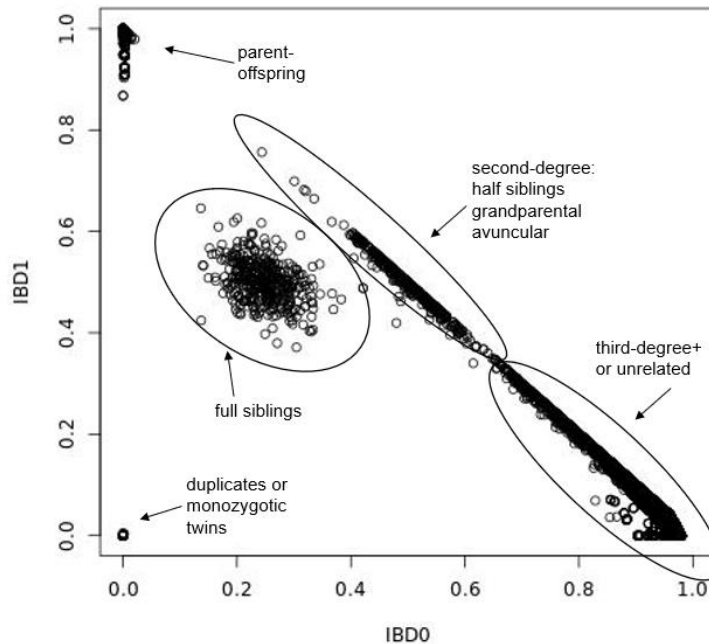


Figure 2.9 Projection of the first three PCAir for CCHC over 1000G individuals for ancestry inference. Nearly all individuals cluster with American ancestry population, which entail Mexican and Native ancestry groups. (Note: CCHC (study group), American (AMR), East Asian (EAS), African (AFR), European (EUR), Southeast Asian (SAS)).

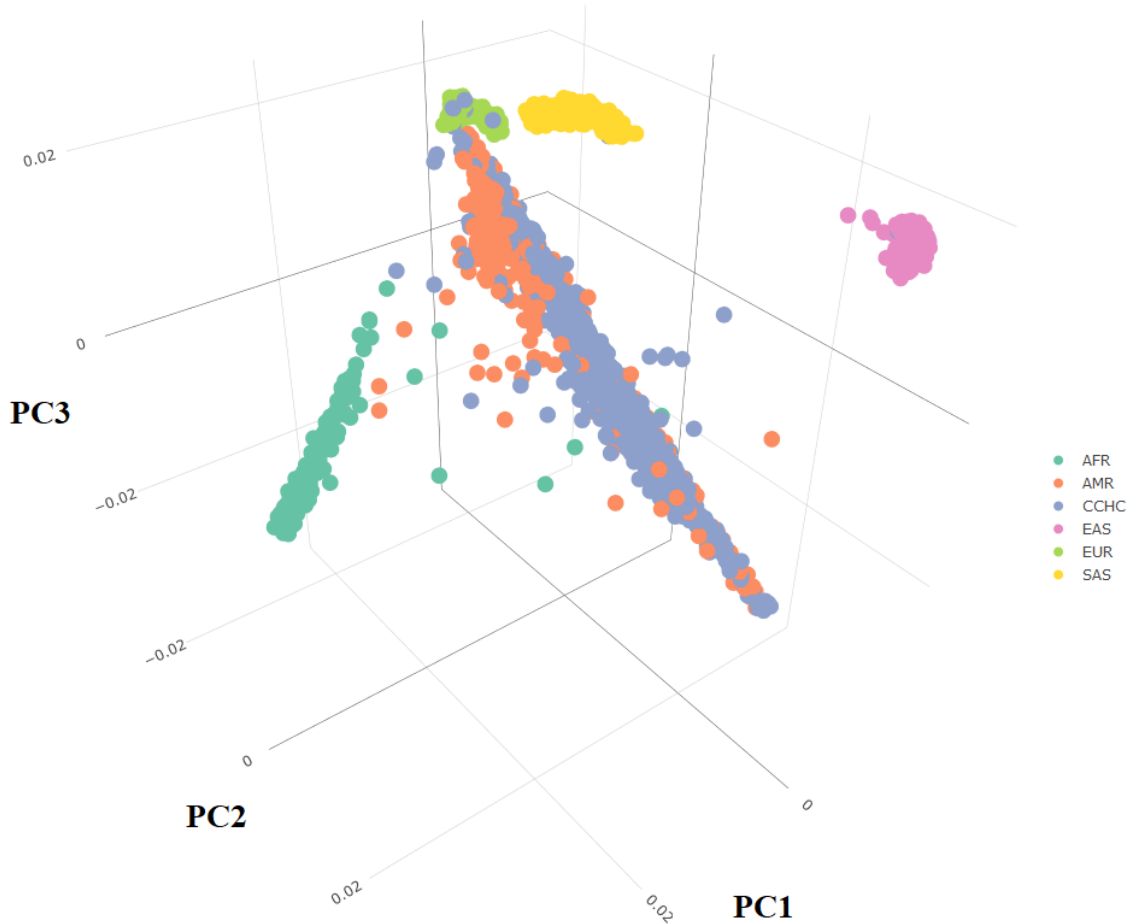


Figure 2.10 Parallel plot of the first 10 PCs. Nearly all individuals cluster with American ancestry population, which entail Mexican and Native ancestry groups. Y-axis values were scaled with “uniminimax’ method with 0 set for minimum observed measure and 1 for maximum. Colors represent ancestry.

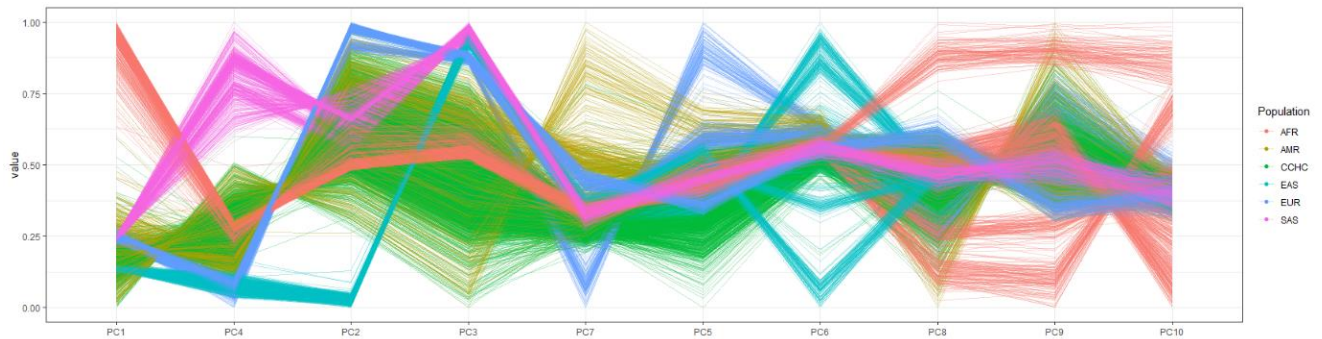


Figure 2.11 Screeplot for PCs. Note that PCs beyond the 3rd do not capture significant variation in the data.

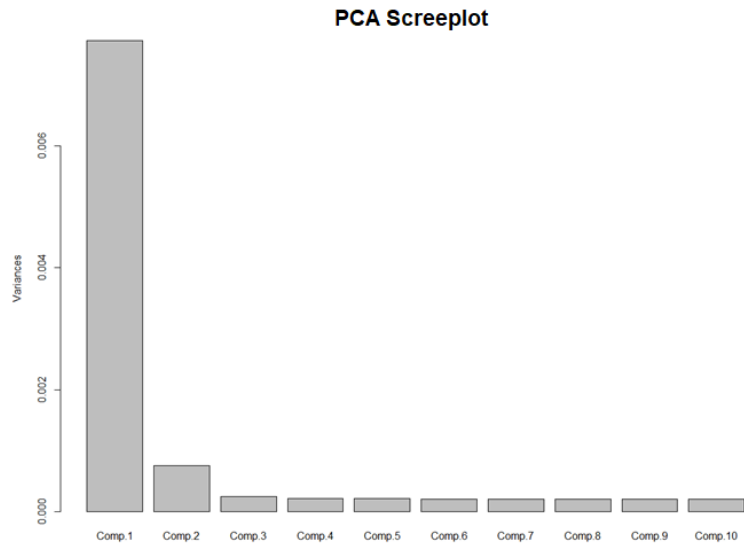


Figure 2.12 Scatter plot of PCAir1 vs PCA1. No practical difference could be observed.

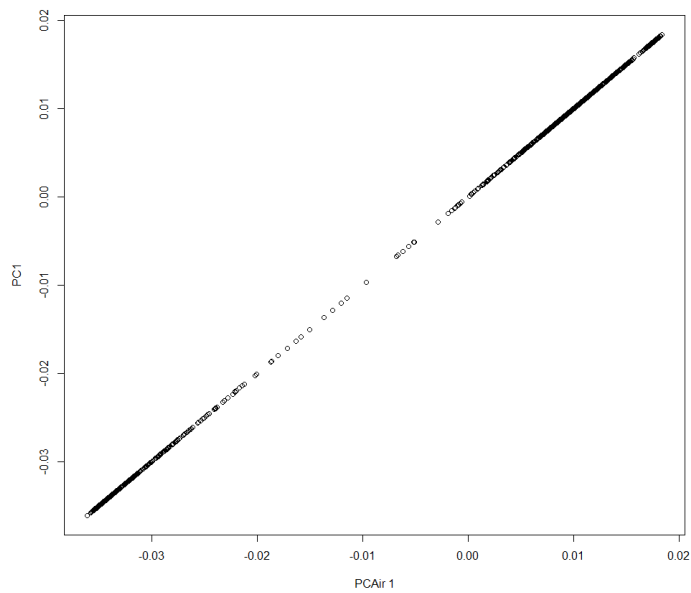


Figure 2.13 Distribution of phenotypic markers using observations from visit 1 & 2. Skewed distributions of inflammatory traits suggest presence of some extreme measures.

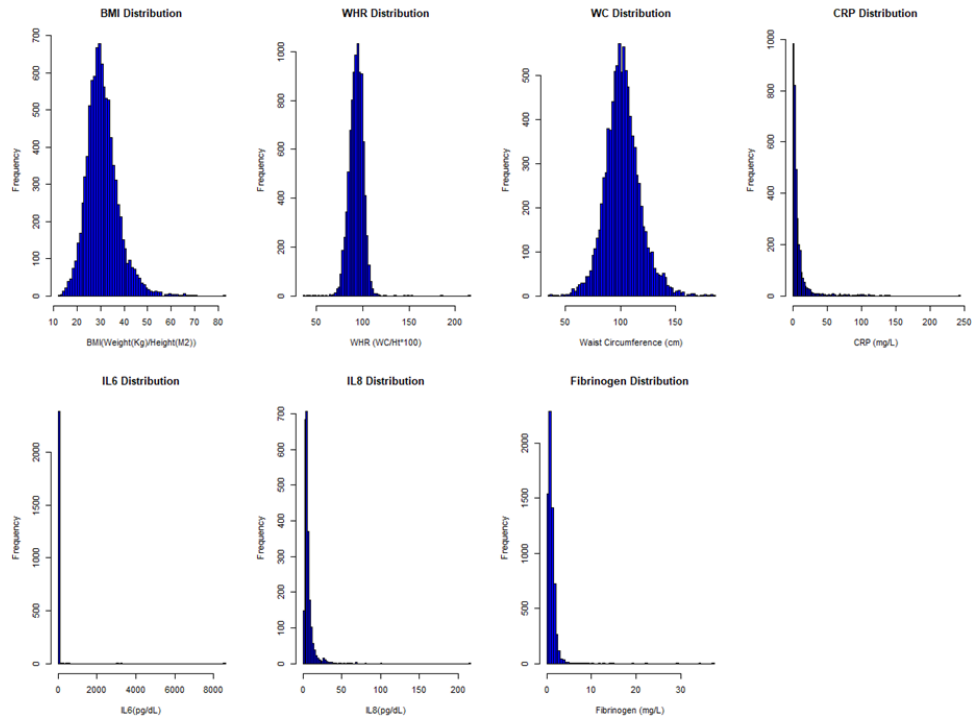
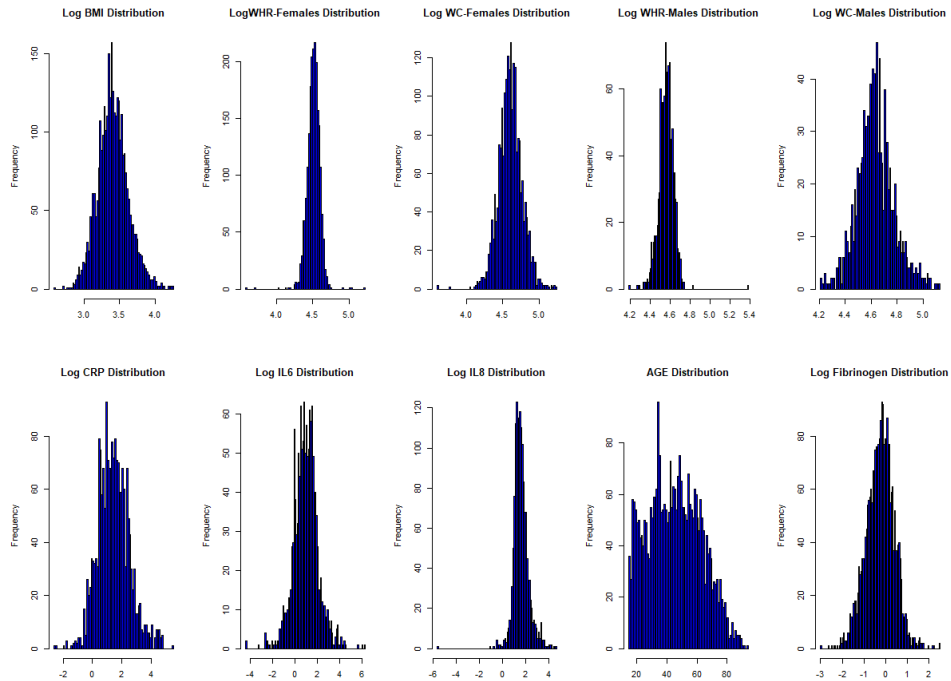


Figure 2.14 Log-transformed distributions of age, anthropometry and inflammation phenotypes.



GWAS AND PLEIOTROPY

A GWAS primary assessment tool used to identify thousands of genetic loci associated with one or multiple biologic and disease traits. In this chapter, mixed models GWAS were used to investigate association of SNPs with 9 anthropometric and inflammatory phenotypes. In the second part, the GWAS summary statistics were utilized to investigate pleiotropy across anthropometry and inflammatory domains. Cross-trait look-ups identified SNPs associated with metabolic, inflammatory and brain disease traits within or in close proximity to 3 genetic regions on chromosome 3, 12 and 18 with suggestions for pleiotropy.

1. Background

Since the early 2000s, GWAS have been extensively used to locate single-nucleotide variants associated with complex phenotypes, including body mass index (BMI)²²⁶, measures of central obesity²²⁷⁻²²⁹, and inflammation^{105,230}. A portion of genetic variants that have been primarily identified in association with anthropometric traits and inflammation have also been noted to display functional relationships with other domains^{159,231}. For instance, obesity-associated variants in the *LEPR*, *APOC1*²³², *GCKR*, *MC4R*¹⁵⁹, *CSN1S*¹¹⁷, *FDFT1* and *PCCB* genes¹⁰⁷ have also been associated with inflammation. Conversely, genetic loci associated with inflammatory markers²³³ also appear to be associated with variations in body weight and adiposity levels^{118,119}.

Besides genetics, other clinical²³⁴, pharmacological²³⁴ and observational²³⁵ studies have also demonstrated synergistic associations between adiposity and inflammation. These overlaps may indicate that anthropometric traits, which are often used as surrogates for adiposity, may share certain biological pathways with the inflammatory processes that underpin cardiovascular diseases. Termed pleiotropic variants¹²⁷, these multi-effect variants have been associated with multiple phenotypes¹⁰⁷.

Despite strong evidence for pleiotropy, existing studies have limited inferential utility because the majority of genetic research has been conducted in populations with European ancestry. This well-attested lack of diversity in genetic studies¹³⁹ not only hinders the discovery of more variants associated with underlying functional pathways, a phenomenon referred to as missing heritability²³⁶, but also poses considerable problems for the functional characterization of known variants due to differences in allelic frequency²³⁷, linkage disequilibrium (LD)²³⁸, and genetic architecture among populations with divergent genetic ancestries from Europeans^{239,240}.

These genetic dissimilarities extend to phenotypic traits including inflammatory markers and adiposity indices. Significant discrepancies have been identified in the distributions of inflammatory markers across various ancestry groups, even after adjusting for differences in adiposity configurations¹⁴². Additionally, BMI-adjusted distributions of adipose tissues suggest distinct ancestry patterns²⁴¹, as Hispanics, in comparison to European ancestry populations, tend to exhibit higher levels of visceral fat²⁴². These observations have suggested that various functional pathways are likely to be affected by ancestry-specific pleiotropic variants. Thus, a genome-wide, systematic, and comprehensive evaluation of pleiotropy in the Hispanic/Latino population is needed.

In this study, using genetic observations obtained from CCHC, GWAS tests were performed to assess the associations between 3 anthropometric traits including BMI, WHR, and WC measurements and 4 inflammatory markers including IL-6, IL-8, fibrinogen, and CRP.

During the second step, using summary statistics from the GWAS analysis, the aSPU method²⁴³ was employed to examine pleiotropy among the phenotypes of interest. Finally, bioinformatic annotation of top signals was conducted.

2. Methods

2.1. Study population

CCHC is a cohort of Mexican Americans residents of Cameron county in the state of Texas. Elevated prevalence of cardiovascular risk factors including type 2 diabetes, dyslipidemia and overweight in this group indicate a population highly susceptible to adverse metabolic events¹⁶⁶.

Out of 5,000 individuals recruited for the cohort, a sub sample of 3313 genotyped individuals are included in this study. Genetic measurements were completed at the Vanderbilt University Medical Center genotyping core facility, VANTAGE, using MEGA-EX Array panel. The median proportion of European ancestry in the cohort is 45.8%, African 11.0%, and Amerindian 42.9%.

2.2. Covariate selection

GWAS studies involving WHR, WC^{244,245}, and inflammatory traits²⁴⁶ often incorporate BMI as a covariate to identify genetic effects that are independent of overall obesity.

Univariate regression was used to examine the associations between the phenotypes included in the present study. Significantly associated variables identified by univariate analysis were incorporated as covariates in the GWAS analysis.

Linear regression models, both with and without adjustments for sex and age, were employed to determine how many ancestry principal components (PCs) were significantly associated with variations in phenotypic measurements and should, therefore, be used as covariates. Age and sex were universally employed as covariates with association tests.

Phenotypes were log-transformed and fit into linear regression models, adjusted for age, sex, BMI, and the first three PCs; BMI was not used during fibrinogen and IL-8 analyses because no association was observed between these variables during the univariate analyses. These residuals were subsequently used as the outcome variables in GWAS tests (supplementary Figure 3.1).

2.3. GWAS test

Covariate-adjusted GWAS typically focuses on associations between single-nucleotide variants and phenotypic measures. In contrast to methods that test a small number of candidate regions, GWAS studies investigate a larger number of genotyped/imputed variants that are spread across the entire genome. Therefore, GWAS is principally a hypothesis free approach²⁴⁷.

Residuals were used to perform GWAS testing (see chapter 2). A score-based, mixed linear regression model was adopted, and a pairwise kinship coefficient was incorporated as a random effect parameter. The kinship threshold was set to a third-degree relationship,

which equates to first-degree cousins. All phenotypes were treated as continuously measured variables.

The tests assumed additive model of association where the SNP effect is proportional to the presence of risk allele. For those samples harboring two risk alleles receive higher scores than those who were heterozygous for the locus, and zero score is assigned if both non-risk increasing alleles are identified²⁴⁸. Because imputed SNPs constituted the dominant share of the variants, the two most common methods of assigning SNP score were the best guess genotype (posteriori) which assigns a discrete count of risk alleles (0,1,2), and allele dosage which is the estimated (expected) fractional counts of effect allele at each SNP for each individual, ranging continuously from 0 to 2. The first method is sensitive to imputation quality while the second approach appears to be effective and efficient to account for the uncertainty in the imputed genotypes²⁴⁹, particularly when expected effect sizes are small. Indeed, regression analysis of a sample ~20K SNPs with BMI phenotype showed a median of 8% [IQR: 3%, 21%] difference in standardized effect sizes using the best guess vs dosage method. Therefore, SNP dosages were used for GWAS tests.

For WC and WHR, GWAS were conducted for both sexes, separately, to account for the inherent sexual dimorphism of these traits¹⁸¹. A total of 9 GWAS were completed including GWAS for BMI, $WHR_{BMI-adj}$ in females, $WHR_{BMI-adj}$ in males, $WC_{BMI-adj}$ in females, $WC_{BMI-adj}$ in males, $CRP_{BMI-adj}$, $IL-6_{BMI-adj}$, $IL-8$ and fibrinogen. Additionally, sex-specific GWAS for WHR and WC were meta-analyzed, using the inverse-variance weighted method, to create sets of sex-combined GWAS results for both traits.

Statistical significance was set at a Bonferroni-corrected level [5×10^{-8}] to adjust for multiple testing. Manhattan plots were used to visualize the GWAS results, in which the negative log of the p-value for each variant-phenotype association test was plotted against each genomic region. Established genetic loci for traits of interest were marked in the Manhattan plots. This step also served as a quality control step to determine whether the GWAS tests successfully replicated known SNPs and could, therefore, be methodologically validated.

GWAS tests were performed using R package GENESIS¹⁷⁴, while visualizations were completed with R package EasyStrata²⁵⁰. WHR and WC GWAS meta-analyses were performed with the METAL package²⁵¹. Computations were completed on the Longleaf server for the University of North Carolina, Chapel Hill.

2.4. Post-GWAS sensitivity analysis

To assess whether the observed signals were driven by less frequent or poorly imputed variants, which can pose a challenge for the replicability of results, a series of filtration criteria were utilized including:

- MAF < 5%
- Minor allele count < 30 (MAF \times N (sample size))
- Imputation quality < 80%
- Effective sample size (EffN) ($2 \times \text{MAF} \times (1 - \text{MAF}) \times \text{Imputation quality} (\%) \times N$) < 30

The threshold for each quality control (QC) criterion was inferred from empirical distributions, as well as a review of the literature^{252,253}. After obtaining different subsets of GWAS results with each filter (or combination of filters) for each trait, QQ-plots and genomic inflation factors (λ)²⁵⁰ were used to assess the type 1 (false-positive) error. An inflation factor (λ) value of >1.05 would be indicative of mild inflation, whereas a value >1.1 would indicate serious inflation based on the observed p-values, which would necessitate the adoption of stricter genomic-control measures.

2.5. Pleiotropy test

To perform an exploratory analysis of genetic correlations between phenotypes, a pairwise genetic distance matrix was constructed based on genetic similarities. The aSPU¹⁹⁹ test was utilized to perform the multivariate assessment of genetic pleiotropy. A major strength of the aSPU test over other methods is that the loss in power is negligible when the number or portion of non-associated variants increases²⁰³, which is a distinguishing feature in the context of this study, in which univariate GWAS results were inferred from smaller sample sizes.

Briefly, the method aggregates information across n phenotypes for a given SNP by taking the sum of its univariate GWAS Z-scores each raised to some power γ , so that a higher γ increases the influence of strongly associated phenotypes on the score. By allowing γ to take one of many competing values $(1, 2, \dots, \infty)$, aSPU selects a maximally efficient scheme to detect combined phenotype effects on the entire group of phenotypes.

A set containing trait-specific GWAS-inferred Z-scores (from all contributing phenotypes) was constructed. Pleiotropy analysis was performed using the JaSPU

(github.com/kaskarn/JaSPU) package, with the number of iterations set to 10^{11} . This program relies on a Markov Chain Monte-Carlo (MCMC) ²⁵⁴ iterative process, which is applied to univariate Z-scores, estimated by inverse-variance-weighted meta-analysis to generate pleiotropic p-values.

Finally, the covariance matrices for pleiotropy results were examined to ensure that no dimension (i.e., phenotype) exhibited considerably small or large eigenvalues; excessively-large eigenvalues are indicative of the disproportionate contribution of a phenotype to a pleiotropic signal, whereas very small values indicate the opposite, and either result would suggest that the underlying hypothesis for pleiotropy would not hold, at least for some phenotypes.

Exploratory genetic correlation analyses [prior to pleiotropy] were completed using GCTA package ²⁵⁵.

2.6. In-depth probe of suggestive pleiotropic signals

All variants with a pleiotropic p-value of $p_{\text{aspu}} < 5 \times 10^{-5}$ were extracted from the output. From this output, a subset of variants with a more stringent significance level ($p_{\text{aspu}} < 5 \times 10^{-6}$), which were robust to post-GWAS quality assessments and harbored clusters of variants with low SNP-specific ($\gamma < 4$) scores were selected. The final criterion provides a mechanism for grouping loci from the most to least likely to be pleiotropic.

Among these select variants, the univariate negative log p-values from the GWAS results were extracted and incorporated into heatmap plots to elucidate which phenotype(s) were the likely drivers or contributors to the pleiotropic signals. Heatmap plots were constructed using *gplots* R package ²⁵⁶.

2.7. Functional evaluation of pleiotropic signals

For selected loci with evidence of pleiotropy, the GWAS catalog was probed to identify whether any variant within the selected 1 Mbp distance on both sides of each genomic region and in LD with variants within the region ($D' > 0.5$) had previously been reported in association with the phenotypes under study or with any other traits/diseases and how commonly these variants were identified. The findings were tallied by phenotypic domains from all suggestive loci. GWAS catalogue (ebi.ac.uk) and NCBI (ldlink.nci.nih.gov/?tab=ldtrait) were used for this assessment.

3. Results

3.1. Descriptive statistics

Descriptive distributions of the study population are shown in Table 3.1. The ages of the participants ranged from 15–94 years, with a median of 45 years. The BMI (weight (kg)/height (m²)) distribution indicates that this cohort is three-quarters overweight (BMI \geq 25), one-half obese (BMI \geq 30), and at least one-quarter moderately-to-severely obese (BMI \geq 35).

The median values for WHR (WC/hip (cm) \times 100) and WC (cm) in women were 91.3 and 99.5, respectively, whereas, in men, these values were 96.0 and 102.5. The difference between the mean and median values of inflammatory markers indicated skewed distributions, with the exception of fibrinogen, with a mean value of 0.92 g/L, which is lower than the average ranges (1.5-4.5) reported in other studies for mixed populations

224,225

3.2. GWAS results

Phenotypic residuals extracted from linear regression analyses were used as dependent variables in mixed model GWAS tests. A total of 9,262,347 variants, covering chromosomes 1-22, passed imputation quality control and were, therefore, included in the GWAS tests. Manhattan plots of the GWAS results for BMI, $WC_{\text{BMI-adj}}$ in women, $WC_{\text{BMI-adj}}$ in men, sex combined $WC_{\text{BMI-adj}}$, $WHR_{\text{BMI-adj}}$ in women, $WHR_{\text{BMI-adj}}$ in men, $CRP_{\text{BMI-adj}}$, $IL6_{\text{BMI-adj}}$, $IL8$, and fibrinogen are shown in supplementary Figures (3.2a–j) respectively, where dots represent variants, x-axes demonstrate genomic region [by chromosome number and base pair position] and y-axes demonstrate negative log of GWAS p-value of each SNP.

In the sex-combined meta-analyses for the $WHR_{\text{BMI-adj}}$ GWAS (Figure 3.1), two GWAS-significant signals ($P < 5.0 \times 10^{-8}$) were observed for $WHR_{\text{BMI-adj}}$, on chromosomes 6 and 8. The signal on chromosome 8 corresponds to a locus previously associated with WHR, whereas the signal on chromosome 6 has not been previously reported in the literature (Table 3.2, Figure 3.1). No GWAS-significant variants were observed in either known loci (denoted by the blue color) or other regions for any other trait.

3.3. Sensitivity analyses

Variants beyond the significance threshold set for the study (i.e., $p\text{-value} < 5 \times 10^{-6}$) for each trait (supplementary Table 3.1) were examined to assess whether the top signals were driven by less frequent or poorly imputed SNPs, which may affect the replicability (i.e., generalizability of observed signals). Supplementary Table 3.1 shows the counts of significant variants remaining after the application of each (or the combination of) QC

criteria for MAF, MAC, imputation score, and effective sample size. Few signals were likely driven by less common variants (i.e., $MAF < 5\%$), no poorly imputed SNPs were identified in the results. Application of MAC & MAF resulted in near maximal removal of less common variants compared to baseline GWAS results (i.e., non-filtered GWAS) and therefore adopted as the post-GWAS filtration criteria.

