GENOMICS AND EPIDEMIOLOGY: DIVERSE MEASURES AND POPULATIONS IN THE TRANS-ETHNIC FINE-MAPPING OF GENETIC LOCI FOR BODY MASS INDEX

Lindsay Fernández-Rhodes

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

Chapel Hill
2016

Approved by:

Nora Franceschini

Penny Gordon-Larsen

Kari E. North

Whitney Robinson

Paul Smokowski

ABSTRACT

Lindsay Fernández-Rhodes: Genomics and Epidemiology: Diverse Measures and Populations in the Trans-Ethnic Fine-Mapping of Genetic Loci for Body Mass Index
(Under the direction of Kari E. North)

Obesity is a global epidemic with concerning disparities in burden across United States (US) racial/ethnic groups. In the absence of measured body weight, self-reports are a commonly used proxy in epidemiologic research. Previous studies have found that self-reported body weight may, on average, underestimate weight. However, this research may not apply to US Hispanics/Latinos, many of whom are recent immigrants from Latin America. We investigated whether self-reported weight was an accurate proxy of measured weight in a sample of Hispanic/Latinos from various Hispanic/Latino backgrounds sampled as part of the baseline examination (2008-2011) of the Hispanic Community Health Study/Study of Latinos. We observed that self-reported weight was an accuracy proxy of measured weight (r^2=0.95, average 0.3 kg over-reporting of weight), but differential patterns of misreporting were evident by age, gender, body mass index (BMI) categories, nativity, study site by background, unit of self-reported weight and end digit preference.

Numerous studies of obesity in primarily non-Hispanic/Latino European descent populations have identified more than >100 BMI loci. However, these loci collectively explain a fraction of the estimated heritability of BMI perhaps, in part, due to the limited racial/ethnic diversity of the previous samples. I addressed this research gap by generalizing nearly a quarter of previously reported SNP-BMI associations and >80% of 36 fine-mapped BMI loci to racially/ethnically diverse US populations and then by trans-ethnically fine-mapping the

underlying functional variants at these loci in a sample of approximately 102,000 African, Hispanic/Latino, Asian, European and American Indian/Alaskan Native descent adults from the Population Architecture using Genomics and Epidemiology Study. These findings will help prioritize the putative functional variants for targeted molecular follow-up and gene-environment interaction studies.

In light of the current mismatch between the mounting body of genetic epidemiologic evidence and the populations most burdened by obesity, this research highlighted the utility of alternative measures, such as self-reported weight, and diverse populations in the search for the underlying functional genetic variants for obesity risk. As such, this work serves as a foundation for a wide-range of future research on the complex genetic and environmental determinants of obesity in US populations, like Hispanic/Latinos.

ACKNOWLEDGEMENTS

I would sincerely like to thank my committee members (Nora Franceschini, Penny Gordon-Larsen, Whitney Robinson, Paul Smokowski) and Chair (Kari E. North) for their unwavering enthusiasm and support as I navigated the many turns of this dissertation. Additionally, Daniela Sotres-Alvarez, served as an unofficial committee member and was instrumental in helping me navigate the Hispanic Community Health Study/Study of Latinos. The funding provided to me by the Cardiovascular Disease Epidemiology Training Grant, American Heart Association, and my research assistantship mentors (Gerardo Heiss, Kari E. North) has given me the protected time and experiences to establish my research trajectory.

I would like to acknowledge the participants of the Hispanic Community Health Study/Study of Latinos and the Population Architecture using Genomics and Epidemiology Study for their devotion of time, bio-specimens and faith in the promise of scientific research.

My academic trajectory can be directly attributed to four amazing teachers to whom I give my heart-felt thanks (Janet Rhodes, Judy Reynolds, Deb McCall-Hardy and Joe Gerzina). I hope to inspire as much passion and conviction in my children (Dalia and Dominic Fernández-Rhodes) and many others.

I would like to acknowledge the 11 million or more United States immigrants, who find the courage to live, work and play in the shadows. Their stories of perseverance are forever intertwined with my own and have intrinsically guided this dissertation.

Lastly, I am grateful to my friends, my family and most importantly, my incredible husband (Juan Luis Fernández Reyes) for the personal sacrifices they have all graciously made to allow me to grow into the researcher that I aspire to be.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Add Health	National Longitudinal Study of Adolescent to Adult Health
ARIC	Atherosclerosis Risk in Communities Study
BF	Bayes Factor
BioME	Mount Sinai Biobank Program
BioVU	Vanderbilt University's Biobank
BMI	Body mass index
CARDIA	Coronary Artery Risk Development in Young Adults
CC	Coordinating Center
CHS	Cardiovascular Health Study
CVD	Cardiovascular disease
EAGLE	Epidemiologic Architecture for Genes Linked to Environment Study
GCTA	Genome-wide complex trait analysis
GEE	Generalized Estimating Equations
GWAS	Genome-wide association studies
GxE	Gene by Environment
HCHS/SOL	Hispanic Community Health Study/Study of Latinos
HyperGEN	Hypertension Genetic Epidemiology Network Study
LD	Linkage disequilibrium
MANTRA	Meta-ANalysis of TRansethnic Association studies
MESA	Multi-Ethnic Study of Atherosclerosis
MEC	Multiethnic Cohort Study
NHANES	National Health and Nutrition Examination Survey
PAGE	Population Architecture using Genomics and Epidemiology
PCA	Principal Component Analysis

SES	Socioeconomic Status
SIGMA	Slim Initiative in Genomic Medicine for the Americas Type 2 Diabetes Consortium
SNP	Single Nucleotide Polymorphism
TaiChi	Taiwan-MetaboChip Study for Cardiovascular Disease
US	United States
WHI	Women’s Health Initiative

CHAPTER I: INTRODUCTION

Obesity is a major cardiovascular disease (CVD) risk factor and public health concern both globally and for the United States (US) [1]. Racial and ethnic disparities in obesity have been described within the context of the US obesity epidemic [2, 3]. Hispanic/Latinos individuals comprise an ethnic group of diverse ancestries and heritages, which as of 2010 represented ~16% of the US population [4]. Yet Hispanic/Latino individuals have a higher prevalence of obesity (39%) than non-Hispanic/Latino individuals of European descent (34%) and as such shoulder what appears to be an increasingly disparate proportion of the epidemic [2, 3, 5]. However, even within this diverse group there is substantial heterogeneity [2, 3, 6], which may in part be influenced by an individual's level of acculturation [7, 8]—the complex process of cultural adaptation that occurs when individuals of different cultures come into contact with each other and experience changes in their cultural practices [9], including the maintenance and adoption of cultural patterns in diet and physical activity [8, 10, 11] .

Recent studies have also suggested an influence of genetic factors in individual susceptibility within the obesity epidemic. Studies indicate that body mass index (BMI, kg/m^2) may be between 40-70% heritable [12, 13]. Even though over 100 loci for BMI have been validated in predominantly European descent samples, the known genetic variants at these established loci explain a small proportion (<3%) of the overall phenotypic variance [14-33].

Two potential explanations of the missing heritability are allelic heterogeneity [34] and the previously neglected gene-environment (GxE) interactions that contextualize inherited genetic susceptibility [35]. Studies of allelic heterogeneity are necessary in

order to obtain true estimates of genetic effect sizes [36, 37], as well as to prioritize strong candidates for future studies, e.g. GxE studies and functional follow-up. Specifically, the temporal and geographic trends in the global obesity epidemic [1, 38] point to the convergence of both thrifty genes and an obesogenic environment [39], and the future promise of targeted GxE studies in unraveling the complex origins of obesity.

The Population Architecture using Genomics and Epidemiology (PAGE) Study, a consortia of several observational studies or sub-consortia, has been created to facilitate the study of the genetic underpinnings of CVD risk factors, as well as address the gap between the most studied populations in genetic epidemiology and the populations with the greatest disease burden [40]. Thus this consortium offers a unique opportunity to investigate the determinants of obesity in US minority populations [41] as it includes several landmark studies of diverse racial and ethnic groups. For example, the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) is the first Hispanic/Latino cohort study of CVD to recruit >16,000 Hispanic/Latino adults (18-74 years at screening) from various diverse backgrounds as well as collect in a culturally- and linguistically-appropriate manner information on anthropometrics, genetics, and acculturation [42, 43].

As Hispanic/Latino individuals shoulder a disproportionate burden of obesity [2, 3, 5], there is a potential for a 'looming' Hispanic/Latino CVD health crisis [44]. Therefore, this dissertation utilizes data on diverse US Hispanic/Latinos collected by the HCHS/SOL and other racial/ethnic groups represented in the studies of PAGE to address important gaps in the public health literature surrounding the application of diverse measures (Aim 1) and populations (Aim 2) to genomic and epidemiologic research intended to directly inform the complex obesity determinants of the US racial/ethnic groups with the greatest burden, such as Hispanic/Latinos.

CHAPTER II: SPECIFIC AIMS

A. Rationale

In order to address the gap between genetic epidemiology and obesity disparities, I have created a conceptual model to depict the complex relationships between acculturation, genetics, and obesity in diverse US populations, such as Hispanic/Latinos (Figure 1). Two interrelated dissertation manuscripts support this broad research agenda.

First, in Aim 1 I established the accuracy of self-reported weight as an alternative to measured weight in HCHS/SOL, a key component of BMI for the measurement of obesity (Figure 1; upper corner). This work will directly inform the use of diverse measurement sources in obesity epidemiology, such as conducted in Aim 2. More generally, in future work I plan to use the self-reported weight histories as proxies of adulthood weight change in HCHS/SOL, which represents a dynamic predominantly immigrant populations that may not have had consistent access to health care or readily accessible information on weight pre- and post-immigrant.

Second, in Aim 2 I fine-mapped the association of common genetic variants with BMI in five diverse ancestral populations of the PAGE Study: African, Hispanic/Latino (including HCHS/SOL), Asian, European and Asian ancestries (Figure 1; right hand corner). This work will directly inform future targeted investigations of the gene-acculturation interactions on BMI in diverse Hispanic/Latino samples of the PAGE Study (Figure 1; center set of arrows) or the functional consequences of the genetic variant.

The two manuscripts of this dissertation will each highlight the utility of diverse measures and populations in current genomics and epidemiology research on obesity. An individual's Hispanic/Latino background, migratory or acculturative history may relate to the accuracy of their self-reported weight (Aim 1), their innate genetic susceptibility (Aim 2), or their exposure to interacting obesogenic environments (Figure 1; center set of arrows).

Acculturation may capture the broader sociocultural context through which Hispanic/Latinos integrate to the US obesogenic environment within and across generations leading to behavioral or cognitive changes relevant to CVD risk. Given that some Hispanic/Latino backgrounds may be encountering an obesogenic environment for the first time in US, and yet others may have emigrated from countries with a rising obesity epidemic [e.g. recent birth cohorts from Mexico, which now has the highest prevalence of adult obesity in the world [45]], the consideration of heterogeneity across Hispanic/Latino backgrounds is of key importance. Future work that considers both acculturation and genetic predisposition to obesity may account for some of the previously unexplained hereditability (or phenotypic variation), and inform our understanding of how the sociocultural context may exacerbate innate genetic susceptibility.

B. Aim 1

The first aim of this dissertation was to establish the accuracy of self-reported current weight as a potential tool for future etiologic studies of self-reported weight or weight histories in US Hispanic/Latinos. After adjusting for the complex sampling design of HCHS/SOL, I described both the unadjusted mean differences between self-reported and measured weight (Aim 1A) and the adjusted estimated differences using a

multivariate model that simultaneously included a number of potential predictors of inaccuracy (Aim 1B).

Throughout Aim 1 I used a quality-controlled dataset that was restricted to HCHS/SOL baseline participants with both self-reported and measured weights as part of the Anthropometry Questionnaire in the fasting block of the HCHS/SOL baseline examination (2008-2011) and who were not at the extremes of either self-reported weight or BMI (calculated from self-reported weight and measured height), not currently pregnant, or did not have a previous limb amputation (n=16,119, aged 18-74 years at screening). The quality control procedures and research methods are described briefly below.

1. Aim 1A

After a staged data cleaning and exclusion protocol, we performed a linear regression of self-reported on measured weight to estimate the correlation between the two and inform unity and modified Bland-Altman plots. We then calculated the difference between self-reported and measured weight in kg. Adjusting for the complex study design of HCHS/SOL, we generated mean differences overall and stratified by potential predictors of inaccuracy including: acculturation (language preference, nativity), demographics (age, education, field center, Hispanic/Latino background, household income, gender), and health/behavioral measures (cancer, diabetes, categories of body mass, self-reported physical activity and current smoking status). Additionally, we used multiple imputation to fill in missing covariate data, which prior to analyses were estimated to represent 5-10% of participants (≥1 missing covariate).

Hypothesis: *I hypothesize that self-reported weight will be a sufficiently accurate proxy of measured weight at baseline across Hispanic/Latino backgrounds, as measured by both the correlation (e.g., r^2>0.9) and the mean difference after accounting for the*

complex study design (Test of difference=0, p≥0.05). I also expect that a Bland-Altman plot will support the observation of agreement of self-reported and measured current weight in HCHS/SOL (p<0.05). Similar to previous studies I anticipate that the mean difference will vary in direction of mis-reporting across a number of previously reported differential predictors of inaccuracy (e.g. age, gender, categories of BMI, language preference, nativity, site by background) [46-51]. I hypothesize that multiple imputation of covariates will increase our precision, but otherwise will not influence the substantive conclusions of the work, thus indicating that predictors of inaccuracy were likely missing at random.

2. Aim 1B

We then used linear regression models to assess how the accuracy of self-reported weight varied across strata of potential predictors of inaccuracy (e.g. age, gender, categories of BMI, language preference, nativity, site by background) after taking all other factors into account.

Hypothesis: *I anticipate that only a subset of the most consistent determinants of accuracy (e.g. age, gender, categories of BMI) will be relevant in this study in a multivariate prediction model (p<0.05). Again I hypothesize that multiple imputation will increase our precision, but otherwise will not influence the conclusions of the work.*

C. Aim 2

After I having garnered a better understanding of the accuracy of self-reported weight in US Hispanic/Latinos, as an alternative measure of body weight for the calculation of BMI, I harnessed the rich data available in the PAGE Study to generalize and fine-map BMI loci, as captured by measured and self-reported weights and heights, in diverse ethnic/racial studies which included 102,514 adults of African (34.7% of total sample), Hispanic/Latino (25.4%), Asian (21.9%), European (17.4%) and American

Indian/Alaskan Native (0.5%) ancestries. Although several large GWAS of BMI have been published to date, no studies have attempted to generalize associations to other racial/ethnic groups or fine-map the established BMI loci with dense genotyping within these multiple racial/ethnic groups (Aim 2A) or trans-ethnically (Aim 2B). For this reason, I have chosen to address this knowledge gap, before continuing research on the other components and pathways of our conceptual diagram (Figure 1). The research methods for Aim 2 are briefly described below.

1. Aim 2A

I utilized the densely-mapped available in the PAGE Study to generalize 170 previously described and validated SNPs from the GWAS literature (or their proxies, $r^2 \geq 0.8$ in the population of discovery). Because these SNPs were identified mostly in individuals of non-Hispanic/Latino European ancestry and are likely correlated with the 'causal' variant, I also fine-map the underlying functional SNPs for increased BMI (captured using measurements and self-reports) at 36 established BMI loci. Each participating study performed linear regression of the natural log of the BMI distribution for each single SNPs while adjusting for relevant confounders [age, gender, population stratification, field center as appropriate, and the complex sampling design and backgrounds of HCHS/SOL in generalized estimating equations (GEEs)]. On the study- and racial/ethnic group- (African, Hispanic/Latino, Asian, European and American Indian/Alaskan Native ancestry) stratified summary results, I performed an inverse variance weighted fixed-effect meta-analysis to generate a single fixed-effect summary per racial/ethnic group. I then assessed the effect heterogeneity across studies. Lastly, I generated forest plots to investigate SNP-associations with evidence of study heterogeneity ($p<0.05/166$ independent previously reported signals).

Hypothesis: *I hypothesized that the majority of the risk alleles of 'index' (previously reported) SNPs would have directionally consistent effects for risk alleles comparing the previous reports and racial/ethnic-specific results, and that this would surpass what we would expect by chance (binominal p<0.05, 170 tests, 0.5 probably of success on each). I anticipated that a smaller proportion (~25%, based on previous reports [52]) of the index SNPs would generalize, defined as being both directionally consistent and statistically significant considering the number of independent tests performed (p<0.05/166), within in each racial/ethnic group. I also anticipated that the number of loci generalized would be the least and greatest in our smallest (American Indian/Alaskan Native) and largest (African American) samples, respectively. I also expected that the additional SNPs available in the 36 fine-mapped BMI loci would improve our ability to select the marker with the lowest p-value ('top' SNP) and that in most cases this top SNP would not be the index SNP (p<0.05/ independent tests per locus, r^2<0.2 in African Americans).*

2. Aim 2B

Using the racial/ethnic-specific summary results from meta-analysis of African, Hispanic/Latino, Asian and European ancestries estimated in Aim 2A, I performed inverse variance weighted fixed-effect meta-analysis to generate a trans-ethnic fixed-effect summary for each SNP, and then assess the evidence of heterogeneity across racial/ethnic groups (p<0.05/166 previously reported independent SNPs). Additionally, I implemented a Bayesian trans-ethnic meta-analysis to cluster the racial/ethnic groups by their allelic frequency differences [53] and within the 36 densely-genotyped BMI loci to construct a 99% credible interval reflective our confidence of observing the causal variant within its bounds. Lastly, approximate conditional analyses were performed to establish the number of independent signals within each locus, and regional plots were

utilized to illustrate the genetic architecture of each of the 36 densely-genotyped loci.