Comparisons of inflation factors for each subset illustrated that no substantial inflation affected the association tests, as all λ values ranged from 0.99–1.03, well below the 1.05 threshold level (supplementary Table 3.2). Therefore, no genomic control application was deemed necessary. However, few Q-Q plots suggested deflation, particularly at the left tail, where fewer highly significant variants were observed than expected, particularly for those GWAS subsets that excluded less common variants (i.e., $MAC > 30$ and $MAF > 5\%$ filtered GWAS subsets) (Figure 3.3a–d).

3.4. Pleiotropy results

The preliminary examination of genetic correlations between phenotypes, inferred from GWAS summary results, illustrated that both inflammation and anthropometry-associated variants clustered with their respective phenotypic domains (supplementary Figure 3.4). However, fibrinogen was a notable exception, which correlated more closely with $WC_{BMI-adj}$ in men than with inflammatory markers.

A Manhattan plot of the pleiotropy results included no significant GWAS variants, either in regions known for associations with any of the study traits (denoted by a blue color) or novel loci (Figure 3.2). However, suggestive signals were identified in several chromosomes, including 1, 3, 12, 18, and 19, which merited further evaluation. The

absence of very small or large eigenvalues in the covariation matrix made it unlikely that the observed pleiotropic outcomes were driven by any solitary trait (supplementary Table 3.3).

The in-depth interrogation of the locus on chromosome 3 revealed that IL-6, BMI, and WC in men were the primary drivers of the pleiotropic signal (Figure 3.3a). For the locus on chromosome 12, GWAS results from CRP, BMI, fibrinogen, WHR, and WC in women were contributors to the signal (Figure 3.3b). However, for the locus on chromosome 18, the pleiotropic association appeared to be disproportionately driven by WHR in women, with fibrinogen playing a minor part (Figure 3.3c). Of the three loci, the region on chromosome 12 appeared to be the most promising for pleiotropic potentiality. The signals on chromosome 1 and 19 only contained solitary variants for each and did not include clusters of associated SNPs.

Finally, the online library of GWAS catalog was probed to explore whether any of the variants in the selected regions had previously been reported for any phenotype or are in LD ($D^{\prime} > 0.5$) with variants identified for those phenotypes. In two of three regions, on chromosomes 12 and 18, several variants were reported in association with energy intake or WHR, although they were not identified often (Table 3.3).

In the target region on chromosome 3 (bp chr3:4120729-4262732, ± 1 Mbp on either directions), a significant number of variants were observed in association with brain function and diseases, lipid levels, inflammation, and metabolite levels; few were identified for body weight and type 2 diabetes. For the locus on chromosome 12 (bp chr12:44476332-44621041, ± 1 Mbp on either directions), SNPs linked to variations in anthropometric traits (WHR), inflammatory diseases, and gut microbiome were more frequently reported. For

the locus on chromosome 18 (bp chr18:61577041-61588083, ± 1 Mbp on either directions), variants associated with blood pressure, energy intake, allergies, autoimmune conditions, and cancer were identified (Table 3.3).

4. Discussion

The study results provided suggestive evidence for at least 3 pleiotropic signals associated with both adiposity and inflammation which indicated a considerable degree of genetic overlap.

Two regions on chromosomes 6 and 8 exceeded the GWAS significance level for sex-combined meta-analyzed WHR results (Table 3.2, Figure 3.1). Although the signal on chromosome 8 has previously been reported, the signal on chromosome 6 has not been previously recorded in the GWAS libraries. GWAS-significant SNPs in the novel locus occur in the region between *LRFN2* and *LOC101929555*. Variants in *LRFN2* have been reported in association with BMI ^{257,258}, type 2 diabetes ²⁵⁹, lipid profiles ²⁶⁰, cognitive abilities ²⁶¹, and WHR ²⁶². *LOC101929555* has not yet been functionally characterized.

A number of loci that have been widely reported and mapped for associations with anthropometric and inflammatory traits, including the SNPs in *IL6* on chromosome 7 (for IL-6) ²⁶³ or *FTO* on chromosome 16 (for BMI) ²⁶⁴, did not exceed the GWAS threshold level. Although larger studies have ascertained the transferability of a sizeable proportion of phenotype-associated common variants across various ancestries ²⁶⁵, including anthropometric ^{266,267} and inflammation-associated variants ²⁶⁷, important ethnic differences were also noted ²⁶⁸⁻²⁷⁰, even in known genomic regions ²⁷¹.

Although these observed patterns could suggest an ancestry-specific association pattern, caution must be exercised because differences in the LD structure²³⁹, study sample sizes²³⁸, and dissimilarities in variant coverage (SNP tagging²⁷²) can all induce differences in the GWAS signals across various study settings or ethnicities²⁷¹.

Although no variants exceeded the GWAS significance level in the pleiotropy analysis, suggestive signals were identified on chromosomes 3, 12, and 18 (Figure 3.2 & Figure 3.3a-c). Multiple phenotypes appeared to contribute to the loci on chromosomes 3 and 12, whereas the signal on chromosome 18 was primarily driven by WHR in women.

Additional variants were also identified: one on chromosome 1 (rs952499) overlapping gene *ABCA4*, and another on chromosome 19 (rs669560) overlapping *ZNF708*, with strong pleiotropy p-values ($< 5 \times 10^{-7}$). However, in both instances, the signals were driven by solitary SNPs that were not in tight [or loose] LD with any other variants that exhibited significant pleiotropy p-values.

A review of the GWAS library to identify suggestive pleiotropic loci showed that an extensive number of variants were already reported for associations with a range of metabolic and inflammatory functions. For the probed region on chromosome 3, a sizeable number of SNPs were shown to be associated with brain function and diseases, liver function, triglyceride, inflammation, type 2 diabetes and body weight (Table 3.3).

Similarly, for the suggestive region on chromosome 12, many variants were associated with anthropometric indices, and inflammatory diseases. The third probed locus on chromosome 18 also contained variants associated with inflammation, WHR and blood pressure indices.

There few methodological considerations. The Q-Q plots for the IL-6 and IL-8 GWAS results suggested deflation at the left tails (supplementary Figure 3.3a-d). Factors including the overcorrection of GWAS test statistics in polygenic phenotypes (such as cytokines)²⁷³ and adjustments for pairwise kinship in mixed-model GWAS tests²⁷⁴ could potentially induce deflation. However, the severity of deflation increased when less common variants (MAF < 5% and MAC < 30) were excluded, which indicated that deflations were likely induced by reduced power (supplementary Figure 3.3c and 3.3d).

The age distribution of the participants suggested a middle-aged cohort that was slightly older than the general Hispanic population (Table 3.1). The prevalence of obesity was higher than the national average level for the Hispanic population, as reported by the Centers for Disease Control and Prevention (CDC) for the 2017–2018 period²⁷⁵ (45% vs. >75% in this study). However, the distributions of WHR and WC suggested that central obesity, likely caused by an accumulation of visceral fat, represents the prevailing form of adiposity, as over three-quarters of both men and women had measures above their respective threshold levels [WHR > 85 in women and WHR > 90 in men²⁷⁶] for obesity. These results, together with the presence of higher levels of inflammatory markers, presented an opportunity for pleiotropy assessment due to the increased statistical power despite a modest study sample size.

Fibrinogen was the only inflammation marker with lower distribution (mean = 0.92 g/L) compared with the average range reported by other studies (1.5–2.5 g/L)^{277,278}. However, this marker is known to exhibit considerable within-sample variability, even in healthy individuals²⁷⁹. Additionally, fibrinogen measures were significantly correlated with WHR, WC, and interleukin levels, which suggested that the observed values may be a function of

the caliber of the assay rather than the reduced distribution of this marker in this study group.

In conclusion, the overlap in SNP associations with similar phenotypic domains for all three regions underscores the complexity and interrelatedness of traits that are otherwise viewed as independent and highlights the need for a greater focus on pleiotropic SNPs that can provide novel insights into the underlying mechanisms that drive these traits. The high prevalence of adiposity and inflammation in this cohort highlighted the need for pleiotropy study in this cohort with high susceptibility to adverse cardiovascular events

5. Tables and Figures

Table 3.1. Distributions of inflammatory and anthropometric phenotypes and age among Cameron County Hispanic Cohort (CCHC) subjects.

Phenotype (Unit)	Min.	1st Qu.	Median	Mean	SD	3rd Qu.	Max.	N
<i>BMI (kg/m²)</i>	16.16	26.30	29.86	30.76	6.53	34.25	57.21	3159
<i>Waist-female (cm)</i>	62.00	90.00	99.50	100.77	15.52	109.50	158.00	2072
<i>Waist-male (cm)</i>	68.30	94.00	102.50	103.95	14.62	112.00	157.00	1089
<i>WHR-Female*</i>	69.89	86.54	91.30	91.17	7.08	96.00	116.39	2077
<i>WHR-Male*</i>	77.87	91.70	96.01	95.94	6.40	100.22	113.58	1084
<i>CRP (mg/L)</i>	0.35	2.00	3.90	6.23	6.93	7.73	50.03	1880
<i>IL-6 (pg/dL)</i>	0.16	1.44	2.68	4.32	5.74	4.87	47.38	1336
<i>IL-8 (pg/dL)</i>	1.10	3.42	4.67	5.67	3.58	6.65	25.56	1448
<i>Fibrinogen (g/L)</i>	0.15	0.51	0.80	0.92	0.56	1.17	4.06	2506
<i>Age(year)</i>	15.00	33.00	45.00	49.92	17.22	59.00	94.00	3161

*reporting unit: Waist/Hip(cm)×100

Table 3.2. GWAS significant variants in sex-combined meta-analyzed WHR_{BMI-adj} GWAS analysis.

Chr.*	Bp Position**	Rsid	Nearest Genes	Z-score	P-value
6	40810472	rs12524956	LRFN2,LOC101929555	-5.61543	1.96E-08
6	40810278	rs12524922	LRFN2,LOC101929555	-5.49004	4.02E-08
8	72499840	rs202053146	EYA1, MSC	5.54286	2.98E-08
8	72524159	rs1424869	EYA1, MSC	5.52103	3.37E-08
8	72488774	rs972738	EYA1, MSC	5.53007	3.20E-08
8	72484519	rs59354633	EYA1, MSC	5.52962	3.21E-08
8	72492249	rs35727416	EYA1, MSC	5.46058	4.75E-08
8	72473729	rs10504510	EYA1, MSC	5.4889	4.04E-08

*Chr=Chromosome, **Bp= base pair

Table 3.3. Functional probe of three loci on chromosome 3, 12 and 18 with suggestion for adiposity-inflammation pleiotropy.

chr3:4120729-4262732	chr12:44476332-44621041	chr18:61577041-61588083
Within \pm 1 Mbp of the region and in LD ($D' > 0.5$) with variants	Within \pm 1 Mbp of the region and in LD ($D' > 0.5$) with variants	Within \pm 1 Mbp of the region and in LD ($D' > 0.5$) with variants
Brain function or diseases	WHR	Allergy, inflammatory, autoimmune diseases
Blood proteins/metabolites	Height	Anthropometric traits (WHR)
Inflammatory diseases/markers	Idiopathic scoliosis	Energy intake
Weight	Immune response	cancer
Type 2 diabetes	Gut microbiome	Blood pressure
Metabolite levels		
Lipid levels (triglyceride)		
Liver function (AST)		
Eye function/diseases		

Figure 3.1. Sex-combined meta-analyzed GWAS for $WHR_{BMI-adj}$. Blue color denotes known loci.

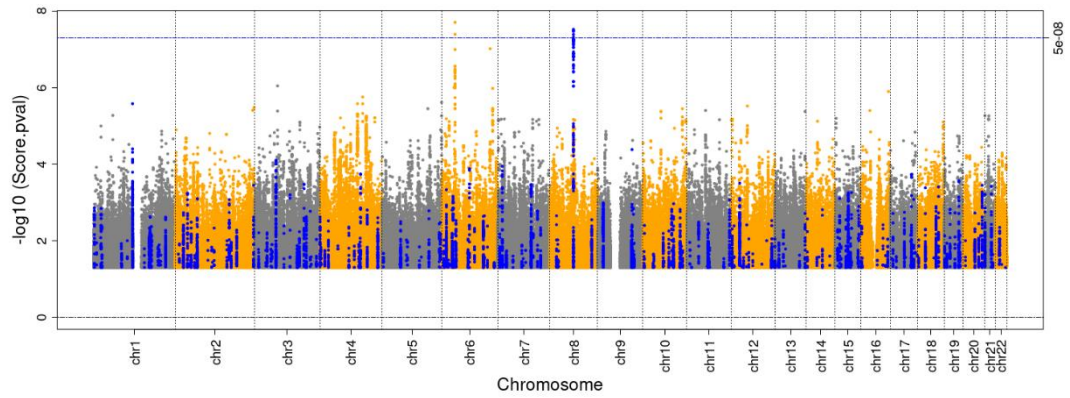
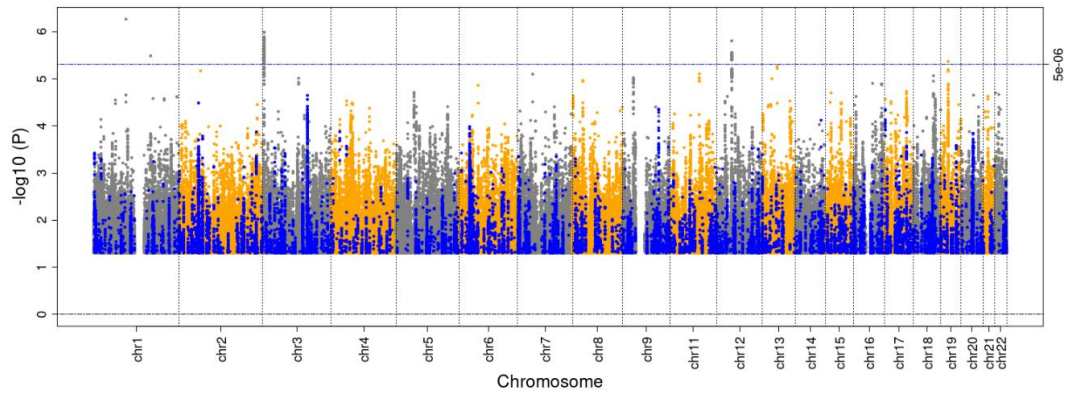
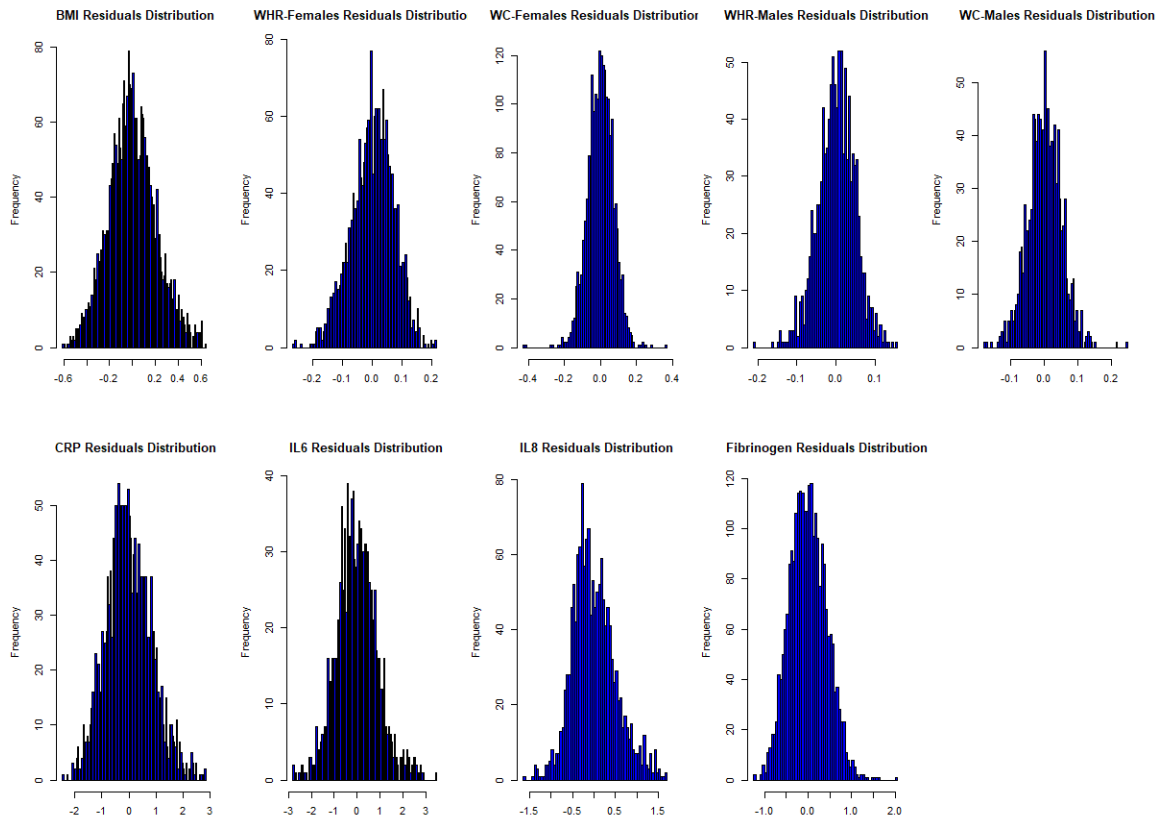


Figure 3.2. Pleiotropy Manhattan plot. Dots represent genomic variants; x-axis indicates chromosome number and base pair position for single nucleotide polymorphisms (SNPs) and y-axis demonstrate pleiotropy p-value for each variant. Blue color denotes known loci associated with phenotypes contributing to pleiotropy.



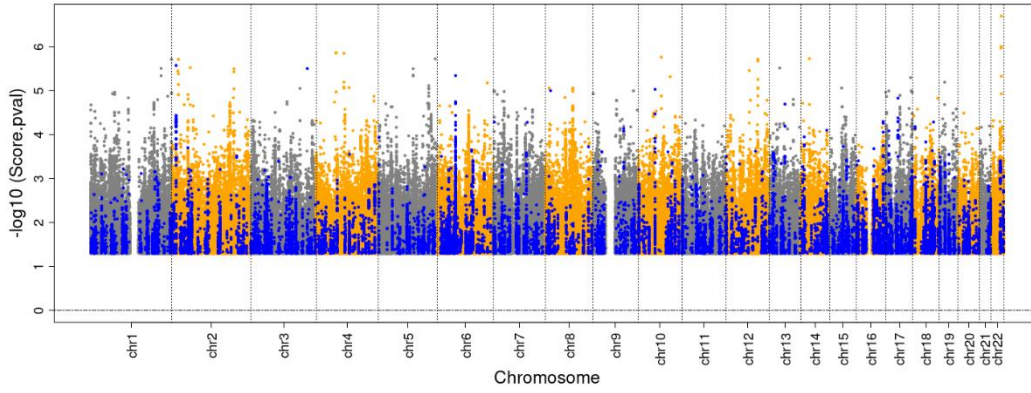
6. Appendices

Supplementary Figure 3.1. Distributions of phenotypic residuals used as dependent variables for GWAS tests; residuals were extracted from linear regression of log-transformed phenotypes, regressed over age, sex, first 3 PCs and BMI (except for IL8 and fibrinogen).

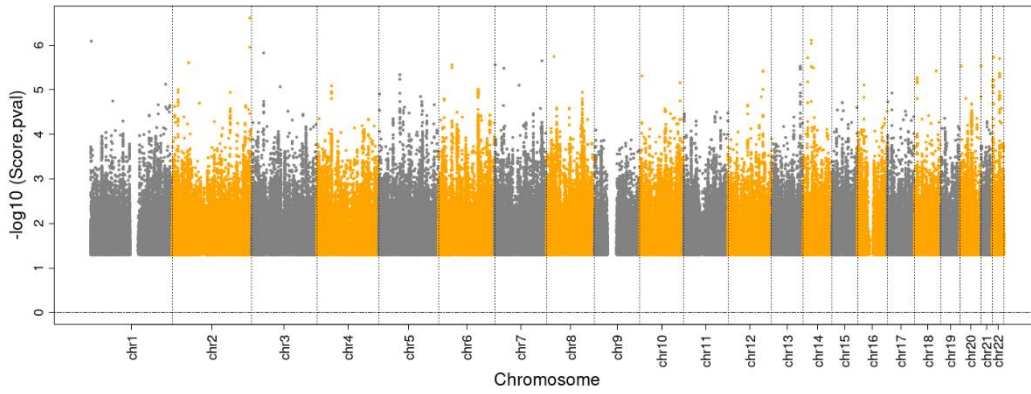


Supplementary Figures 3.2a-i:

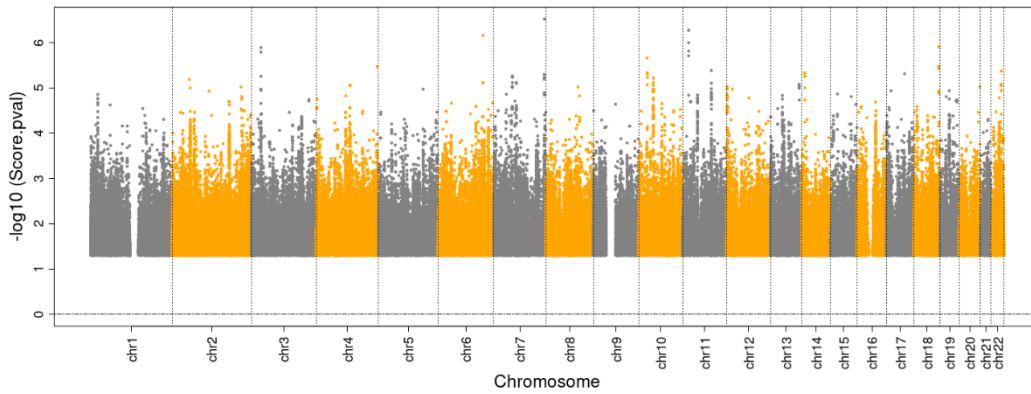
a) BMI Manhatten plot



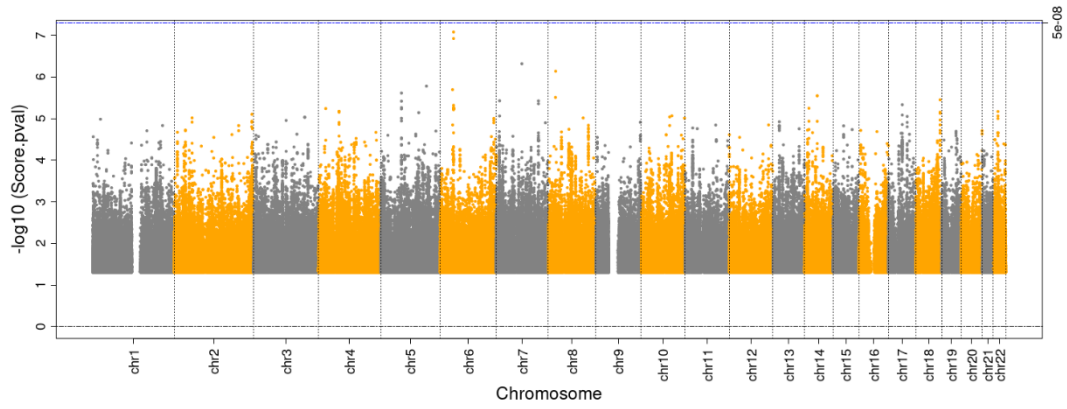
b) WC-Female Manhatten plot (Adjusted for BMI)



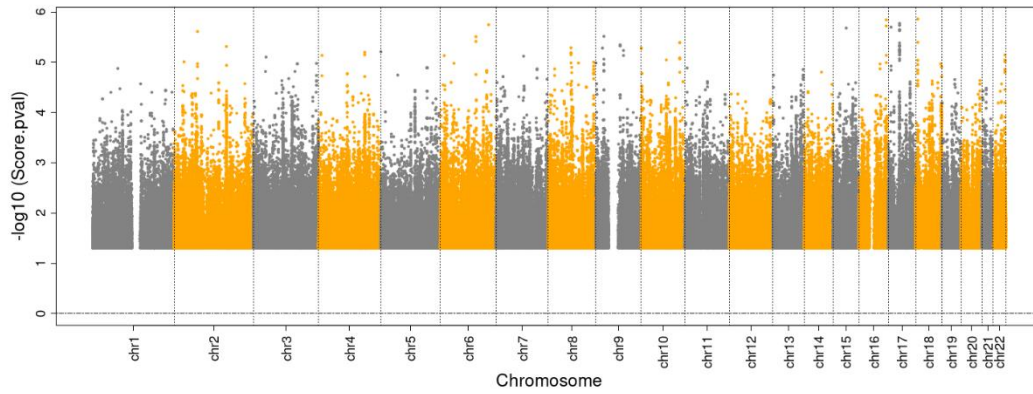
c) WC-Male Manhatten plot (Adjusted for BMI)



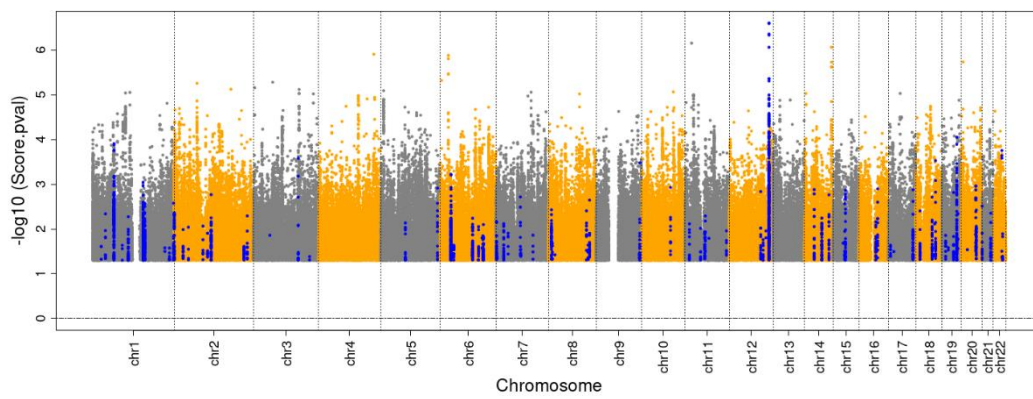
d) WHR-Female Manhatten plot (Adjusted for BMI)



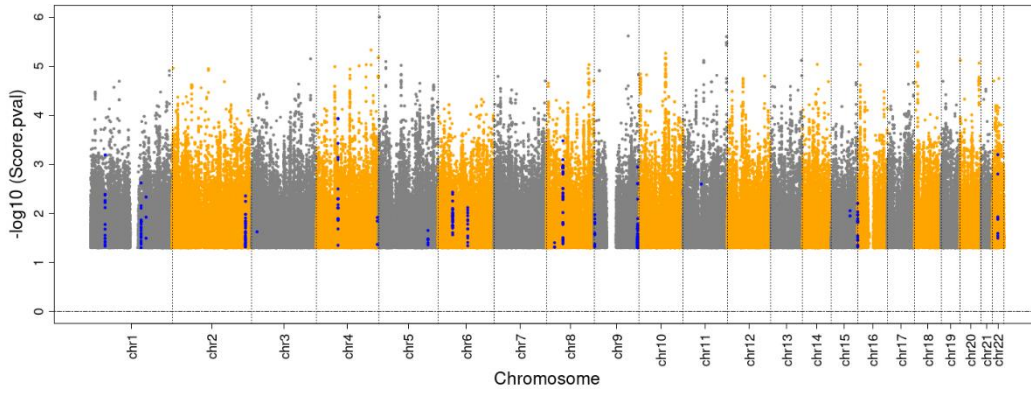
e) WHR-Male Manhatten plot (Adjusted for BMI)