Hypothesis: *Similar to Aim 2A, I hypothesized that the majority of the risk allele of the 'index' (previously reported) SNPs in our trans-ethnic analyses would be directional consistent with the risk allele in the previous reports (binominal p<0.05, 170 tests, 0.5 probably of success on each). I anticipated that we would see an even greater proportion of the loci generalize in the trans-ethnic fixed-effect analyses (>25%, based on previous reports [52]; defined as directional consistency and statistically significance, p<0.05/166), due to the increased sample size. However, I hypothesized that a handful of these loci may have evidence of racial/ethnic group heterogeneity; therefore the Bayesian trans-ethnic meta-analysis would relax the assumption of fixed-effects and, at times, may note a different 'top' trans-ethnic SNP. I expected that the 99% credible set would help narrow the interval of putative interest for most loci (>16 loci). Lastly, based on previous reports in a large sample of European descent [33], I thought that conditional analyses would replicate/reveal a small number of (≤5) secondary or tertiary independent associations at these loci (p<0.05/independent tests per locus, r^2<0.2 in African Americans), providing additional information about the genetic architecture of BMI loci in diverse populations.*

D. Public Health Implications

Several US racial/ethnic groups, including Hispanic/Latinos, are oversimplified with respect to their diversity or understudied altogether in public health and genetics [40]. In contrast, the above dissertation aims were conceived with diversity in mind. Furthermore, my dissertation manuscripts fill two gaps in the epidemiologic literature on the implementation of diverse alternative measures of body weight (Aim 1), and diverse ancestral populations in genetic obesity research (Aim 2). I describe the public health implications of each aim separately below.

The diversity of Hispanic/Latino backgrounds captured across the study sites of the HCHS/SOL is unprecedented and reflect multiple subgroups based on self-reported Hispanic/Latino heritages and country of origin. With regards to Aim 1, this work adds novel knowledge to the literature by describing the accuracy of self-reported weight among multiple backgrounds of US Hispanic/Latinos, which has not been investigated previously. Alternative measures of body weight can be used to calculate BMI, a commonly used proxy measure of obesity in epidemiology, or can be used to track changes in self-reported weight in epidemiologic studies of dynamic US immigrant populations, such as Hispanic/Latinos. For example, the genetic analyses conducted as part of Aim 2 included studies of BMI with both measured and self-reported weights. Knowledge of the overall direction of misreporting (under versus over) and the predictors thereof are imperative for the design of future epidemiologic studies and etiologic analyses, such as the one conducted in Aim 2.

Hispanic/Latinos are now the largest US minority group, but there are limited data on their genetic risk factors (addressed in Aim 2) and the environmental origins of their obesity disparities. The generalization and fine-mapping of obesity loci in Aim 2 are an important first step towards prioritization of strong candidates SNPs for future epidemiologic studies in diverse populations in addition to follow-up and targeted functional follow-up. An improved understanding of the genetic architecture and its population diversity in complex traits such as BMI will undoubtedly improve our ability to tailor interventions for diverse populations such as Hispanic/Latinos, as we attempt to translate statistical associations to public health interventions.

My two dissertation manuscripts provide a strong basis for future research on the complex interplay between the sociocultural and genetic determinants of obesity (see conceptual diagram in Figure 1) and the future study of how US Hispanic/Latinos assimilate to the obesogenic patterns of diet and physical activity in the US, which in turn

may have a detrimental effect on their health. In summary it is my hope that this large body of work in genetics will result in a deeper understanding of the etiology of obesity, how obesogenic environments are embodied, and what might be potential modifiable targets for public health interventions for obesity among US Hispanic/Latinos.

E. Supporting Figure

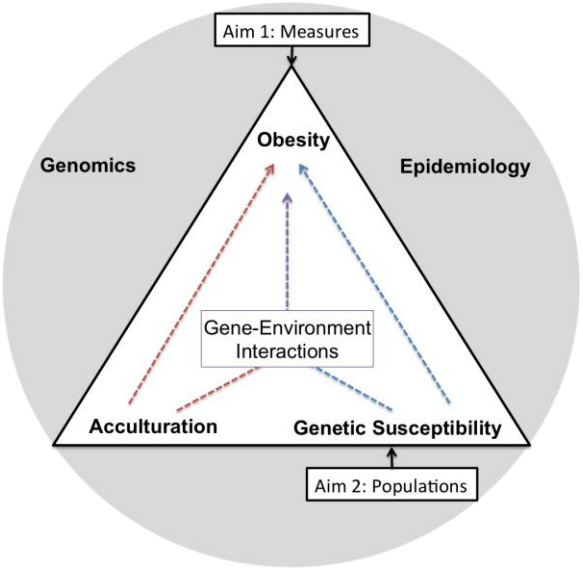


Figure 1. Conceptual diagram of alternative measures (Aim 1) and diverse populations (Aim 2) in genomics and epidemiology research on the complex interconnections between genetics, acculturation and obesity.

CHAPTER III: BACKGROUND AND SIGNIFICANCE

A. Overview

In this next chapter I will provide the background and significance for this dissertation. First, I will describe the global obesity epidemic in terms of how it is defined, its trends, and then in terms of its established risk factors. Second, I will give a brief summary of the diversity of US Hispanic/Latinos and their obesity disparities as compared to non-Hispanic/Latino European Americans as well as across the varied Hispanic/Latino heritages represented in the US. Third, I will describe self-reported weight as a measure of obesity commonly available in large epidemiologic studies, and its accuracy in Hispanic/Latinos. Fourth, I will focus on how diverse measures and populations have been applied to the study of obesity and the common genetic determinants. Lastly, I will present our conceptual model supporting future work on the complex interactions between the underlying determinants of obesity, genetic and environmental, in diverse populations.

B. The Global Obesity Epidemic

Obesity is a global health epidemic and major CVD risk factor [1]. As described in more detail below, BMI (the ratio of weight to height squared) is a perfect measures of body mass, but an imperfect index of body fatness as it unable to differentiate between fat-free and fat mass [54]. This index was originally described by Quetelet, a Belgium mathematician and astronomer, for its proportional relationship between weight and height squared among adults [55]. It then re-emerged in the early 20th century to describe to “desirable” body weights for the Metropolitan Life Insurance Company [56].

The limitations of initial recommendations based on these tables are well documented (e.g. data collected for these tables utilized both measured and self-reported heights and weights, included only non-Hispanic individuals from the United States and Canada 25-59 years without chronic conditions, and included smokers). However in spite of these initial limitations, BMI has since been validated with other more detailed measurements of fat mass, mortality, and CVD diseases [55], which I will discuss in more detail below.

The components of BMI (weight and height) are easily reported or measured in a variety of settings (e.g. home, clinic) as they does not require special equipment. Therefore weight and height have been included and BMI calculated in many large and nationally representative epidemiologic studies [55]. If the rankings of BMI and other measures of body fatness are similar [54], quantitative differences between BMIs of individuals at the same time or changes within individual in BMI over time may still be informative in studying both trends in the overall obesity epidemic and individual health trajectories. For this reason, surveillance of the obesity epidemic has historically been conducted by tracking the temporal and geographic trends in the distributions of BMI or prevalence of under- ($<20 kg/m^2$ among adults, $<5^{th}$ percentile among children/adolescents), normal- (20 to $<25 kg/m^2$, 5^{th} to $<85^{th}$ percentile), over-weight (25 to $<30 kg/m^2$, 85^{th} to $<95^{th}$ percentile) and obese ($\geq 30 kg/m^2$, $\geq 95^{th}$ percentile) individuals, which can be further divided into obesity classes of increasing severity [56, 57]. In order to understand the obesity epidemic, both the temporal and geographic trends should be reviewed.

1. Temporal Trends

The prevalence of obesity remained around 10% in the 1970s and early 1980s, but began notably increasing in the US in the mid-1980s [2, 58]. As of 2010 the age-adjusted prevalence of obesity among US civilian non-institutionalized adults 20 years or

older was 36%, with another one-third of the population being overweight. Likewise nearly one-third of US children and adolescents (2-19 years) were obese or overweight in 2010 [3, 59]. There is some preliminary evidence to support a plateauing of the epidemic in some subgroups of the US [2, 3]. Yet other work indicates that the epidemic, in particular the most severe forms of obesity, may still be growing among US children and adolescents [60, 61].

2. Geographic Trends

As Westernization had been exported across the globe, so too has the obesity epidemic. As of 2008, 10% of the world population was obese [62]. Traditionally, developed countries like the US have been at the forefront of the epidemic. However, more recently, developing countries have undergone transitions from under- to over-nutrition, especially in urban settings. For example, in 2008, Mexico surpassed the age-adjusted prevalence of obesity in United States (33% versus 32% among adults 20 years or older) [45].

Although globally childhood/adolescent obesity prevalence has historically remained lower than adult prevalence, concerning increases have occurred [3, 63], especially in developing countries with emerging economies where the prevalence can exceed 30% among preschool aged children [62].

C. Risk Factors for Obesity

There is a wealth of CVD literature on the determinants of obesity. Using a broad frame of reference, I have grouped risk factors into six categories (age, period and birth cohort; energy balance; sex and gender; smoking; genetics; and psychosocial, socioeconomic and –cultural factors) to highlight the interconnections between several sets of factors. Intentionally I have focused on the obesity risk factors that are most relevant to this dissertation. These broad categories relate to both inter- and intra-

individual characteristics and environments, further indicating how adult obesity is a multifactorial public health concern that requires methodologies that can study multiple contributors (risk factors) simultaneously.

1. Age, Period and Birth Cohort Effects

Excessive weight gain beyond what is required for normal development and maturation at any time point in the life course is concerning and can have metabolic implications. Weight gain early in life is a public health concern as overweight or obese children and adolescents are at extremely heightened increased risk for overweight or obese in adulthood and therefore may remain at higher risk for CVD throughout their lives [64-67].

Similarly, weight gain in adulthood may lead to the development of overweight/obesity for the first time in adulthood or the intensification of obesity-related CVD risk. For example, longitudinal studies of weight gain have shown that it predicts a number of CVD risk factors throughout adulthood in a number of populations including Mexican Americans [68, 69]. With respect to outcomes, weight gain since 18 years old (based on self-report) has been shown to a risk factor for all-cause mortality later in life (specifically CVD-related deaths and non-smoking-related cancers) in the Nurses' Health Study [70]. This association was independent of one's early adulthood body mass (self-reported BMI at 18 years of age). Importantly, weight gain in adulthood is more likely to lead to the deposition of extra weight in the abdomen, where fat is most metabolically detrimental [71].

Adults have been at the forefront of the obesity epidemic as compared to their children/adolescents [58]. Yet in addition to age-related disparities, there have been marked period effects in the obesity epidemic since the 1980s. The interaction of these two time-related factors (age and period) produces susceptible cohorts. For example,

individuals born in the US in the midst of the obesity epidemic (after 1980) carry a larger burden of obesity than seen in other cohorts [72]. Although the exact genetic and environmental etiologic mechanisms are unclear, weight gain during adulthood may exacerbate the public health burden experience by birth cohorts from the obesity epidemic [71].

2. Energy Balance

Presumably the current obesity epidemics seen in developed nations like the US are due to a continued imbalance between energy intake and expenditure [58]. In developing countries transition from a tradition to a Westernized diet, the adaptation of an urbanized lifestyle or other cultural practices ('reverse acculturation') has been associated with a nutrition transition leading to population-level increases in obesity. However, on an individual level poor diet or physical activity have been challenging to quantify and change in epidemiology and public health.

3. Sex and Gender

Women tend to have higher percentages of body fat than men, which they store preferentially in subcutaneous rather than visceral depots. As such at a given BMI (e.g. 30 kg/m^2) women on average have higher percentages body fat and less fat free or bone mass than men [58]. Among parous women, both gestational weight gain and post-partum weight retention are considered risk factors for obesity [73]. Moreover, independent of aging effects menopause has been documented to associate with increased weight gain and shifts in body composition towards more central adiposity [74, 75].

In addition to evidence of a biological basis of differences in obesity prevalence between men and women, women's weight appears to vary more across SES conditions, racial/ethnic groups, and nativity than men's weight. This indicates that body

weight and obesity may also be engendered through complex social processes or cultural roles.

4. Smoking

Generally current smoking and tobacco use has been associated with lower average body weight and therefore less overweight and obesity [58]. Yet this generalization does not hold for all current smokers. While it has been demonstrated that weight is generally lower among adult smokers (ages 25-44), and higher among former adult smokers, this trend has not been found in younger smokers (ages 16-24) [76]. In addition, suggested weight control effects of smoking may dissipate over time, as long-term smokers (20 years and older) have been shown to be heavier than never or former smokers, and heavy smokers are more likely to be obese, and have greater abdominal obesity, than both other smokers and non-smokers [77, 78]. Additionally, an individual's weight may increase after smoking cessation [58]. Evidence supports the role of nicotine in metabolic pathways, although the biologic mechanisms are still unclear.

5. Psychosocial, Socio-Economic and -Cultural Factors

Psychosocial stress [79] and depression [80] are risk factors for adult obesity. More specifically a variety of sources of perceived psychosocial stressors, such as work and caregiving stress, childhood adversity, and financial insecurity have been associated with modest increases in obesity over time [81]. Previous reports also describe increasing obesity prevalence with lower SES; yet the mechanism is likely complex and may be bidirectional in nature (i.e. obesity may influence one's ability for SES advances through the life course) [82]. When studies have jointly assessed the influence of socio-economic and -cultural factors on BMI and obesity, variability in SES-gradients is evident by racial/ethnic group. Furthermore, the differences in BMI across racial/ethnic groups are not completely accounted for by SES differences. This observation highlights the

distinct role that sociocultural factors may play in determining one's body image and access to health care, which could impact the accuracy of self-reported measures (e.g. Aim 1 of body weight), and in determining an individual's burden of obesity (e.g. gene-environment interactions).

Acculturation is a multi-dimensional process of cultural adaptation that begins when individuals from more than one culture come into continuous contact with each other, which results in the maintenance and development of cultural practices with one or both cultures [9, 83, 84]. Work by Berry has outlined four potential strategies during the process of acculturation: integration, assimilation, separation and marginalization [9, 84].

Yet the role of acculturation in public health disparity research and its operationalization has been widely debated [85-89]. The concept of acculturation has been criticized in the field of public health for not involving the structural factors that might also account for health disparities [90]. Additionally, uni-directional and uni-dimensional measures of acculturation are based on the assumption that the host culture is static and that any observed change in the individual is due to their assimilation in the host culture. Such measures therefore have been criticized for their ability to differentiate between the varied components of acculturation (e.g. individuals that have successfully integrated the cultural influences of both cultures to their benefit, and those who have become equally marginalized from both cultures) [84].

6. Genetics

Above I have already described some of the environmental risk factors for obesity. Yet obesity is a multifactorial disease due to a combination of environmental and genetic influences [91, 92]. Estimates of the heritability of obesity range between 40-70%, with single gene disturbances (monogenic) accounting for less than 5% of severe

cases of obesity. Monogenic forms of extreme obesity can be grouped into three main etiologic categories [91]. First, several occur in the genes involved in the hypothalamic leptin-melanocortin system that regulates energy balance. Second, the genes involved in the neurodevelopment of the hypothalamus can also be distributed to produce extreme obesity. Lastly, extreme obesity has been associated with a handful of pleiotropic syndromes due to the dysfunction of the primary cilium. Given that monogenic forms of extreme obesity explain a small portion of the heritability of obesity, most of the heritability may be explained by less impactful but more common genetic variations. Below I will describe the current state of evidence on the population-level genetic influences on obesity, which are the exposures of interest in this dissertation.

D. United States Hispanic/Latinos

Hispanic/Latinos comprise the largest US minority group. I will now describe their diversity with respect to geographic location, Hispanic/Latino background, immigration status, language, culture, and genetic ancestry. Lastly I will summarize how US Hispanic/Latinos have been included in epidemiologic research to date and the potential for innovation in the field of Hispanic/Latino health.

i. The Term ‘Hispanic/Latino’ Defined

The ethnic group Hispanic/Latino can include individuals with ancestry or heritage from Latin America or Spain [4]. Whereas ‘Hispanic’ may refer either broadly to the peoples from Hispania (or Iberian Peninsula, including both Spain and Portugal) or more narrowly to only the Spanish speaking peoples from Spain or Latin America (Figure 2, countries highlighted in green in panels C and B, respectively), ‘Latino’ refers only to the peoples from Latin America (Figure 2, panel A), regardless of language [93].

The US government uses ‘Hispanic or Latino’ interchangeably and defines it as a descriptor of a person of “Cuban, Mexican, Puerto Rican, South or Central American, or

other Spanish culture or origin regardless of race" [4]. Thus as shown in Figure 2, the union of these two terms in the US can encompass individuals living in the highlighted countries in Panel B, as well as those living in the US in Panel D [93]. Not surprisingly peoples from other countries conceptualize the individual terms 'Hispanic' and 'Latino' as well as their union differently [94], and individuals from certain regions of the US may prefer one term over another.

As the US government has defined Hispanic/Latinos, individuals with ancestry or heritage from Portugal, Belize, Haiti, Brazil or other predominantly non-Spanish speaking countries in Latin America are often not officially considered to be Hispanic/Latino by the US government [93]. Yet individuals with ancestry or heritage in these countries may still self-identify as Hispanic/Latino in their home countries or in the US even though they would be or are not included in government enumerations and initiatives. In spite of the debate around the term Hispanic/Latino, in this dissertation I have operationalized it as the US government and its funded public health projects have done prior (e.g. HCHS/SOL).

ii. Population Growth

After nearly a six-fold increase in the Hispanic/Latino population since 1970 [95], US Hispanics/Latinos represented 16% of the total population in 2010 [4] and more recent US estimates place Hispanic/Latinos at 17% of the US population [96]. According to the US Census, the most rapid Hispanic population growth occurred between 2000-2010, when 43% of the total growth occurred [4]. Although historically the growth in the Hispanic/Latino population has been driven by waves of immigration (in particular from Mexico), between 2000 and 2012 Hispanic/Latinos US births accounted for 60% of the population growth [95]. In some regions of the US the population growth of the Hispanic/Latino population is several times that of the general population. The five states

with the largest estimated growth of the Hispanic/Latino population are Tennessee, South Carolina, Alabama, Kentucky, and South Dakota (163-132% percent change 2000-2012). It is estimated that nearly one in three Americans will be Hispanic/Latino by 2060. I will now complete my description US Hispanic/Latinos by acknowledging the complexity of the geographical distribution of US Hispanic/Latinos of varying backgrounds, their immigration statuses as well as their linguistic, cultural, and ancestral heritages.

1. Diversity

Just as there are notable regional differences across and within Hispanic/Latino Spain and Latin America, so does this diversity resurface in the US in a conglomerate ethnic group like Hispanic/Latinos. In light of the global obesity epidemic and the potential for disparities in such epidemics, it is important to revisit the intra-ethnic group diversity in geographic location, Hispanic/Latino background, immigration status, language, culture, and ancestry that characterizes US Hispanic/Latinos. Each aspect will be discussed in turn below.

i. Geographic Location, Immigration Status and Background

The majority of Hispanic/Latinos reside in the West (41%) and the South (36%). The US states of California, Texas and Florida contain 75% of the US Hispanic/Latino population, with an additional 6% living in the border states of Arizona and New Mexico [4]. However, Hispanic/Latinos can be found in every state with notable pockets of ethnic enclaves in Washington, Kansas, Idaho, Oklahoma, Nebraska, Colorado, around Chicago, IL, and along the East coast from New York to Virginia. Residential segregation has been documented among US Hispanic/Latinos, which may in part explain the origins of immigrant enclaves within the US [97, 98]. For example, residential segregation may occur [97] as undocumented immigrants settle in regions of US where they have safety

nets of family and friends (who may also be undocumented) or easily accessible work opportunities [99].

The term ‘US Hispanic/Latino’ combines individual of multiple backgrounds. For example, the majority of Hispanic/Latinos captured by the 2010 US Census have ancestry (or heritage) from Mexico (63%), followed by much smaller populations of individuals of Puerto Rican (9%), Central (8%) or South American (6%) ancestries [4]. Mexican and Central Americans primarily reside in California, while Cuban and South Americans reside mostly in Florida or other states of the US South. Puerto Ricans (not living in Puerto Rico) and US Dominicans mostly reside in the areas around New York City, NY. Salvadorans were the largest group in and around the nation’s capital.

Perhaps not surprisingly many communities in Latin America have sister communities in the US where a substantial fraction of their former inhabitants now reside. For example, as shown in Figure 3 studies of remittances sent from the West Coast of the US to Mexico by the World Bank Group shows that most Mexican Americans living in US states along the US-Mexico border send remittances to Northern Mexico, whereas Mexican Americans living in more northern locations may have ties to central and southern Mexico [100]. The same pattern is replicated for Mexican Americans residing in other regions of the US and speaks generally to the non-random nature of where US Hispanic/Latinos reside in the US. As such there is an inherent ‘confounding by geography,’ wherein it is difficult to disentangle the influences of geography and Hispanic/Latino background [101]. For this reason, community-based studies of Hispanic/Latino studies, such as HCHS/SOL, are limited in their ability to disentangle the two influences on CVD.

In addition to geography and background differences in the US, there are notable differences across Hispanic/Latinos with respect to immigration status, which can determine an individual’s opportunities, resources and therefore their perception of

discrimination or individual rights [102]. The Pew Hispanic Center estimates that currently there are more than 11 million undocumented immigrants living in the US, the majority of whom are from recent waves of immigration from Mexico or other Latin American countries [103]. Furthermore, they estimate that a subset of undocumented Hispanic/Latinos, perhaps as much as 15%, did not participate in the 2010 US Census and therefore the current population estimates are an underestimate. There are a variety of ways immigrants can become 'undocumented', the majority have either overstayed their visas or entered without prior authorization across the US-Mexico or US-Canada border [103].

The US-Mexico border is the most violent border in the world between two countries not at war [104]. Every year while trying to cross innumerable undocumented immigrants experience trauma or worse and thousands experience death [104] along the US-Mexico border from imprisonment, kidnappings, assault, robbery, torture, dehydration, snake bites, and exhaustion [105]. Traumatic migratory experiences of foreign-born Hispanic/Latinos shape their health and can cause a higher burden of both morbidity and mortality after arrival to the US [105]. This potential health gap is further widened because undocumented immigrants suffer from a host of constrained choices and environments [106]. In addition to residential segregation, undocumented immigrant are less likely to have health insurance, obtain/use legal representation, assistance from federal or state assistance programs, or pursue a higher education [107, 108]. They tend to work in unskilled or manual labor occupations, live in poverty, and experience less upward mobility. They are more likely to be paid less for the same work [109] and perhaps be exploited by their more permissive employers.

Migratory experiences of documented or undocumented immigrants, as well as Hispanic/Latinos families and communities may shape how Hispanic/Latinos thrive in the US and impact the findings of research in Hispanic/Latino health. Given the sensitivity of

the matter, many large epidemiologic studies of Hispanic/Latinos such as HCHS/SOL have opted not to inquire with participants about their immigration status. On an aggregate level, undocumented immigration may drive differences between US Hispanic/Latino backgrounds groups that are not explained by traditiona CVD risk factors, SES, etc. (e.g. Puerto Ricans who are all US citizens versus Mexican Americans, many of whom may have immigrated to the US without authorization).

ii. Linguistic and Cultural Identities

As mentioned above the operationalization of the ethnic group Hispanic/Latino in the US relies on “Spanish culture or origin” and is thus limited to individuals with heritage in one or more Spanish-speaking countries [4]. Based on this operationalization one might assume that all Hispanic/Latinos would be linguistically and culturally homogenous. Yet US Hispanic/Latinos are very linguistically and culturally diverse.

With regards to linguistics, not all Hispanic/Latino immigrants may have spoken Spanish as their first language in their Latin American home country prior to emigrating. For example, the 2010 Mexican Census estimated that 7% of the Mexican population spoke one of 60 indigenous languages [110]. Additionally, cultural diversity is a byproduct of centuries of mixing of indigenous, Spanish, and other cultures in Latin America. According to a survey by the Pew Hispanic Center, less than a third of US Hispanic/Latinos consider the designation ‘Hispanic/Latino’ to represent a common underlying culture [94]. This can be seen in the creation of mixed identities, such as the identity of ‘Mestizo’ in Mexico, which captures both the cultural and ancestral mixing of Spanish and native cultures in their history [111].

iii. Ancestry

There is a substantial amount of genetic diversity in US Hispanic/Latinos. Three primary ancestral groups, American Indian, Europeans, and Africans, are known to

contribute to the current genetic diversity seen in Latin American countries [112-116]. However, the genetic diversity within each of these three ancestral groups has yet to be fully appreciated in studies of Hispanic/Latinos [111, 117]. Although remarkable genetic diversity across some American Indian populations has been described [117], no systematic survey has informed the breadth of this diversity across the Americas. From historical records we know that prior to Columbus' arrival in the New World in 1492, the Iberian Peninsula was very diverse and contained Iberians, Celts, Greeks, Romans, Sephardic Jews, Arabs, Gypsies alike [111]. Although African slaves being brought to the New World were classified according to their port of departure [111], they likely originated from other ancestrally diverse locations in Africa [118].

Many Hispanic/Latinos do not identify with a particular race categorization. For example, in the 2010 US Census, 31% of Hispanic/Latino respondents identified themselves as 'some other race' or chose to not respond, resulting in a non-response rate for race of 13%, nearly three times that of the general US population [119]. In the Multi-Ethnic Study of Atherosclerosis (MESA) study, 29% of self-identified Hispanic/Latinos had their self-reported "racial/ethnic classification" reassigned by principal component analysis (PCA)-based racial groupings, which were constructed using genotypes from 96 ancestry informative markers, Ward's minimum variance method, and *K*-means clustering algorithm to identify four clusters of ancestrally related individuals. Reclassification was highest for Hispanic/Latinos than any of the other racial/ethnic group [European (12% of sample reassigned), African (11%), or Chinese American (<1%)] [120].

Although, racial classification into discrete groupings is problematic in general, this MESA study and other related data demonstrate the heightened difficulty of racial identification in the highly admixed ethnic group designated as 'Hispanic/Latino' as one's self-conception of race are often grounded in linguistic and cultural factors in addition to

knowledge about their ancestral genetic background. Although the federally funded HCHS/SOL had been mandated to collect information on race, similarly to what was done on the US Census, HCHS/SOL investigators have subsequently chosen to not include race in their descriptive and association analyses of CVD. In the next section I will continue to discuss how generally diverse Hispanic/Latinos have been incorporated into public health research to date and focus on the opportunities to improve upon the current practices in Hispanic/Latino health research.

iv. Inclusion in Epidemiologic Research

Future epidemiologic research on Hispanic/Latinos should appropriately acknowledge or account for the various origins and/or diversity in US Hispanic/Latinos. In spite of the current body of research on Hispanic/Latinos, studies designed to investigate and highlight the diversity Hispanic/Latinos will yield the greatest benefit in epidemiology [111]. Stratification of results by self-identified Hispanic/Latino background group may allow researchers to assess the impact of background on observed heterogeneity, but it may only capture portions of the other components of diversity described above, which include geographic, immigration, linguistic, cultural and ancestral factors.

In the field of genetic epidemiology, it is important to account for ancestral differences given that genetic diversity may be confounded by the aforementioned other aspects of diversity and confound observed health associations. Although it is common practice in genetic epidemiology to account for this confounding by ancestral diversity (population stratification), it is also possible that the other sources of diversity may act as potential effect modifiers. With this in mind, this dissertation aims to explore the role that diverse ancestry may play in determining genetic susceptibility to increased BMI (Aim 2).

2. Disparities in Obesity

Having described the trends in the obesity epidemic as well as the diversity of US Hispanic/Latinos, here I will describe the current disparities in obesity and the current understanding of its underlying causes.

i. Adulthood

Within the US there are striking disparities in obesity prevalence, which are masked by looking at just overall national estimates as I had done in my description of the temporal and geographic trends in the obesity epidemic. For example, as of 2010 non-Hispanic/Latino White adults 20 years or older were estimated to have the lowest age-adjusted prevalence of obesity (34%, civilian non-institutionalized) [2]. Hispanic/Latinos (39%) and non-Hispanic/Latino Black (50%) adults had a higher burden of obesity. The Hispanic Health and Nutritional Examination Survey in 1982-1984 was the first to show that the burden of obesity may not be similar across all background groups of US Hispanic/Latinos [121, 122]. Restricting the 2010 estimates of obesity prevalence to just Mexican Americans demonstrated a slightly higher proportion were obese for this background group than overall for Hispanic/Latinos (40%). More recent nationally-representative estimates of obesity across the US Hispanic/Latino backgrounds are currently lacking.

In this regard, community-based studies, like HCHS/SOL, may be a helpful snapshot of heterogeneity in the burden of obesity in US Hispanic/Latinos. In the HCHS/SOL communities a slightly smaller proportion of Hispanic/Latinos were obese (37%) than in contemporary national estimates [6]. Yet the burden of obesity across Hispanic/Latino backgrounds was indeed highly variable—with South American (27%) and Puerto Rican (41%) adults representing the ends of the spectrum in obesity prevalence [42, 43]. Nonetheless the causes of heterogeneity remain unclear due, in

part, to the confounding by geography mentioned above of Hispanic/Latinos in the US [42, 43].

ii. Childhood and Adolescence

Unfortunately obesity disparities are even more pronounced among US Hispanic/Latino children and adolescents than in adults [2, 3]. Whereas in 2010 14% of non-Hispanic/Latino White children and adolescents (2-19 years) were obese, a higher burden of obesity was shouldered by Hispanic/Latino (21%) and non-Hispanic/Latino Black (24%) children and adolescents. Although considering all of childhood and adolescence it appears that the prevalence of obesity in Mexican Americans is similar to the large Hispanic/Latino designation, an alarming trend towards obesity can be seen in particular among Mexican American adolescent boys (12-19 years). This observation is supported by recent work in the National Longitudinal Study of Adolescent Health, which among the Hispanic/Latino backgrounds saw the largest gains in BMI between 12-32 years in adolescents of Mexican (males) or Puerto Rican (females) ancestry or heritage [123]. Adolescents of Central/South American, Cuban or other backgrounds gained body mass at or below the non-Hispanic/Latino White adolescents in the study.

iii. Non-Genetic Determinants of Obesity

This brings us to the question of what might be key underlying determinants of obesity disparities for US Hispanic/Latinos, as well as what are potential sources of heterogeneity across this ethnic group's diverse backgrounds. One possible component is the high proportion of Hispanic/Latinos who do not have health insurance and therefore affordable access to health care services [124, 125]. Although the Affordable Healthcare Act was enacted to equalize access to health care in the US, it does not include undocumented immigrants who are estimated to collectively amount to more than 11 million individuals [103] and may increase barriers to health care and further

marginalize this vulnerable population [105]. In 2010, 34% of all US Hispanics <65 years of age did not have health insurance and 45% of US Hispanic/Latinos in families earning <200% of the poverty line were uninsured [126]. At the same time 14% of non-Hispanic Whites and 21% of non-Hispanic Blacks were uninsured. Lack of insurance varies substantially by Hispanic/Latino background groups (e.g. from 50% of Hondurans to 15% of Puerto Ricans were uninsured in 2010) [127], and tends to be highest among background groups that have the highest proportions of undocumented immigrants such as immigrants from Mexico (34%) and Central America (41-50%) [103, 127]. When lacking adequate clinical monitoring and management in roughly half of all US Hispanic/Latinos <65 years old [124, 126], inequitably some individuals may be subjected to an array of adverse environmental and lifestyle factors as they assimilate to the US resulting in poor population-level health outcomes [44]. A systematic review of mortality disparities in US Hispanic/Latinos (as compared to the general US population) revealed that although the greatest disparity is seen in diabetes-related mortality, Hispanic/Latinos also suffer from mortality disparities in a number of other conditions including some cancers, liver disease, HIV, homicide, and work-related injuries [125]. Diversity across Hispanic/Latino background groups may related to barriers to health care (e.g. citizenship or legal resident requirements for Medicaid/Medicare and Affordable Care Act, type and location of employment opportunities, language preference) and result in both lower seeking and receipt of healthcare services among Hispanic/Latinos.

Another determinant of the obesity epidemic and the observed disparities relates to the social determinants of health, which can include poverty, trauma, stress, discrimination, unskilled or unreliable employment [125]. These health determinants have been linked to allostatic loads and hypothesized by Marmot to relate to a ‘status syndrome’, characterized by lower participation in and sense of control over their

surroundings [128]. Low socioeconomic status (SES) has been linked to rates of obesity, metabolic syndrome, and mortality [125].

Additionally, the sociocultural environment and an individual's strategy of acculturation [9, 84] may be key determinants in the patterning of diet, physical activity, obesity disparities. In the absence of the time and the resources to measure acculturation using detailed scales, proxies of acculturation have become common in epidemiologic studies of Hispanic/Latinos [9, 83]. In a systematic review of the public health literature by Thomson and Hoffman-Goetz, nearly a third of studies of Hispanic/Latinos relied on one or a combination of proxies of acculturation (language preference, nativity, time in the US, language preference, etc.). The need to balance practicality with validity is important to the study of the effects of acculturation. According to their review of the literature when both proxies of acculturation and detailed scales have been assessed in the same study, the correlations varied across the scales (r=0.17-0.76).

A number of cross-sectional studies have investigated acculturation and obesity among US Hispanics/Latinos and have shown positive associations between acculturation and measures of adiposity, which vary by background [122, 129-138]. In the cross-sectional literature on this topic, time living in the US is a consistent cross-sectional predictor of increasing weight status, independent of age, and shows evidence of a threshold effect after 10 years in the US [8]. A number of other measures of acculturation, such as age at immigration, generational status, language preference at examination, nativity, and the Short Acculturation Scale for Hispanics, have previously shown the most acculturated Hispanic/Latinos to carry the largest obesity burden; however, the results from studies using these measures have generally been less consistent [9, 83].