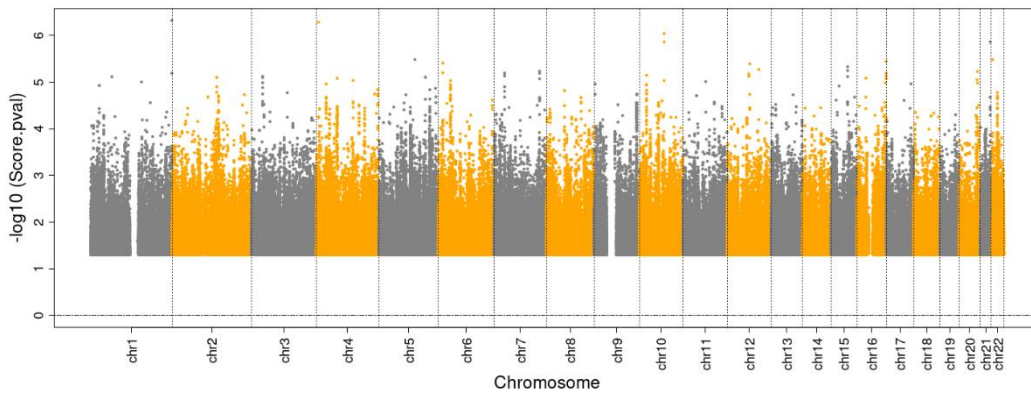
f) C-Reactive Protein Manhatten plot (Adjusted for BMI)



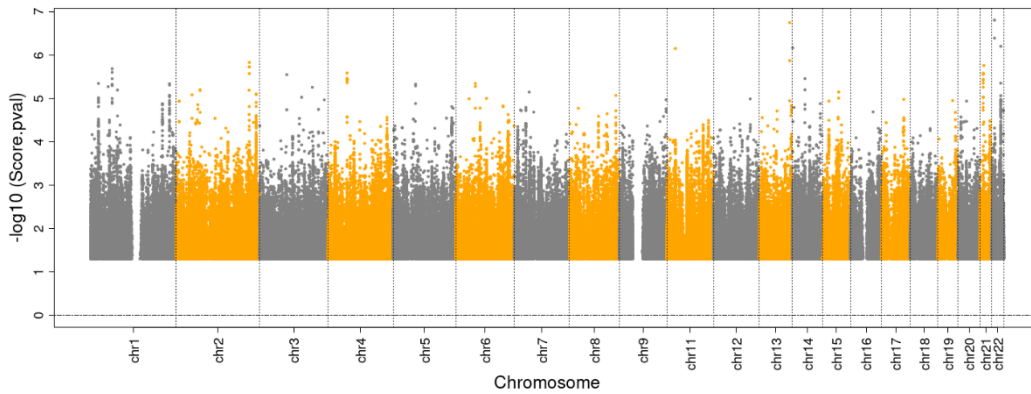
g) Interleukin 6 Manhatten plot (Adjusted for BMI)



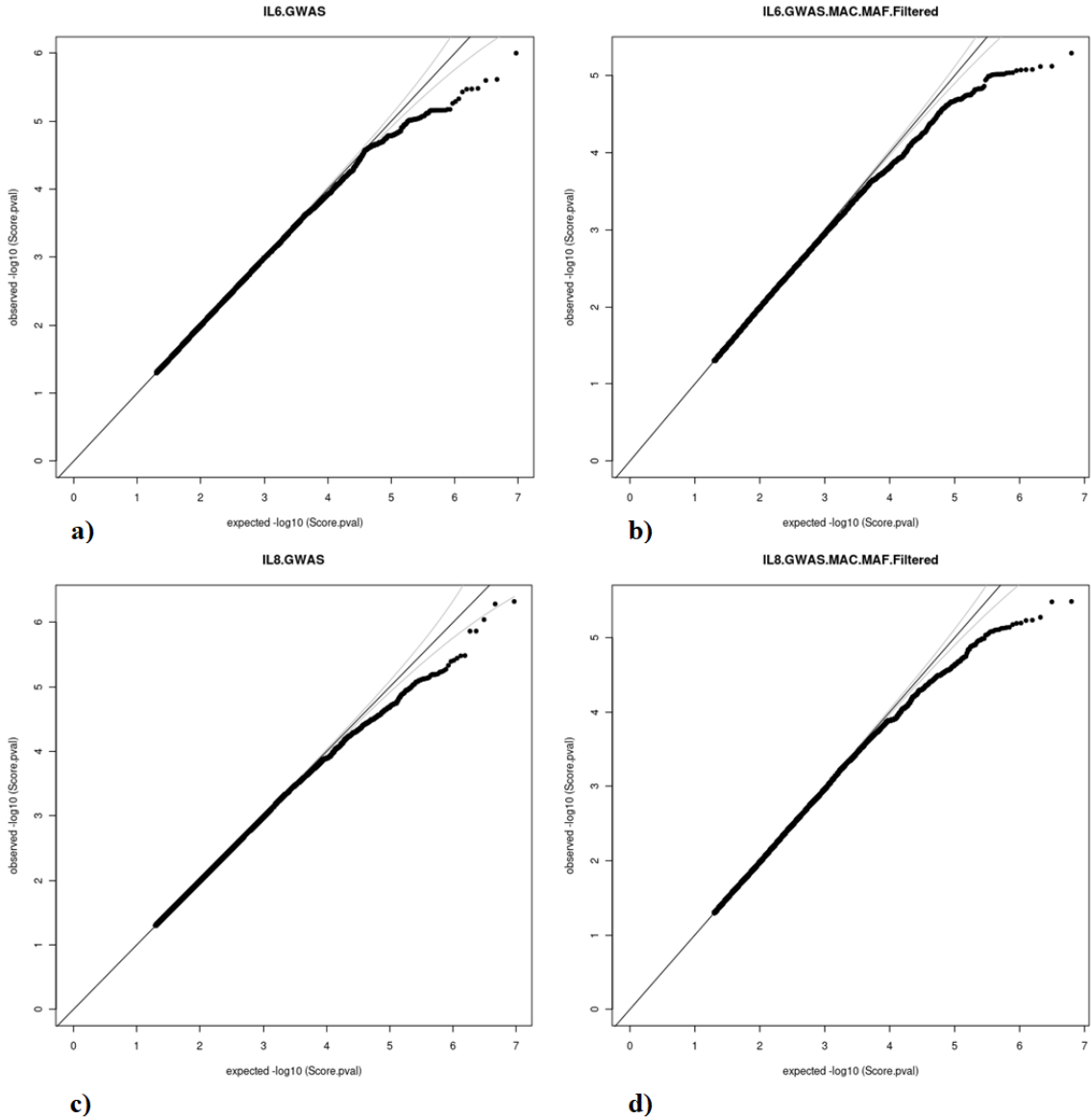
h) Interleukin 8 Manhatten plot



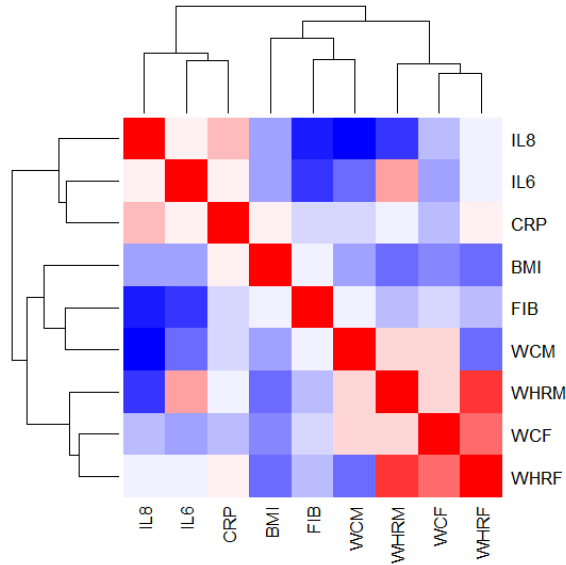
i) Fibrinogen Manhatten plot



Supplementary Figures 3.3a-d. Q-Q plots for IL6 and IL8 GWAS. Note the deflation with the left tails which suggest fewer highly significant variants were observed than expected, particularly for GWAS sets with reduced number of variants after exclusion of less common variants (i.e., minor allele frequency (MAF<%5) & minor allele count (MAC<30)). (Figures b & d for IL6 and IL8 respectively).



Supplementary Figures 3.4. Genetic correlation between anthropometry and inflammatory phenotypes. All phenotypes except IL8, fibrinogen (FIB) and BMI were adjusted for BMI, age and sex. Blue color in the heatmap represents negative and red a positive pairwise correlation, while color intensity demonstrates degree of pairwise correlation. Clusters are based on genetic distance (i.e., overlap of associated genetic markers).



FINE-MAPPING PLEIOTROPIC SIGNALS

In this chapter, the results of fine-mapping analysis to distinguish the set of variants with high posterior causal probability is presented. Three genomic loci on chromosome 3, 12 and 18 with suggestion for genetic pleiotropy between adiposity and inflammation traits were probed. The locus on chromosome 12 likely harbors two independent signals. The functional evaluation showed the majority of fine-mapped variants were SNPs with regulatory features associated with several tissues including inflammation-related cells; several variants also exhibited regulatory features associated with dipocytes. Results underscored potential contribution of regulatory variants in pleiotropy.

1. Background

Both GWAS and pleiotropy studies have provided essential, initial evidence linking genetic variants with biologic traits or diseases. However, GWAS and, by extension, pleiotropy assessments, which primarily rely on summary statistics obtained from GWAS studies, cannot reveal the functional consequence of significant variations.

GWAS-identified signals are typically represented by a ‘lead’ SNP, which has the strongest (i.e., lowest) p-value in the suspected region. This lead variant may not directly affect phenotypic expression but, rather, may be a proxy for a ‘true’ causal variant due to LD, which induces strong correlations between SNPs. Overlap or proximity with a coding region does not necessarily imply causality²⁸⁰. Further complicating the picture, multiple

causal and apparently independent variants within a single region can act through joint allelic effects on the same haplotype¹²⁵.

The biological interpretations of associations between likely causal variants are highly dependent on the genomic region in which the variants are located. Over 90% of the GWAS variants are in non-coding locus and do not directly affect messenger RNAs (mRNA)²⁸¹ transcription. The functional characteristics of these variants can vary greatly when operating under different cell-tissue environments²⁸².

Such genetic architecture complexities create challenges when attempting to functionally interpret GWAS and pleiotropy results. Localized analyses of risk-inducing loci must be performed with considerations for both allelic structures and functional annotations.

The fine-mapping approach assigns a statistical causality probability to candidate variants located within the GWAS-identified regions²⁸³. In other words, this technique is used to filter GWAS-identified variants and distinguish SNPs that are most likely to be causal in each genomic region.

Recently, a number of fine-mapping tools have been developed that utilize iterative algorithms to perform causal variation identification, including Markov Chain Monte Carlo (MCMC)²⁸⁴, exhaustive¹²⁵ and stochastic searches²⁸⁵. Most of these tools assume the presence of a single causal variant for a given region²⁸⁶. However, this assumption may not reflect true biological reality, particularly when examining complex traits, such as adiposity and inflammation. Additionally, many fine-mapping packages lack rigorous statistical framework to incorporate tissue or cell-specific functionality^{286,287}, which could impact predictive accuracy.

Thus, two complementary methods, PAINTOR²⁰⁶ and FINEMAP¹⁶², were used to perform the fine-mapping of adiposity–inflammation pleiotropy-identified regions from CCHC. Although FINEMAP is unable to account for functionality, this program uses an efficient algorithm that accommodates the identification of multiple causal variants, with no substantial computational cost¹⁶². After they were identified, the fine-mapped prioritized variants were subjected to functional assessments to determine biologic effect potentials.

2. Methods

2.1. Genome-wide pleiotropy summary statistics

The dataset consisted of pleiotropy summary statistics for approximately 6 million SNPs. Pleiotropy statistics reflect the outcomes of univariate GWAS and pleiotropy assessments performed during a prior round.

In brief, using genomic measurements from $N = 3,313$ participants from the CCHC study, a mixed-model, univariate GWAS were conducted, including anthropometric traits (BMI, $WHR_{BMI-adj}$, and $WC_{BMI-adj}$) and inflammation markers ($CRP_{BMI-adj}$, fibrinogen, $IL-6_{BMI-adj}$ and IL8; $WHR_{BMI-adj}$ and $WC_{BMI-adj}$ -GWAS were completed in males and females separately.

Pleiotropy test was subsequently performed using the trait-specific GWAS summary statistics. Pleiotropy results were converted from newer genome build GRCh38 to older GRCh37 for compatibility with fine-mapping packages.

2.2. Statistical framework

A traditional approach used to refine association signals is conditional analysis, a stepwise process that conditions on the variants with the lowest p-values in association tests until no additional SNPs are identified that achieve the pre-determined p-value threshold. Although this approach can be informative regarding the potential number of signal sources for a specific locus, it does not deliver probabilistic measures of causality for variants.

Bayesian framework has been increasingly being adopted as a way to overcome this problem²⁸⁸. However, a Bayesian approach requires both genotypic and phenotypic inputs, which become impractical when the sample size or number of variants increases. Therefore, fine-mapping methods have been extended to use only the summary statistics from genetic tests, together with SNP correlations obtained from reference panels¹²⁵.

PAINTOR represents a fine-mapping method that uses the LD structure and Z-scores from either univariate GWAS or pleiotropy results²⁸⁹. A distinguishing feature of this approach is that it leverages functional annotations of individual variants to estimate causal posterior probabilities. This approach is also robust when multiple, distinct causal variants appear to underlie functionality and in the context of heterogeneous effect sizes.

In addition to PAINTOR, the FINEMAP package¹⁶² was utilized. Pleiotropy suggests functional overlap, interrelatedness, and complexity in the underlying mechanisms, which can all increase the possibility that multiple causal variants can be found in a given region. Given the high density of tested variants, the incorporation of functional annotations with a fine-mapping probe will exponentially increase the computational costs.

PAINTOR is only efficient for the probing of a maximum of 3 causal variants for variant-dense regions. In contrast, FINEMAP package uses a Shotgun Stochastic Search (SSS) algorithm²⁰⁷ to explore various causal configurations by concentrating its efforts on configurations with non-negligible probabilities¹⁶². FINEMAP computes probabilities using summary statistics obtained from GWAS (or pleiotropy Z-scores) and pairwise LD estimates to facilitate the probing of an increased number of causal variants without computational penalty. Both packages were used complementarily to maximize the set of variants identified with high posterior causality probabilities.

2.3. Defining fine-mapping loci

From the pleiotropy summary statistics, a subset of variants with a pleiotropic significance level of $P_{\text{aspu}} < 5 \times 10^{-6}$ was subtracted. Next, the variants were further subdivided into independent groups based on their chromosome and base pair numbers. In each locus, variants were ranked by significance level, and the top SNP with the lowest P_{aspu} value was selected to serve as the index for the region.

All variants within 500 Kbps in both directions of the index were extracted from the pleiotropy summary statistics. For each SNP, the reference SNP cluster ID (rsid) number, chromosome, base pair (bp) number, pleiotropy Z-score, P_{aspu} -value, and effective and alternative alleles were extracted. For the estimation of SNP pleiotropic effect sizes and standard errors, the following formulas were utilized:

$$\beta = \frac{z}{\sqrt{2p(1-p)(n+z^2)}} \dots (1)$$

$$Se = \frac{1}{\sqrt{2p(1-p)(n+z^2)}} \dots (2)$$

where β is the effect size, z is the Z-score, p is the allele frequency, and n is the sample size for the variants.

2.4. Linkage disequilibrium estimation

Pairwise Spearman's rank correlation coefficients were estimated using the subset of unrelated individuals in the CCHC, in which each calculated value could vary between $[1,-1]$ for any given pair. For LD calculations, LDStore2 package²⁹⁰ was utilized, which uses the binary BGEN file format²⁹¹, rather than the variant call format (VCF)²⁹², and is computationally more efficient than similar platforms.

2.5. Functional annotation

Annotation libraries included:

- a) enhancer elements from Functional Annotation of the Mammalian Genome project (FANTOM5)²⁹³,
- b) various functional annotations from ENCylopedia of Dna Elements (ENCODE) project²⁹⁴ including transcription factors²⁹⁵, enhancers²⁹⁶ and promoters²⁹⁷,
- c) protein coding elements from gene-ENCODE (GENCODE)²⁹⁸,
- d) super enhancers²⁹⁹,
- e) DNase I hypersensitive sites (DHSs)²⁸¹ regulatory elements.

The functional annotation of selected regions was performed with the PAINTOR package in two steps.

During the first step, all annotations associated with the functionality of the liver, brain, blood, muscle, adipose tissues, and exon variants were used to annotate variants in the target loci.

Because a proportion of annotations could overlap and induce multicollinearity, correlation matrices were created using annotation scores from the annotated loci, such that each value in the matrices represented a pairwise correlation between a pair of annotations. One pair of annotations, with correlation values > 0.2 , was set to missing.

The removal process was repeated several times until the remaining set of annotations was roughly uncorrelated (i.e., all pairwise correlations were below the established threshold). The remaining annotation set was then utilized for the target regions annotation.

2.6. Fine-mapping process

Fine-mapping analysis was completed using PAINITOR at each locus, separately, incorporating LD scores and annotations files and assuming up to 3 causal variants per locus. With FINEMAP, analyses were performed by assuming up to 10 causal variants per region. All variants with posterior causality probabilities > 0.5 were extracted from the outcomes of both packages to maximize the number of likely causal variants identified at each locus. These subsets were used for the downstream functional assessments.

2.7. Functional evaluation

Several bioinformatic platforms were utilized

i) Ensemble Variant Effect Predictor (VEP), from Ensemble³⁰⁰ (grch37.ensembl.org/), was used to determine the effect of each likely causal variant on genes, transcripts, protein

sequences, and regulatory regions. VEP was also used to perform the Combined Annotation-Dependent Depletion (CADD) coding score estimation³⁰¹; this score is widely used to measure variant deleteriousness. A negative log-transformed CADD score > 10 suggests a high probability of an effect on protein coding.

ii) Enhanced Accuracy in Predicting the Functional Consequences of Non-coding and Coding SNV (FATHMM-XF)³⁰² is a method for predicting point mutations within the human genome the probable effect of the predicted point mutation on protein coding: a score > 0.5 (of a maximum of 1) indicates a likely coding effect, whereas a score > 0.9 indicates a highly pathogenic or protein-altering mutation.

iii) Haploreg is a tool for evaluating the annotation of a noncoding genome at variants on haplotype blocks, such as candidate regulatory SNPs at disease-associated loci³⁰³. This annotation tools is use to assess the effects of SNPs on regulatory motifs and expression Quantitative Trait Loci (eQTL), which are genomic regions associated with expression levels of messenger RNA (mRNA).

iv) ClinVar is a PubMed-supported and publicly available tool for assessing links between variants and human phenotypes.

2.8. Independence of signals

Lastly, pairwise Spearman's rank correlation coefficients were calculated between likely causal variants in each region. A cut-off value of $R^2 < 0.1$ was used to determine independence, as correlation values below this threshold suggest that the pair of variants likely represent different functional signals.

2.9. Visualization of pleiotropic signals.

To perform a more localized visualization of those loci with evidence for pleiotropy, locuszoom plots³⁰⁴ were employed. For each locus, the variant with the lowest p_{aspu} was selected as the index, and all variants within a distance of 1 megabase pairs (Mbp) were plotted on either side of the index SNP.

LD between the index SNP and all other variants within the probed region were calculated using the set of unrelated individuals from the CCHC (i.e., individuals with a third-degree kinship relationship or more distantly related). The populations used for the genotyping reference were Hispanics from the Population Architecture using Genomics and Epidemiology (PAGE II) cohort (pagestudy.org). The fine-mapped variants with high causal posterior probabilities were marked with distinguishing shape.

3. Results

The analysis of pleiotropy results indicated three regions on chromosomes 3, 12, and 18 with potential for pleiotropic associations. Out of 7,236 variants assessed, a total of 22 variants showed a posterior probability of > 0.5 for all three regions combined. In addition to 9 variants suggest by PAINTOR, 13 SNPs were distinguished by FINEMAP. The results in Tables 4.1a–c show the variants with the highest posterior probability of causality in each locus.

For the locus on chromosome 3 (bp range: 3,653,147–4,650,039), a total of 9 SNPs were identified with posterior probability values > 0.5 (out of 3,234 SNPs assessed, Figure 4.1). As noted, all variants are intronic, nonsense-mediated mRNA decay (NMD) SNPs. Both

the FATHMM-FX and negative-log CADD scores were considerably below their respective threshold values of 0.5 and 10, which suggested low probabilities that these SNPs were protein-altering variants.

However, several variants showed either extensive regulatory annotations with immunity, skin, pancreas, spleen, gastrointestinal, and lung tissues or were in tight LD (i.e., $R^2 > 0.8$) with SNPs featuring regulatory annotations associated with those tissues (Table 4.1a). One SNP, rs60505812, featured a regulatory annotation associated with adiposity tissue, although it was not in high LD with any other regulatory variant.

All variants overlapped with the gene encoding sulfatase-modifying factor 1 (*SUMF1*)-NMD transcript on the reverse strand (Ensemble ID: ENSG00000144455). Estimates from pairwise LD analysis (Supplementary Table 4.1, Figure 4.1) indicated that likely one causal SNP exists in this region because the pairwise correlation estimates were generally above the threshold level of $R^2 > 0.1$.

For the locus on chromosome 12 (bp range: 43,981,459–44,980,751), 8 variants had posterior probabilities > 0.5 (out of 1,501 SNPs examined, Table 4.1b, Figure 4.2). Similar to results for chromosome 3, most SNPs featured low coding scores and were intronic and NMD, except rs440389 and rs277221, which are intergenic with no significant regulatory features associated with any tissue. The remaining 6 SNPs have variable regulatory annotations associated with immunity, skeletal muscle, adipose, blood, vessels, lung, and pancreas tissues.

The most notable variant from this list is rs73093474, which features multiple regulatory annotations with adipose, immunity, brain, and pancreas tissues, and is also located in a

region replete with regulatory SNPs. The two variants at the top of the list overlapped with the gene encoding interleukin 1 receptor-associated kinase 4 (*IRAK4*) (ENSG00000198001); these two variants may represent a second, independent signal within the region given their negligible pairwise LD with the other likely causal variants in the region (Supplementary Table 4.2, Figure 4.2). The remaining fine-mapped SNPs overlapped with the gene encoding transmembrane 117 (*TMEM117*) (ENSG00000139173).

Finally, for the locus on chromosome 18 (bp range: 61,079,369–62,075,853), 5 variants exhibited posterior probability values > 0.5 (out of 1,301 SNPs evaluated). The second and third variants on this list are intergenic, whereas the remaining variants are intronic and downstream/upstream to regulatory SNPs. The last two variants (rs28483202 and rs631815) did not exhibit any regulatory annotations with the bioinformatic tools that were used.

The first and fourth variants overlapped with the genes encoding serpin family B members 10 (*SERPINB10*) and 2 (*SERPINB2*). The last variant on the list, rs631815, overlapped with both the genes encoding histocompatibility minor serpin domain-containing (*HMSD*) and the AC009802.1 region, which has not yet been functionally characterized.

All variants appeared to point to the same signal, as the pairwise LDs were above the threshold levels (Supplementary Table 4.3, Figure 4.3).

4. Discussion

In this study, summary statistics from pleiotropy analysis, information on pairwise correlations between SNPs inferred from a set of unrelated individuals in the study, and functional annotations for regulatory and coding variants from highly referenced libraries, including ENCODE²⁹⁴ and FANTOM5²⁹³, were incorporated to determine the genomic loci with the highest probability of exerting a causal role on genetic pleiotropy between adiposity and inflammation.

Functional assessments were intended to expand upon the statistical findings to narrow the results to those SNPs that may have the most biological relevance. Because adiposity and inflammation are complex traits^{305,306} that are highly polygenic³⁰⁷⁻³¹⁰, a higher probability exists for the identification of multiple, independent causal signals, even within relatively small regions³¹¹. Therefore, the fine-mapping analysis using PAINTOR, which allows the incorporation of functionality, was complemented with FINEMAP outcomes, both to expand on the credible set of causal variants for each signal and to increase the probability of identifying additional signals in each fine-mapped region, assuming that such signals exist.

Based on summary statistics from the pleiotropy analysis, three loci on chromosomes 3, 12, and 18 were identified based on suggestive evidence for contributions to pleiotropy between adiposity and inflammatory traits in a Hispanic population. After the application of the statistical approach, 22 variants showed posterior causality probabilities greater than 50% from the 7,236 variants assessed across all 3 regions.

The index SNPs identified in each region were associated with the most significant p-values for pleiotropy, including rs200671707 on chromosome 3, rs60939199 on chromosome 12, and rs4609952 on chromosome 18; however, these three SNPs did not have high posterior causality probability values. The fine-mapping algorithm may not have been able to differentiate causal variants when the underlying structure is correlated, and several significant variants exist.

Highly significant variants from the pleiotropy analysis results may also represent proxies of ‘true’ causal SNPs that are not presented in the dataset and, therefore, the observed associations are biased toward the tagged (observed) SNPs³¹². Evidence exists to support a role played by topologically associated domains (TADS), which refer to distant loci that interact with each other³¹³ and are believed to have significant impacts on the functional genome³¹⁴. Therefore the GWAS and pleiotropy signals used for this analysis may have resided in different genomic regions that interacted with the observed locus.

Finally, this lack of causality associated with highly significant SNPs may indicate the presence of multiple, independent signals; however, the pairwise LD estimates (Supplementary Table 4.1-3) suggest that only the fine-mapped region on chromosome 12 may harbor an additional independent signal if any such signal exists.

In the fine-mapped region on chromosome 3, a total of 9 variants had high causal probability values. Two SNPs, including rs60505812 and rs10514654, are notable. The first variant has regulatory features associated with adipose tissue, whereas the second variant exhibited regulatory annotations associated with immunity, skin, and intestinal tissues. A relatively elevated coding score (CADD = 6.8) suggested that the mutation in rs60505812 may potentially affect mRNA transcription, although the score remains below

the cut-off value of 10 that denotes mutational significance³¹⁵. Therefore, in the context of the protein-coding process, it is less likely to be a high-impact SNP.

All suggestive causal variants in this region overlapped with *SUMF1*, which encodes sulfatase-modifying factor 1 enzyme, a protein that catalyzes glycosaminoglycans, sulfolipids, and steroid sulfates³¹⁶. Defects in this gene are known to cause multiple sulfatase deficiency³¹⁶, a lysosomal storage disease that contributes to the deterioration of tissues in the nervous system, resulting in motor dysfunction, seizures, and developmental delays³¹⁷. However, variants in this gene have also been linked to inflammation in chronic obstructive pulmonary disease (COPD)³¹⁸. Furthermore, steroid sulfates play functional roles in energy homeostasis and inflammation³¹⁹.

The analysis of the target region on chromosome 12 showed that 8 variants had high causality probability values. Pairwise LD estimates (Table 4.2b) indicated two different signals. The two variants at the top of the list overlapped with *IRAK4*, whereas the remaining variants were associated with *TMEM117*. Although the first variant is intergenic, its only regulatory feature was associated with trophoblast tissue, whereas the second variant (rs4251527) was linked with regulatory activities in blood and immune cells.

IRAK4 encodes a protein kinase, which is essential during the innate immune response to Toll-like receptors (TLRs), and its deficiency increases the susceptibility to infections³²⁰. Studies of obese mice showed that the overexpression of TLRs in enlarged adipose tissue might play an important role in obesity-induced inflammation³²¹.

Two variants in the same region (on chromosome 12) that overlap with *TMEM117*, including rs73093474 and rs10506239, merit further interrogation. The first variant has

regulatory features associated with an extensive number of tissues, including adipose, lung, brain, lung, and pancreas. The second variant shows regulatory annotation with brain tissue (Table 4.1b). *TMEM117* encodes transmembrane protein 117 and has been associated with endoplasmic reticulum-stress (ER-stress) and mitochondria-mediated cell death³²².

Lastly, after probing the locus of interest on chromosome 18, a total of 5 variants showed high causality probability values. The first variant (rs6567401) overlaps with *SERPINB10*, which encodes a protein that controls the regulation of protease functions during hematopoiesis (blood production) and has been associated with allergic inflammation^{323,324}. The second variant (rs28483202) is intronic for *SERPINB2*, which encodes plasminogen activator inhibitor 2 (PAI2), a protein known to play a role in the manipulation of immunity³²⁵ and is upregulated during allergic responses³²⁶. The third variant (rs631815) is upstream of *HMSD*, which is primarily expressed in cells of myeloid lineage³²⁷. The remaining two variants are intergenic, although the closest functional gene is *SERPINB2*.

Generally, all but a few of the identified variants with high causality probability values across all three regions were intronic and NMD SNPs. NMD variants are considered to be regulatory, quality control variants that reduce gene expression errors by eliminating mRNA transcripts that contain premature termination codons (PTCs)³²⁸.

Growing evidence suggests that variations and tissue-specific differences in NMDs can alter the underlying pathologies of genetic diseases³²⁹⁻³³¹. Notably, interactions between NMDs and protein-coding variants have been linked to an increased risk of developing obesity and type 2 diabetes, as reported in one recent trans-ancestry cohort³³². The clinical study of a patient with the early onset of severe obesity was associated with an NMD

alteration affecting proopiomelanocortin³³³, which is involved in a wide range of activities, including body weight regulation.