These cross-sectional studies have led to a number of competing hypotheses about the underlying pathway between acculturation and obesity. Hispanic/Latino immigrants have been documented to be healthier than their US-born peers, in what has been described as the 'healthy immigrant effect.' Explanations of this pattern are similar to the 'Hispanic paradox' described above [125] and have revolved in part around the selective migration of the healthiest individuals from the sending countries as well as retention of protective cultural practices such as a healthier diet and physically active lifestyle [86, 139]. Others have pointed out that return migration due to immigration enforcement, retirement, or health concerns (i.e. 'salmon bias') [125] may create a reverse selection bias that could mask or accentuate observed differences between foreign- and US-born Hispanic/Latinos in cross-sectional and longitudinal studies alike [140, 141].

'Social adaptation' provides another alternative explanation of disparities and has been cast in both a positive and negative light in the current literature [125]. In contrast the 'unhealthy assimilation' hypothesis has been posited as an explanation for the cross-sectional observation of the effect of increasing duration in the US [138, 141, 142]. If recent Hispanic/Latino immigrants are exposed to obesogenic environments as they assimilate to the US then they would gain weight faster than native-born Hispanic/Latinos until their weights converged. In a more positive light, 'divergence' describes the possibility that more recent immigrants may be more likely to maintain cultural practices, which support healthy lifestyles with regards to diet and physical activity, as they negotiate the process of establishing new relationships in the US.

A handful of longitudinal studies [142-145] and repeated cross-sectional studies [140, 141] (Table 1) have sought to test the hypotheses of an 'healthy immigrant effect' followed by 'unhealthy assimilation' to the US by testing for baseline differences in obesity by generational status and then assessing the rates of weight gain between

foreign- and US-born Hispanic/Latinos. However, the findings have been far more mixed than seen previously in the cross-sectional literature [122, 129-138].

Within the body of literature some studies have noted incomplete mediation of the cross-sectional [139, 146, 147] or longitudinal [148] effect of acculturation on obesity by diet or physical activity, which may indicate that other sociocultural or environmental factors other than diet or physical activity may mediate the influence of acculturation (assimilation to US society) on obesity [8, 147]. Specifically, the association may be mediated by coping strategies for the stress related to immigration, discrimination, or other characteristic of being Hispanic/Latino in the US [149]. Other researchers propose that 'segmented' (unequal) assimilation [150] or structural factors not captured by the individual concept of acculturation [90] may in turn compromise the health of certain segments of the Hispanic/Latino population [145].

In summary restricted or inconsistent access to health care or SES disadvantage may interact with geographic, Hispanic/Latino background, immigration status, linguistic, sociocultural and ancestral diversity to determine the patterning of obesogenic environments in the US. This could in turn yield the complex picture of Hispanic/Latino health and health disparities in obesity we see currently in the US [2, 125].

E. Classification of Obesity

In this section I will describe how excess fat mass (or adiposity) is measured in obesity research, with a particular focus on self-reported weight and BMI, a proxy measure of obesity. For each I will present the current evidence on the validity and reliability in Hispanic/Latinos. In the absence of studies in Hispanic/Latinos I will summarize what trends in validity and reliability are seen in other US populations.

Excess fat mass is the hallmark of obesity but can be measured in a variety of manners [54, 55]. An individual's body composition can be measured using costly

precise assessments or using more inexpensive proxies [56]. Each type of measurement plays an important role in public health research.

Detailed reference measurements are often performed in small samples as “gold standards” for body composition and include densitometry, hydrometry, computed tomography, magnetic resonance imagining, dual-energy X-ray absorptiometry, and potassium counting [55]. Often the findings from these small samples are used as references by which to calibrate and validate how less expensive indexes at capturing the components of body composition. Measures based on two-compartment models estimate fat mass and fat free mass, whereas multi-compartment models (e.g. computed tomography, magnetic resonance imagining, dual-energy X-ray absorptiometry, potassium counting) are able to decompose fat-free mass into further subdivisions including body water, protein, or bone mineral. Both types of models have been used to establish the accuracy of less expensive indexes of obesity to be discussed below. As the technological sophistication of these valid and reliable reference methods increases it becomes particularly important to consider their limitations. They are more costly, time-consuming, and often are not easily portable, which impedes their wide-spread use in large epidemiologic studies [54, 55].

On the other hand, there are a number of more accessible anthropometric measures and proxies of obesity, which include weight, BMI, waist circumference, waist-to-hip ratio, skinfold thickness, and bioimpedance [54, 55]. For example, weight changes during a period of adulthood when height is not expected to change substantially tracks well with changes in fat mass [55]. Weight and height are reported or measured in a variety of settings as they do not require additional equipment beyond what is normally found in the home or clinic. BMI can then be easily calculated from an individual’s weight and height, and given this utility has already been integrated into US public health applications [57]. BMI has also been included by many public health researchers in most

large or nationally representative epidemiologic studies [55], many of which also implement standardized measurement or self-reporting protocols [55, 151].

In addition self-reported weight and height have been used in the absence of measured indexes, however with additional error discussed further below. Self-reported or recalled weights and heights may be particularly helpful when information on obesity or its change over time is otherwise unavailable to the researcher. Therefore many large epidemiologic studies (e.g. Atherosclerosis Risk in Communities Study or ARIC, HCHS/SOL, WHI, MEC, MESA, CHS) also capture self-reported and recalled weights at a number of epochs prior to study enrollment in order to capture weight maintenance, gain, or loss across an individual's life course [152]. Yet these self-reported measures may have questionable significance across populations or variable measurement in a variety of settings, making the study of their validity and reliability particularly relevant for epidemiology research.

I will begin a summary of the validity and reliability of self-reported weight (Aim 1), and then describe the evidence for the validity and reliability of BMI, calculated using measured weights and heights (Aim 2).

1. Validity of Self-Reported Weight and its Change

I will first summarize the validity of self-reported weight, which I propose to assess in the first aim of this dissertation. The validity of self-reported weight at both a single time point and multiple time points has been investigated using four perspectives (Table 2). The first two perspectives each isolate two different potential sources of bias when self-reported weights are used in place of measured weights by comparing i) contemporaneous self-reported and measured weight as an estimate of self-report bias, and ii) current self-reported and recalled weight at the same age as an estimate of recall bias. The third perspective results in a combination of self-report and recall bias by

contrasting iii) measured weight and recalled weight at the same age. The final perspective harnesses multiple measures of self-reported and measured weights or weight changes to investigate the difference between iv) self-reported and measured weight over time. Although each perspective may be informative for informing 'validity' in particular study designs/datasets and is summarized below, we will focus on the first perspective, contemporaneous self-report and measured weight, as our preferred approach to isolate the magnitude of self-report bias in diverse US Hispanic/Latinos in Aim 1.

i. Contemporaneous Self-Reported versus Measured Weight

According to a review on the self-report bias in current self-reported weight (Table 2), most studies have described a tendency towards modest under-reporting of current weight (0.1 to 1.2 kg), but there is a large amount of variability in individual reporting [151]. In samples containing Hispanic/Latinos, categories of body mass index [48, 49, 51], aging [47, 49, 51, 153, 154], gender [47, 51], reproductive factors (parity and menopause) [75], household income [51], education, employment and nativity [49] have been described as predictors of self-reported inaccuracy. Yet an important limitation of this literature is that only a handful of validity studies have been completed in Hispanic/Latin American countries, or in US among Hispanic/Latinos. I will now focus on these studies in an effort to compare and contrast the body of evidence most relevant to US Hispanic/Latinos.

From Latin America, a study of a nationally-representative sample of Mexican citizens born before 1951 noted modest over-reporting of weight (0.6kg) and a good correlation (r^2>0.8) between self-reported and measured weights in Mexican adults (r^2=0.84) [47]. Another study of Mexican asthma cases and controls noted under-reporting of weight, but did not describe this observation quantitatively [155]. A

population-based study from Spain documented a trend towards under-reporting of weight and stronger correlations between self-reported and measured weight (r^2=0.96) [46].

In the US, the most recent estimates from the National Health and Nutrition Examination Survey (NHANES, 2007-2008) reveal that the difference between measured and self-reported current weight among non-institutionalized Hispanic/Latinos ≥20 years old appears to be more consistent across gender than observed in other ethnic/racial groups [156]. Specifically, in NHANES researchers found that both Hispanic/Latino men and women tend to under-report their current weight (0.09 and 0.59kg, men and women) and over-report their current height (0.89 and 1.56 cm, men and women). Because men do this to a lesser extent their calculated BMI values are less markedly under-reported than for women (0.29 and 0.79 kg/m^2, men and women). Even though predictors of under-reporting in weight were not directly assessed, predictors of under-reporting in BMI included being overweight and obese, elderly (≥60 years), and college educated, which could distort the population distribution of BMI by underestimating the largest BMIs but may not impact the classification of obesity.

Earlier national estimates among Mexican Americans (1988-1994) indicated that the correlation between self-reported and measured weight was generally good (r^2=0.96) [50] and that there may be more under-reporting of weight among foreign-born as compared to US-born Mexican American women [49]. Although Mexican American men tended to over-report their weight, this was invariant to their nativity. This observation indicates that previous studies of Hispanic/Latino adults in their home countries may not entirely represent the accuracy of self-reported weight we would expect to observe in a sample of both foreign- and US-born Hispanic/Latinos [47]. This may be due to the varying sociocultural influences of diverse Hispanic/Latino backgrounds as well as post-immigration changes in these influences, which may in turn shape an individual's body

image [157]. For example, Mexican American women describe that they have had to transition from Mexican culture that considers a full figure to be desirable to an American standard, which they describe as “extremely thin” and questionably healthy.

Due to the predominance of Mexican Americans in NHANES, the validity of self-reported weight by other Hispanic/Latino backgrounds is still largely unknown. However, another study of post-menopausal women (45-60 years) in Miami-Dade County, Florida found that the under-reporting of current weight (mean under-reporting 1.55 kg, 95% confidence interval 1.25 to 1.85) [75] was greater than national estimates of under-reporting from NHANES for Hispanic/Latina women of multiple backgrounds (difference in means of 0.59kg) [156], but more similar to under-reporting of weight captured by earlier national estimates Mexican American women (mean difference of 1.37kg) [50]. This may be due post-menopausal weight fluctuations. Self-reporting bias was also more marked among the Hispanic/Latina women in their sample (mean under-reporting of weight 1.55 versus 1.51 kg, and of BMI 1.41 versus 1.05 kg/m^2, respectively) than among non-Hispanic/Latina women due to the greater over-reporting of height in Hispanic/Latina women (mean over-reporting of height 2.48 versus 1.62cm, respectively) [75]. The investigators did not provide any information about the Hispanic/Latino background(s) of these women, but given the Hispanic/Latino background distribution in HCHS/SOL [43] one might assume the majority of women were of Cuban heritage. This report indicates further that there is the potential for heterogeneity in self-reporting bias by diverse Hispanic/Latino backgrounds, which could be masked in previous combined estimates of accuracy in Hispanic/Latinos [156].

ii. Current Self-Reported Weight versus Recalled Weight at the Same Age

As a measure of recall bias, the next perspective on the validity of self-reported weight comes from the comparison of self-reported current weights and recalled weight

at the same age (Table 2). Any differences between these two self-reports should be due to the fact that one is made with some time lag (e.g. a self-reported current weight when an individual was 21 years of age, versus the recalled 21-year old weight of the same individual at 40 years old). Estimates of this recall bias are particularly informative for studies of self-reported weight histories, which in HCHS/SOL for example could have be recalled up to 55 years prior to baseline.

Not many US-based studies have two self-reported weights (at the same age) collected as part of multiple waves of data collection across a long period of time in adulthood [158]. I am not aware of any such study that includes samples of Hispanic/Latinos. In a large US report from the Adventist Health Study of older non-Hispanic/Latinos, the correlation of these two self-reports (current weight and recalled weight for an age 26 years ago) was good ($r \geq 0.83$) for self-reported weights for ages 28-72 years [158]. This corresponded to an average under-reporting of self-reported weight (recall bias) of 0.67 kg over a 26-year period, which for example varied from 0.01 kg over-reporting of recalled weights for ages 28-32 years (n=691) to 1.43 kg under-reported of recalled weight for ages 58-62 years (n=384). Likewise the recalled weights for participants 30-64 years old at the second wave of data collection recalled weights were over-estimated by 0.16 kg, whereas participants 65-74 years old tended to under-report their recalled weights by 0.84 kg. Recalled weights were also under-reported more substantially among the obese ($\geq 30 kg/m^2$, versus $<30 kg/m^2$) and ever smokers (versus never smokers). This study from a 26-year period of time indicates that in this population the average recall bias was roughly 0.25 kg per decade of recalled time, but that recalled weights earlier in the adult life course may be more accurate than those later in the adulthood and those made by younger participants may be more precise than those for older participants. Based on their results the authors conclude that the accuracy of recalled weights 20-28 years prior to the inception of a study (e.g. as part of

HCHS/SOL weight histories) may also be minimally biased by recall, but they do recommend that future studies of weight histories consider accounting for this potential bias.

iii. Measured Weight versus Recalled Weight at the Same Age

The third perspective on the validity of self-reported weight compares an individual's actual measured weight and recalled weight for the same age (Table 2). Because any differences between these two measurements (measured and recalled weight) could be due to either self-report or recall bias, this perspective only roughly helps us gauge how much larger these differences are combined than individually (described sections i and ii above), i.e. through an interaction of biases. Again this area of research has been conducted exclusively among non-Hispanic populations [159-161]. In these studies, correlations between measured and self-reported recalled weight at a given age were high ($r>0.8$) and have been reported for recalled weights up to 28 years ago [160, 161]. One study in the bi-racial elderly cohort of the Charleston Heart Study used two periods of recall (28-years and 4-years in the past) indicates that the accuracy of recalled weights as compared to measured weights at the same age may vary across populations. For example, African American men under-reported their weights 28-years in the past less than White men (difference in means, 1.2 versus 1.5 kg, respectively), whereas African American women under-reported these recalled weights more than White women (1.8 versus 2.0 kg). Both African American men and women had more bias in their weights recalled from 4-years prior as compared to White men and women (0.3 and <0.1 kg over-reporting, respectively). However, African American men over-reported their weights 4-years prior by 0.95 kg and African American women under-reported their weights 4-years prior by 1.15 kg.

More recent work from the repeated cross-sections in NHANES I Epidemiology Follow-Up Study (1982-1992) has indicated that the inaccuracy between measured weight and recalled weight at the same age was modest (1.77 kg under-report) over an average period of repeated cross-sections of 19 years [159]. These estimates were generated based on calculating the mean discrepancy between measured weight and estimated recalled weight for the same age based on the predicted recalled weight for the same age as the NHANES I examination (1971-1975) and a best fit line per individual of the recalled adult (20-65 years) weights prior to and after the NHANES I Survey (on average 19 years prior to the follow-up). This report also noted heterogeneity by racial groups in the estimated intercept and slope effects of discrepancy between measured weight and recalled weight at the same age but did not explore them systematically due, which warrants further research in diverse cohorts. Additionally across the range of periods of recall in this sample, this combination of self-report and recall bias was observed to increase with each additional decade of recalled time by approximately 2 kg (estimated range of recall periods, 7-21 years) [159], which is substantially larger than previous estimates of recall bias alone of 0.25 kg per decade [158]. This indicates that self-report bias may be a stronger influence in the study of self-reported weight histories in HCHS/SOL.

iv. Self-Reported versus Measured Weight over Time

The final perspective on the validity of self-reported weight allows us to directly assess its utility in capturing changes over time; however again it represents a less-straight forward method to investigate self-report and recall bias. In addition to comparing the accuracy of self-reported weights at two waves of data collection, the National Longitudinal Study of Adolescent to Adult Health (Add Health) investigators used multiple self-reported and measured current weights of adolescents and young

adults to calculated weight change over a 5 or 6-year period of time. Calculations of weight change based on self-reported weights were on average 1.0-1.3 kg less than the calculated weight change based on measured weights (female versus males respectively, Waves II, 1996 and III, 2001-2002) [48]. Hispanic/Latino ethnicity and other known predictors of true weight change were not related to the discrepancy between self-reported and measured weight change ($p \geq 0.07$). In contrast with the reports of the accuracy of self-reported weight at one-time point, obese participants under-reported their individual weights more consistently than non-obese individuals, which resulted in more negligible under-estimation using calculated change among this group.

Additional work including the Add Health Wave IV (2007-2008) indicated that the under-estimation of weight increased as adolescent girls transitioned into young adulthood, whereas self-reporting bias was not seen among adolescent or young adult boys. Race/ethnicity did not predict changes in self-report bias in either gender group [162]. This study indicates that with some systematic under-estimation of changes, multiple self-reported current weights can be used to capture weight changes in adolescence/early adulthood stably over a range of body mass categories. The calculation of weight trajectories based on a combination of recalled and self-reported current weights, such as in HCHS/SOL, may also underestimate population-level weights, but be a useful tool for studying weight trajectories in otherwise hard to study dynamic populations of Hispanic/Latinos.

2. Validity of Measured Body Mass Index

I will now shift my focus to the validity of measured BMI as a proxy measure of obesity and its downstream health consequences. As previously indicated, BMI is a convenient measure of body mass in large epidemiologic studies, and was initially described and studied in populations of non-Hispanic European descent [55]. The

validity of BMI has been studied in primarily non-Hispanic populations (although some studies of Hispanic/Latinos exist) and related to reference measurements of overall adiposity, mortality, and other outcomes such as CVD. I will describe each line of evidence separately below.

i. Measured BMI versus Obesity

First, BMI correlates with other proxy measures of obesity in a number of populations, but the interpretation varies somewhat by sex, age, and race/ethnicity [55]. For example, in the bi-racial ARIC Study, BMI strongly correlated with weight, waist, and hip circumferences ($r \geq 0.85$) [54]. Separately in a sample of Blacks and Whites from New York City, potentially including some individuals that could have had Hispanic/Latino ancestry, BMI correlated strongly with percent body fat calculated from a four-component model including body weight and other measurements from densitometry, hydrometry, and dual-energy X-ray absorptiometry in both racial groups and in men and women ($r \geq 0.58$) [163]. Sex was a significant modifier of the correlation between BMI and percent body fat consistently over the life course (at 23kg/m^2, 8.7-12.7% difference between women and men). Age also modified this association but far less strongly; older individuals had a modestly increased percent body fat for the same BMI (e.g. 23 kg/m^2) than their younger counterparts—estimated to correspond to an average difference of 0.7 to 1.1 percent per decade.

A number of validity studies of BMI have also investigated racial/ethnic differences in the distribution of percent body fat and fat-free mass [164]. The observations that compared to Whites with the same percent body fat, individuals in other racial/ethnic groups can on average have a BMI between 4.6kg/m^2 lower (Ethiopians) to 1.3kg/m^2 higher (African Americans) BMI values has questioned the utility

of this measure of obesity across diverse populations [165, 166]. Moreover, there is substantial variability within racial groups such as Asians.

It has also been noted that at certain BMI values, US Hispanic/Latinos have lower percentages of body fat than European Americans, and at other BMIs they have higher percentages of body fat [167, 168]. Nonetheless, most recently the correlation between BMI and adiposity was shown in a small but diverse sample from New York City to be comparable between US Hispanic/Latinos and Whites (r^2=0.48-0.60 versus 0.46-0.47) [169], the population in which BMI was first described and categorized. Therefore in order to avoid potential differences in racial/ethnic groups in the adiposity captured by the common categorizations of BMI into overweight (25 to <30kg/m^2) and obesity (≥30kg/m^2) categories, I have opted to study the continuous variation in BMI among Hispanic/Latinos as an indicator of the population-level variability in body mass, a useful tool in identifying obesity (Aim 2).

ii. Measured BMI versus Mortality

Second, a number of studies of mortality have described a J- or U-shaped dose response relationship between BMI and mortality with a nadir around a BMI at 25kg/m^2 [58]. The observation of a non-monotonic dose response curves relates to how BMI combines the influences of two components of body composition: fat mass (a risk factor at high percentages) and fat-free mass (a risk factor at low percentages) [54, 170]. Moreover, the observed dose-response relationship is also sensitive to the age, sex, racial/ethnic and smoking distribution of the population being studied. The association of BMI, a measure of overall adiposity, on all-cause mortality is independent of a variety of measures of central adiposity [171]. The obesity epidemiology literature suggests that excess fat mass or obesity, BMI ≥30kg/m^2, may be a valid indicator of elevated mortality risk due to mechanisms independent of central adiposity [58].

In Hispanic/Latinos mortality studies of BMI have been primarily conducted in Mexican descent populations. One study has shown that the absolute reduction in life expectancy at 60 years of age from severe obesity is slightly larger among Mexicans living in Mexico (7.7 years) than among the general US population or among a sample of Americans of European descent (6.7 and 5-6 years, respectively) [172]. Yet Mexican Americans residing in the US have historically had comparable or less mortality burden from obesity as compared to Americans of European descent [173], which has been described as part of the 'Hispanic paradox' [174]. A recent pooled analysis of 16,798 Hispanic/Latino adults (primarily samples from Puerto Rico and Mexican Americans living in the US Southwest) showed no significant relative increase in mortality with overweight or obesity [175]. Although, substantial heterogeneity has been described in the obesity-related impacts on mortality across Hispanic/Latino background groups [125].

The origins of the Hispanic paradox are unclear [125]. Inherent differences in the populations (e.g. foreign and US-born Hispanic/Latinos and other US-born racial/ethnic groups, SES, age, and/or gender) may mask differences in mortality [176], or the average percentage of body fat for a given BMI may be less impactful among Hispanic/Latinos [167, 168]. It is also possible that the process of immigration to the US selects for a population of healthy immigrants or the improved living conditions in the US as compared to the sending country provide a health advantage for the foreign-born Hispanic/Latinos as compared their US-born counterparts. Another hypothesis contends that reverse selection ('salmon bias') may be the driving factor, wherein US Hispanic/Latino immigrants with health concerns may rejoin with family and receive health care in their country of origin and are subsequently undercounted. Others have pointed out that there may be an under-estimate of mortality rates if Hispanic/Latinos are systematically misclassified as non-Hispanic/Latino. Yet no single theory singly or jointly

appears to accounts for the differentials in health within Hispanic/Latinos, which appear to track with “life-course epidemiologic factors in both sending and receiving nations.” In contrast, Vega and colleagues propose [125] that social adaptation, characterized by “social learning in contexts that supply opportunities, environmental conditions, and psychological reinforcement for health-degrading behaviors that increase in prevalence between generations after immigration,” is a useful paradigm to understand Hispanic/Latino health disparities and these paradoxical findings. Future studies of obesity and mortality in Hispanic/Latinos should consider diverse Hispanic/Latino backgrounds and their inherent characteristics of the populations being studied in order to determine if the paradox is an artifact of confounding factors, or how it may relate to Hispanic/Latino health research [176].

iii. Measured BMI versus Cardiovascular Disease

Lastly, there exists a wealth of literature regarding obesity as a risk factor for future CVD [170]. Briefly the influence of obesity on future CVD is dependent on both the degree of excess adiposity and the duration [177] and partially mediated by obesity’s influence on other CVD risk factors [58]. Given the multiple measurements of BMI, researchers have previously inquired if BMI is indeed the best measurement to capture future risk of CVD. In a study from the United Kingdom, the strength of association between overall (BMI), central (waist circumference, waist-hip ratio) or more direct measures of adiposity (bioimpedance, skinfold thickness with either CVD risk factors or incident coronary heart disease were similar, indicating that BMI performs at least equally well at capturing adiposity-related CVD risk [178].

The most contemporary and diverse estimates of BMI and CVD risk factors in Hispanic/Latinos come from the HCHS/SOL [179]. At the baseline examination BMI categorizations of overweight and three subclasses of obesity (25-<30 and ≥30 kg/m^2)

associated positively with a higher prevalence of CVD risk factors (hypertension, diabetes, low high-density lipoprotein cholesterol, high C-reactive protein and triglycerides) across a wide age range of diverse Hispanic/Latinos. At high BMI levels (≥40kg/m^2) men and younger individuals were more likely to have unfavorable CVD risk factor profiles, including both traditional CVD risk factors as well as emergent risk factors (e.g. C-reactive protein and triglycerides).

This observation is concerning, given that Hispanic/Latinos have both a high burden of obesity and a predisposition to obesity-related disorders, such as diabetes and low high-density lipoprotein cholesterol [179]. Recent preliminary research from NHANES Mexican Americans indicates that with respect to diabetes this may be due in part to the stronger influence of overall adiposity, as measured by BMI, on diabetes as compared to other US racial/ethnic groups [180]. Next I will describe the reliability of the measures of obesity I propose to use in this dissertation (self-reported weight and its change, measured BMI).

3. Reliability of Self-Reported Weight and its Change

Above I have described the validity of the measures of obesity I propose to use in this dissertation. In contrast, here I will describe the reliability (overall consistency over at least two assessments) of weight and its change. With regards to self-reported weight, to my knowledge primarily non-Hispanic European studies have investigated the reliability of self-reported current or past weight using two time points within one year and found that measurement error was generally good [181, 182]. Specially, a study from Potsdam, Germany reported that greater than 75% of middle-aged adults reported a weight that was ±3kg of their previously reported weight (at either 25 or 40 years of age), regardless of gender, education, current age, weight or current BMI [181]. In a large US study of the sisters of women with breast cancer (including a small subset of 322 Hispanic/Latina

women, 1.8% of sample), the reliability of self-reported weight was found to be high (80% of self-reports within 1.36 kg of each other, r=0.99) when it was assessed using two different self-reporting methods (computer-assisted telephone interview and questionnaire) [183].

4. Reliability of Measured Body Mass Index

Now I will discuss the reliability of measured BMI, which I propose to use in Aim 2 of this dissertation. As mentioned above standardized protocols have been recommended for the measurement of weight and height, which are used to calculate BMI in large and national epidemiologic studies [55]. For example in HCHS/SOL, weights and heights were measured on participants in a scrub suit or examination gown without their shoes, using a digital scale (Tanita Body Composition Analyzer, TBF 300A, Tanita Corporation) and standiometer (SECA 222, Perspective Enterprises, Inc.) by trained staff and following a standardized clinic protocol [43]. Additional efforts were made to increase the reliability of the measurements in HCHS/SOL, which resulted in reliability estimates of 0.97 and 0.94 for measured weight and height in a random subset of participants (n=565 and 570, respectively).

Under such standardized conditions the reliability of measured weight and height is generally considered to be very good [184, 185]. In particular weight is measured with less imprecision and technical error than other anthropometric measures such as waist circumference and skinfold thickness. In NHANES II the inter-observer reliabilities were also highest for weight and height ($r \geq 0.97$) than for the other anthropometric measures collected [185].

5. Strengths and Limitations of Current Knowledge

Here I have presented the various perspectives on validity and reliability of self-reported weight and its change and measured BMI in Hispanic/Latinos. With respect to

the validity of self-reported weight and its change, several important determinants were described in Hispanic/Latinos (or in their absence a non-Hispanic/Latino population), and include age, sex, overweight/obese status, nativity, and smoking status. The validity of measured BMI at capturing the underlying construct of excess adiposity, mortality and CVD were sensitive to differences in a number of individual characteristics: age, sex, racial/ethnic groups, nativity, and smoking. Reliability estimates for both self-reported weight and measured BMI were generally good when standardized protocols were followed. I plan to leverage this body of evidence to inform the design of this dissertation as well as the interpretation of my findings.

Even though there is wealth of existent evidence to support the use of these measures in obesity research, not all supporting studies included Hispanic/Latinos and if they did the diversity in Hispanic/Latino samples were often restricted to one or two Hispanic/Latino background groups. This is a limitation of the current literature as the ability of BMI to serve as an index of fat mass may vary across racial/ethnic groups [164, 166] and US Hispanic/Latinos are extremely diverse with respect to a number of factors including their ancestral backgrounds.

6. Opportunities for Research

This limited data on obesity research in Hispanic/Latinos presents an interesting opportunity for future research, which I plan to address in Aim 1 of this dissertation. There are a handful of studies in the literature that have assessed the self-report bias of current self-reported weight in Hispanic/Latinos. Yet none represent the full diversity of US Hispanic/Latinos. In this respect, an assessment of the validity of self-reported and measured weight at baseline among across multiple Hispanic/Latino backgrounds could contribute to the current literature and my understanding of the self-report and recall biases inherent in existing repeated measures of body mass in diverse Hispanic/Latinos,

such as in the self-reported weight histories of HCHS/SOL. In light of this future opportunity for contribution to the scientific literature, I will next describe the current state of the knowledge on the risk factors of adult obesity.

F. Genetic Epidemiology of Common Complex Obesity

Above I have given a brief overview of genetics as a risk factor for obesity, which is a multi-factorial trait with heritability estimates that are only partially accounted for by rare monogenic forms of extreme obesity [91]. I will devote this next section to reviewing the current state of knowledge on the genetic determinants of obesity that are more common in the general population (≥1% minor allele frequency) and as such may account for some of the remaining heritability of obesity.

Within the past decade genome-wide association studies (GWAS) have led to the discovery of over 100 adult BMI single nucleotide polymorphisms (SNPs), which explain less than 3% of the overall variability in the trait [14-33]. Further analyses have shown that the genetic architecture of BMI is similar between men and women [186]. In summary, these discoveries highlight the importance of genetic pathways in the central nervous system, as well as pathways involved in synaptic function, glutamate signaling, lipids and insulin, energy metabolism, and adipogenesis [33]. Yet the exact functional variants underlying most of these BMI loci or their allelic heterogeneity across diverse racial/ethnic populations remain unknown. Furthermore the single-SNP design of genetic discovery studies (e.g. GWAS) may preclude the accurate estimation of genetic effects in the context of the genome (i.e. jointly). This inaccuracy is exacerbated by the ‘‘winner’s curse’’ phenomenon, whereby initial assessments of risk may be overestimated, rendering subsequent studies of similar sizes underpowered [187].

As a substantial portion of the overall heritability in obesity has not been yet been explained, future studies that search for secondary or rare variant effects in known loci,

or contextualize heritable effects (e.g. GxE interactions as shown in our conceptual diagram Figure 1, epigenetic modification, etc.) will be useful in accounting for the remaining heritability of traits like BMI. For example, a number of loci previously known to be involved in monogenic forms of obesity have also been described in agnostic GWAS, indicating the potential for a shared underlying mechanisms between monogenic obesity disorders and more common etiologies of obesity [92].

An example of the added value of fine-mapping and conditional analysis comes from Yang *et al.*, who used genome-wide complex trait analysis (GCTA) [188] to map 49 additional signals in 36 established height loci. The additional variants explained 1.3% of the heritability, nearly doubling the heritability explained in the initially identified SNPs in these 36 loci in a discovery sample of over 133,000 European descent samples [34]. The same group performed a similar experiment for BMI (n=123,865), but perhaps due to its lower overall accounted heritability no conditional signals were found at the time. Yet more recently using a larger sample of up to 322,154 primarily European descent individuals, investigators were able to describe two additional independent SNPs in *MC4R* and another independent SNP in *BDNF*, explaining an estimated 0.5% of the phenotype variation (for a total estimated phenotypic variation of 2.7% across 97 loci) [33]. Two of the newly identified SNPs had lower frequency compared to the initially identified SNPs (minor allele frequency, MAF, in European sample <5%). Therefore, it stands to reason that additional signals of varying frequencies (rare; <1%; low frequency: 1-5%;) could be identified at these >100 established loci previously tagged by more common genetic variants (>5%) given a larger sample size or greater ancestral diversity in functional variation.

1. Meta-Analysis Methodologies

In light of the need to combine across multiple studies and provide summary genetic results (i.e. distributed data model), meta-analysis is an important tool in the field of genetic epidemiology. In particular, because single genetic effects in complex diseases are often modest to weak, meta-analysis can improve the validity of the findings (i.e. improve power and reduce the false positive rate) [189]. The most common meta-analysis method in the field of genetic epidemiology is inverse variance weighted fixed-effect analysis. However, random-effects and other implementations along the continuum of K=1 to K=number of studies [e.g. Meta-ANalysis of TRansethnic Association studies (MANTRA) by Morris [53], a Modified Random-Effects by Han and Eskin [190]] are becoming increasingly popular for trans-ethnic analysis. The Bayesian implementation of the trans-ethnic meta-analysis approach [53] is described in more detail as part of manuscript 2, below.

Additionally methodologies have been generated to allow for investigators working under a distributed data-model to perform approximate conditional (described in more detail as part of manuscript 2) or joint additional analyses on study-specific results [34], or perform cross-phenotype meta-analyses of interrelated traits of interest [191]. In the case of the approximate conditional analysis, this is a substantial improvement over exact conditional approaches, which often require several months to request and received study-specific conditional results and may thereby limit investigators' ability to test the presence of tertiary or further independent signals [34], such as those seen in recent BMI GWAS [33].

2. Limited Diversity

As of 2009, only 4% of current published GWAS involved samples of non-European populations [40]. Similarly, few non-European descent GWAS studies have

been published for obesity [14, 20, 25, 26, 30, 32, 192-195]. Three GWAS of exclusively East Asian descent populations have described 9 genome-wide significant loci and generalized over 10 additional BMI loci to Asian populations [25, 26, 195]. One recent GWAS of African descent populations has described two novel genome-wide significant loci and generalized four previously described BMI loci to African descent populations [30]. No GWAS have been conducted of exclusively Hispanic/Latinos to date. However, two studies included Hispanic/Latinos either in the discovery or replication samples. A GWAS has described three genome-wide significant findings at previously known loci using a trans-ethnic discovery sample including Hispanic/Latinos [32]. A study of African descent populations has included Afro-Caribbean individuals in their replication sample [30]. Yet the direct replication of candidate SNPs from previous genetic studies of predominantly European descent populations, may be inappropriate given the distinct linkage disequilibrium patterns across ancestral populations and the potential for effect dilution from less correlated markers, and spurious findings [196, 197]. Furthermore genome-wide interaction studies have been criticized for their inability to differentiate chance findings from biologic interactions and limited ability to account for all of the heritability at previous identified BMI loci or studying GxE interactions [198].

The lack of a published Hispanic/Latino GWAS to date can be attributed in part to the unavailability of large epidemiologic samples of Hispanic/Latinos, but also exacerbated by the current limited understanding of Hispanic/Latino genetic structure. A handful of studies have described the broad ancestral populations (i.e. West African, European, and American Indian) that contributed to the current genetic diversity in Latin American countries [112-116]. Although anthropological geneticists have reconstructed the migrations across and peopling of the Americas [199, 200], to date we are aware of only one population-based study that has described this important component of diversity and its impact on health for US Hispanic/Latinos [201]. Interestingly, in this

study the investigators noted that the residual variability in a number of traits, including BMI, was substantially decreased if both global measures of ancestry and genetic ancestry groupings (which incorporated self-reported Hispanic/Latino backgrounds as well as multi-dimensional genetic clustering) in a linear mixed model in HCHS/SOL. Similarly, the statistical models (e.g. GEE) used by the PAGE Study have all adjusted for self-reported backgrounds in HCHS/SOL [202].