Overwhelming (>90%) portion of the GWAS variants were discovered in noncoding genomic regions²⁸¹. The results from this study underscore the significance of these regulatory variants and their effects on pathophysiologic processes upstream of physiologic phenotypes and clinical traits.

5. Tables and Figures

Table 4.1a Characteristics of variants with high posterior causal probability in the target region on chromosome 3 (base pair range: 3,653,147–4,650,039).

Rsid	PP*	Base pair position	Consequence	Biotype	-Log CADD	FATHMM	Nearest or overlapping gene	Regulatory annotation	In tight LD with other regulatory annotated SNPs
rs35441937	1	4155769	1-Intron 2- NMD** transcript	1-NMD 2-PC [§]	1.4	0.03	SUMF1	Immunity skin	modest numbers
rs10510288	1	4157417	1-Intron 2- NMD transcript	1-NMD 2-PC	0.5	0.02	SUMF1	Pancreas spleen	modest numbers
rs6808934	1	4158758	1-Intron 2- NMD transcript	1-NMD 2-PC	1	0.02	SUMF1	Immunity skin	modest numbers
rs2874822	0.87	4159892	1-Intron 2- NMD transcript	1-NMD 2-PC	5.5	0.04	SUMF1	Immunity Neural cells	modest numbers
rs60505812	1	4174307	1-Intron 2- NMD transcript	1-NMD 2-PC	6.8	0.03	SUMF1	Adipose spleen	Solitary
rs952644	1	4181750	1-Intron 2- NMD transcript	1-NMD 2-PC	6.8	0.02	SUMF1	Immunity Skin Intestine carcinoma	few numbers
rs2600116	0.94	4224731	1-Intron 2- NMD transcript	1-NMD 2-PC	4.5	0.03	SUMF1	Immunity Skin GI lung	extensive numbers
rs10514654	0.95	4233925	1-Intron 2- NMD transcript	1-NMD 2-PC	0.2	0.03	SUMF1	Immunity Skin intestine	solitary
rs6777465	0.95	4237036	1-Intron 2- NMD transcript	1-NMD 2-PC	0.7	0.02	SUMF1	Intestine lung	few numbers

*PP= posterior probability, **NMD= Non-sense mediated decay, [§]PC= protein coding.

Table 4.1b Characteristics of variants with high posterior causal probability in the target region on chromosome 12 (base pair range: 43,981,459–44,980,751).

Rsid	PP*	Base pair Position	Consequence	Biotype	-Log CADD	FATHMM	Nearest or overlapping gene	Regulatory annotation	In tight LD with other regulatory SNPs
rs440389	0.67	44042139	intergenic	-	1.4	0.25	IRAK4	trophoblast	modest numbers
Rs4251527	0.84	44176863	1-Intron 2- NMD** transcript	1-NMD 2- PC [§]	3.7	0.01	IRAK4	Immunity blood	modest numbers
Rs277221	0.99	43986008	Intergenic	-	8.4	0.11	TMEM117	-	solitary
rs60995491	1	44496655	1-Intron 2- NMD transcript	1-NMD 2- PC	2.7	0.02	TMEM117	skeletal muscle	extensive numbers
rs11182419	1	44500239	1-Intron 2- NMD transcript	1-NMD 2- PC	1.6	0.03	TMEM117	-	solitary
rs73093474	1	44511120	1-Intron 2- NMD transcript	1-NMD 2- PC	1.4	0.3	TMEM117	Adipose Skin Brain Muscle Bessels Blood	extensive numbers
rs10506239	1	44522900	1-Intron 2- NMD transcript	1-NMD 2- PC	1.3	0.2	TMEM117	Immunity Skin Brain Atrium Lung Pancreas	extensive numbers
rs73094803	0.96	44524961	1-Intron 2- NMD transcript	1-NMD 2- PC	3.8	0.3	TMEM117	Brain Skin Muscle	extensive numbers

*PP= posterior probability, **NMD= Non-sense mediated decay, [§]PC= protein coding.

Table 4.1c Characteristics of variants with high posterior causal probability in the target region on chromosome 18 (base pair range: 61,079,369–62,075,853).

RsId	PP*	Base pair position	consequence	biotype	-Log CADD	FATHMM	Nearest or overlapping gene	Regulatory annotation	In tight LD with other regulatory SNPs
rs6567401	1	61574833	1-Intron 2- Upstream /Downstream regulatory	1-PC** 2-PF	2.9	0.02	SERPINB10 SERPINB2	Immunity skin/blood	modest numbers
rs1243064	0.99	61513975	1-Intergenic	-	5.7	0.16	SERPINB2	Breast Colon liver	solitary
rs2689416	0.99	61520361	1- Intergenic	-	2.6	0.06	SERPINB2	Immunity skin	few numbers
rs28483202	1	61541093	1-Intron	1-PC	5.5	0.01	SERPINB2		extensive numbers
rs631815	0.99	61611972	1- Upstream variant	1-PC	0.9	0.06	HMSD		solitary

*PP= posterior probability, **PC= protein coding, \$PF= promoter flanker.

Figure 4.1 Regional zoomplot of the suspect pleiotropic locus on chromosome 3. The blue colored rotated square in the middle represents index variant while straight squares show fine-mapped prioritized variant with high posterior causal probabilities (>50%). The Rsid label shows the variant with regulatory feature associated with adipose tissue. Colors show degree of linkage disequilibrium (LD) with lead (i.e., index) variant in the region. Blue lines demonstrate recombination rate inferences from Population Architecture using Genomics and Epidemiology (PAGE II) Hispanic population. The horizontal bar under x-axis show genes overlapping or close to target region. The variant highlighted here is Non-sense mediated decay (NMD) variant for sulfatase modifying factor 1 (SUMF1) gene on the reverse strand (range of the SUMF1 gene transcript Chromosome 3: 3,742,498-4,508,965).

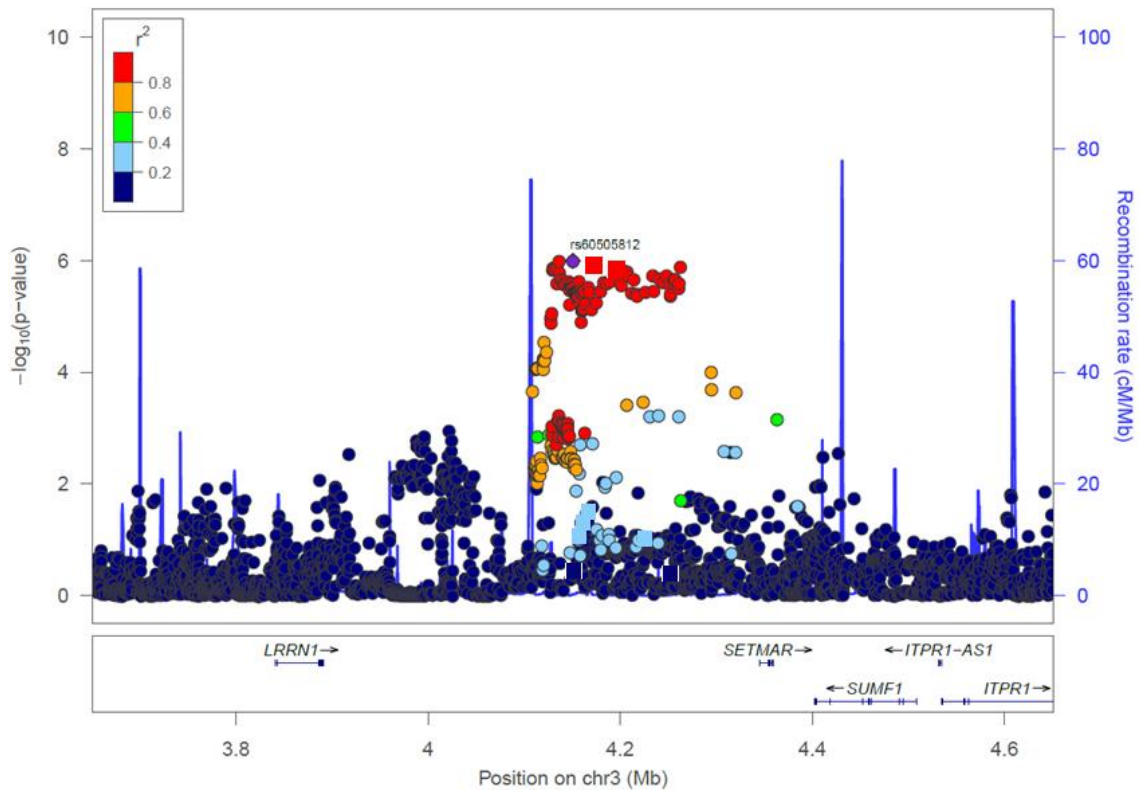


Figure 4.2. Regional zoomplot of the suspect pleiotropic locus on chromosome 12. The blue colored rotated square in the middle represents index variant while strait squares show fine-mapped prioritized variant with high posterior causal probabilities (>50%). The Rsid label shows the variant with regulatory feature associated with adipose tissue. The horizontal bar under x-axis show genes overlapping or close to target region. The variant highlighted here is Non-sense mediated decay (NMD) variant for transmembrane 117 (TMEM117) gene.

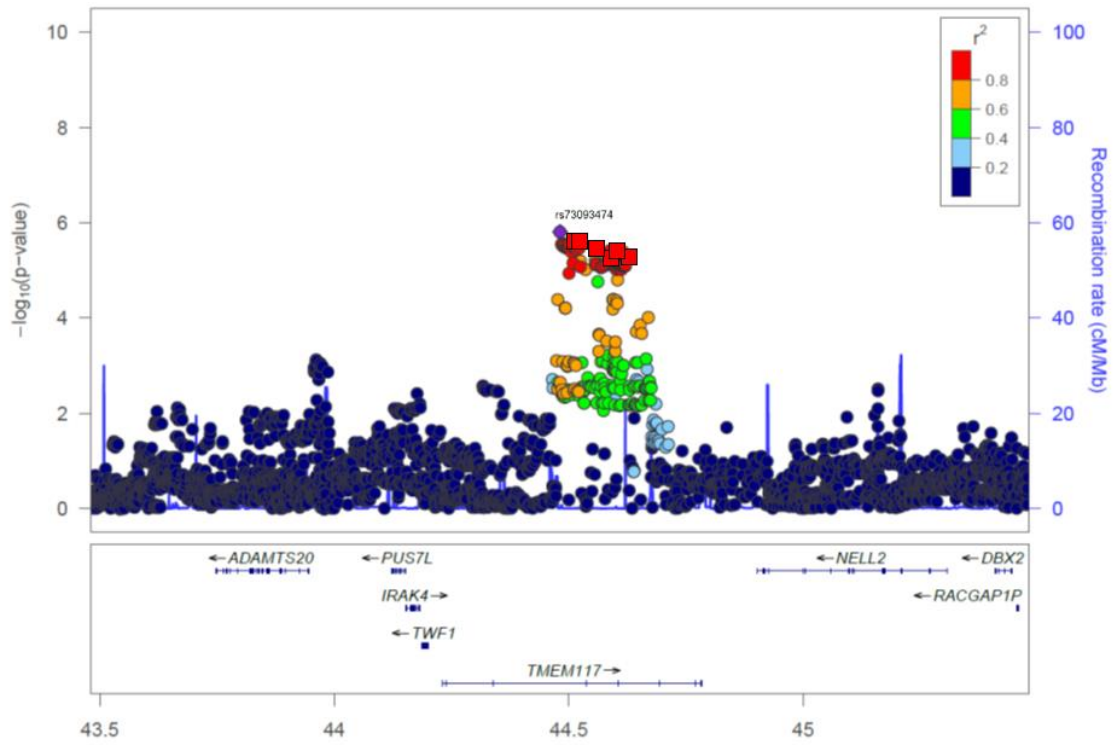
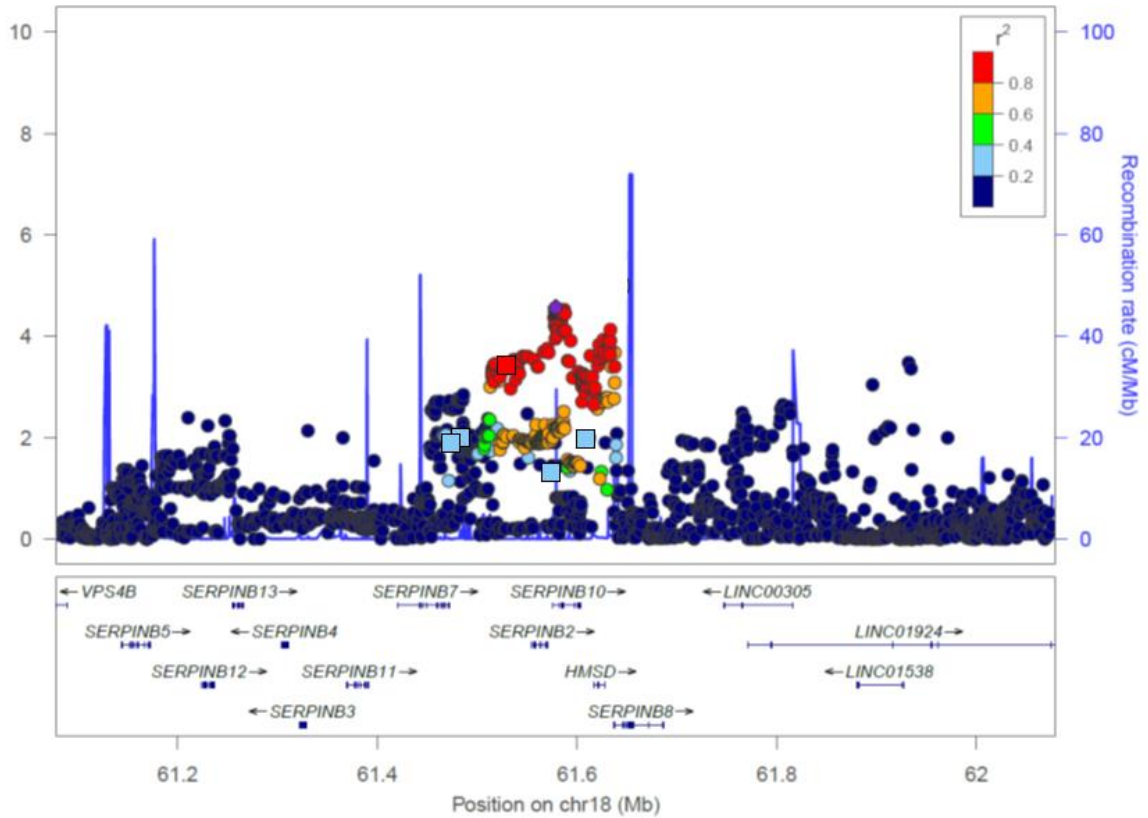


Figure 4.3. Regional zoomplot of the suspect pleiotropic locus on chromosome 18. The blue colored rotated square in the middle represents index variant while straight squares show fine-mapped prioritized variant with high posterior causal probabilities (>50%). The Rsid label shows the variant with regulatory feature associated with adipose tissue. The horizontal bar under x-axis show genes overlapping or close to target region. Fine-mapping prioritized variants mostly overlap with serpin family B members 10 (SERPINB10), SERPINB7 and SERPINB2 genes.



6. Appendices

Supplementary Table 4.1. Pairwise linkage disequilibrium (LD) among variants with high posterior causal probability in the target region on chromosome 3. LD can range from $R^2 \approx 0$ [no LD] to >0.99 [very tight LD]. Values indicate variants belong to the same signal [using the $R^2 > 0.1$ cutpoint].

SNP	rs2600116	rs35441937	rs2874822	rs10510288	rs6808934	rs60505812	rs952644	rs10514654	rs6777465
rs2600116	1	0.46	0.34	0.76	0.09	0.34	0.39	0.09	0.69
rs35441937	0.46	1	0.81	0.32	0.20	0.76	0.05	0.21	0.26
rs2874822	0.34	0.81	1	0.45	0.19	0.64	0.02	0.20	0.18
rs10510288	0.76	0.32	0.45	1	0.09	0.26	0.29	0.09	0.56
rs6808934	0.09	0.20	0.19	0.09	1	0.1911	0.20	0.90	0.11
rs60505812	0.34	0.76	0.64	0.26	0.19	1	0.10	0.20	0.38
rs952644	0.39	0.05	0.02	0.29	0.20	0.10	1	0.22	0.56
rs10514654	0.09	0.21	0.20	0.087	0.90	0.20	0.22	1	0.11
rs6777465	0.69	0.26	0.18	0.56	0.11	0.38	0.56	0.11	1

Supplementary Table 4.2. Pairwise linkage disequilibrium (LD) among variants with high posterior causal probability in the target region on chromosome 12. This locus may harbor two independent signals [using the $R^2 > 0.1$ cutpoint].

	rs440389	rs4251527	rs277221	rs60995491	rs11182419	rs73093474	rs10506239	rs73094803
rs440389	1.00	0.20	0.15	0.00	0.01	0.00	0.00	0.00
rs4251527	0.20	1.00	0.08	0.02	0.02	0.02	0.02	0.01
rs277221	0.15	0.08	1.00	0.01	0.01	0.01	0.00	0.01
rs60995491	0.00	0.02	0.01	1.00	0.01	1.00	0.93	0.88
rs11182419	0.01	0.02	0.01	0.01	1.00	0.01	0.01	0.01
rs73093474	0.00	0.02	0.01	1.00	0.01	1.00	0.93	0.88
rs10506239	0.00	0.02	0.00	0.93	0.01	0.93	1.00	0.83
rs73094803	0.00	0.01	0.01	0.88	0.01	0.88	0.83	1.00

Supplementary Table 4.3. Pairwise linkage disequilibrium (LD) among variants with high posterior causal probability in the target region on chromosome 18. Values indicate variants belong to the same signal [using the $R^2 > 0.1$ cutpoint].

	rs6567401	rs1243064	rs2689416	rs28483202	rs631815
rs6567401	1.00	0.08	0.22	0.22	0.09
rs1243064	0.08	1.00	0.34	0.73	0.82
rs2689416	0.22	0.34	1.00	0.29	0.32
rs28483202	0.22	0.73	0.29	1.00	0.73
rs631815	0.09	0.82	0.32	0.73	1.00

CAUSAL MEDIATION PATHWAY ANALYSIS

In this chapter, the results from causal mediation analysis are presented. Based on univariate regression analysis, tighter LD, more significant pleiotropic association and functional annotation of likely causal variants in 3 loci with evidence for pleiotropic associations with adiposity and inflammation, two variants including rs60505812 on chromosome 3 and rs73093474 on chromosome 12 were selected for in-depth causal mediation analysis to distinguish variants with direct (biologic) from indirect (mediated) pleiotropic effects. Results suggested rs60505812 may have direct and biologic association with both adiposity and inflammation traits. The other variant, rs73093474, exhibit both biologic and mediated pleiotropic effect on inflammation (via adiposity).

1. Background

Although GWAS have identified thousands of genetic variants associated with phenotypic traits¹²⁶, a variety of methodological issues^{281,334-338} have made understanding the impacts of these findings in a biological context somewhat challenging³³⁹⁻³⁴¹. The difficulty identifying causal variants²⁸⁸ and pathways that mediate genetic effects³⁴² have been the two key barriers to the meaningful interpretation of genetic studies because only a very small fraction of identified variants are causal³¹¹.

Despite the increased availability of fine-mapping tools for performing the localized probing of genomic regions^{206,285,286}, which has greatly contributed to pinpointing likely causal variants, the functional follow-up necessary to determine the underpinning

mechanisms that connect these causal variants with phenotypic expression lags considerably behind³⁴³.

This gap between genetic association studies and the identification of the related biological effects is an important issue, particularly when characterizing multi-potent variants³³⁹. The added complexity of studying these prevalent³⁴⁴ but less well-characterized ‘pleiotropic’ variants¹²⁷ is due to the existence of several models of associations³⁴⁵ that each implies different types of functional effects.

As mentioned previously, based on the concept of “biologic” pleiotropy, a genomic locus that is independently associated with several phenotypes would suggest that the genetic effect is transmitted through a common pathway that is upstream of all associated phenotypes¹²⁷; however, pleiotropic effects can also be transmitted indirectly due to relationships between the outcome phenotype and a “mediating” phenotype¹²⁷. Elucidating these complexities could provide novel biological insights to better understand the biological pathways that underpin genotype-phenotype associations³⁴⁶, as well as disease traits¹²⁷, creating opportunities for improved clinical classification and drug repurposing.

In this study, a causal mediation pathway analysis²¹¹ of likely causal pleiotropic variants was performed based on evidence of an association between adiposity and inflammation. Mediation analysis provides a framework for distinguishing biologic from mediated pleiotropy and for quantifying the proportion of total genetic effects on outcome phenotypes³⁴⁷. This method was previously utilized to dissect the direct and indirect genetic associations between SNPs and chronic obstructive disease (COPD)³⁴⁸, metabolic syndromes³⁴⁹, mineral density³⁵⁰, and bladder cancer³⁵¹.

The aims of these analyses were to describe the associations between fine-mapped variants identified on chromosomes 3, 12, and 18 with anthropometry and inflammation-related phenotypes and distinguish whether the likely causal variants act directly (i.e., irrespective of any hypothesized mediator) or whether the association is predominately due to correlations, with particular attention paid to the prevailing understanding of causal inflammation and adiposity pathways^{352,353}. To achieve these aims, a set of phenotyped and genotyped subjects from the CCHC were examined.

2. Methods

2.1. Candidate variant set

A set of variants with posterior causal probability values > 0.5 , located at three separate loci on chromosomes 3, 12, and 18, were considered for the causal mediation analysis. Variants were identified through a fine-mapping analysis performed on loci associated with suggestive evidence for pleiotropy between anthropometry traits, including BMI, WHR, and WC, and inflammation-related phenotypes CRP, fibrinogen, IL-6 and IL-8.

Pleiotropy assessments were completed using GWAS summary statistics during prior steps, based on genomic measurements obtained from the participants in the CCHC study. Additionally, the index variants for each locus, which exhibited the lowest pleiotropic p-values and were used to configure the fine-mapping target regions, were also included in the set.

There was a total of 25 variants likely causal variants across all 3 regions. SNP allelic dosage for these selected variants were extracted from variant calling format (VCF)²⁹² files. Allele dosage is the estimated (expected) fractional and continuous counts of effect allele

at each SNP for each individual, ranging continuously from 0 to 2, where a value of 2 for a given variant indicated that the subject harbors both phenotype-increasing alleles, whereas a value of 0 implies the complete absence of the risk allele.

2.2. Phenotypes

The last non-missing phenotypic recorded observations for BMI, WHR, WC, CRP, IL-6, and IL-8 were extracted for all CCHC individuals whose genetic measurements were used in the univariate GWAS and pleiotropy assessments analyzing the association between anthropometry and inflammation. GWAS analyses similarly incorporated last non-missing phenotypic recorded observation for each trait, as well as corresponding covariates' values for age and sex. Details on phenotype data collection methods were presented in prior steps.

2.3. Covariate selection

A review of prior studies showed that several factors are likely to play significant roles in obesity/adiposity and inflammation. In addition to age and sex³⁵⁴⁻³⁵⁶, variables such as alcohol consumption and smoking^{357,358}, physical activity^{359,360}, hypertension³⁶¹, lipid profile^{362,363}, and insulin sensitivity³⁶⁴ have been reported.

It is arguable that few of the enlisted factors, including lipid levels and blood pressure, could act as the mediators of the association between adiposity and inflammatory outcomes^{365,366} and not as the confounders. In the case of insulin resistance, although studies have indicated causal associations, with adiposity serving as an outcome predictor^{367,368}, other assessments have indicated that the accumulation of visceral adipose tissue could be a consequence of increased insulin resistance rather than a cause^{20,369}.

Therefore, in this study, the following covariates were incorporated into the mediation models: homeostatic model assessment of insulin resistance (HOMA-IR), age (years), sex (binary), alcohol consumption (oz/week), moderate to vigorous level of physical activity (binary), current smoking status (binary), and ancestry principal components (to adjust for population stratification). The alcohol level was obtained by tallying the average number of alcoholic beverages consumed within a week, which was multiplied by 0.6 (or 0.48 for wine, based on the National Institute of Alcohol Abuse and Alcoholism (*niaaa.nih.gov*) formulation for the level of alcohol content in different beverages). A moderate to vigorous level of physical activity was defined as 150 minutes of sweat-inducing activities in a week.

The values for covariates were taken from the same exam where primary phenotypes were taken.

2.4. Statistical analysis

The list of fine-mapped-prioritized SNPs and index variants in the 3 target regions with evidence for pleiotropy and their SNP dosage values were tested for univariate associations with the study phenotypes (BMI, WC, WHR, CRP, IL-6, and IL-8) using linear regression assuming an additive genetic effect.

Pairwise correlations were calculated between the index SNPs and variants with high posterior causality probability values (with *PLINK* package³⁷⁰) to assess whether the fine-mapped variants are in LD with their respective index SNPs. Variants with the strongest significant associations with phenotypes, in the tightest LD with the index SNPs, and sharing functional features associated with adiposity- and inflammation-related tissues were subsequently selected for causal mediation analysis.

Analyses were limited to samples with non-missing observations for outcome phenotypes. An exploratory analysis of the phenotypes showed the skewed distributions of inflammatory markers; therefore, inflammation phenotypes were log-transformed (supplementary Figure 6.1) because mediation analysis assumes the normal distribution of outcomes. Some covariates, including physical activity, alcohol consumption, and smoking status, were characterized by considerable missingness. Because mediation analysis only uses complete cases, missing covariates were imputed using the predictive mean matching (PMM)³⁷¹ method. Subsequent mediation analyses were conducted with both an imputed dataset (assuming missingness at random) and a reduced set that only included empirically complete cases.

Directed acyclic graphs (DAGs) were used to conceptualize the biologic pathways of associations (Supplementary Figure 6.2). Causal mediation analysis requires that the causal model (i.e., the likely temporal sequence of phenotypes) be pre-specified. Causal models were based on the prevailing knowledge of adiposity and inflammation associations (causal directions) to determine the mediator and outcome phenotypes from each mediation analysis.

2.5. Causal mediation pathway analysis

Under the biologic pleiotropy model, a likely causal variant is associated with multiple phenotypes through distinct pathways (Supplementary Figure 5.3a). Therefore, the directional association between the SNP and the outcome phenotype should be significant (outside of the 95% error levels), once the association has been adjusted for the mediator (phenotype) and the covariate effects; a lack of significant associations may suggest mediated pleiotropy (Supplementary Figure 5.3b).

For this study, analyses were performed using the causal inference approach to mediation analysis²⁰⁹, which allows for the assessment of whether an SNP has a direct effect on the phenotype of interest or whether the SNP only acts through a secondary (correlated) phenotype, in a sequential manner. Distinguishing the pleiotropic operational mode assists in the identification of the biological pathways through which a genomic variant affects multiple phenotypes.

The R package Mediation³⁷² was used for analysis, which conducts a three-step process. The total effect between an SNP and an outcome phenotype is assumed to be the sum of the average causal mediation effects (ACMEs) and the average causal direct effect (ADE). The first step includes estimating the distribution of the mediator phenotype as a function of the SNP, adjusted for ancestry genetics (e.g., anthropometry ~ genotype + PCs). The next step involves estimating the distribution of the outcome phenotype as a function of the mediator phenotype, genotype, and covariates (e.g., inflammation marker ~ mediator + genotype + PCs + covariates). The final step combines the fitted models in the mediation equation, providing estimates and p-values for the direct (e.g., inflammation ~ SNP) and indirect (via adiposity) associations.

Mediation analysis involved the stepwise selection of covariates during the second step. The initial formulation (Model 1) included no covariates except ancestry PCs (e.g., inflammation marker ~ genotype + mediator + PCs). In subsequent models, age and sex (Model 2), alcohol and physical activity (Model 3), smoking status (Model 4), and insulin resistance (Model 5) were progressively added to assess the confounding effects of covariates on the observed associations. Because no significant differences were observed between the different adjusted models, only the final adjusted model was included in this

report. No meaningful difference was observed between the results inferred from the imputed dataset and those based on the empirical observations. Therefore, models using the multiple-imputation set were presented in the results section, and the results from the empirical set can be found in the supplementary materials.

The number of simulations was set to 1,000, and models were calculated with robust standard errors.