In light of these challenges it is not surprising that to date most genetic studies of BMI among US Hispanic/Latinos are generalization studies [193, 203], which investigate the loci discovered and described in other populations (e.g. of European descent) as candidates and may also utilize high-dimensional data to narrow in on the underlying functional variants. For example, I have participated in two previous projects in the PAGE study [204] and its substudies [193] that have indicated that fine-mapping may be a useful tool to generalize the established BMI signal and narrow the putative interval of interest for future follow up.

In a subset of the Hispanic/Latino WHI sample, my collaborators and I have investigated the generalizability of known BMI genetic loci among 3,587 female US Hispanic/Latinas [193]. Of the 32 BMI loci tested using genotyped GWAS data, 9 loci showed evidence for generalization at the same previously described signal or at an independent signal at the same locus. This study provides some insight that a fine-mapping approach to investigate known BMI loci among Hispanic/Latinos is valid given sufficient sample size and densely genotyped platforms, such as the MetaboChip (Illumina, Inc.; San Diego, CA).

More recent fine-mapping work in a subset of the African American sample in the PAGE Study has shown that 8 of 21 established BMI loci at the time and fine-mapped on the MetaboChip array generalize to African Americans at $p<5.8x10^{-5}$ and *GNPDA2* exhibited evidence of an independent signal in exact conditional analyses [204]. To date

we are unaware of any published fine-mapping studies that have done this trans-ethnically or with a Bayesian trans-ethnic mapping approach to appropriately cluster the racial/ethnic groups, as we have done in Aim 2.

3. Opportunities for Research

Even though many SNPs do not generalize directly to all populations, most loci do appear to be relevant to multiple populations and may even contain multiple causal variants as marked by multiple independent ('secondary') signals [197]. This lack of SNP-specific generalization is likely primarily due to differences in linkage disequilibrium patterns and genetic architecture between diverse ancestral and admixed populations, which if harnessed correctly, can ultimately help identify the underlying functional variant. Therefore, it is important to fine-map established loci in ancestrally diverse and admixed populations as it can further aid in this localization among these arguably under studied racial and ethnic groups. In Aim 2 I address this research gap directly by fine-mapping 36 BMI Loci on the MetaboChip array (Illumina, Inc., San Diego, CA) in more than 100,000 diverse samples of the PAGE Study.

G. Gene-Environment Interactions

GxE interactions are believed to play an important role in CVD, like obesity [36, 196]. For example given the pace of the global obesity epidemic [1] and the apparent obesity disparities across racial/ethnic groups in the US for example [2, 3], it is unlikely that recent changes in population genetics have driven the current obesity epidemic. Instead, it is more likely that obesogenic environments are conferring susceptibility to obesity—in concert with thrifty genes that evolved to store/preserve fat in times of scarcity [39, 205]. Further it is possible that changes in obesogenic environments with migration and social adaptation may increase innate genetic susceptibility and contribute to the apparent Hispanic/Latino obesity disparity [125, 206].

Given the difficultly in measuring physical activity and dietary changes individually [58], GxE interactions using measures of broad social adaptations such as acculturation may be particularly helpful in advancing the field of GxE research.

Now that I have described the motivating rationale for studying GxE interactions in obesity, I will describe the current methods that are commonly applied to the study of GxE interactions. Then I will conclude with what I believe are future opportunities for contributions to the field of GxE interactions in obesity.

1. Methodologies

Currently in the literature there are a number of methods to support meta-analysis, but only a few designed specifically for GxE interactions. One approach to the meta-analysis of GxE studies includes either fixed- or random-effect meta-analysis of the estimated GxE coefficient (i.e. a '1df' test).

A more powerful approach focuses on meta-analyzing the combination of genetic and GxE interaction effects (i.e. a '2df' test) [207, 208]. Consortia like the Genetic Investigation of Anthropometric Traits (GIANT) Consortium have begun to try to reduce residual variation by employing genome-wide GxE models with '2df' tests to account for known determinants of the phenotype of study [186]. Although a '2df' test can be derived from an interaction model and then meta-analyzed [207], it can alternatively be made by stratifying the association models by the environmental variable of interest and then meta-analyzing the stratum-specific results [208]. As illustrated by Randall et al. such a '2df' test corresponds to a joint test of the genetic and GxE interaction effects, or in other words the difference in the genetic effect between two or more strata [209].

Although the two approaches (interaction model and stratified analysis) are deemed to be equivalent mathematically and currently the method selection strongly depends "on researchers' prior beliefs regarding the likely form of any true gene-

environment interaction pattern" [208], no systematic comparison of type I or II error between the two methods has been published to date [207, 208]. Future GxE model selection will be informed by a simulation study I conducted for the Genetic Analysis Workshop 19 (Fernández-Rhodes *et al.,* in press). In this analysis we compared the performance of an interaction model incorporating an interaction term to capture the gene-medication interaction on systolic blood pressure and the stratified analysis by medication status for both '1df and 2df tests' of GxE effects [207-209]. In this specific simulation I observed more stable false positive proportions (Figure 4) and slightly more conservative true positive estimation (Figure 5A-I) across a range of minor allele frequencies using the stratified analysis as compared to the interaction term model. As the SNP of interest becomes less common, the performance of both methods deteriorated.

2. Previous Applications and Limitations

Despite the recent methodological innovations to support GxE studies, they continue to face challenges to the fulfillment of their public health relevance for CVD [35, 36, 196]. According to a recent review, there are >200 GxE studies in the literature, but usually are subject to concerns about data quality (environmental and phenotypic), are not significant after multiple testing penalties, or do not include independent replication [196]. Moreover, few include diverse racial/ethnic groups.

There is a large and growing body of GxE literature [196], which may elucidate how "common diseases result from common exposures to which I are all susceptible, albeit in varying degrees" [206]. Additionally, GxE studies can help explain some of the missing heritability and contextualize estimates of genetic effects, which previously were estimated without respect the impact of the environments. However, according to a recent review, many of these studies are limited due to concerns about data quality

(environmental and phenotypic), are not significant after multiple testing penalties, or do not include independent replication [196]. Furthermore, the generalizability of current GxE studies is hampered when few include diverse populations. For example, between 2011-2014 I identified 18 published GxE studies on BMI/obesity [102, 205, 210-225]. Only three of these studies reported findings from cohorts with Hispanic/Latino individuals (Add Health, the WHI-SNP Health Association Resource) [210, 213, 218], now the largest US ethnic minority [4]. In addition only five of these studies included upstream environmental or psychosocial characteristics to capture aspects of the obesity epidemic, which are more contextual [205, 213, 219-221].

As the state of the science moves from *a priori* hypotheses to agnostic interrogation of genome-wide GxE effects, concerns remain for GxE studies around their multiple testing burden and the distinction between statistical and biologic interactions [198, 226]. Arguably an ideal GxE investigation would be ancestrally diverse and utilize a design, which could 1- integrate *a priori* information to inform the environmental factor of interest, 2- limit the number of associations tested (using an '*a priori* genetic profile' [198]) or explore independent replication, 3- involve intensive variable harmonization across multiple studies, and 4- test for the presence of multiplicative interactions more likely to be biologic in nature.

3. Future Promise in Epidemiology

GxE interaction studies are considered to be integral to the future of genetic epidemiology and its public health applications for the following three reasons. First, many environmental risk factors for common, complex human diseases such as CVD have been revealed by epidemiologic studies, but how variants at specific loci modulate the effect of environmental risk factors is largely unknown [36]. The use of GxE methodologies in epidemiologic studies may decrease the residual variation in traits like

BMI and may thereby account of unexplained heritability. The majority of GWAS have not included environmental information to account for this 'missing heritability.' This integrative approach may inform the impact of future interventions for obesity and its CVD consequences, in a way that cannot be studied using experimental designs.

Second, the design of genetic discovery studies (e.g. GWAS) generally precludes the accurate estimation of genetic effect size. This inaccuracy is due to the ''winner's curse'' phenomenon, whereby initial assessments of risk may be overestimated, rendering subsequent studies of similar sizes underpowered [187]. One could counteract this by estimating the effects of associations discovered in GWAS in a diverse population. Moreover, such a study could reveal the true population impact of genetic variants by assessing their interaction with environmental factors. Yet as the genetic architecture of obesity to be complex, a constellation of SNPs may explain varying amounts of heritability in adiposity depending on the particular population or environment of study.

Lastly, hypotheses about social and cultural environments and genetic susceptibility cannot be interrogated using animal models or experimental designs. For example, acculturation is a complex social phenomenon unique to humans, which among other things is hypothesized to capture exposure to obesogenic behavioral and lifestyle factors [8, 86] for which experimental designs may be unethical. By integrating sociocultural information into an analysis of the genetic determinants of obesity among US Hispanic/Latinos, such findings would gain real world applicability, which could not be obtained without using an observational study design.

I am confident that as more reference samples are included in the 1000 Genomes project [227] or sequenced as part of the second phase of the PAGE Study, as genome-wide discovery efforts like the Hispanic/Latino Anthropometry (HISLA) Consortium or the fine-mapping of established BMI loci in diverse populations (Aim 2)

are completed, the basis for future research on the genetic architecture of diverse US populations, such as Hispanic/Latinos, and gene-environment interactions in obesity will be dramatically strengthen (see conceptual diagram, Figure 1). Although few genetic studies have been conducted among US Hispanics/Latinos, an ethnic group with roughly half of the foreign-born immigrants and notable obesity disparities [125], they hold great promise in genetic epidemiology in particular with regards to their multiple aspects of diversity and their unique migratory and sociocultural histories [125, 206].

With respect to the components of an ideal GxE study design, as described above there is a wealth of longitudinal information on acculturation and obesity risk in public health (Table 1), as well as cross-sectional studies in HCHS/SOL [138]. Moreover, the fine-mapping of established BMI loci in the PAGE Study would be the perfect supporting work for a targeted gene-acculturation study in HCHS/SOL of the best marker, or functional, SNPs on BMI (untransformed, in order to interrogate multiplicative interactions). Through my work with the HISLA Consortium of studies of Hispanic/Latinos from across the US and Latin America, I have already identified several independent studies with acculturation, genetic and obesity information for possible replication.

H. Supporting Tables and Figures

Table 1. Summary of the longitudinal or repeated cross-sectional literature on acculturation and measures of obesity.

First Author, Year (Ref)	Study (Total N)	Exposure	Outcome	Model	Follow Up	Age	Baseline Differences by Exposure?	Evidence of change over time by Exposure?	Additional Analyses/Considerations?	Diversity?
LONGITUDINAL										
Balistreri, 2009 [143]	ECLS-K (n=12,696)	Generation (Children of immigrants v natives)	Measured BMI	GC	~5.5 years	~5-10 years	Yes (S, children of immigrants higher than of natives)	Divergence (faster BMI growth among children of immigrants)	SES Gradients, Sampling weights	Hispanic/Latino (majority Mexican), non-Hispanic/Latino White
Harris, 2009 [144]	Add Health (n<=20,745 Waves I-III)	Generation (1st, 2nd, or 3rd or native)	SR (Wave I) and measured (Waves II,III) BMI	GC	~7-9 years	11-28 years old	Yes (S)	Divergence	Sampling weights	Hispanic/Latino, non-Hispanic/Latino Asian, Black, and White
Jackson, 2011 [142]	Add Health (n= 15,601, Waves I-III)	Generation (1st, 2nd, or 3rd or native)	SR (Wave I) and measured (Waves II,III) BMI	GC	~7-9 years	11-28 years	Yes (S)	Divergence	Sampling Weights	Hispanic (Mexican, Non-Mexican), East Asian, Other Asian, non-Hispanic/Latino Black and White
Albrecht, 2014 [145]	Add Health (n=13,701, Waves I and IV)	SES/Generation (1st, 2nd, 3rd plus)	Measured BMI (Wave IV)	Linear models	~12 years	12-33 years old	Yes (S)	Divergence	Sampling Weights, M (IN): language spoken at home (Wave I, English or other)	Hispanic/Latino, non-Hispanic/Latino Asian, Black and White
Albrecht, 2013 [148]	MESA (n=2,288)	Nativity (US, F born)/Time in US (<15, 15-30, >30 years)	Measured WC, BMI	LMM	5 years	45-84 years old	Yes (S)	Convergence (Mexican), Stability (Non-Mexican, Chinese)	M (IN, NS): smoking, alcohol, physical activity, diet	Mexican, Non-Mexican, Chinese
Ullmann, 2013 [228]	L.A. FANS (n=975)	Generation (1st v 2nd plus)	Annual SR weight change	LMM	5-8 years	18-85 years old	No (NS)	Divergence (S total and women only)	Environment using 2000 census-tract information (primarily NS): Population density, composition, foreign-born %, and average BMI, socioeconomic disadvantage, collective efficacy (S in women only), neighborhood safety Adjusted for baseline weight and height	Hispanic/Latino (majority Mexican), White, Black, Asian/Pacific Islander
REPEATED CROSS-SECTIONS										
Antecol, 2006 [10]	NHIS (n=47,006, 1986-1996)	Nativity (US, F)/Time in US (0-4, 5-9, 10-14, >14years)/Arrival cohort (>1981, 1981-1985, 1986-1990, 1991-1995)	Regression calibrated SR weight and height (based on NHANES III data) used to calculate BMI, overweight and obesity (yes, no)		NA, Panel study	20-64 years	Yes	Convergence (BMI most strongly)	Arrival cohorts assessed separately	Hispanic/Latino (majority Mexican although not stated)
Park, 2009 [141]	NHIS (n=17,300, 1994-1996 and 2004-2006)	Nativity (US, F)/Time in US (<5 years in 1995, 10-14 years in 2005)	SR weight and height used to calculate obesity (yes, no)	Logistic models	NA, Panel study	18+ years	Yes (S 18-54 year olds only)	Divergence	Decomposition of age, period, cohort, and duration of time in US effects.	Hispanic/Latino (majority Mexican although not stated)
Albrecht, 2013 [140]	NHANES (pooled NHANES III 1988-1994 n=3,175 and continuous 1999-2004 n=3,037 and 2005-2008 n=1,937)	Nativity (US, F born)/Time in US (<10, ≥10 years)	Measured WC, BMI	Linear models	NA, Panel study	20-64 years	Yes (S all outcomes among men, S only with WC 2005-2008 among women)	NA	Sampling Weights, No Secular trends (NS) in nativity/time in US differentials on WC or BMI	Mexican American

Abbreviations: Add Health=National Longitudinal Study of Adolescent to Adult Health, AIC=Alkaike Information Criterion, BMI=body mass index, ECLS-K=Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999, GC=Growth curves implemented using mixed models, IN=Incomplete mediation, L.A. FANS=Los Angeles Family and Neighborhood Survey, LMM=Linear mixed model, M=Mediation analysis through adjustment in model, MESA=Multi-ethnic Study of Atherosclerosis, NA=Not applicable, NHANES=National Health and Nutrition Survey, NHIS=National Health Interview Survey, NS=Non-significant $p \geq 0.05$, Ref=Reference, S=Significant $p < 0.05$, SES=Socieconomic Status, SR=Self-reported, WC=waist circumference

Table 2. Summary of literature on accuracy of self-reported weight and weight change in Hispanic/Latinos (when available).

First Author, Year (Ref)	Study (Total N)	Age	Measure	Timing of Comparison	Bias Tested (Estimate**)	Hispanic/Latino Samples Noted (Diversity)	Comparison of Accuracy to non-Hispanic/Latinos
SELF-REPORTED VERSUS MEASURED CURRENT WEIGHT FOR THE SAME AGE							
Alvarez-Torices, 1993 [46]*	León, Spain Study (n=572)	18-? years	Weight	Current	SR Bias (-0.6kg)	Yes (Spanish Nationals)	NA
Santillan, 2003 [155]*	Monterrey, Mexico Asthma Case-Control Study (n=961)	18-? years	Weight	Current	SR Bias (-?kg)	Yes (Mexican Nationals)	NA
Avila-Funes, 2004 [47]*	Mexican National Health and Aging Study (n=1,707)	24-95 years	Weight	Current	SR Bias (0.6kg)	Yes (Mexican Nationals)	NA
Merrill, 2009 [51]	NHANES Continuous 2001-20006 (n=16,814)	16-? years	Weight	Current	SR Bias (0.10 to -1.16kg)	Yes (Combined across background groups)	NA
Griebeler, 2011 [75]	Soy Phytoestrogens as Replacement Estrogen Study (n=428)	45-60 years	Weight	Current	SR Bias (-1.55kg)	Yes (Combined across background groups)	Yes, Hispanic/Latino NS predictor
Wen, 2012 [156]	NHANES Continuous 2007-2008 (n=5,343)	20-? years	Weight	Current	SR Bias (-0.1 to -0.6kg)	Yes (Combined across background groups)	Yes, Hispanic/Latino ethnicity NS predictor after covariate adjustments
SELF-REPORTED CURRENT WEIGHT VERSUS RECALLED WEIGHT FOR THE SAME AGE							
Kyulo, 2012 [158]	Adventist Health Study 2 (n=2,727)	48-100 years?	Weight	Past (20-28 years prior)	Recall Bias (-0.67kg)	No	NA
MEASURED CURRENT WEIGHT VERSUS RECALLED WEIGHT FOR THE SAME AGE							
Stevens, 1990 [160]	Charleston Heart Study (n=703)	62-100 years	Weight	Past (4 and 28 years prior)	SR and Recall Bias (-1.2 to 0.3kg at 4 years, -1.2 to -2.0kg at 28 years)	No	NA
Troy, 1995 [161]	Nurses' Health Study II (n=118)	25-42 years	Weight	Past (at 18 years)	SR and Recall Bias (-1.4kg)	No	NA
Kovalchik, 2009 [159]	NHANES I Epidemiological Follow-Up Study (n=6,101)	20-65 years	Weight	Past (Recalled Weight at NHANES I Exam Imputed from Linear Change Across SR Weight History)	SR and Recall Bias (-1.77kg)	No (Nationally representative sample from 1971-1975 may have included)	NA
SELF-REPORTED VERSUS MEASURED WEIGHT OVER TIME							
Field, 2007 [48]	Add Health Wave II-III (n=?)	16-26 years across waves	Weight change	Past versus Current (5-6 years after)	Recall Bias, Under assumptions about SR bias stability over time (-1.0 to -1.3kg)	Yes (Combined across background groups)	Yes, Hispanic/Latino ethnicity NS predictor
Clarke, 2014 [162]	Add Health Waves II-IV (n=19,238 with at least Wave II and another wave)	13-32 years across waves	Weight	Current	SR Bias, between 13-32 years of age (-0.86kg adolescent girls at 13 years, increased with aging, Boys NS at all time-points)	Yes (Combined across background groups)	Yes, Hispanic/Latino ethnicity NS predictor of SR bias change over time

Abbreviations: Add Health=National Longitudinal Study of Adolescent to Adult Health, NA=not applicable, NHANES=National Health and Nutrition Examination Survey, NS=Non-significant, Ref=Reference, SR=Self-report.