2.6. Sensitivity analysis

The empirical distributions of the covariates showed the presence of extreme observations. Comparative analyses were conducted both with and without outliers to assess whether the outcomes would differ.

For causal mediation analysis, the proposed sensitivity test in the Mediation R package was used. This test relies on sequential ignorability²¹², which implies that after accounting for an observed confounder, no unmeasured variable may confound any of the following: 1) genotype-mediator association, 2) genotype-outcome phenotype association, and 3) mediator-outcome association.

The package passes the results from mediation analysis to *medsens* function. The function chooses as the sensitivity parameter the correlation ρ between the residuals of the mediator and outcome regressions. If there exist unmeasured confounder which affect either the mediator or the outcome, it is expected that the sequential ignorability assumption is violated and ρ is no longer zero. The sensitivity analysis is conducted by varying the value of ρ and examining how the estimated ADE and/or ACME changes.

The number of simulations used for the sensitivity test was set at 100, and correlation parameter ρ increment was set at 0.1.

3. Results

The exploratory analysis of covariates (Table 5.1) showed elevated distribution of insulin resistance [median =2.9 μ IU/mL, interquartile range (IQR): 1.7, 4.8], a low proportion of individuals reporting moderate to vigorous physical activity (approximately 24% of participants), and current smokers being the majority of the cohort (55%). The empirical distribution of insulin resistance values showed the presence of a few extreme measures (Figure 5.1). The distributions of anthropometric measures indicated a high level of obesity (>75% with BMI > 30 and WHR > 88) and high levels of inflammation markers (IL-6 median =2.7pg/dL, IQR:1.4, 4.8).

The set of variants with a high posterior probability of causality, including their respective index variants, were fit into a linear regression with phenotypes as the outcome. The results of the univariate analysis and pairwise LD-score with the lead pleiotropic (index) variant for each region (Table 5.2) showed inconsistent patterns of associations across phenotypes. In the target region on chromosome 3, only 5 out 10 variants were associated with BMI, WC and WHR, 4 with CRP, 3 with IL-6, 2 with IL-8, and none with fibrinogen, including the results from index SNP. Only 2 variants in the locus, rs605058102 and rs10514654, were significantly associated with 5 out of 6 phenotypes.

For the target region on chromosome 12, 5 of 9 variants were significantly associated with BMI, WC and CRP, 6 with WHR, 1 with IL-6, 2 with fibrinogen, and none with IL-8, including the results from the index SNP. However, 5 of 9 variants (rs60939199,

rs60995491, rs73093474, rs10506239, and rs73094803) appeared to be associated with multiple phenotypes, including BMI, WC, WHR, and CRP. For the target region on chromosome 18, the results did not show any significant associations between phenotypes and likely causal pleiotropic variants, except for the index SNP (rs4609952), which was significantly associated with WHR (Table 5.2).

Based on the significant associations with multiple phenotypes identified in the univariate regressions (Table 5.2), two variants were selected for in-depth causal mediation pathway analysis due to tighter LD with their respective index SNPs, lower pleiotropic p-values (P_{aspu}), and adipose tissue regulatory features (Table 5.1a-b, prior chapter): rs60505812, on chromosome 3 and rs73093474 on chromosome 12.

To distinguish biological pleiotropy from mediated pleiotropy, causal mediation analyses were performed. Using DAGs, the regression results, and existing biological knowledge, separate but methodologically identical models were used to assess the total effects between the likely causal variants and the phenotypes of interest (Supplementary Figure 5.4).

For rs60505812 on chromosome 3, BMI was assumed to be the mediating phenotype, and IL-6 was chosen as the outcome. The selection of associated phenotypes was guided by univariate GWAS p-values. For this variant, no meaningful mediated effect was observed [estimate = 0.01 (pg/dL) per risk allele, 95% confidence interval (95% CI): -0.01, 0.04; Table 5.3], but the direct effect was significant (estimate = 0.13 (pg/dL) per risk allele, 95% CI: 0.05, 0.22) after adjusting for age, sex, the first 2 PCs, alcohol consumption, physical activity, current smoking status, and insulin resistance level.

For the variant rs73093474 on chromosome 12, WHR was the mediator, and CRP was chosen as the outcome phenotype. Evidence supported both a mediated effect (estimate = -0.02 (mg/L) per risk allele, 95% CI: -0.04 , -0.00) and a direct effect between the SNP and CRP (estimate = -0.16 (mg/L) per risk allele, 95% CI: -0.25 , -0.06). This result implies that the variant exhibits both biologic and mediated (via WHR) pleiotropy. These results suggested that these two variants contributed to inflammation, independently of adiposity.

The sensitivity analysis for unmeasured confounders (Figure 6.3) showed that unmeasured cofounders were unlikely to unduly confound direct SNP-outcome-phenotype associations for both variants, as the estimates were outside of the 95% CI for sensitivity parameters. Finally, mediation analyses were performed, excluding samples that had extremely high insulin resistance values. No tangible difference was observed; therefore, no data points were excluded from this report.

4. Discussion

Pleiotropy assessments performed during prior steps identified loci with suggestive evidence for multi-phenotype associations among CCHC participants. The functional evaluation of likely causal variants, identified through fine-mapping in the prior steps, revealed that the majority were nonsense-mediated (NMD) regulatory SNPs, underscoring their potential for pleiotropic effects, particularly in the context of adiposity and inflammation.

However, fine-mapping analysis is restricted to the assignment of causal probability based on observed associations and functional annotations. Mediation analysis is an

approach that can be utilized to model causal associations between variants with a high probability of causality and phenotypic outcomes³⁷³. The characterization of pleiotropy associations facilitates biological interpretations.

A total of 25 SNPs of likely causal variants, residing in three different regions, with evidence of inflammation-adiposity pleiotropy effects, were considered for mediation analysis. This set included three 3 index SNPs (i.e., the variants with the most significant pleiotropic p-values in each target locus) and 22 variants that demonstrated high probability of causality values during the fine-mapping analysis. LD analyses suggested that the examined variants consisted of 4 distinct signals, two of which were located on chromosome 12.

Based on the consistency of associations across multiple phenotypes, adiposity tissue regulatory features, and tight LD with respective index SNPs, two variants, including rs60505812 on chromosome 3 and rs73093474 on chromosome 12, merited in-depth examination. None of the SNPs in the third region on chromosome 18 were associated with any of the traits under the study.

Mediation analysis indicated that rs60505812 was independently associated with both IL-6 and BMI, which are surrogates of inflammation and adiposity, respectively, even after adjusting for age, sex, physical activity level, alcohol consumption, and current smoking status (Table 6.3), which is consistent with biologic pleiotropy. However, no indirect association was identified between the SNP and IL-6 [via BMI]; therefore, the model failed to provide evidence for mediated pleiotropy.

As indicated by functional analyses using bioinformatic tools in prior steps, rs60505812 is an NMD variant. NMDs, in general, play important roles in the reduction of gene expression error by eliminating mRNAs that contain premature stop codons³²⁸. This variant is located in an intronic region of the sulfatase modifying factor 1 (*SUMF1*) gene, which is known to encode a similarly named enzyme that catalyzes steroid sulfates³¹⁶. Although no direct association between *SUMF1* and metabolic traits has previously been reported, variants in this gene have been suggested to be associated with inflammation³¹⁸. Additionally, steroid sulfates play a functional role in energy homeostasis and inflammation³¹⁹.

The functional assessment of rs60505812 with Haploreg demonstrated an annotation with histone H3 lysine 4 trimethylated (H3K4me3)³⁷⁴ in fetal adipose cell nuclei. H3K4me3 promotes gene expression by altering the accessibility of genes to initiate mRNA transcription³⁷⁵. Whether this regulatory association and concurrent synergistic phenotypic variation [with fat uptake and increased pro-inflammatory activities] is promoted by common pathways at molecular, cellular or tissue levels require further assessments. However, regulatory annotation with fetal nuclei would suggest functional association with a wider range of metabolic processes, and therefore involving sub-cellular ‘foundational’ pathways.

For the variant rs73093474, on chromosome 12, mediation analysis indicated both a direct association with CRP (Table 6.3) and an indirect association (via WHR) in the covariate-adjusted models, which demonstrated concurrent biologic and mediated pleiotropy. In contrast to rs60505812, this variant has a regulatory annotation with the

mono-methylated H3K4me1-Enhancer in adipose cell nucleus, in addition to skeletal muscle and brain tissues.

H3K4me1 is the primary modification used to predict whether an SNP is a potential enhancer (booster) of the gene expression³⁷⁶. Therefore, the associated regulatory effect of rs73093474 [or the signal it represents] may be biologically more specific than rs60505812, and less likely to involve multipotent pathways that underpin a greater range of phenotypic traits. Caution must be exercised because a concrete biological inference necessitates both the replication of genetic associations in larger samples and experimental assessment.

This SNP is also an NMD variant, intronic to transmembrane 117 (*TMEM117*), which encodes a similarly named protein. This gene has been associated with endoplasmic reticulum stress (ER stress) and mitochondria-mediated apoptotic cell death³²². Obesity-induced ER stress has been shown to induce inflammation in adipose tissue³⁷⁷.

In conclusion, mediation analysis showed that the observed pleiotropic associations in two of three target regions are potentially biologic associations, in which the SNP affects a common pathway upstream of both inflammation- and adiposity-related phenotypes. However, evidence for mediated pleiotropy was also identified for rs73093474, which suggested that this variant also affects inflammatory phenotypes via the promotion of adiposity.

5. Tables and Figures

Table 5.1. Distributions of the phenotypes and covariates in Cameron County Hispanic Cohort (CCHC) among participants with non-missing values for either C-reactive Protein (CRP) or interleukin (IL)-6.

Variable (unit)	Median [IQR]	Sample Size
Age (year)	47.0[34.0, 60.0]	2247
Female, N(%)	1492(66)	2247
IL6 (pg/dL)	2.7[1.4, 4.8]	1316
CRP (mg/L)	3.9 [2.0, 7.7]	1860
BMI (kg/height ²)	30.1[26.6, 34.3]	2237
Waist Circumference (cm)	100.7 [92.5, 110.5]	2247
WHR (Waist/Height x 100)	92.9[88.0, 97.9]	2245
Alcohol (Oz/week)	0.0 [0.0, 0.6]	1196
Moderate to vigorous physical activity, N(%)	252(23.9)	1054
Current smoker, N(%)	578(55.0)	1050
Homa-IR* (uIU/mL)	2.9[1.7, 4.8]	2123

* homeostatic insulin resistance

Table 5.2. Univariate regression results with genetic variants as the outcomes and phenotypes as predictors.

Locus	Variant	P_{aspu}	R^{2*}	d'	BMI	CRP	IL6	IL8	FIB	WHR	WC
chr3:4150897	rs200671707	1.02E-06			2.25	2.86	2.82	1.70	0.28	3.07	2.45
	rs35441937	3.28E-01	0.10	0.96	0.52	0.52	-1.15	0.37	-0.57	0.26	0.68
	rs10510288	6.61E-03	0.24	0.99	2.14	1.98	1.15	1.64	-0.30	2.57	2.48
	rs6808934	9.34E-02	0.23	0.99	1.14	1.18	1.59	1.24	-0.79	1.86	1.36
	rs2874822	1.35E-01	0.10	0.96	-0.62	-0.75	-0.57	-0.44	-0.62	-0.62	-0.57
	rs60505812	1.28E-06	0.86	0.98	2.73	2.42	2.65	2.34	0.59	3.02	2.59
	rs952644	9.20E-03	0.19	0.92	2.10	1.68	1.39	1.75	-0.19	2.46	2.49
	rs2600116	9.48E-02	0.21	0.95	0.33	0.25	1.67	0.19	0.17	0.80	0.02
	rs10514654	1.89E-06	0.86	0.95	2.75	2.60	2.96	2.20	0.11	2.83	2.73
	rs6777465	7.47E-01	0.13	0.98	0.56	0.38	-0.91	-0.27	-0.19	0.19	0.91
chr12:44481380	rs60939199	1.56E-06			-2.50	-2.79	-1.25	-0.60	-0.87	-2.62	-3.28
	rs277221	1.33E-03	0.00	0.08	0.33	0.16	-0.48	0.56	-1.58	2.89	0.20
	rs440389	1.63E-01	0.02	0.49	-0.53	0.60	-1.95	0.04	-1.28	1.54	0.26
	rs4251527	3.87E-03	0.00	0.07	-0.68	-0.93	0.99	-0.68	2.01	-1.53	-0.16
	rs60995491	2.81E-06	0.87	0.94	-2.49	-2.85	-0.33	-0.79	-0.56	-3.32	-3.53
	rs11182419	6.56E-01	0.00	1.00	-0.60	0.26	-0.71	1.83	2.34	-0.67	-0.83
	rs73093474	2.81E-06	0.87	0.94	-2.50	-2.86	-0.33	-0.78	-0.56	-3.32	-3.53
	rs10506239	3.35E-06	0.93	0.99	-2.42	-2.84	-1.16	-0.42	-0.57	-2.87	-3.36
	rs73094803	6.26E-06	0.77	0.93	-2.51	-2.65	0.19	-1.06	-0.12	-3.19	-3.51
	chr18:61578750	rs4609952	2.67E-05			0.73	0.02	-0.20	-1.14	0.31	2.35
rs1243064		1.03E-02	0.21	0.97	-0.21	0.63	-0.50	0.54	-0.99	-0.81	-0.39
rs2689416		6.65E-03	0.30	0.99	0.25	0.18	0.05	-0.77	-0.12	1.31	0.03
rs28483202		4.48E-04	0.88	0.99	1.06	0.34	0.18	-1.40	0.38	1.87	0.01
rs6567401		4.40E-02	0.27	0.99	0.60	0.78	-0.26	-1.15	-0.32	1.53	0.41
rs631815		9.91E-03	0.32	1.00	0.75	0.23	0.45	-0.29	-0.47	1.06	0.21

* R -squared and d' are measures of pairwise linkage disequilibrium (LD) with index SNP in locus [highlighted with light gray]. Variants in tighter LD with their respective index single nucleotide variants (SNV), having significant pleiotropy p -value (P_{aspu}), significant association with multiple phenotypes and regulatory annotation features associated with adipose tissue were selected for in-depth causal mediation analysis [denoted by light blue color].

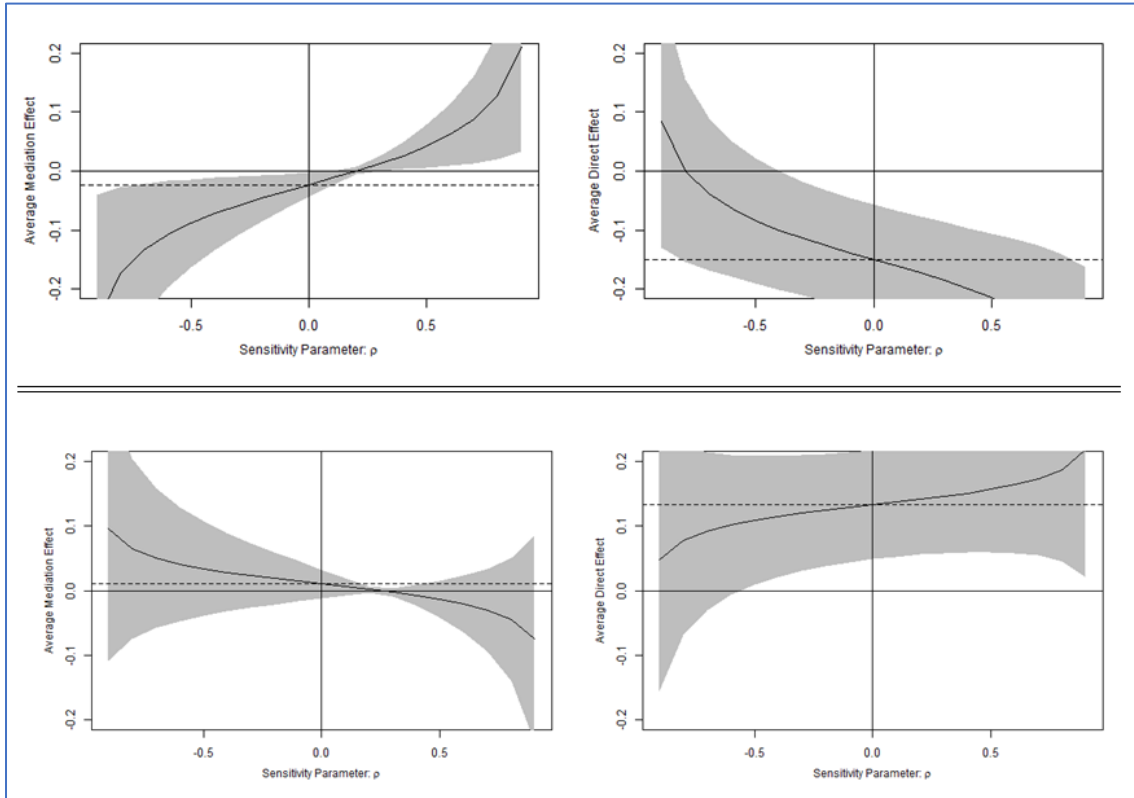
Table 5.3. Causal mediation analysis for rs60505812 and rs73093474.

Estimate/variant	rs60505812		rs73093474	
	<i>Adjusted Model</i>		<i>Adjusted Model</i>	
	<i>Estimate</i>	<i>95% CI</i>	<i>Estimate</i>	<i>95% CI</i>
Mediated Effect (ACME)*	0.01	(-0.01, 0.04)	-0.02	(-0.04, -0.00)
Direct Effect (ADE)**	0.13	(0.05, 0.22)	-0.16	(-0.25, -0.06)
Total Effect (TE)	0.14	(0.06, 0.23)	-0.18	(-0.27, -0.08)
Proportion Mediated	0.07	(-0.11, 0.29)	0.12	(0.01, 0.32)

*ACME=average causal mediated effect. For the variant rs60505812, this pertains to the statistical association between the SNP and the outcome phenotype (IL6), mediated through BMI. For the variant rs73093474, the ACME indicates the association between the SNP and the outcome phenotype (CRP), mediated through WHR.

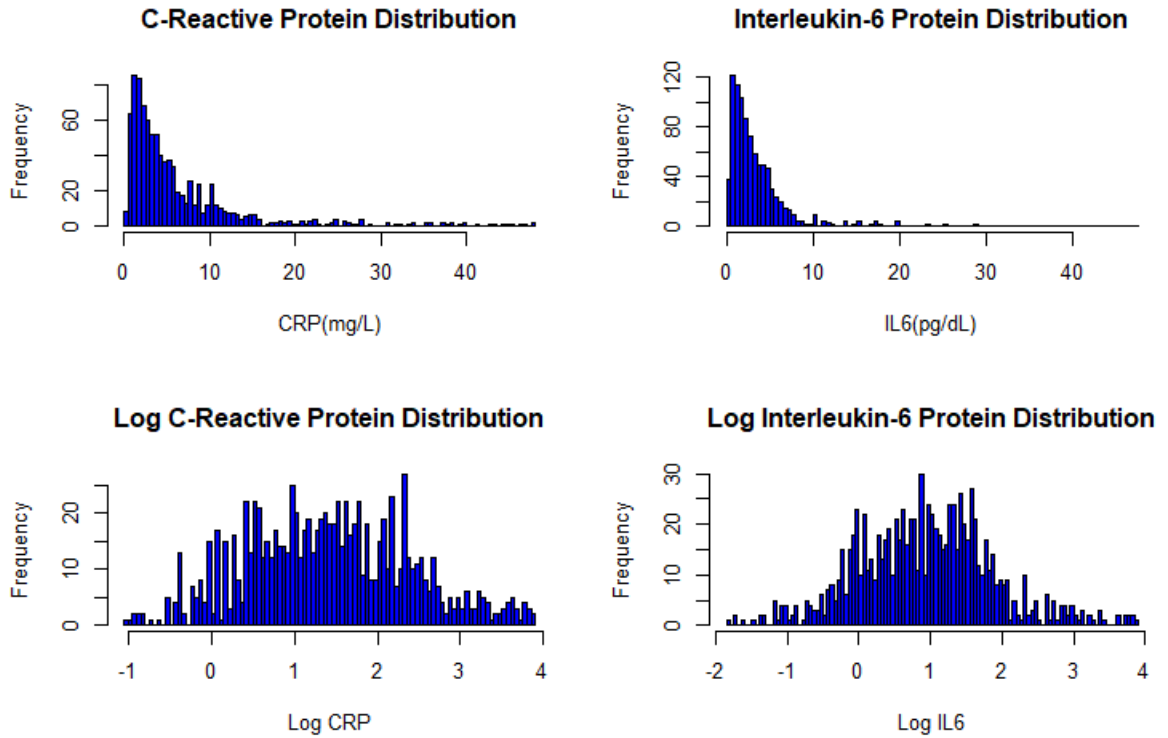
**ADE=average direct effect. For the variant rs60505812, this pertains to direct association between the SNP and the outcome phenotype (IL6), after adjusting for the mediator (BMI). For the variant rs73093474, this indicates the direct association between the SNP and CRP after accounting for the mediator effect (WHR). ADE is a measure for biologic pleiotropy.

Figure 5.1. Sensitivity analysis of observed average mediation effect and average direct effect for variants rs60505812 (lower row) and rs73093474 (top row). The 95% confidence interval for average mediation effect with rs60505812 includes 0 value for sensitivity parameter ρ but the confidence interval for average direct effect does not. For rs73093474, both direct and mediated effects' confidence intervals do not contain 0 value, suggesting robustness of associations to unmeasured confounders.

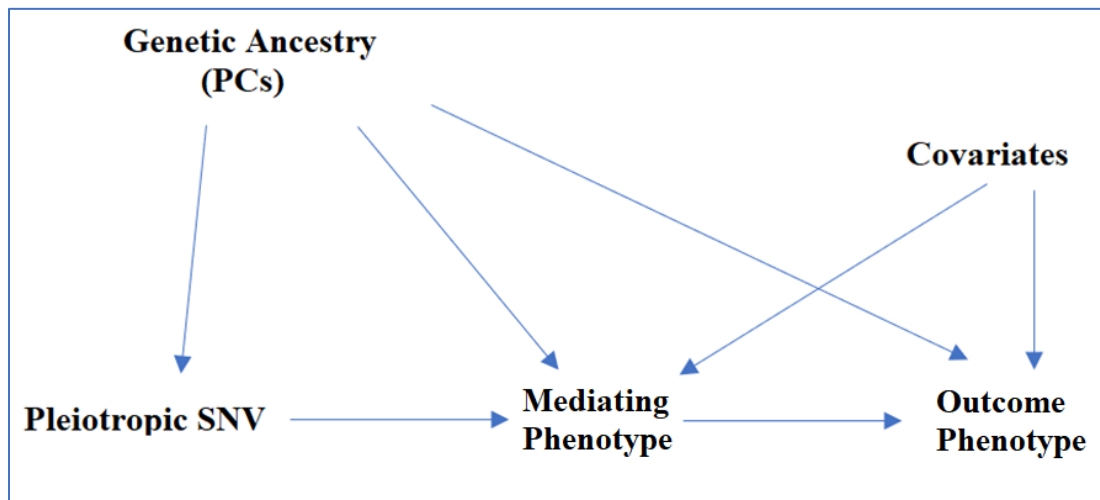


6. Appendices

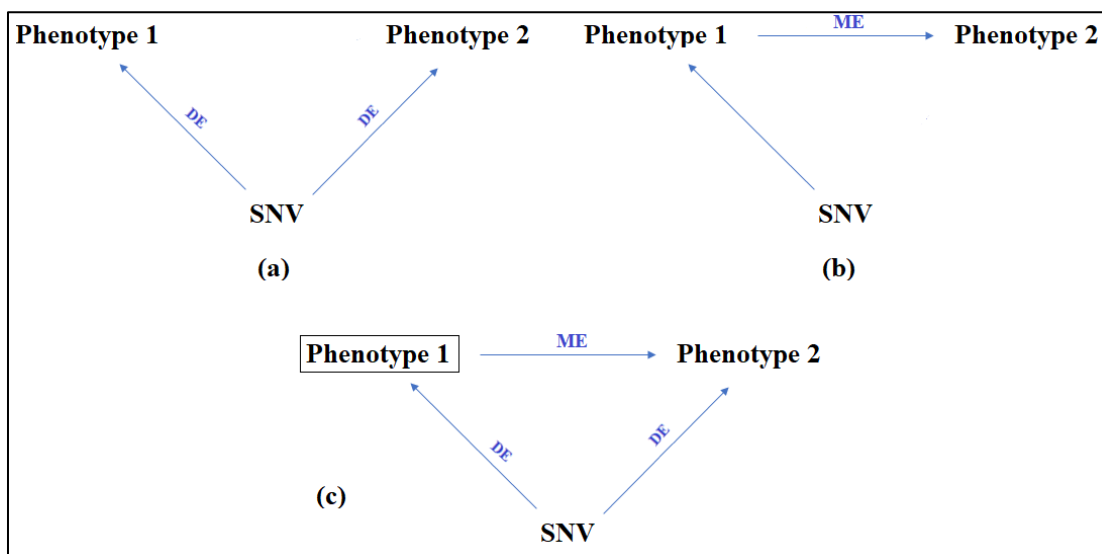
Supplementary Figure 5.1. Empirical and log-transformed distributions of inflammation markers.



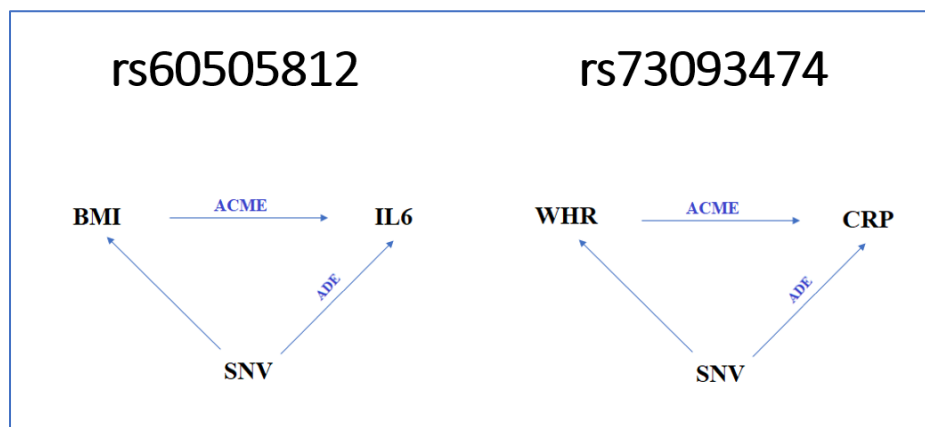
Supplementary Figure 5.2. Conceptual directed acyclic graph (DAG) used as the framework for causal mediation analysis. This approach is based on general mediation model (VanderWeele, T.J. & Vansteelandt)²⁰⁹.



Supplementary Figure 5.3. Biologic pleiotropy indicates direct association with multiple phenotypes (Figure a), while biologic pleiotropy is induced by phenotypic correlation (Figure b); multi-mode pleiotropy is also plausible (Figure c). DE: direct effect, ME: mediated effect, SNV: single nucleotide variant.



Supplementary Figure 5.4. Directed acyclic graphs (DAGs) used to formulate causal mediation analysis for variants rs60505812 and rs73093474. For variant rs60505812, BMI is proposed as the mediating phenotype and interleukin(IL)-6 as the outcome. For the remaining variant, rs60505812, waist to hip ratio (WHR) was proposed as the mediating phenotype and C-reactive protein (CRP) as the outcome.



DISCUSSION

Obesity is one of the major public health crises of our time^{378,379} and has steadily increased worldwide⁶, affecting urban and rural dwellers⁹, children and adults¹⁰, and men and women alike⁹. Many factors have been associated with the increased incidence of obesity, including increased access to cheap, high-calorie processed foods³⁸⁰, reduced physical activity³⁸¹, changes in the environment³⁸², and increased exposure to chemicals³⁸³, among other potential contributors. Increased obesity, particularly abdominal obesity¹⁹, has been associated with the increased incidence of various adverse metabolic disorders^{17,18,384} including insulin resistance²⁰, hypertensive blood pressure²¹, elevated lipid levels²², and hormonal dysregulation¹⁸. These metabolic disorders, in turn, have been associated with the increased susceptibility to cardiovascular diseases²³.

1. Inflammation-adiposity association

The underlying mechanisms that link obesity to cardiovascular diseases remain subject to debate⁵⁰⁻⁵², and certain hypotheses, including insulin resistance-induced arterial stiffness⁵³ and kidney-promoted hemodynamic imbalances⁵⁵, have received wider attention than others. However, changes in the local distribution of nutrition-sensitive⁶² adipose tissue (due to the increased accumulation of visceral fat)⁵⁹ have been associated with altered morphologic and hormonal profiles^{60,61}, and the resulting upregulation in inflammation^{68,69} is thought to be a root cause of a large proportion of metabolic disorders. An increased number (hyperplastic) of enlarged (hypertrophic) adipocytes promotes the

increased production of cytokines, including IL-6 and IL-8^{78,79}. Cytokines attract macrophages^{81,82}, which, in turn, can activate the intracellular signals responsible for upregulating the innate immune response⁸⁷ and trigger the inception of atheromatic processes that seed arterial lipid plaque formation.

In contrast, compelling evidence has also suggested that insulin resistance^{88,89} and increased circulatory⁹¹ and hypothalamic inflammatory responses^{92,93} might precede the development of obesity. An increase in inflammatory mediators has been shown to predict future obesity outcomes³⁸⁵. These seemingly contradictory observations necessitate a more nuanced exploration of the relationships between obesity and immunity.

2. Rule of genetics and ancestry in inflammation-adiposity association

Although some of these findings might merely reflect circular feedback mechanisms, the complex natures of both phenotypic domains might also indicate the involvement of other mechanisms. Changes in both adiposity and inflammation may be mediated by common pathways, a proposition that is supported by the observed overlap between genetic variants associated with both adiposity and inflammatory markers^{110,111,114-116}. The observations of increased inflammatory responses following the upregulation of adiposity gene expression^{112,113} and weight-gain associated inflammatory gene activation^{118,119} further support the idea that these pathways may share common mediators. The existence of pervasive¹²⁴ and multi-potent ‘pleiotropic genetic variants’¹²² have been hypothesized, which could provide insight into the shared biology and pathophysiological processes that underpin adverse cardiovascular events¹⁰⁷ and inform phenotypical and clinical classifications.

However, potential pleiotropic variants have not yet been sufficiently explored. Methodological issues, including difficulties distinguishing biologic pleiotropy (in which a genetic locus is independently associated with multiple phenotypes) from mediated pleiotropy (in which a genetic locus is correlated with a phenotype but may not be causal)¹²⁷, ascertainment bias due to co-occurring traits¹³⁵, phenotypic misclassifications¹³⁶, variations in allelic structures, or difficulty to identify shared controls¹³⁷, and ambiguity when a causal variant is identified as a regulatory, non-coding SNP¹³⁸, can all pose significant barriers to identifying shared pleiotropic regulators. The most important challenge has been the limited inferential utility of variants identified through GWAS for pleiotropy testing. A disproportionately large proportion of existing GWAS have been performed on largely European ancestry populations^{139,140}.

The significance of genetic architecture can be further highlighted by the differential distributions of adiposity and inflammation among distinct genetic ancestry groups that are otherwise comparable¹⁴²⁻¹⁴⁴. Hispanics/Latino (HLA) populations, in particular, have been noted to present an increased incidence of cardiometabolic abnormalities¹⁴⁵, with a diabetes prevalence that is 2.5-fold the level reported among non-Hispanic Whites (NHW)²⁵, and 2 in 3 HLA adults present with abnormal lipid levels²⁶.

Although some of this discrepancy could be explained by an increased prevalence of the obesity^{11,14}, differences in inflammation levels have persisted, even in adiposity-adjusted analyses^{142,146}. These disparities extend to the “Hispanic paradox”, in which the cardiovascular mortality rate among this population is lower than other populations, despite elevated metabolic risk factors^{27,28} and have not yet been sufficiently explained by other

contributing factors^{30,33}. These findings underscore the need to re-focus on the important but sparsely explored contribution of genetics to obesity and inflammation interaction³⁴.

3. Study object

The aims of this study were: to identify genetic loci with potential evidence of pleiotropic effects across inflammatory factors, including CRP, IL-6, IL8, and fibrinogen, overall obesity (BMI), and central adiposity (WC & WHR) domains, in an HLA population (Aim 1); to fine-map potential pleiotropic loci and distinguish likely causal variants (Aim 2); and to evaluate whether observed signals demonstrate independent associations with both phenotypic domains (i.e., biologic pleiotropy) or whether the association is induced via phenotypic correlations (i.e., mediated pleiotropy, Aim 3).

4. Cameron County Hispanic Cohort

The population examined in this study was the CCHC. Most of these participants self-identify as Mexican Americans. This cohort is characterized by a high prevalence of obesity and cardiometabolic abnormalities^{166,167}, with increased susceptibility to adverse health events. The prevalence of obesity in this population was higher than the national HLA average, as reported by the CDC for the 2017–2018 period²⁷⁵ (45% nationally vs. >75% in this study).

A total of 3,313 samples were examined in this study. The number of SNPs identified for genetic testing was >9 million. Analyses of ancestry PCs showed that the majority of PCs clustered predominantly with Mexican populations, with fewer individuals aligning with other ethnic groups.

5. Genetic signals

The GWAS analyses showed that two regions on chromosomes 6 and 8 exceeded the GWAS significance level ($p\text{-value} < 5 \times 10^{-8}$) for sex-combined meta-analyzed WHR results; the GWAS-identified significant SNPs in the novel locus are located in regions near *LRFN2* and *LOC101929555*, and have not been previously recorded in GWAS libraries. Variants in *LRFN2* have been positively associated with BMI^{257,258}, type 2 diabetes²⁵⁹, lipid profiles²⁶⁰, cognitive abilities²⁶¹, and WHR²⁶².

The absence of GWAS-identified significant associations in regions that have been widely reported to be associated with anthropometric and inflammatory traits^{263,264} may be the result of ancestry-specific patterns, similar to findings from other studies that have reported ethnic differences in genetic associations²⁶⁸⁻²⁷¹. However, caution must be exercised when making any inferences because differences in the LD structure²³⁹, statistical power²³⁸, and SNP coverage²⁷² among studies can all induce variations in the GWAS signal.

6. Evidence for pleiotropy

Pleiotropy assessments were performed using the aSPU¹⁹⁹ testing method. Although no pleiotropic signals were observed at the GWAS significance level, at least five signals, on chromosomes 1, 3, 12, 18 and 19, were identified that suggested the presence of SNPs with strong pleiotropic p-values ($< 5 \times 10^{-6}$). Although the signals on chromosomes 1 and 19 were driven by solitary variants in each instance, the suggestive loci in the remaining three chromosomes were robust to sensitivity analyses and contained clusters of SNPs with strong pleiotropy p-values.

A review of the GWAS catalogue revealed an extensive number of variants with varying degree of LD with SNPs within the three suspected pleiotropic regions that were previously reported in association with a range of metabolic traits. All three regions contained variants associated with inflammatory diseases, anthropometric measures, and energy intake. Additionally, for the loci identified on chromosome 3, a considerable number of variants were in LD with variant in the region that were previously identified for brain functions and neurological diseases.

The fine-mapping of these three suspected pleiotropic regions (on chromosomes 3, 12, and 18) was performed, incorporating the summary statistics from pleiotropy analyses, LD structures, and functional annotations with regulatory and coding features^{293,294}, which distinguished 22 variants with posterior causality probabilities greater than 50% among the 7,326 total variants assessed across all 3 regions.

None of the index SNPs, which had the most significant pleiotropy p-values, were identified among the credible causal variant sets in each region, which could be due to several possibilities, including the high correlation with several significant variants, the index SNP serving as a proxy for the true ‘causal variant’³¹², interactions with distant topologically associated domain (TAD)³¹³ variants, or the presence of multiple signals. However, a potential second (independent) signal was only observed for the suspected pleiotropic locus identified on chromosome 18.

7. Characteristics of likely pleiotropic signals

The clear majority of likely causal variants identified in each of the three regions were intronic, nonsense-mediated decay (NMD) variants, which overlapped with six different

genes, including sulfatase-modifying factor 1 (*SUMF1*) on chromosome 3, interleukin 1-associated receptor kinase (*IRAK4*) and transmembrane 117 (*TMEM117*) on chromosome 12, and serpin family B 10 (*SERPINB10*), *SERPINB2*, and histocompatibility minor serpin domain-containing (*HMSD*) on chromosome 18.

NMD variants play regulatory roles, quality-checking gene expression errors by eliminating messenger ribonucleotide (mRNA) molecules that contain premature termination codons³²⁸. Growing evidence has suggested that variations and tissue-specific differences in NMDs can alter the underlying pathologies of genetic diseases³²⁹⁻³³¹. Notably, interactions between NMDs and protein-coding variants have been linked with an increased risk of developing obesity and type 2 diabetes, as reported in one recent trans-ancestry cohort³³². For instance, in a clinical study, a patient with early-onset obesity was characterized with a mutation-induced malfunction of the NMD surveillance pathway, which affected Pro-opiomelanocortin (POMC) gene regulation; this gene is involved in a wide range of activities, including body-weight regulation³³³. The results from this study further underscored the importance of NMD variants in the context of pleiotropy.

Pleiotropy association studies cannot distinguish whether an observed association between a genetic variant and a phenotype occurs due to direct (i.e., biologic pleiotropy) or indirect effects (i.e., mediated pleiotropy) due to phenotypic correlations with the primary trait. Mediation analysis refers to a type of conditional analysis, in which the total effect between a predictor and an outcome is divided among several factors that contribute to the association and relies on counterfactual principles²⁰⁸. Using this framework, the causal effects can be viewed as the difference between two potential phenotypic outcomes, depending on whether an individual harbors the risk alleles (at the same time).

In addition to the 3 identified index SNPs, which featured the lowest pleiotropic p-values for each locus with evidence for pleiotropy, all 22 fine-mapped variants with high posterior causality probabilities were fit into univariate regression models using the phenotypes as predictors. Those variants in tighter LD with their respective index SNPs exhibited consistent and significant associations with multiple phenotypes in these univariate regressions, and those with the lowest pleiotropic p-values and regulatory features associated with adipose tissue were selected for in-depth causal mediation analysis.

The mediation analysis indicated that rs60505812 was independently associated with both an inflammatory marker (IL-6) and an adiposity measure (BMI), even after adjusting for potential confounders like age, sex, physical activity level, alcohol consumption, and current smoking status, which is consistent with the biologic pleiotropy model.

Although no direct association between *SUMF1* and metabolic traits has previously been reported, variants in this gene have been suggested to be associated with inflammatory diseases, including chronic obstructive pulmonary disease (COPD)³¹⁸. Additionally, steroid sulfates play functional roles in energy homeostasis and inflammation³¹⁹. The overexpression of human steroid sulfatase in the adipose tissue of transgenic male mice and female mice that underwent ovariectomy exacerbated metabolic phenotypes, resulting in weight gain and the accumulation of fat mass, worsened insulin sensitivity, and altered energy expenditures³⁸⁶.

8. The functional evaluation of likely pleiotropic variants

The functional assessment of rs60505812 with Haploreg demonstrated an annotation associated with histone 3 lysine 4 trimethylated (H3K4me3)³⁷⁴ in fetal adipose cell nuclei.

H3K4me3 regulates gene expression by altering the accessibility of genes to promote mRNA transcription³⁷⁵ and has generally been associated with actively transcribed genes in animal models³⁸⁷. H3K4me3 is also known to play a significant role in the genetic regulation of stem cell potency³⁸⁷. H3K4me3 has been extensively observed on nucleosomes at the 5' end of eukaryotic genes undergoing active transcription by RNA polymerase II³⁸⁸. A regulatory annotation in fetal nuclei would suggest functional associations with pathway(s) upstream to a wide range of metabolic traits, resulting in pleiotropic properties at protein or variant levels. However, further assessments remain necessary to establish the precise level at which the suggestive signal operates.

For the variant rs73093474, on chromosome 12, mediation analysis indicated both a direct association with CRP (Table 6.3) and an indirect association (via WHR) in the covariate-adjusted models, which demonstrated concurrent biologic and mediated pleiotropy. In contrast with rs60505812, rs73094374 has a regulatory annotation with the mono-methylated H3K4me1-enhancer in adipose cell nuclei, in addition to skeletal muscle and brain tissues. H3K4me1 is the primary modification used to predict whether an identified SNP is a potential enhancer (booster) of gene expression³⁷⁶.

This SNP was also an NMD variant, intronic to transmembrane 117 (*TMEM117*), which encodes a similarly named protein. This gene has been associated with endoplasmic reticulum stress (ER-stress) and mitochondria-mediated apoptotic cell death³²². Obesity-induced ER-stress has been shown to induce inflammation in adipose tissue³⁷⁷. In experimental studies examining mice, sustained fat accumulation interrupted ER homeostasis and induced the unfolded protein response (UPR)³⁸⁹. Unfolded proteins within the ER lumen activate the ER stress pathway³⁹⁰, and lipid-sensitive transmembrane

proteins act as sensors for unfolded proteins^{391,392}. Elevated levels of UPR activation have been observed in adipose tissue from obese individuals³⁹³ and has been suggested to serve as a potential mechanism that might link obesity with inflammation and insulin resistance^{394,395}.

The regulatory annotation associated with H3K4me1 combined with the suggested association with UPR-induced ER-stress activity in adipocytes suggest that the biological effects of rs73093474 (or the signals it represents) may be relatively tissue-specific, compared with those for rs60505812, and less likely to act upon multipotent pathways. Further genetic and functional studies in larger settings and varied populations are warranted before concrete biological inferences are made.

9. Study strengths and limitations

The present study had notable limitations. The sample size used for this study was small. Adiposity and inflammation traits are highly polygenic and, therefore, the effect of any given genetic variant on the overall phenotype is small, and adequate statistical power is necessary to isolate signals. GWAS performed in larger samples present considerably improved detection power compared with smaller samples³⁹⁶. The general absence of significant GWAS results in this study can largely be attributed to insufficient detection power. However, the study incorporated a pleiotropy test that served the dual aims of evaluating pleiotropy among correlated phenotypes and increasing the statistical power for locus discovery. Previous studies have illustrated^{133,397-399} that harnessing pleiotropy could successfully increase statistical power.

Although this study only included a discovery cohort, the analysis did account for multiple factors that could affect the generalizability of observed signals⁴⁰⁰, including each variant's effective sample size and ancestry architecture⁴⁰¹. Furthermore, low frequency and rare variants with MAF <5% , were excluded from the analysis to improve the likelihood of replicability.

The study could not distinguish whether suggestive pleiotropic signals acted at specific levels or were associated with multi-potent pathways that affected a wider range of phenotypic traits. Generally, trans-regulatory SNPs act at the variant level, whereas cis-regulatory and coding SNPs act at the protein (i.e., gene) and tissue levels⁴⁰². The precise characterization of pleiotropic signals requires insight from expression quantitative trait (eQTL) analysis which seeks to examine SNPs that affect the mRNA expression levels, or more relevantly candidate SNP-mRNA study. However, functional annotation evaluation was able to provide suggestive evidence for possible functional roles associated with likely causal variants. For instance, the fetal nuclei H3K4me3 annotation for rs60505812 was suggestive of potential pleiotropic effects at either variant or protein levels.

The present study also had major strengths. This study is the first comprehensive assessment of pleiotropy relationships between adiposity and inflammation traits in an HLA or any population. The results provided suggestive evidence for the regulatory effects of identified genetic variants on metabolic pathways and highlighted the complexity and interrelatedness of seemingly independent phenotypic traits.

The intentional selection of a genetically tri-admixed HLA population¹⁵¹ with shorter haplotypes compared with European ancestry (EA) populations¹⁵⁵ presented opportunities

for precise fine-mapping¹⁵⁴, which greatly reduced the number of SNPs identified for causal evaluations^{157,158}.

Rigorous association tests, which incorporated pairwise kinship structures, improved genetic associations' estimates. The integrative analytical framework that was used for this study is innovative; and the combination of various steps, including GWAS, pleiotropy assessments, fine-mapping, functional evaluations, and causal mediation analysis collectively improved the biological interpretation of the results.

10. Study summary and conclusion

The characterization of overlapping genetic pleiotropy between inflammatory and adiposity pathways is essential for improving our understanding of the etiology of the various metabolic pathways that regulate cardiovascular disease development. This research provided a systematic assessment of potential variants that exert pleiotropic effects on inflammatory and adiposity traits. The analysis distinguished a novel locus on chromosome 6 associated with WHR outcomes in sex-combined results.

The examination of 3 loci with suggested pleiotropy resulted in the isolation of 22 variants with high posterior causal probabilities, including 2 SNPs on chromosomes 3 (rs60505812) and 12 (rs73093474) that exhibited regulatory features associated with adipose tissue. The identification of likely pleiotropic variants indicated that 1) a considerable degree of overlapping genetic pleiotropy exists between adiposity and inflammation, and 2) evidence exists to support both the direct and independent pleiotropic effects of likely causal variants on both phenotypic domains, in addition to the potential mediated pleiotropy.

The results showed the potential of these genetic variants to provide biological insight, which might inform the development of preventive and clinical strategies intended to improve the cardiovascular outcomes of HLA populations, and by extension, improve targeted healthcare for all ethnic group in this era of precision medicine.

REFERENCES

1. Hatton, T.J. & Bray, B.E. Long run trends in the heights of European men, 19th–20th centuries. *Economics & Human Biology* **8**, 405-413 (2010).
2. Cole, T.J. The secular trend in human physical growth: a biological view. *Economics & Human Biology* **1**, 161-168 (2003).
3. Komlos, J. & Brabec, M. The trend of BMI values of US adults by deciles, birth cohorts 1882–1986 stratified by gender and ethnicity. *Economics & Human Biology* **9**, 234-250 (2011).
4. Gardner, G., Halweil, B. & Peterson, J.A. Underfed and overfed: the global epidemic of malnutrition. (2000).
5. Wolfenden, L., Ezzati, M., Larijani, B. & Dietz, W. The challenge for global health systems in preventing and managing obesity. *Obesity Reviews* **20**, 185-193 (2019).
6. Jaacks, L.M. *et al.* The obesity transition: stages of the global epidemic. *The lancet Diabetes & endocrinology* **7**, 231-240 (2019).
7. Finucane, M.M. *et al.* National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9· 1 million participants. *The lancet* **377**, 557-567 (2011).
8. Ng, M. *et al.* Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The lancet* **384**, 766-781 (2014).
9. Collaboration, N.R.F. Rising rural body-mass index is the main driver of the global obesity epidemic in adults. *Nature* **569**, 260 (2019).
10. Abarca-Gómez, L. *et al.* 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. *The Lancet* **390**, 2627-2642 (2017).
11. Ogden, C.L., Carroll, M.D., Kit, B.K. & Flegal, K.M. Prevalence of childhood and adult obesity in the United States, 2011-2012. *Jama* **311**, 806-814 (2014).
12. Conway, B.N. *et al.* The obesity epidemic and rising diabetes incidence in a low-income racially diverse southern US cohort. *PLoS one* **13**, e0190993 (2018).

13. Kelley, E.A. *et al.* Geography, race/ethnicity, and obesity among men in the United States. *American Journal of Men's Health* **10**, 228-236 (2016).
14. Ogden, C.L. *et al.* Trends in obesity prevalence by race and hispanic origin—1999-2000 to 2017-2018. *JAMA* **324**, 1208-1210 (2020).
15. Jiwani, S.S. *et al.* The shift of obesity burden by socioeconomic status between 1998 and 2017 in Latin America and the Caribbean: a cross-sectional series study. *The Lancet Global Health* **7**, e1644-e1654 (2019).
16. Mujica, O.J. & Victora, C.G. Obesity inequality among adults in Latin America and the Caribbean. *The Lancet Global Health* **7**, e1589-e1590 (2019).
17. Segula, D. Complications of obesity in adults: a short review of the literature. *Malawi Medical Journal* **26**, 20-24 (2014).
18. Chikunguwo, S. *et al.* Influence of obesity and surgical weight loss on thyroid hormone levels. *Surgery for Obesity and Related Diseases* **3**, 631-635 (2007).
19. Yusuf, S. *et al.* Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *The lancet* **364**, 937-952 (2004).
20. Kahn, B.B. & Flier, J.S. Obesity and insulin resistance. *The Journal of clinical investigation* **106**, 473-481 (2000).
21. Pausova, Z. From big fat cells to high blood pressure: a pathway to obesity-associated hypertension. *Current opinion in nephrology and hypertension* **15**, 173-178 (2006).
22. Howard, B.V., Ruotolo, G. & Robbins, D.C. Obesity and dyslipidemia. *Endocrinology and metabolism clinics of North America* **32**, 855 (2003).
23. Kannel, W.B. *et al.* Regional obesity and risk of cardiovascular disease; the Framingham Study. *Journal of clinical epidemiology* **44**, 183-190 (1991).
24. Fryar, C.D. *Hypertension, high serum total cholesterol, and diabetes: racial and ethnic prevalence differences in US adults, 1999-2006*, (US Department of Health and Human Services, Centers for Disease Control and ..., 2010).
25. Kanchi, R. *et al.* Gender and race disparities in cardiovascular disease risk factors among New York City adults: New York City Health and Nutrition Examination Survey (NYC HANES) 2013–2014. *Journal of Urban Health* **95**, 801-812 (2018).
26. Daviglius, M.L. *et al.* Prevalence of major cardiovascular risk factors and cardiovascular diseases among Hispanic/Latino individuals of diverse backgrounds in the United States. *Jama* **308**, 1775-1784 (2012).
27. Iribarren, C., Darbinian, J.A., Fireman, B.H. & Burchard, E.G. Birthplace and mortality among insured Latinos: the paradox revisited. *Ethnicity & disease* **19**, 185 (2009).

28. Willey, J.Z. *et al.* Coronary death and myocardial infarction among Hispanics in the Northern Manhattan Study: exploring the Hispanic paradox. *Annals of epidemiology* **22**, 303-309 (2012).
29. Beydoun, M. *et al.* Racial disparities in adult all-cause and cause-specific mortality among us adults: mediating and moderating factors. *BMC public health* **16**, 1113 (2016).
30. Abraido-Lanza, A.F., Dohrenwend, B.P., Ng-Mak, D.S. & Turner, J.B. The Latino mortality paradox: a test of the "salmon bias" and healthy migrant hypotheses. *American journal of public health* **89**, 1543-1548 (1999).
31. Medina-Inojosa, J., Jean, N., Cortes-Bergoderi, M. & Lopez-Jimenez, F. The Hispanic paradox in cardiovascular disease and total mortality. *Progress in cardiovascular diseases* **57**, 286-292 (2014).
32. Valles, S.A. The challenges of choosing and explaining a phenomenon in epidemiological research on the "Hispanic Paradox". *Theoretical Medicine and Bioethics* **37**, 129-148 (2016).
33. Beltrán-Sánchez, H., Palloni, A., Riosmena, F. & Wong, R. SES gradients among Mexicans in the United States and in Mexico: A new twist to the hispanic paradox? *Demography* **53**, 1555-1581 (2016).
34. Qi, L. & Campos, H. Genetic predictors for cardiovascular disease in hispanics. *Trends in cardiovascular medicine* **21**, 15-20 (2011).
35. Ferrannini, E. Physiological and metabolic consequences of obesity. *Metabolism* **44**, 15-17 (1995).
36. Lamvu, G., Zolnoun, D., Boggess, J. & Steege, J.F. Obesity: physiologic changes and challenges during laparoscopy. *American journal of obstetrics and gynecology* **191**, 669-674 (2004).
37. Parameswaran, K., Todd, D.C. & Soth, M. Altered respiratory physiology in obesity. *Canadian respiratory journal* **13**(2006).
38. Hall, J.E. *et al.* Obesity-induced hypertension: role of sympathetic nervous system, leptin, and melanocortins. *Journal of Biological Chemistry* **285**, 17271-17276 (2010).
39. Mertens, I.L. & Van Gaal, L.F. Overweight, obesity, and blood pressure: the effects of modest weight reduction. *Obesity research* **8**, 270-278 (2000).
40. Chagnac, A. *et al.* Obesity-induced glomerular hyperfiltration: its involvement in the pathogenesis of tubular sodium reabsorption. *Nephrology Dialysis Transplantation* **23**, 3946-3952 (2008).
41. Marti, A., Marcos, A. & Martinez, J.A. Obesity and immune function relationships. *Obesity reviews* **2**, 131-140 (2001).

42. Forny-Germano, L., De Felice, F.G. & Vieira, M.N.d.N. The role of leptin and adiponectin in obesity-associated cognitive decline and Alzheimer's disease. *Frontiers in neuroscience* **12**, 1027 (2019).
43. Meldrum, D.R. Introduction: Obesity and reproduction. *Fertility and Sterility* **107**, 831-832 (2017).
44. Fabbrini, E., Sullivan, S. & Klein, S. Obesity and nonalcoholic fatty liver disease: biochemical, metabolic, and clinical implications. *Hepatology* **51**, 679-689 (2010).
45. Hall, J.E. Pathophysiology of obesity hypertension. *Current hypertension reports* **2**, 139-147 (2000).
46. Wilson, P.W., D'Agostino, R.B., Sullivan, L., Parise, H. & Kannel, W.B. Overweight and obesity as determinants of cardiovascular risk: the Framingham experience. *Archives of internal medicine* **162**, 1867-1872 (2002).
47. Jiang, J., Ahn, J., Huang, W.-Y. & Hayes, R.B. Association of obesity with cardiovascular disease mortality in the PLCO trial. *Preventive medicine* **57**, 60-64 (2013).
48. Elagizi, A. *et al.* An overview and update on obesity and the obesity paradox in cardiovascular diseases. *Progress in cardiovascular diseases* **61**, 142-150 (2018).
49. Heron, M.P. Deaths: leading causes for 2017. (2019).
50. Van Gaal, L.F., Mertens, I.L. & Christophe, E. Mechanisms linking obesity with cardiovascular disease. *Nature* **444**, 875 (2006).
51. Kim, S.H., Després, J.-P. & Koh, K.K. Obesity and cardiovascular disease: friend or foe? *European heart journal* **37**, 3560-3568 (2015).
52. Neeland, I.J., Poirier, P. & Després, J.-P. Cardiovascular and metabolic heterogeneity of obesity: clinical challenges and implications for management. *Circulation* **137**, 1391-1406 (2018).
53. Shah, A.S. *et al.* Insulin sensitivity and arterial stiffness in youth with type 1 diabetes: the SEARCH CVD study. *Journal of Diabetes and its Complications* **29**, 512-516 (2015).
54. Ekelund, U. & Mellander, S. Endogenous nitric oxide as a physiological regulator of vascular tone in cat skeletal muscle during haemorrhage. *Acta physiologica scandinavica* **157**, 471-479 (1996).
55. Hall, J.E. The kidney, hypertension, and obesity. *Hypertension* **41**, 625-633 (2003).
56. Hall, J.E., Mizelle, H.L., Hildebrandt, D.A. & Brands, M.W. Abnormal pressure natriuresis. A cause or a consequence of hypertension? *Hypertension* **15**, 547-559 (1990).
57. Hall, J.E. Mechanisms of abnormal renal sodium handling in obesity hypertension. *American journal of hypertension* **10**, 49S-55S (1997).

58. Hall, J.E., Brands, M.W., Henegar, J.R. & Shek, E.W. Abnormal kidney function as a cause and a consequence of obesity hypertension. *Clinical and experimental pharmacology and physiology* **25**, 58-64 (1998).
59. Peiris, A.N. *et al.* Adiposity, fat distribution, and cardiovascular risk. *Annals of Internal Medicine* **110**, 867-872 (1989).
60. Benoit, S.C., Clegg, D.J., Seeley, R.J. & Woods, S.C. Insulin and leptin as adiposity signals. *Recent progress in hormone research* **59**, 267-286 (2004).
61. Katagiri, H., Yamada, T. & Oka, Y. Adiposity and cardiovascular disorders: disturbance of the regulatory system consisting of humoral and neuronal signals. *Circulation research* **101**, 27-39 (2007).
62. Longo, M. *et al.* Adipose tissue dysfunction as determinant of obesity-associated metabolic complications. *International Journal of molecular sciences* **20**, 2358 (2019).
63. Abe, T., Sakurai, T., Kurata, J., Kawakami, Y. & Fukunaga, T. Subcutaneous and visceral fat distribution and daily physical activity: comparison between young and middle aged women. *British journal of sports medicine* **30**, 297-300 (1996).
64. Schleinitz, D., Böttcher, Y., Blüher, M. & Kovacs, P. The genetics of fat distribution. *Diabetologia* **57**, 1276-1286 (2014).
65. Frugé, A.D., Cases, M.G., Schildkraut, J.M. & Demark-Wahnefried, W. Associations between obesity, body fat distribution, weight loss and weight cycling on serum pesticide concentrations. *Journal of food & nutritional disorders* **5**(2016).
66. Unger, R.H. Longevity, lipotoxicity and leptin: the adipocyte defense against feasting and famine. *Biochimie* **87**, 57-64 (2005).
67. Virtue, S. & Vidal-Puig, A. Adipose tissue expandability, lipotoxicity and the metabolic syndrome—an allostatic perspective. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* **1801**, 338-349 (2010).
68. Schmidt, M.I. *et al.* Clustering of dyslipidemia, hyperuricemia, diabetes, and hypertension and its association with fasting insulin and central and overall obesity in a general population. *Metabolism* **45**, 699-706 (1996).
69. Jafar, T.H., Chaturvedi, N. & Pappas, G. Prevalence of overweight and obesity and their association with hypertension and diabetes mellitus in an Indo-Asian population. *Cmaj* **175**, 1071-1077 (2006).
70. WINEGRAD, A.I., SHAW, W.N., Lukens, F. & STADIE, W.C. Lipogenesis in adipose tissue. *The American Journal of Clinical Nutrition* **8**, 651-665 (1960).
71. Jo, J. *et al.* Hypertrophy and/or hyperplasia: dynamics of adipose tissue growth. *PLoS Comput Biol* **5**, e1000324 (2009).