*These three studies of Hispanic/Latino samples were referenced as part of the review by Connor Gorber *et al.* [151].

**Estimate given as mean difference (=self-report minus measurement).

Figure 2. Definitions for “Hispanic” and “Latino.” One definition for the term *Latino* refers to persons whose origin or ancestries are from countries of Latin America (A). The US Office of Management and Budget uses the terms *Hispanic* and *Latino* interchangeably to refer to persons who indicated that their origin is Mexican, Puerto Rican, Cuban, Central and South American, or other Spanish culture or Spanish-speaking country or origin, regardless of race (B). Other definitions for the term *Hispanic* include individuals whose origin or ancestry comes from Hispania, the former name for the Iberian Peninsula (C), and Spanish-speaking persons of Latin American descent living in the United States (D).[1]

[1]Figure and legend reproduced with permission from *JAMA Dermatology.* 2013;149(3):274-275. doi:10.1001/jamadermatol.2013.1304. Copyright © (2013) American Medical Association. All rights reserved. [93]

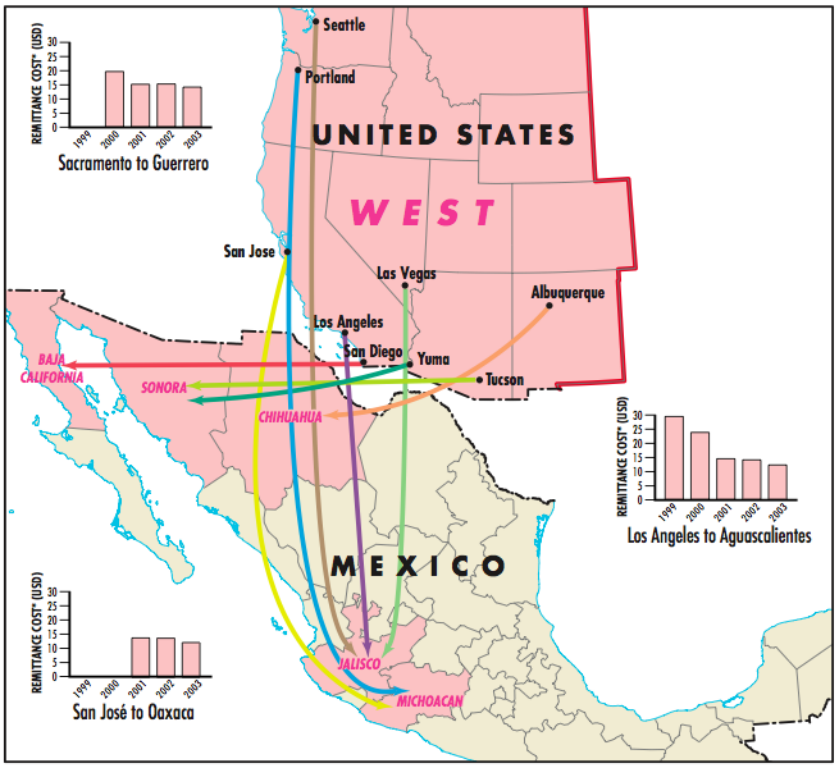


Figure 3. Illustration of Inverted Remittance Corridors from the West Coast of the US to Mexico, 1999-2003.[2]

[2]Figure adapted from Hernández-Coss, Raúl. 2005*. The U.S.–Mexico Remittance Corridor: Lessons on Shifting from Informal to Formal Transfer Systems*. *World Bank Working Paper No. 47.* © World Bank. https://openknowledge.worldbank.org/handle/10986/7322 License: Creative Commons Attribution License (CC BY 3.0 IGO). [100]

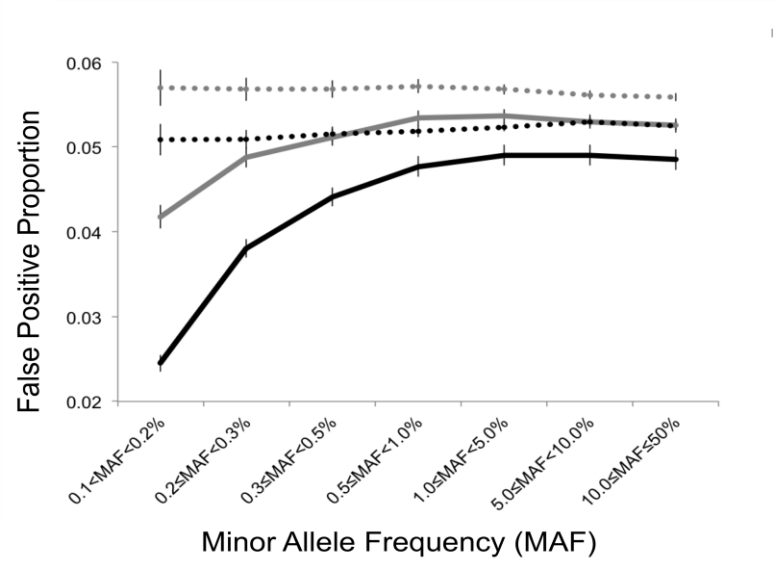


Figure 4. Comparison of the false positive proportions (FPP) at the *CYP3A43* locus (chr7:98957518-99957518) and 95% confidence intervals for the interaction model (solid line) and med-diff approaches (dashed line) and 200 replicates of true negative findings on the odd-numbered chromosomes using the 1 degrees of freedom, df (gray), and 2df tests (black) and their variability across bins of minor allele frequency (>0.1% to 50%).[3]

[3]Reproduced from Fernández-Rhodes *et al.,* in press with *BMC Proceedings*. License: Creative Commons Attribution License 4.0.

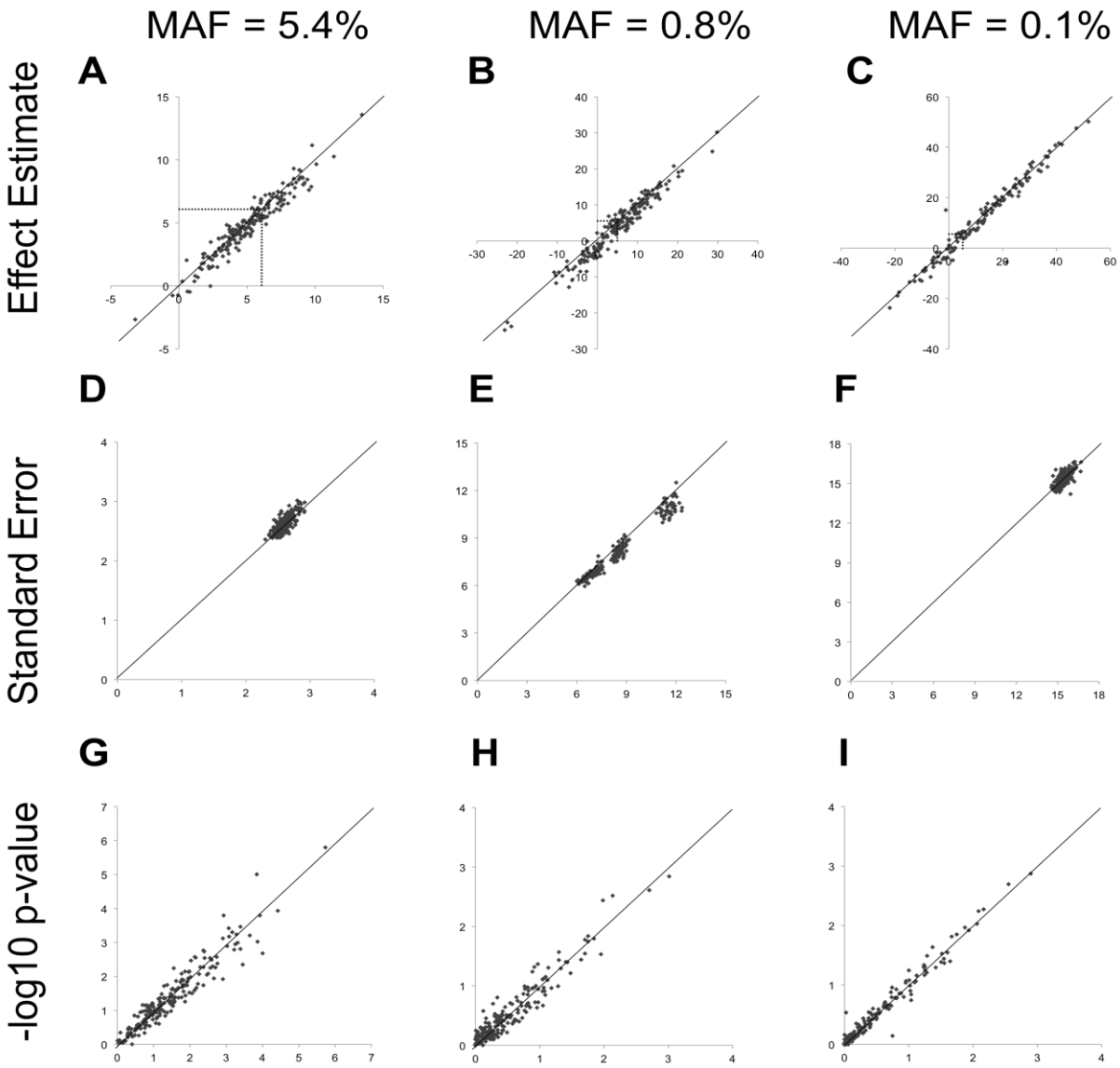


Figure 5A-I. Comparison of the estimated effects (A-C), standard errors (D-F), and –log of the 1f p-values (G-I) on SBP from the interaction model (X-axis) and med-diff (Y-axis) approaches in up to 200 replicates of simulated gene-medication interactions at three SNPs at *CYP3A43* (6.2mmHg, dashed line in A-C) of varying minor allele frequencies (MAF).[4]

[4]Reproduced from Fernández-Rhodes *et al.,* in press with *BMC Proceedings*. License: Creative Commons Attribution License 4.0.

CHAPTER IV: METHODS

A. Overview

In this chapter I will attempt to outline the specific steps required to conduct my dissertation research. First, I will begin by summarizing the study design and available data in HCHS/SOL and the PAGE Study for a secondary-analysis (Sections B-C). Then I will describe the data analyses to address each aim in turn: definition of pertinent variables, statistical analyses, power calculations, as well as the strengths and limitations. Lastly, I will describe my compliance with the guidelines on ethical human subjects research from the Internal Review Board in the Office of Human Research Ethics at the University of North Carolina at Chapel Hill.

B. Hispanic Community Health Study/Study of Latinos (HCHS/SOL)

The HCHS/SOL is the first population-based cohort study of CVD to recruit more than 16,000 self-identified Hispanic/Latino adults (18-74 years at screening) from diverse Hispanic/Latino backgrounds (Central American, Cuban, Dominican, Mexican, South American, and Puerto Rican), resident in four US urban communities (Bronx, NY; Chicago, IL; Miami, FL; San Diego, CA) between 2008-2011 [42, 43]. The two main analytic objectives of HCHS/SOL were to 1- collect data that could support estimates of prevalence of CVD risk factors in the HCHS/SOL communities at baseline (2008-2011), and 2- estimate the association between CVD risk factors and disease outcomes over the course of follow-up (2009 to present).

1. Study Design

Due to limited time and resources allocated to the tasks of studying the prevalence and CVD incidence in Hispanic/Latino residents of the four HCHS/SOL study sites, it was identified early in the design of HCHS/SOL that probability sampling would be necessary [42]. First, census block groups were sampled within strata of SES (percent of population 25 years or older with at least a high school education) and concentration of Hispanic/Latino households as per the 2000 US Census, in order to proportionately sample low and high SES areas and disproportionately select census block groups with a high concentration of Hispanic/Latino households. Second, using lists of postal addresses and Hispanic surnames within the selected census block groups (primary sampling units), households with Hispanic surnames were over-selected within block groups. Third, once contact with a household member was made (in either Spanish or English) a roster of household members was taken and a digital hand-held device [229] was used to determine eligibility and the probability of individual selection in one of two ways [42]. The first method sampled whole households, where households with all Hispanic/Latino adult household members 45-74 years of age are selected with certainty (=1) and all other households are selected with a lower probability (<1), but was only implemented in the first 6 months of the study due to its low proportion of selected households. The second method implemented after the first 6 months of recruitment divided each household into sub-clusters: Hispanic/Latino adults 45-76 years selected with certainly (=1), and 18-44 years selected with a lower probability (<1). Furthermore during screening individuals were deemed ineligible for the HCHS/SOL examination if they were on active military duty, not currently living at home, unable to give informed consent, travel to the field center, complete the study questionnaires, or had plans to move from area in the next 3 years. Pregnant women were rescheduled for baseline

interviews ~3 months postpartum as physiologic changes in cardiometabolic risk factors may change during and immediately after pregnancy.

Even though all HCHS/SOL study materials were available in Spanish or English and the study engaged Hispanic community partners at each study site, the sample of Hispanic/Latinos who attended the baseline examination may not be fully representative of all Hispanic/Latino adults living in these communities. For example, the US Census was used to inform the sampling procedures but may be an imperfect tool for identifying the location of marginalized populations in community-based studies. The Pew Hispanic Center estimates that perhaps as much as 15% of undocumented Hispanic/Latino adults did not participate in the 2010 US Census, which could result in an inaccurate estimate of the true distribution of Hispanic/Latinos in the US [103]. This bias could vary across the four communities in HCHS/SOL (e.g. San Diego, CA may have a higher population of undocumented Hispanic/Latinos due to its proximity to the US/Mexico border), by Hispanic/Latino background group, or other characteristics. For example, despite study-wide efforts to engender trust with their participants, to exclude the discussion of sensitive information (e.g. legal status) or to minimize the length of the baseline exam to under 8 hours, it is still possible that consenting and eligible participants were more likely to be US legal residents or citizens or to work in the formal sector or have flexible work schedules.

Throughout this dissertation I applied the sampling weights recommended by the HCHS/SOL Coordinating Center (CC). The use of sampling weights yields appropriate estimates of population characteristics and the corresponding standard errors using a sandwich variance estimator [230-232]. Failure to account for the sampling in this study could produce biased parameter estimation. Specifically, these weights are inverse participant selection probabilities that were adjusted for non-response (at both the household and individual level), windsorized within each field center, calibrated to 2010

US Census population estimates with respect to the Hispanic/Latino background and age/gender distributions and then normalized based on the mean and standard deviation of the resulting weights of the entire HCHS/SOL sample. Although it is impossible to determine the effect that selection bias could have on the generalizability of the results, the use of non-response adjusted sampling weights is an important step towards minimizing this potential bias while increasing the generalizability of inferences to the HCHS/SOL communities.

In addition to their primary analytic objectives mentioned above, HCHS/SOL investigators wanted to collect information that would inform secondary analyses of less commonly studied aspects of CVD. Therefore, they also collected detailed information on anthropometrics (including self-reported weight histories), sociocultural factors, and genetics in a culturally- and linguistically-appropriate manner, which is described in more detail below.

In general, trained HCHS/SOL study personnel assisted participants during the baseline exam in the language of their preference (Spanish or English, noted at the beginning of the exam) with filling out the various study questionnaires. Participants could choose not to respond to any question as part of the examination. If participants were unable to complete the entire exam in one day, in an effort to minimize missing data, they were encouraged to complete the missing portions at a later date and this was noted.