72. Shimizu, H. & Mori, M. The brain–adipose axis: a review of involvement of molecules. *Nutritional neuroscience* **8**, 7-20 (2005).
73. Laurencikiene, J. *et al.* Regulation of lipolysis in small and large fat cells of the same subject. *The Journal of Clinical Endocrinology & Metabolism* **96**, E2045-E2049 (2011).
74. Wang, Z. *et al.* Leptin resistance of adipocytes in obesity: role of suppressors of cytokine signaling. *Biochemical and biophysical research communications* **277**, 20-26 (2000).
75. Brennan, A.M. & Mantzoros, C.S. Drug Insight: the role of leptin in human physiology and pathophysiology—emerging clinical applications. *Nature clinical practice Endocrinology & metabolism* **2**, 318-327 (2006).
76. Koh, E.H. *et al.* 11b-HSD1 reduces metabolic efficacy and adiponectin synthesis in hypertrophic adipocytes. *Journal of Endocrinology* **225**, 147-158 (2015).
77. Samad, F. & Ruf, W. Inflammation, obesity, and thrombosis. *Blood, The Journal of the American Society of Hematology* **122**, 3415-3422 (2013).
78. Blackburn, P. *et al.* Postprandial variations of plasma inflammatory markers in abdominally obese men. *Obesity* **14**, 1747-1754 (2006).
79. Gustafson, B. Adipose tissue, inflammation and atherosclerosis. *Journal of atherosclerosis and thrombosis* **17**, 332-341 (2010).
80. Xu, L.L., Warren, M.K., Rose, W.L., Gong, W. & Wang, J.M. Human recombinant monocyte chemotactic protein and other C-C chemokines bind and induce directional migration of dendritic cells in vitro. *Journal of leukocyte biology* **60**, 365-371 (1996).
81. Nishimura, S. *et al.* CD8+ effector T cells contribute to macrophage recruitment and adipose tissue inflammation in obesity. *Nature medicine* **15**, 914-920 (2009).
82. Weisberg, S.P. *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *The Journal of clinical investigation* **112**, 1796-1808 (2003).
83. Mathieu, P., Lemieux, I. & Després, J.P. Obesity, inflammation, and cardiovascular risk. *Clinical Pharmacology & Therapeutics* **87**, 407-416 (2010).
84. Abeywardena, M.Y., Leifert, W.R., Warnes, K.E., Varghese, J.N. & Head, R.J. Cardiovascular biology of interleukin-6. *Current pharmaceutical design* **15**, 1809-1821 (2009).
85. Wang, A. *et al.* Cumulative Exposure to High-Sensitivity C-Reactive Protein Predicts the Risk of Cardiovascular Disease. *Journal of the American Heart Association* **6**, e005610 (2017).
86. Mathieu, P., Pibarot, P. & Després, J.-P. Metabolic syndrome: the danger signal in atherosclerosis. *Vascular health and risk management* **2**, 285 (2006).

87. Vaure, C. & Liu, Y. A comparative review of toll-like receptor 4 expression and functionality in different animal species. *Frontiers in immunology* **5**, 316 (2014).
88. Blüher, M. Adipose tissue inflammation: a cause or consequence of obesity-related insulin resistance? *Clinical science* **130**, 1603-1614 (2016).
89. Goodpaster, B.H., Katsiaras, A. & Kelley, D.E. Enhanced fat oxidation through physical activity is associated with improvements in insulin sensitivity in obesity. *Diabetes* **52**, 2191-2197 (2003).
90. Reilly, S.M. & Saltiel, A.R. Adapting to obesity with adipose tissue inflammation. *Nature Reviews Endocrinology* **13**, 633-643 (2017).
91. Engström, G. *et al.* Inflammation-sensitive plasma proteins are associated with future weight gain. *Diabetes* **52**, 2097-2101 (2003).
92. De Git, K. & Adan, R. Leptin resistance in diet-induced obesity: the role of hypothalamic inflammation. *Obesity reviews* **16**, 207-224 (2015).
93. Scarpace, P.J. & Zhang, Y. Leptin resistance: a predisposing factor for diet-induced obesity. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **296**, R493-R500 (2009).
94. Kim, K.-A., Gu, W., Lee, I.-A., Joh, E.-H. & Kim, D.-H. High fat diet-induced gut microbiota exacerbates inflammation and obesity in mice via the TLR4 signaling pathway. *PloS one* **7**, e47713 (2012).
95. Ding, S. *et al.* High-fat diet: bacteria interactions promote intestinal inflammation which precedes and correlates with obesity and insulin resistance in mouse. *PloS one* **5**, e12191 (2010).
96. Liu, L., Mei, M., Yang, S. & Li, Q. Roles of chronic low-grade inflammation in the development of ectopic fat deposition. *Mediators of inflammation* **2014**(2014).
97. de Heredia, F.P., Gómez-Martínez, S. & Marcos, A. Obesity, inflammation and the immune system. *Proceedings of the Nutrition Society* **71**, 332-338 (2012).
98. Rick, C.M. Tomato paste: a concentrated review of genetic highlights from the beginnings to the advent of molecular genetics. *Genetics* **128**, 1 (1991).
99. Lees, C., Barrett, J., Parkes, M. & Satsangi, J. New IBD genetics: common pathways with other diseases. *Gut* **60**, 1739-1753 (2011).
100. Pearson, T.A. & Manolio, T.A. How to interpret a genome-wide association study. *Jama* **299**, 1335-1344 (2008).
101. Chang, D. *et al.* Accounting for eXentricities: analysis of the X chromosome in GWAS reveals X-linked genes implicated in autoimmune diseases. *PloS one* **9**, e113684 (2014).

102. Lindgren, C.M. *et al.* Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS genetics* **5**, e1000508 (2009).
103. Yoneyama, S. *et al.* Gene-centric meta-analyses for central adiposity traits in up to 57 412 individuals of European descent confirm known loci and reveal several novel associations. *Human molecular genetics* **23**, 2498-2510 (2013).
104. Heard-Costa, N.L. *et al.* NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. *PLoS genetics* **5**, e1000539 (2009).
105. Ligthart, S. *et al.* Genome analyses of > 200,000 individuals identify 58 loci for chronic inflammation and highlight pathways that link inflammation and complex disorders. *The American Journal of Human Genetics* **103**, 691-706 (2018).
106. Plourde, M. *et al.* A variant in the LRRFIP1 gene is associated with adiposity and inflammation. *Obesity* **21**, 185-192 (2013).
107. Ligthart, S. *et al.* Bivariate genome-wide association study identifies novel pleiotropic loci for lipids and inflammation. *BMC genomics* **17**, 443 (2016).
108. Naitza, S. *et al.* A genome-wide association scan on the levels of markers of inflammation in Sardinians reveals associations that underpin its complex regulation. *PLoS genetics* **8**, e1002480 (2012).
109. Sleiman, P.M. *et al.* GWAS identifies four novel eosinophilic esophagitis loci. *Nature communications* **5**, 5593 (2014).
110. Hallstrand, T.S. *et al.* Genetic pleiotropy between asthma and obesity in a community-based sample of twins. *Journal of Allergy and Clinical Immunology* **116**, 1235-1241 (2005).
111. Park, E.J. *et al.* Dietary and genetic obesity promote liver inflammation and tumorigenesis by enhancing IL-6 and TNF expression. *Cell* **140**, 197-208 (2010).
112. Soukas, A., Cohen, P., Socci, N.D. & Friedman, J.M. Leptin-specific patterns of gene expression in white adipose tissue. *Genes & development* **14**, 963-980 (2000).
113. Way, J.M. *et al.* Comprehensive messenger ribonucleic acid profiling reveals that peroxisome proliferator-activated receptor γ activation has coordinate effects on gene expression in multiple insulin-sensitive tissues. *Endocrinology* **142**, 1269-1277 (2001).
114. Bego, T. *et al.* Association of FTO gene variant (rs8050136) with type 2 diabetes and markers of obesity, glycaemic control and inflammation. *Journal of Medical Biochemistry* **38**, 153-163 (2019).
115. Fisher, E. *et al.* Association of the FTO rs9939609 single nucleotide polymorphism with C-reactive protein levels. *Obesity* **17**, 330-334 (2009).

116. Chung, S.J. *et al.* ADIPOQ/adiponectin induces cytotoxic autophagy in breast cancer cells through STK11/LKB1-mediated activation of the AMPK-ULK1 axis. *Autophagy* **13**, 1386-1403 (2017).
117. Vordenbäumen, S. *et al.* Human casein alpha s1 induces proinflammatory cytokine expression in monocytic cells by TLR4 signaling. *Molecular nutrition & food research* **60**, 1079-1089 (2016).
118. Popko, K., Gorska, E. & Demkow, U. Influence of interleukin-6 and G174C polymorphism in IL-6 gene on obesity and energy balance. *European Journal of Medical Research* **15**, 1-5 (2010).
119. Joffe, Y.T., Collins, M. & Goedecke, J.H. The relationship between dietary fatty acids and inflammatory genes on the obese phenotype and serum lipids. *Nutrients* **5**, 1672-1705 (2013).
120. Huang, R.-C. *et al.* Synergy between adiposity, insulin resistance, metabolic risk factors, and inflammation in adolescents. *Diabetes Care* **32**, 695-701 (2009).
121. Lessard, A. *et al.* Adiposity and pulmonary function: relationship with body fat distribution and systemic inflammation. *Clinical and Investigative Medicine*, E64-E70 (2011).
122. Stearns, F.W. One hundred years of pleiotropy: a retrospective. *Genetics* **186**, 767-773 (2010).
123. Hill, W.G. & Zhang, X.-S. On the pleiotropic structure of the genotype–phenotype map and the evolvability of complex organisms. *Genetics* **190**, 1131-1137 (2012).
124. Gratten, J. & Visscher, P.M. Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. *Genome medicine* **8**, 78 (2016).
125. Hormozdiari, F., Kostem, E., Kang, E.Y., Pasaniuc, B. & Eskin, E. Identifying causal variants at loci with multiple signals of association. *Genetics* **198**, 497-508 (2014).
126. Visscher, P.M. *et al.* 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics* **101**, 5-22 (2017).
127. Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M. & Smoller, J.W. Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics* **14**, 483-495 (2013).
128. Wong, K.-C. *Big data analytics in genomics*, (Springer, 2016).
129. Eichler, E.E. *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews Genetics* **11**, 446-450 (2010).
130. Vattikuti, S., Guo, J. & Chow, C.C. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLoS Genet* **8**, e1002637 (2012).

131. Berndt, S.I. *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nature genetics* **45**, 501-512 (2013).
132. Andreassen, O.A. *et al.* Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *The American Journal of Human Genetics* **92**, 197-209 (2013).
133. Chung, D., Yang, C., Li, C., Gelernter, J. & Zhao, H. GPA: a statistical approach to prioritizing GWAS results by integrating pleiotropy and annotation. *PLoS Genet* **10**, e1004787 (2014).
134. Price, A.L., Spencer, C.C. & Donnelly, P. Progress and promise in understanding the genetic basis of common diseases. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20151684 (2015).
135. Smoller, J.W., Lunetta, K.L. & Robins, J. Implications of comorbidity and ascertainment bias for identifying disease genes. *American journal of medical genetics* **96**, 817-822 (2000).
136. Wray, N.R., Lee, S.H. & Kendler, K.S. Impact of diagnostic misclassification on estimation of genetic correlations using genome-wide genotypes. *European Journal of Human Genetics* **20**, 668-674 (2012).
137. Lin, D.-Y. & Sullivan, P.F. Meta-analysis of genome-wide association studies with overlapping subjects. *The American Journal of Human Genetics* **85**, 862-872 (2009).
138. Broekema, R., Bakker, O. & Jonkers, I. A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. *Open biology* **10**, 190221 (2020).
139. Popejoy, A.B. & Fullerton, S.M. Genomics is failing on diversity. *Nature News* **538**, 161 (2016).
140. Landry, L.G., Ali, N., Williams, D.R., Rehm, H.L. & Bonham, V.L. Lack of diversity in genomic databases is a barrier to translating precision medicine research into practice. *Health Affairs* **37**, 780-785 (2018).
141. Wojcik, G.L. *et al.* Genetic diversity turns a new PAGE in our understanding of complex traits. *BioRxiv*, 188094 (2017).
142. Khan, U.I. *et al.* Race–ethnic differences in adipokine levels: the Study of Women's Health Across the Nation (SWAN). *Metabolism* **61**, 1261-1269 (2012).
143. Nazare, J.-A. *et al.* Ethnic influences on the relations between abdominal subcutaneous and visceral adiposity, liver fat, and cardiometabolic risk profile: the International Study of Prediction of Intra-Abdominal Adiposity and Its Relationship With Cardiometabolic Risk/Intra-Abdominal Adiposity. *The American journal of clinical nutrition* **96**, 714-726 (2012).
144. Lê, K.-A. *et al.* Ethnic differences in pancreatic fat accumulation and its relationship with other fat depots and inflammatory markers. *Diabetes care* **34**, 485-490 (2011).

145. Frank, A.T. *et al.* Racial/ethnic differences in dyslipidemia patterns. *Circulation* **129**, 570-579 (2014).
146. Achilike, I. The University of Texas School of Public Health (2013).
147. Lu, Y. *et al.* New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nature communications* **7**, 1-15 (2016).
148. Evans, J. *et al.* Depot-and ethnic-specific differences in the relationship between adipose tissue inflammation and insulin sensitivity. *Clinical endocrinology* **74**, 51-59 (2011).
149. Vashi, N. *et al.* Genetic markers of inflammation may not contribute to metabolic traits in Mexican children. *PeerJ* **4**, e2090 (2016).
150. Comuzzie, A.G. *et al.* Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PloS one* **7**, e51954 (2012).
151. Conomos, M.P. *et al.* Genetic diversity and association studies in US Hispanic/Latino populations: applications in the Hispanic Community Health Study/Study of Latinos. *The American Journal of Human Genetics* **98**, 165-184 (2016).
152. Swenson, B.R. *et al.* GWAS of QRS duration identifies new loci specific to Hispanic/Latino populations. *PloS one* **14**, e0217796 (2019).
153. Schick, U.M. *et al.* Genome-wide association study of platelet count identifies ancestry-specific loci in Hispanic/Latino Americans. *The American Journal of Human Genetics* **98**, 229-242 (2016).
154. Wu, Y. *et al.* Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. *PLoS Genet* **9**, e1003379 (2013).
155. Consortium, G.P. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
156. Asimit, J.L., Hatzikotoulas, K., McCarthy, M., Morris, A.P. & Zeggini, E. Trans-ethnic study design approaches for fine-mapping. *European journal of human genetics* **24**, 1330-1336 (2016).
157. Van de Bunt, M. *et al.* Evaluating the performance of fine-mapping strategies at common variant GWAS loci. *PLoS Genet* **11**, e1005535 (2015).
158. Kichaev, G. & Pasaniuc, B. Leveraging functional-annotation data in trans-ethnic fine-mapping studies. *The American Journal of Human Genetics* **97**, 260-271 (2015).
159. Kraja, A.T. *et al.* Pleiotropic genes for metabolic syndrome and inflammation. *Molecular genetics and metabolism* **112**, 317-338 (2014).

160. Kim, J., Bai, Y. & Pan, W. An adaptive association test for multiple phenotypes with GWAS summary statistics. *Genetic epidemiology* **39**, 651-663 (2015).
161. Kichaev, G. *et al.* Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics* (2016).
162. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493-1501 (2016).
163. Colby, S.L. & Ortman, J.M. Projections of the Size and Composition of the U.S. Population: 2014 to 2060. (United States Census Bureau, 2015).
164. AR, S.J., Pérez, A., Brown, H. & Reininger, B. Socioeconomic Status and Prevalence of Obesity and Diabetes in a Mexican American Community, Cameron County, Texas, 2004-2007. *Center for Disease Control and Prevention* **7**(2010).
165. Watt, G.P. *et al.* Peer Reviewed: The Precarious Health of Young Mexican American Men in South Texas, Cameron County Hispanic Cohort, 2004–2015. *Preventing chronic disease* **13**(2016).
166. Fisher-Hoch, S.P., Vatcheva, K.P., Rahbar, M.H. & McCormick, J.B. Undiagnosed diabetes and pre-diabetes in health disparities. *PLoS One* **10**, e0133135 (2015).
167. Laing, S.T. *et al.* High Prevalence of Subclinical Atherosclerosis by Carotid Ultrasound among Mexican Americans: Discordance with 10-Year Risk Assessment using the Framingham Risk Score. *Echocardiography* **29**, 1224-1232 (2012).
168. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, s13742-015-0047-8 (2015).
169. Leutenegger, A.-L. *et al.* Estimation of the inbreeding coefficient through use of genomic data. *The American Journal of Human Genetics* **73**, 516-523 (2003).
170. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-2873 (2010).
171. Staples, J., Nickerson, D.A. & Below, J.E. Utilizing graph theory to select the largest set of unrelated individuals for genetic analysis. *Genetic epidemiology* **37**, 136-141 (2013).
172. Turner, S. *et al.* Quality control procedures for genome-wide association studies. *Current protocols in human genetics* **68**, 1.19. 1-1.19. 18 (2011).
173. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS genet* **2**, e190 (2006).
174. Conomos, M.P. & Thornton, T. GENetic ESTimation and inference in structured samples (GENESIS): statistical methods for analyzing genetic data from samples with population structure and/or relatedness. *R package version* **2**(2016).

175. Conomos, M.P., Miller, M.B. & Thornton, T.A. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genetic epidemiology* **39**, 276-293 (2015).
176. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nature genetics* **48**, 1284-1287 (2016).
177. Das, S., Abecasis, G. & Fuchsberger, C. Minimac4: A next generation imputation tool for mega reference panels;(Abstract# 1278W). in *the 65th Annual Meeting of the American Society of Human Genetics* (2015).
178. Loh, P.-R., Palamara, P.F. & Price, A.L. Fast and accurate long-range phasing in a UK Biobank cohort. *Nature genetics* **48**, 811-816 (2016).
179. Consortium, G.R. Genome Reference Consortium Human Build 38 patch release 13 (GRCh38. p13). NCBI https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405 **39**(2019).
180. Manco, M. & Dallapiccola, B. Genetics of pediatric obesity. *Pediatrics* **130**, 123-133 (2012).
181. Randall, J.C. *et al.* Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet* **9**, e1003500 (2013).
182. Landry, A., Docherty, P., Ouellette, S. & Cartier, L.J. Causes and outcomes of markedly elevated C-reactive protein levels. *Canadian Family Physician* **63**, e316-e323 (2017).
183. da Silva Krause, M. *et al.* Physiological concentrations of interleukin-6 directly promote insulin secretion, signal transduction, nitric oxide release, and redox status in a clonal pancreatic-cell line and mouse islets. *Journal of Endocrinology* **214**, 301 (2012).
184. Beharka, A.A. *et al.* Interleukin-6 production does not increase with age. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **56**, B81-B88 (2001).
185. Klok, A.-M. *et al.* Elevated serum IL-8 levels are associated with disease activity in idiopathic intermediate uveitis. *British journal of ophthalmology* **82**, 871-874 (1998).
186. Arican, O., Aral, M., Sasmaz, S. & Ciragil, P. Serum levels of TNF- α , IFN- γ , IL-6, IL-8, IL-12, IL-17, and IL-18 in patients with active psoriasis and correlation with disease severity. *Mediators of inflammation* **2005**(2005).
187. Zhang, J. & Bai, C. Elevated serum interleukin-8 level as a preferable biomarker for identifying uncontrolled asthma and glucocorticosteroid responsiveness. *Tanaffos* **16**, 260 (2017).
188. Beasley, T.M., Erickson, S. & Allison, D.B. Rank-based inverse normal transformations are increasingly used, but are they merited? *Behavior genetics* **39**, 580 (2009).
189. McCullagh, P. *Generalized linear models*, (Routledge, 2018).

190. Gogarten, S.M. *et al.* Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics* **35**, 5346-5348 (2019).
191. Kang, H.M. *et al.* Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709-1723 (2008).
192. Aulchenko, Y.S., De Koning, D.-J. & Haley, C. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* **177**, 577-585 (2007).
193. Kang, H.M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nature genetics* **42**, 348-354 (2010).
194. Engle, R.F. Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. *Handbook of econometrics* **2**, 775-826 (1984).
195. Gilmour, A.R., Thompson, R. & Cullis, B.R. Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics*, 1440-1450 (1995).
196. Conomos, M.P., Reiner, A.P., Weir, B.S. & Thornton, T.A. Model-free estimation of recent genetic relatedness. *The American Journal of Human Genetics* **98**, 127-148 (2016).
197. Hackinger, S. & Zeggini, E. Statistical methods to detect pleiotropy in human complex traits. *Open biology* **7**, 170125 (2017).
198. Schaid, D.J. *et al.* Statistical methods for testing genetic pleiotropy. *Genetics* **204**, 483-497 (2016).
199. Xu, Z., Pan, W. & Initiative, A.s.D.N. Approximate score-based testing with application to multivariate trait association analysis. *Genetic epidemiology* **39**, 469-479 (2015).
200. Conneely, K.N. & Boehnke, M. So many correlated tests, so little time! Rapid adjustment of P values for multiple correlated tests. *The American Journal of Human Genetics* **81**, 1158-1168 (2007).
201. Basu, S. & Pan, W. Comparison of statistical tests for disease association with rare variants. *Genetic epidemiology* **35**, 606-619 (2011).
202. Pan, W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society* **33**, 497-507 (2009).
203. Pan, W., Kim, J., Zhang, Y., Shen, X. & Wei, P. A powerful and adaptive association test for rare variants. *Genetics* **197**, 1081-1095 (2014).
204. Han, F. & Pan, W. A data-adaptive sum test for disease association with multiple common or rare variants. *Human heredity* **70**, 42-54 (2010).

205. Gusev, A. *et al.* Quantifying missing heritability at known GWAS loci. *PLoS Genet* **9**, e1003993 (2013).
206. Kichaev, G. *et al.* Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet* **10**, e1004722 (2014).
207. Hans, C., Dobra, A. & West, M. Shotgun stochastic search for “large p” regression. *Journal of the American Statistical Association* **102**, 507-516 (2007).
208. Pearl, J. Probabilities of causation: three counterfactual interpretations and their identification. *Synthese* **121**, 93-149 (1999).
209. VanderWeele, T.J. & Vansteelandt, S. Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface* **2**, 457-468 (2009).
210. Robins, J.M. & Greenland, S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 143-155 (1992).
211. Imai, K., Keele, L. & Tingley, D. A general approach to causal mediation analysis. *Psychological methods* **15**, 309 (2010).
212. Forastiere, L., Mattei, A. & Ding, P. Principal ignorability in mediation analysis: through and beyond sequential ignorability. *Biometrika* **105**, 979-986 (2018).
213. Rana, J.S. *et al.* Inflammatory biomarkers, physical activity, waist circumference, and risk of future coronary heart disease in healthy men and women. *European Heart Journal* **32**, 336-344 (2011).
214. Sokol, A. *et al.* Association between the dietary inflammatory index, waist-to-hip ratio and metabolic syndrome. *Nutrition research* **36**, 1298-1303 (2016).
215. Brooks, G.C., Blaha, M.J. & Blumenthal, R.S. Relation of C-reactive protein to abdominal adiposity. *The American journal of cardiology* **106**, 56-61 (2010).
216. Ligthart, S. *et al.* Pleiotropy among common genetic loci identified for cardiometabolic disorders and C-reactive protein. *PLoS one* **10**, e0118859 (2015).
217. Qi, L., Zhang, C., van Dam, R.M. & Hu, F.B. Interleukin-6 genetic variability and adiposity: associations in two prospective cohorts and systematic review in 26,944 individuals. *The Journal of Clinical Endocrinology & Metabolism* **92**, 3618-3625 (2007).
218. Bruun, J.M. *et al.* Higher production of IL-8 in visceral vs. subcutaneous adipose tissue. Implication of nonadipose cells in adipose tissue. *American Journal of Physiology-Endocrinology and Metabolism* **286**, E8-E13 (2004).
219. Meilahn, E.N. *et al.* Association of sex hormones and adiposity with plasma levels of fibrinogen and PAI-1 in postmenopausal women. *American journal of epidemiology* **143**, 159-166 (1996).

220. Luna, J.M. *et al.* High-sensitivity C-reactive protein and interleukin-6–dominant inflammation and ischemic stroke risk: the Northern Manhattan Study. *Stroke* **45**, 979-987 (2014).
221. Vella, C.A. *et al.* Associations of insulin resistance with cardiovascular risk factors and inflammatory cytokines in normal-weight Hispanic women. *Diabetes Care* **36**, 1377-1383 (2013).
222. Kayser, B.D., Toledo-Corral, C.M., Alderete, T.L., Weigensberg, M.J. & Goran, M.I. Temporal relationships between adipocytokines and diabetes risk in Hispanic adolescents with obesity. *Obesity* **23**, 1479-1485 (2015).
223. Mirza, S. *et al.* Type 2-diabetes is associated with elevated levels of TNF-alpha, IL-6 and adiponectin and low levels of leptin in a population of Mexican Americans: a cross-sectional study. *Cytokine* **57**, 136-142 (2012).
224. Kritchevsky, S.B., Bush, A.J., Pahor, M. & Gross, M.D. Serum carotenoids and markers of inflammation in nonsmokers. *American journal of epidemiology* **152**, 1065-1071 (2000).
225. McBane II, R.D., Hardison, R.M., Sobel, B.E. & Group, B.D.S. Comparison of plasminogen activator inhibitor-1, tissue type plasminogen activator antigen, fibrinogen, and D-dimer levels in various age decades in patients with type 2 diabetes mellitus and stable coronary artery disease (from the BARI 2D trial). *The American journal of cardiology* **105**, 17-24 (2010).
226. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~ 700000 individuals of European ancestry. *Human molecular genetics* **27**, 3641-3649 (2018).
227. Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature genetics* **42**, 949-960 (2010).
228. Wen, W. *et al.* Genome-wide association studies in East Asians identify new loci for waist-hip ratio and waist circumference. *Scientific reports* **6**, 17958 (2016).
229. Liu, C.-T. *et al.* Genome-wide association of body fat distribution in African ancestry populations suggests new loci. *PLoS Genet* **9**, e1003681 (2013).
230. Liu, J.Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature genetics* **47**, 979-986 (2015).
231. Saucedo, R. *et al.* Gene variants in the FTO gene are associated with adiponectin and TNF-alpha levels in gestational diabetes mellitus. *Diabetology & Metabolic Syndrome* **9**, 32 (2017).
232. Dehghan, A. *et al.* Meta-analysis of genome-wide association studies in > 80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation* **123**, 731-738 (2011).

233. Di Renzo, L. *et al.* Interleukin-1 (IL-1) receptor antagonist gene polymorphism in normal weight obese syndrome: relationship to body composition and IL-1 α and β plasma levels. *Pharmacological Research* **55**, 131-138 (2007).
234. Nunes, P.R.P. *et al.* Effect of resistance training on muscular strength and indicators of abdominal adiposity, metabolic risk, and inflammation in postmenopausal women: controlled and randomized clinical trial of efficacy of training volume. *Age* **38**, 40 (2016).
235. Browning, L.M., Krebs, J.D., Magee, E.C., Frühbeck, G. & Jebb, S.A. Circulating markers of inflammation and their link to indices of adiposity. *Obesity facts* **1**, 259-265 (2008).
236. Sirugo, G., Williams, S.M. & Tishkoff, S.A. The missing diversity in human genetic studies. *Cell* **177**, 26-31 (2019).
237. Li, M. *et al.* Allelic differences between Europeans and Chinese for CREB1 SNPs and their implications in gene expression regulation, hippocampal structure and function, and bipolar disorder susceptibility. *Molecular psychiatry* **19**, 452-461 (2014).
238. Grinde, K.E. *et al.* Generalizing polygenic risk scores from Europeans to Hispanics/Latinos. *Genetic epidemiology* **43**, 50-62 (2019).
239. Carlson, C.S. *et al.* Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS Biol* **11**, e1001661 (2013).
240. Martin, A.R. *et al.* Human demographic history impacts genetic risk prediction across diverse populations. *The American Journal of Human Genetics* **100**, 635-649 (2017).
241. Dugas, L.R., Cao, G., Luke, A.H. & Durazo-Arvizu, R.A. Adiposity is not equal in a multi-race/ethnic adolescent population: NHANES 1999–2004. *Obesity* **19**, 2099-2101 (2011).
242. Carroll, J.F. *et al.* Visceral fat, waist circumference, and BMI: impact of race/ethnicity. *Obesity* **16**, 600-607 (2008).
243. Kim, J., Zhang, Y. & Pan, W. Powerful and adaptive testing for multi-trait and multi-SNP associations with GWAS and sequencing data. *Genetics* **203**, 715-731 (2016).
244. Khan, S.S. *et al.* Association of body mass index with lifetime risk of cardiovascular disease and compression of morbidity. *JAMA cardiology* **3**, 280-287 (2018).
245. Cho, Y.S. *et al.* A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nature genetics* **41**, 527-534 (2009).
246. Reiner, A.P. *et al.* Genome-wide association and population genetic analysis of C-reactive protein in African American and Hispanic American women. *The American Journal of Human Genetics* **91**, 502-512 (2012).
247. Manolio, T.A. Genomewide association studies and assessment of the risk of disease. *New England journal of medicine* **363**, 166-176 (2010).

248. Gaye, A. & Davis, S.K. Genetic model misspecification in genetic association studies. *BMC research notes* **10**, 569 (2017).
249. Zheng, J., Li, Y., Abecasis, G.R. & Scheet, P. A comparison of approaches to account for uncertainty in analysis of imputed genotypes. *Genetic epidemiology* **35**, 102-110 (2011).
250. Winkler, T.W. *et al.* EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics* **31**, 259-261 (2015).
251. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).
252. Bomba, L., Walter, K. & Soranzo, N. The impact of rare and low-frequency genetic variants in common disease. *Genome biology* **18**, 77 (2017).
253. Hong, E.P. & Park, J.W. Sample size and statistical power calculation in genetic association studies. *Genomics & informatics* **10**, 117 (2012).
254. Geyer, C.J. Practical markov chain monte carlo. *Statistical science*, 473-483 (1992).
255. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* **88**, 76-82 (2011).
256. Warnes, M.G.R., Bolker, B., Bonebakker, L., Gentleman, R. & Huber, W. Package 'gplots'. *Various R Programming Tools for Plotting Data* (2016).
257. Pulit, S.L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Human molecular genetics* **28**, 166-174 (2019).
258. Zhu, Z. *et al.* Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *Journal of Allergy and Clinical Immunology* **145**, 537-549 (2020).
259. Vujkovic, M. *et al.* Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nature genetics* **52**, 680-691 (2020).
260. Deng, X. *et al.* Genome-wide association study for multiple phenotype analysis. in *BMC proceedings* Vol. 12 55 (Springer, 2018).
261. Lee, J.J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics* **50**, 1112-1121 (2018).
262. Tachmazidou, I. *et al.* Whole-genome sequencing coupled to imputation discovers genetic signals for anthropometric traits. *The American Journal of Human Genetics* **100**, 865-884 (2017).

263. Fishman, D. *et al.* The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. *The Journal of clinical investigation* **102**, 1369-1376 (1998).
264. Hunt, S.C. *et al.* Association of the FTO gene with BMI. *Obesity* **16**, 902-904 (2008).
265. Marigorta, U.M. & Navarro, A. High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS Genet* **9**, e1003566 (2013).
266. Kuchenbaecker, K. *et al.* The transferability of lipid loci across African, Asian and European cohorts. *Nature communications* **10**(2019).
267. Turcotte, M. *et al.* Genetic contribution to waist-to-hip ratio in Mexican children and adolescents based on 12 loci validated in European adults. *International journal of obesity* **43**, 13-22 (2019).
268. Hoffmann, S.C. *et al.* Ethnicity greatly influences cytokine gene polymorphism distribution. *American Journal of Transplantation* **2**, 560-567 (2002).
269. Ivanova, M. *et al.* IL-6 SNP diversity among four ethnic groups as revealed by bead-based liquid array profiling. *International journal of immunogenetics* **38**, 17-20 (2011).
270. Wing, M. *et al.* Analysis of FTO gene variants with obesity and glucose homeostasis measures in the multiethnic Insulin Resistance Atherosclerosis Study cohort. *International journal of obesity* **35**, 1173-1182 (2011).
271. Stryjecki, C., Alyass, A. & Meyre, D. Ethnic and population differences in the genetic predisposition to human obesity. *Obesity Reviews* **19**, 62-80 (2018).
272. Schork, A.J. *et al.* All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet* **9**, e1003449 (2013).
273. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics* **19**, 807-812 (2011).
274. Franklin, C.S. *et al.* Conjunctival fibrosis and the innate barriers to Chlamydia trachomatis intracellular infection: a genome wide association study. *Scientific reports* **5**, 17447 (2015).
275. Hales, C.M., Carroll, M.D., Fryar, C.D. & Ogden, C.L. Prevalence of obesity and severe obesity among adults: United States, 2017–2018. (2020).
276. Berber, A., Gomez-Santos, R., Fanghänel, G. & Sanchez-Reyes, L. Anthropometric indexes in the prediction of type 2 diabetes mellitus, hypertension and dyslipidaemia in a Mexican population. *International journal of obesity* **25**, 1794-1799 (2001).
277. Oswald, M., Hunt, H. & Lazarchick, J. Normal range of plasma fibrinogen. *The American Journal of Medical Technology* **49**, 57-59 (1983).

278. Schöch, H., Forster, L., Woidke, R., Solomon, C. & Voelckel, W. Use of rotation thromboelastometry (ROTEM®) to achieve successful treatment of polytrauma with fibrinogen concentrate and prothrombin complex concentrate. *Anaesthesia* **65**, 199-203 (2010).
279. Rosenson, R.S., Mosca, L., Staffileno, B.A. & Tangney, C.C. Variability in fibrinogen measurements: an obstacle to cardiovascular risk stratification. *Atherosclerosis* **159**, 225-230 (2001).
280. Petersen, A., Alvarez, C., DeClaire, S. & Tintle, N.L. Assessing methods for assigning SNPs to genes in gene-based tests of association using common variants. *PLoS One* **8**, e62161 (2013).
281. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-1195 (2012).
282. Fu, J. *et al.* Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. *PLoS Genet* **8**, e1002431 (2012).
283. Spain, S.L. & Barrett, J.C. Strategies for fine-mapping complex traits. *Human molecular genetics* **24**, R111-R119 (2015).
284. Fang, M. & Georges, M. BayesFM: a software program to fine-map multiple causative variants in GWAS identified risk loci. *bioRxiv*, 067801 (2016).
285. Greenbaum, J. & Deng, H.W. A statistical approach to fine mapping for the identification of potential causal variants related to bone mineral density. *Journal of Bone and Mineral Research* **32**, 1651-1658 (2017).
286. Chen, W. *et al.* Fine mapping causal variants with an approximate Bayesian method using marginal test statistics. *Genetics* **200**, 719-736 (2015).
287. Hormozdiari, F., Kichaev, G., Yang, W.-Y., Pasaniuc, B. & Eskin, E. Identification of causal genes for complex traits. *Bioinformatics* **31**, i206-i213 (2015).
288. Schaid, D.J., Chen, W. & Larson, N.B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature Reviews Genetics* **19**, 491-504 (2018).
289. Kichaev, G. *et al.* Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics* **33**, 248-255 (2017).
290. Benner, C. *et al.* Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. *The American Journal of Human Genetics* **101**, 539-551 (2017).
291. Band, G. & Marchini, J. BGEN: a binary file format for imputed genotype and haplotype data. *BioRxiv*, 308296 (2018).

292. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).
293. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-461 (2014).
294. Consortium, E.P. The ENCODE (ENCyclopedia of DNA elements) project. *Science* **306**, 636-640 (2004).
295. Auerbach, R.K., Chen, B. & Butte, A.J. Relating genes to function: identifying enriched transcription factors using the ENCODE ChIP-Seq significance tool. *Bioinformatics* **29**, 1922-1924 (2013).
296. Inoue, F. & Ahituv, N. Decoding enhancers using massively parallel reporter assays. *Genomics* **106**, 159-164 (2015).
297. Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83-90 (2012).
298. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research* **22**, 1760-1774 (2012).
299. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319 (2013).
300. McLaren, W. *et al.* The ensembl variant effect predictor. *Genome biology* **17**, 122 (2016).
301. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research* **47**, D886-D894 (2019).
302. Rogers, M.F. *et al.* FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* **34**, 511-513 (2018).
303. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic acids research* **40**, D930-D934 (2012).
304. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-2337 (2010).
305. Apostolopoulos, V. *et al.* The complex immunological and inflammatory network of adipose tissue in obesity. *Molecular nutrition & food research* **60**, 43-57 (2016).
306. Gualillo, O. Mediators of inflammation in obesity and its comorbidities. (Hindawi, 2010).
307. Xia, Q. & Grant, S.F. The genetics of human obesity. *Annals of the New York Academy of Sciences* **1281**, 178 (2013).

308. Liu, L. *et al.* A genomewide integrative analysis of GWAS and eQTLs data identifies multiple genes and gene sets associated with obesity. *BioMed research international* **2018**(2018).
309. Zwicker, A. *et al.* Genetic disposition to inflammation and response to antidepressants in major depressive disorder. *Journal of psychiatric research* **105**, 17-22 (2018).
310. Ahola-Olli, A.V. *et al.* Genome-wide association study identifies 27 loci influencing concentrations of circulating cytokines and growth factors. *The American Journal of Human Genetics* **100**, 40-50 (2017).
311. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nature genetics* **51**, 1339-1348 (2019).
312. Faye, L.L., Machiela, M.J., Kraft, P., Bull, S.B. & Sun, L. Re-ranking sequencing variants in the post-GWAS era for accurate causal variant identification. *PLoS Genet* **9**, e1003609 (2013).
313. Chen, C.-y., Chang, I.-S., Hsiung, C.A. & Wasserman, W.W. On the identification of potential regulatory variants within genome wide association candidate SNP sets. *BMC medical genomics* **7**, 34 (2014).
314. Sexton, T. & Cavalli, G. The role of chromosome domains in shaping the functional genome. *Cell* **160**, 1049-1059 (2015).
315. Itan, Y. *et al.* The mutation significance cutoff: gene-level thresholds for variant predictions. *Nature methods* **13**, 109-110 (2016).
316. Cosma, M.P. *et al.* The multiple sulfatase deficiency gene encodes an essential and limiting factor for the activity of sulfatases. *Cell* **113**, 445-456 (2003).
317. Schlotawa, L., Adang, L., De Castro, M. & Ahrens-Nicklas, R. Multiple sulfatase deficiency. *GeneReviews ((R)); Adam, MP, Ardinger, HH, Pagon, RA, Wallace, SE, Bean, LJH, Stephens, K., Amemiya, A., Eds* (2019).
318. Weidner, J. *et al.* Sulfatase modifying factor 1 (SUMF1) is associated with Chronic Obstructive Pulmonary Disease. *Respiratory research* **18**, 77 (2017).
319. Jiang, M. University of Pittsburgh (2014).
320. Day, N. *et al.* Interleukin receptor-associated kinase (IRAK-4) deficiency associated with bacterial infections and failure to sustain antibody responses. *The Journal of pediatrics* **144**, 524-526 (2004).
321. Kim, S.-J., Choi, Y., Choi, Y.-H. & Park, T. Obesity activates toll-like receptor-mediated proinflammatory signaling cascades in the adipose tissue of mice. *The Journal of nutritional biochemistry* **23**, 113-122 (2012).

322. Tamaki, T. *et al.* A novel transmembrane protein defines the endoplasmic reticulum stress-induced cell death pathway. *Biochemical and biophysical research communications* **486**, 149-155 (2017).
323. Mo, Y. *et al.* Epithelial SERPINB10, a novel marker of airway eosinophilia in asthma, contributes to allergic airway inflammation. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **316**, L245-L254 (2019).
324. Marenholz, I. *et al.* Genome-wide association study identifies the SERPINB gene cluster as a susceptibility locus for food allergy. *Nature communications* **8**, 1-10 (2017).
325. Schroder, W.A., Major, L. & Suhrbier, A. The role of SerpinB2 in immunity. *Critical Reviews™ in Immunology* **31**(2011).
326. ELBadawy, N.E., Abdel-Latif, R.S. & El-Hady, H.A. Association between SERPINB2 gene expression by real time PCR in respiratory epithelial cells and atopic bronchial asthma severity. *Egypt J Immunol* **24**, 165-181 (2017).
327. Kawase, T. *et al.* Alternative splicing due to an intronic SNP in HMSD generates a novel minor histocompatibility antigen. *Blood* **110**, 1055-1063 (2007).
328. Baker, K.E. & Parker, R. Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Current opinion in cell biology* **16**, 293-299 (2004).
329. Dyle, M.C., Kolakada, D., Cortazar, M.A. & Jagannathan, S. How to get away with nonsense: Mechanisms and consequences of escape from nonsense-mediated RNA decay. *Wiley Interdisciplinary Reviews: RNA* **11**, e1560 (2020).
330. Miller, J.N. & Pearce, D.A. Nonsense-mediated decay in genetic disease: friend or foe? *Mutation Research/Reviews in Mutation Research* **762**, 52-64 (2014).
331. Khajavi, M., Inoue, K. & Lupski, J.R. Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. *European journal of human genetics* **14**, 1074-1081 (2006).
332. Grarup, N. *et al.* Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes. *Nature genetics* **50**, 172-174 (2018).
333. Cirillo, G. *et al.* Lack of red hair phenotype in a North-African obese child homozygous for a novel POMC null mutation: nonsense-mediated decay RNA evaluation and hair pigment chemical analysis. *British Journal of Dermatology* **167**, 1393-1395 (2012).
334. Pulit, S.L., de With, S.A. & de Bakker, P.I. Resetting the bar: Statistical significance in whole-genome sequencing-based association studies of global populations. *Genetic epidemiology* **41**, 145-151 (2017).
335. Wei, W.-H., Hemani, G. & Haley, C.S. Detecting epistasis in human complex traits. *Nature Reviews Genetics* **15**, 722 (2014).

336. McClellan, J. & King, M.-C. Genetic heterogeneity in human disease. *Cell* **141**, 210-217 (2010).
337. Rosenberg, N.A. *et al.* Genome-wide association studies in diverse populations. *Nature Reviews Genetics* **11**, 356-366 (2010).
338. Altshuler, D., Daly, M.J. & Lander, E.S. Genetic mapping in human disease. *science* **322**, 881-888 (2008).
339. Hermosilla, M.I. & Lemus, J.A. Therapeutic Translation of Genomic Science: Opportunities and limitations of GWAS. (National Bureau of Economic Research, 2017).
340. Du, Y., Xie, J., Chang, W., Han, Y. & Cao, G. Genome-wide association studies: inherent limitations and future challenges. *Frontiers of medicine* **6**, 444-450 (2012).
341. Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics* **20**, 467-484 (2019).
342. Giacomini, K.M. *et al.* The pharmacogenetics research network: from SNP discovery to clinical drug response. *Clinical Pharmacology & Therapeutics* **81**, 328-345 (2007).
343. Gallagher, M.D. & Chen-Plotkin, A.S. The post-GWAS era: from association to function. *The American Journal of Human Genetics* **102**, 717-730 (2018).
344. Sivakumaran, S. *et al.* Abundant pleiotropy in human complex diseases and traits. *The American Journal of Human Genetics* **89**, 607-618 (2011).
345. Hodgkin, J. Seven types of pleiotropy. *International Journal of Developmental Biology* **42**, 501-505 (2002).
346. Torday, J.S. Pleiotropy, the physiologic basis for biologic fields. *Progress in biophysics and molecular biology* **136**, 37-39 (2018).
347. Lutz, S.M. & Hokanson, J.E. Mediation analysis in genome-wide association studies: current perspectives. *Open Bioinforma J* **7**, 1-5 (2015).
348. Siedlinski, M. *et al.* Dissecting direct and indirect genetic effects on chronic obstructive pulmonary disease (COPD) susceptibility. *Human genetics* **132**, 431-441 (2013).
349. Teng, M.-S. *et al.* Association of CDH13 genotypes/haplotypes with circulating adiponectin levels, metabolic syndrome, and related metabolic phenotypes: the role of the suppression effect. *PloS one* **10**, e0122664 (2015).
350. Sritara, C. *et al.* Causal relationship between the AHSG gene and BMD through fetuin-A and BMI: multiple mediation analysis. *Osteoporosis International* **25**, 1555-1562 (2014).
351. Gu, J. *et al.* A genome-wide association study identifies a locus on chromosome 14q21 as a predictor of leukocyte telomere length and as a marker of susceptibility for bladder cancer. *Cancer prevention research* **4**, 514-521 (2011).

352. Ferrante Jr, A. Obesity-induced inflammation: a metabolic dialogue in the language of inflammation. *Journal of internal medicine* **262**, 408-414 (2007).
353. Berg, A.H. & Scherer, P.E. Adipose tissue, inflammation, and cardiovascular disease. *Circulation research* **96**, 939-949 (2005).
354. Kawamoto, R., Kusunoki, T., Abe, M., Kohara, K. & Miki, T. An association between body mass index and high-sensitivity C-reactive protein concentrations is influenced by age in community-dwelling persons. *Annals of Clinical Biochemistry* **50**, 457-464 (2013).
355. Cartier, A. *et al.* Sex differences in inflammatory markers: what is the contribution of visceral adiposity? *The American journal of clinical nutrition* **89**, 1307-1314 (2009).
356. Bochud, M. *et al.* Association between C-reactive protein and adiposity in women. *The Journal of Clinical Endocrinology & Metabolism* **94**, 3969-3977 (2009).
357. Santos, A., Lopes, C., Guimaraes, J. & Barros, H. Central obesity as a major determinant of increased high-sensitivity C-reactive protein in metabolic syndrome. *International journal of obesity* **29**, 1452-1456 (2005).
358. Wee, C.C. *et al.* Obesity and C-reactive protein levels among white, black, and Hispanic US adults. *Obesity* **16**, 875-880 (2008).
359. Valentine, R.J. *et al.* Sex differences in the relationship between obesity, C-reactive protein, physical activity, depression, sleep quality and fatigue in older adults. *Brain, behavior, and immunity* **23**, 643-648 (2009).
360. Marques-Vidal, P. *et al.* Association between inflammatory and obesity markers in a Swiss population-based sample (CoLaus Study). *Obesity facts* **5**, 734-744 (2012).
361. Stępień, M. *et al.* Obesity indices and inflammatory markers in obese non-diabetic normo- and hypertensive patients: a comparative pilot study. *Lipids in health and disease* **13**, 29 (2014).
362. McLaughlin, T. *et al.* Differentiation between obesity and insulin resistance in the association with C-reactive protein. *Circulation* **106**, 2908-2912 (2002).
363. Aronson, D. *et al.* Obesity is the major determinant of elevated C-reactive protein in subjects with the metabolic syndrome. *International journal of obesity* **28**, 674-679 (2004).
364. Winer, J.C. *et al.* Adiponectin in childhood and adolescent obesity and its association with inflammatory markers and components of the metabolic syndrome. *The Journal of Clinical Endocrinology & Metabolism* **91**, 4415-4423 (2006).
365. Chandra, A. *et al.* The relationship of body mass and fat distribution with incident hypertension: observations from the Dallas Heart Study. *Journal of the American College of Cardiology* **64**, 997-1002 (2014).

366. MIETTINEN, T.A. Cholesterol production in obesity. *Circulation* **44**, 842-850 (1971).
367. Oh, J.Y., Sung, Y.A. & Lee, H.J. The visceral adiposity index as a predictor of insulin resistance in young women with polycystic ovary syndrome. *Obesity* **21**, 1690-1694 (2013).
368. Steinberger, J., Moran, A., Hong, C.-P., Jacobs Jr, D.R. & Sinaiko, A.R. Adiposity in childhood predicts obesity and insulin resistance in young adulthood. *The Journal of pediatrics* **138**, 469-473 (2001).
369. Boyko, E.J., Leonetti, D.L., Bergstrom, R.W., Newell-Morris, L. & Fujimoto, W.Y. Low insulin secretion and high fasting insulin and C-peptide levels predict increased visceral adiposity: 5-year follow-up among initially nondiabetic Japanese-American men. *Diabetes* **45**, 1010-1015 (1996).
370. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics* **81**, 559-575 (2007).
371. Morris, T.P., White, I.R. & Royston, P. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC medical research methodology* **14**, 75 (2014).
372. Imai, K., Keele, L., Tingley, D. & Yamamoto, T. Causal mediation analysis using R. in *Advances in social science research using R* 129-154 (Springer, 2010).
373. Parker, M.M. *et al.* Assessing pleiotropy and mediation in genetic loci associated with chronic obstructive pulmonary disease. *Genetic epidemiology* **43**, 318-329 (2019).
374. Sims III, R.J., Nishioka, K. & Reinberg, D. Histone lysine methylation: a signature for chromatin function. *TRENDS in Genetics* **19**, 629-639 (2003).
375. Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86-90 (2006).
376. Gerasimova, A. *et al.* Predicting cell types and genetic variations contributing to disease by combining GWAS and epigenetic data. *PLoS one* **8**, e54359 (2013).
377. Kawasaki, N., Asada, R., Saito, A., Kanemoto, S. & Imaizumi, K. Obesity-induced endoplasmic reticulum stress causes chronic inflammation in adipose tissue. *Scientific reports* **2**, 799 (2012).
378. Friedrich, M. Global obesity epidemic worsening. *Jama* **318**, 603-603 (2017).
379. Apovian, C.M. & Riffenburt, K.M. Perspectives on the global obesity epidemic. *Current Opinion in Endocrinology, Diabetes and Obesity* **24**, 307-309 (2017).
380. Hall, K.D. Did the food environment cause the obesity epidemic? *Obesity* **26**, 11-13 (2018).
381. Church, T. & Martin, C.K. The obesity epidemic: a consequence of reduced energy expenditure and the uncoupling of energy intake? *Obesity* **26**, 14-16 (2018).

382. An, R., Ji, M. & Zhang, S. Global warming and obesity: a systematic review. *Obesity Reviews* **19**, 150-163 (2018).
383. Lichtveld, K., Thomas, K. & Tulve, N.S. Chemical and non-chemical stressors affecting childhood obesity: a systematic scoping review. *Journal of exposure science & environmental epidemiology* **28**, 1-12 (2018).
384. Gregg, E.W. & Shaw, J.E. Global health effects of overweight and obesity. (Mass Medical Soc, 2017).
385. Dandona, P., Aljada, A. & Bandyopadhyay, A. Inflammation: the link between insulin resistance, obesity and diabetes. *Trends in immunology* **25**, 4-7 (2004).
386. Bi, Y. *et al.* Sex-Dimorphic and Sex Hormone-Dependent Role of Steroid Sulfatase in Adipose Inflammation and Energy Homeostasis. *Endocrinology* **159**, 3365-3377 (2018).
387. Bernstein, B.E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326 (2006).
388. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837 (2007).
389. Deldicque, L., Cani, P.D., Delzenne, N.M., Baar, K. & Francaux, M. Endurance training in mice increases the unfolded protein response induced by a high-fat diet. *Journal of physiology and biochemistry* **69**, 215-225 (2013).
390. Schröder, M. & Kaufman, R.J. ER stress and the unfolded protein response. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **569**, 29-63 (2005).
391. Credle, J.J., Finer-Moore, J.S., Papa, F.R., Stroud, R.M. & Walter, P. On the mechanism of sensing unfolded protein in the endoplasmic reticulum. *Proceedings of the National Academy of Sciences* **102**, 18773-18784 (2005).
392. Volmer, R., van der Ploeg, K. & Ron, D. Membrane lipid saturation activates endoplasmic reticulum unfolded protein response transducers through their transmembrane domains. *Proceedings of the National Academy of Sciences* **110**, 4628-4633 (2013).
393. Gregor, M.F. *et al.* Endoplasmic reticulum stress is reduced in tissues of obese subjects after weight loss. *Diabetes* **58**, 693-700 (2009).
394. Boden, G. *et al.* Increase in endoplasmic reticulum stress-related proteins and genes in adipose tissue of obese, insulin-resistant individuals. *Diabetes* **57**, 2438-2444 (2008).
395. Fu, S. *et al.* Aberrant lipid metabolism disrupts calcium homeostasis causing liver endoplasmic reticulum stress in obesity. *Nature* **473**, 528-531 (2011).
396. Korte, A. & Farlow, A. The advantages and limitations of trait analysis with GWAS: a review. *Plant methods* **9**, 1-9 (2013).

397. Avery, C.L. *et al.* A phenomics-based strategy identifies loci on APOC1, BRAP, and PLCG1 associated with metabolic syndrome phenotype domains. *PLoS Genet* **7**, e1002322 (2011).
398. He, Q., Avery, C.L. & Lin, D.Y. A general framework for association tests with multivariate traits in large-scale genomics studies. *Genet Epidemiol* **37**, 759-67 (2013).
399. Andreassen, O.A. *et al.* Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *Am J Hum Genet* **92**, 197-209 (2013).
400. Marigorta, U.M., Rodríguez, J.A., Gibson, G. & Navarro, A. Replicability and prediction: lessons and challenges from GWAS. *Trends in Genetics* **34**, 504-517 (2018).
401. Palmer, C. & Pe'er, I. Statistical correction of the Winner's Curse explains replication variability in quantitative trait genome-wide association studies. *PLoS genetics* **13**, e1006916 (2017).
402. Shikov, A.E., Skitchenko, R.K., Predeus, A.V. & Barbitoff, Y.A. Phenome-wide functional dissection of pleiotropic effects highlights key molecular pathways for human complex traits. *Scientific reports* **10**, 1-10 (2020).

CURRICULUM VITAE

NAME Mohammad Yaser (Anwar)

ADDRESS Room No. 227, 485 E. Gray St., Louisville, KY 40202

DOB: 09/06/1984

EDUCATION & TRAINING: PhD. in Epidemiology
University of Louisville, Louisville-KY, USA
2020

Master of Public Health
Johns Hopkins Bloomberg School of Public Health,
Baltimore-MD, USA
2014

Medical Doctorate
Herat University-Herat, Herat, Afghanistan
2008

PUBLICATION: Anwar MY, Joshua L. Warren, Pitzer V: Diarrhea
Patterns and Climate: a spatiotemporal analysis of
diarrheal disease in Afghanistan. *American Journal of
Tropical Medicine and Hygiene* 2019, DOI:
10.4269/ajtmh.18-0735

Anwar MY, Lewnard J, Parikh S, Pitzer V: Time series
analysis of malaria in Afghanistan: using ARIMA
models to predict future trends in incidence. *Malaria
Journal* 2016, 15. DOI: [10.1186/s12936-016-1602-1](https://doi.org/10.1186/s12936-016-1602-1)

Anwar MY, Burnham Gilbert: Trends in infectious
disease incidence among children in Afghanistan at a
time of public health services expansion. *East
Mediterranean Health J.* 2017 Feb 1; 22(11):778-785.

Longitudinal Assessment of Alcohol Intake and
Variation in Fecundability During the Luteal and
Follicular Windows in Mount Sinai Study of Women
Office Workers Cohort (pre-print *Human Reproduction
Journal*).

Anthropometric Polygenic Risk Scores and Adiposity
Traits in African Americans: Inferences from

Observational Assessment of Jackson Heart Study Cohort (pre-print copy available from *medRxiv*).

PROFESSIONAL SOCIETIES:

American Heart Association (2020)

American Society of Human Genetics (2020)

PRESENTATIONS:

Poster presentation of Jackson Heart adiposity polygenic risk score study in American Society of Human Genetics (ASHG) annual conference (October 15 to 19 2019, Houston, TX).

Presentation of “*Assessing efficiency of fine-mapping anthropometric associated variants through leveraging ancestry architecture and functional annotation*” in PAGE genetic working group meeting.

Presentation on of mixed-model GWAS with GENESIS, Gillings School of Global Public Health, University of North Carolina, Chapel Hill.