2. Genetic Data

Consent for participation in the HCHS/SOL baseline examination was obtained separately from the consent for genetic testing and the sharing of this data on dbGap (n=12,472). Genomic characterization in HCHS/SOL occurred in the context of the PAGE Study; using the MetaboChip genotyping array (Illumina, Inc., San Diego, CA)

[41]. The 196,725 SNPS genotyped on the MetaboChip were chosen to fine-map 257 loci that had been validated with cardiometabolic traits as of 2009 [233]. After excluding duplicates, individuals with <95% call rates or reporting a sex that was discordant (phenotype) with the genotypic sex (n=355), 12,117 individuals (74% of total HCHS/SOL cohort) had available genotype information for analysis. More information about the MetaboChip genotyping completed as part of the PAGE Study and derived genetic variables considered as part of this dissertation is presented in the following section about the PAGE Study.

3. Anthropometric Data

The HCHS/SOL anthropometric data were collected from two sources: the Anthropometry (ANTA) and Weight History (WHEA) Questionnaires (English versions in Appendices). In general, both the absolute number completed and completion rate were higher for the ANTA (n≤16,388) than the WHEA Questionnaire (n≤15,279), Whereas the WHEA Questionnaire relied exclusively on recalled self-reports of individuals at least 21 years of age, the ANTA Questionnaire was a combination of self-reported and measured anthropometric information on the entire cohort. Importantly the ANTA Questionnaire asked participants about their weight before measurements were made, to ensure that the self-reports (ANTA3A, to the whole kilogram, kg, or pound, lb) were made without the knowledge of the actual measured value from a digital scale (ANTA4, Tanita Body Composition Analyzer, TBF 300, Japan), which then transcribing it into an electronic data entry system by the study personnel. The ANTA questionnaire and measurements were collected during the fasting block of time immediately after a urine collection [43] on participants who were able to stand on both feet while wearing a scrub suit or examination gown and no shoes [179].

In HCHS/SOL great efforts were made to minimize the measurement error of the anthropometrics captured by the ANTA Questionnaire. First, during the measurements participants wore a scrub suit or examination gown without shoes [234]. Weight in kg was measured using a scale and height in centimeters using a stadiometer. Second, quality control measures were put into place by the HCHS/SOL study to ensure that all personnel were collecting data in a similar manner such as the provision of trainings, certifications (requiring a minimum of 5 practice subjects with 0.5 kg weight and 0.5 cm height agreement between a trainee and expert), and periodic observations. Third, HCHS/SOL anthropometric equipment was calibrated frequently to ensure quality measurements (e.g. scales zero balanced daily and calibrated weekly, and stadiometer inspected daily). Lastly, HCHS/SOL assessed inter-technician agreeability in HCHS/SOL by randomly selecting 3-5% of baseline participants for retest and the HCHS/SOL Quality Control Committee then analyzed and uses these agreements during the recertification process. Additionally, inter-rater reliability was assessed by randomly selecting 3-5% of participants for retest by a second trained technician immediately after the initial anthropometric exam during the same baseline visit. Self-reported and measured weights (n=565) differed on average between two technicians by 0.46 kg (95% confidence interval, CI: -0.12, 1.03 kg) and 0.16 kg (95% CI: -0.18, 0.50 kg), respectively. This resulted in good reliabilities of 0.93 and 0.97 and relatively low coefficients of variation (the within-specimen variation expressed as a percentage of the mean) of 6.3 and 3.7%.

C. Population Architecture using Genomics and Epidemiology (PAGE) Study

Since the start of the obesity epidemic, a number of consortia of observational studies have been created to facilitate the study the population-level changes in CVD risk factors. For example the PAGE Study offers a unique opportunity to investigate the

genetic and environmental contributors to obesity in minority US populations such as African, Hispanic/Latinos, Asian, and Alaskan Natives/American Indians [41].

1. Consortium Design

The Population Architecture using Genomics and Epidemiology (PAGE) Study has a coordinating center and four large study sites/consortia that include a number of diverse observational studies including: the ARIC Study, the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) Study accessing the Vanderbilt University BioBank (BioVU), Coronary Artery Risk Development in Young Adults (CARDIA), Cardiovascular Health Study (CHS), the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), Multiethnic Cohort (MEC), the Women's Health Initiative (WHI) [41]. Additional studies collaborating in this analysis also included: The GenNet Network (GenNet), the Hypertension Genetic Epidemiology Network (HyperGEN) Study, the MEC-Slim Initiative in Genomic Medicine for the Americas Type 2 Diabetes Consortium (SIGMA) Type 2 Diabetes Consortium, the Mount Sinai School of Medicine's Biobank (BioME), and the Taiwan-MetaboChip Study for Cardiovascular Disease (TaiChi) study. A detailed description of each study can be found in our Appendices (Supplemental Materials for Manuscript 2).

2. Genetic Data

The MetaboChip was designed to fine-map 257 loci with 196,725 genotyped SNPs. These loci had been validated with cardiometabolic traits as of 2009 [233] and include 36 validated BMI loci [14-33]. The genotyping of the MetaboChip was performed for the PAGE Study at certified research genomics laboratories: the Human Genetics Center of the University of Texas-Houston (Houston, TX), University of Southern California Genomics Core (Los Angeles, CA), Translational Genomics Research Institute (Phoenix, AZ), and Vanderbilt Technologies for Advanced Genomics (Nashville, TN)

[235]. HapMap control samples were also genotyped independently by each site to allow for an internal control and both used Illumnia GenomeStudio with the GenCall 2.0 algorithm to make genotypic calls. SNPs with a Gentrain score <0.6, cluster separation <0.4, more than 1 Mendelian inconsistency, discordant duplicate calls, GenoSNP score >3.3%, call rate <95%, ambiguous mapping to the genome, or that were discordant with GWAS SNP in PAGE pilot studies were excluded by the PAGE Study CC at Rutgers University (n=13,808). There were then 182,917 SNPs left in final set of SNPs in the analytic set provided by the PAGE Study CC, and an additional 2,646 SNPs were excluded for ambiguous intensity plots in later quality control analyses. Intentional and unintentional duplicates, individuals with low call rates (<95%), excess heterozygozity, phenotype-genotype sex discordance, or ancestry outliers were also excluded from the quality controlled data set as potential sample errors [235]. Any samples with an inbreeding coefficient (F) >0.15 were also excluded [236].

Given the potential for population stratification in ancestrally diverse Hispanic/Latinos, the PAGE Study CC did not apply an overall Hardy-Weinberg Equilibrium filter for Hispanic/Latinos, but instead have chosen to perform this test stratified by Hispanic/Latino background in the sub-study with the most diversity, HCHS/SOL. If a SNP had a Hardy-Weinberg Equilibrium p-value<$1x10^{-6}$ in at least one background group as well as in the overall sample, the PAGE Study CC recommends its flagging/exclusion as a potential genotyping error.

Imputation of MetaboChip SNPs was conducted in MEC-SIGMA (Hispanic/Latinos only), BioME (African and Hispanic/Latino American), and WHI (representing 54% of WHI African American women, and all of the WHI European descent women) [237] using 1000 Genomes phase 1 reference samples and filtered on imputation quality.

Due the design of HCHS/SOL, the cohort included related individuals and unequal sampling probabilities that must be accounted for in genetic analyses [238]. This influences the sample selection, statistical models and adjustments for ancestry that have been recommended for genetic analyses in HCHS/SOL as compared to the other three studies. I will describe each below.

First, in most population-based genetic epidemiologic studies the member of each first degree related pair based on identity-by-descent statistics (e.g. probability of sharing 0, 1, or 2 alleles at a given marker [239]) with the lowest call rate is excluded from statistical models that assume independence. The PAGE Study CC assessed relatedness within and across the studies with Hispanic/Latinos using PLINK to generate identity-by-descent statistics [240], and all duplicate or 1st degree relative pairs with other studies than HCHS/SOL were broken by preferentially including the pair from HCHS/SOL.

Second, within the HCHS/SOL cohort there is large variability in the number of participants per household in HCHS/SOL and the distribution is skewed to the right (range of 1-14, median of 1, mean of 1.7) [238]. Although most population-based genetic analyses exclude 1st degree relatives to obtain an 'independent' set of observations, alternatives exists to account for empirical estimates of relatedness between individuals and retain all observations in the statistical modeling. Given the variability in shared households and the potential for relatedness, the PAGE Study CC recommends modeling shared household and estimates of relatedness in all analyses. As such flags have been distributed to cluster individuals in the subset of HCHS/SOL with MetabocChip data into groupings based on shared households or high amounts of identity-by-descent relatedness (defined as a 1st degree relatives with $0.35<\Phi<0.98$ estimated identity-by-descent allele sharing). These flags result in 6,899, 761, 96, and 35 extended families of sizes 1-2, 3-4, 5-6, and ≥ 7, respectively. Although variance-

component models [241-245] perform well in instances of cryptic relatedness (i.e. 3rd or more degrees) and can be used for binary, ordinal, and age-at onset traits [246-248], GEEs provide a more robust and less computational intensive alternative to appropriately account for the complex sampling design of HCHS/SOL [230]. GEEs are also applicable to binary, ordinal, age-of-onset and quantitative traits, such as many CVD traits [249]. For this reason the PAGE Study CC recommends the use of GEEs in HCHS/SOL to account for both relatedness, aspects of the study design, and Hispanic/Latino backgrounds and have provided a statistical package, SUGEN, in C++ to facilitate this analysis [238].

Lastly, given that the frequency of various genetic markers and the trait of interest may both depend on an individual's ancestral background, a great deal of attention in genetic epidemiology has been given to controlling for this potential source of confounding (i.e. population stratification) [250, 251]. PCA is commonly used in genetic epidemiology to condense the multiple marker observations within a given study into their ancestral components, but PCA is not valid in samples with relatedness due to their inter-dependence. Imputation can be used to infer the global ancestry of one relative based on the estimates of the other, but this is not appropriate in samples with substantial amounts of relatedness. Therefore, HCHS/SOL investigators have chosen an alternative strategy that allows for the estimation of global ancestry in the entire sample. Briefly, this was accomplished by using the 1000 Genomes publically available reference populations (CEU, YRI, MXL, PUR, CLM, CHB) to train the principal component space (i.e. create 20 eigenvectors of genotypes) and then project the entire HCHS/SOL sample along each of these components based on their observed genotypes. These analyses have been performed in Eigensoft software and have been shared across the PAGE Study for investigators doing genetic association studies [250, 252].

Unlike HCHS/SOL, the other participating PAGE Study studies with Hispanic/Latino samples include far smaller proportions of related individuals. Therefore the decisions regarding sample selection, statistical models, and ancestry adjustments are more straightforward. After estimating relatedness using PLINK [240], all duplicate or 1st degree relative pairs were broken across studies by either including the pair with the highest call rate [235] (unless one pair was from HCHS/SOL). Within-study 1st degree relative pairs were broken between the PAGE studies by including the pair with the highest call rate. Linear regression was then performed using PLINK [240]. The PAGE Study CC has distributed PCA estimates calculated separately in an unrelated subset (as defined as 2nd degree relatives or beyond, ≤0.35 estimated identity-by-descent allele sharing) for use in genetic association analyses.

3. Anthropometric Data

One of the advantages to a consortium such as the PAGE is that the data is cleaned and harmonized centrally (Rutgers University, Piscataway, NJ) before it is distributed to the investigators of the participating sub-studies [41]. The harmonization of the anthropometric variables such as BMI in the PAGE Study has been described previously [203, 204]. Both weight and height were measured in HCHS/SOL [43] and WHI at examination by trained staff following a standardized protocol [253]. However, in MEC these anthropometric values were self-reported [203, 254]. In EAGLE BioVU and BioME, biobanks of electronic medical records, height and weight were measured in the clinic as part of a patient's examination [30, 33, 255, 256].

D. Research Plan

In this section I will briefly describe my methodological approach to establish the accuracy of self-reported weight as a tool for future studies of trajectories of self-reported weight in HCHS/SOL (Aim 1). Then I will outline the approaches I took to study the

generalization and fine-mapping of BMI loci to diverse populations in the PAGE Study (Aim 2). Additionally information on the methods of each manuscript can be located in the Results sections below.

1. Aim 1

I will now describe the approach I took to examine the accuracy of self-reported weight in HCHS/SOL before and after adjustment for potential predictors of inaccuracy (Aims 1A and 1B, respectively). I next describe how I constructed a quality controlled analytic dataset of self-reported and measured weight among the 16,119 non-pregnant participants 18-74 years at screening, without a limb amputation at the baseline examination, and without extreme self-reported (or BMI values, calculated using measured height at baseline).

i. Data Quality Control

Further details on my data cleaning and exclusions protocol for the self-reported and measured weights are provided in Manuscript 1 (below). Based on *a priori* knowledge of Hispanic/Latino migratory patterns and exploratory analyses of the data, we were concerned about the potential for unit confusion in the self-report (kg or lb) and sought to compare our range of difference between self-reported and measured weight with previous reports. As such we flagged absolute differences between self-reported and measured weight ≥15kg as possible data errors based on a previous study from Mexico [47].

We then applied a detailed data quality control protocol (described in detail in manuscript 1), to all 16,203 participants with data on both self-reported weight and measured weight (98.7% of entire sample) to: 1) address the flagged calculated differences between self-reported and measured weight as potential data errors (42 individuals excluded), and 2) exclude currently pregnant women (who reported not being

pregnant during the screening but reported being pregnant as part of their medical history), or individuals with limb amputations (≥45 years, not otherwise affecting their ability to stand on both feet), or a body mass index <16 or >70 kg/m^2 (42 individuals excluded). Unless indicated otherwise, all results presented below pertain to the sample with both self-reported and measured weight that remained after applying this quality control protocol (n=16,119).

ii. Statistical Analyses

In Aim 1A I first described the frequency of unit (kg or lb) and end digit preference. This description culminated in the calculation of an End Digit Preference Score (DPS) [257]. Using an *a priori* criterion for end digit preference, we interpreted DPS>20 as supportive of digit preference (i.e. heterogeneity across end digits). Second, we described 1) the mean difference between contemporaneous self-reported and measured weights, 2) the mean percentage difference relative to mean measured weight, and 3) then stratified the mean differences by factors hypothesized to influence the accuracy of self-reported weight. Because persons from specific Hispanic/Latino backgrounds tend to concentrate in specific geographic areas, not all Hispanic/Latino backgrounds were represented at each study center, creating confounding between background and center. Therefore we considered the cross-classification of Hispanic/Latino background by field site to construct meaningful contrasts of the differences within either background groups or sites. All Hispanic/Latino backgrounds with <100 participants at a given study site were pooled with individuals at the same site self-identifying as being of 'Mixed' or 'Other' backgrounds. Third, we used an unadjusted linear regression of measured weight on self-reported weight to estimate the overall correlation coefficient (r^2) using *a priori* criteria of good model fit of r^2>0.9.

Given that stratified means and correlation coefficients do not capture the complex differential sources of under- or over-reporting, in Aim1B I applied multivariate linear models to assess the joint influence of potential predictors of inaccuracy on the differences between self-report and measured weights using disjoint indicator variables. Whereas stratified mean differences between self-reported and measured weight reflect the observed misreporting (kg) for all individuals in a given stratum, multivariate effect estimates represent the estimated change (kg) in the difference between self-report and measured weight (henceforth referred to as 'change in difference') for a given stratum compared to the difference observed for the referent, after holding all other potential predictors constant.

In both Aims 1A and 1B I performed multiple imputation [258] to fill in missing predictor information (8.5% of the sample missing ≥1 predictors) and generate 25 stacked datasets for use in the multivariate analyses (20 burn in period). All statistical analyses accounted for the complex sampling design and sampling weights of HCHS/SOL in SAS 9.4 (SAS Institute, Research Triangle Park, NC).

iii. Strengths and Limitations

This current work was strengthened by our data quality control protocol. In contrast data quality control approaches that are based on a single criterion (e.g. >4 standard deviations), although straightforward, may compromise the representativeness of the analytic sample and artificially inflate estimates of accuracy. As such our approach to data quality control may be useful for future accuracy studies. This analysis constitutes the largest and most diverse accuracy study of Hispanic/Latinos to date. Additionally we utilized multiple imputation to account for missing predictors of inaccuracy and retain the full data set (n=16,119) in our stratified estimates (Aim 1A) and in our multivariate models (Aim 1B).

This work was not without limitations, however. Although the assessment of contemporaneous self-reported and measured weight isolates the impact of self-report bias, it does not shed light on the magnitude of recall bias or its interaction with self-report bias in the HCHS/SOL weight history data. Nonetheless good accuracy of current self-report weight in HCHS/SOL suggests that self-reported weight histories could be a valid tool for studying dynamic immigrant populations, if recall bias were to also be minimal. As discussed previously, there is an inherent confounding by geography in the design of HCHS/SOL, which relates to the non-random distribution of varied Hispanic/Latino backgrounds across the US. Therefore we assessed the cross-classification of background and site and created contrasts both within background and site for strata with ≥100 individuals to assure positivity. Yet we were unable to fully decompose the effects of site and background in our predictive modeling, given that both components had within group variability. Although the HCHS/SOL baseline design and data collection were extensive, we were unable to fully explore all conditions that might lead to large weight fluctuations in adulthood or frequent doctor visits (e.g. Auto-Immune Deficiency Syndrome, Bariatric Surgery, etc.). Lastly our results are not generalizable beyond the communities sampled in HCHS/SOL. Yet these communities give us important insights into the range of misreporting of self-reported weight that we might see in similarly diverse samples of US Hispanic/Latino adults.

2. Aim 2

To fully describe the components of a fine-mapping study of diverse populations in the PAGE Study I will begin by defining my exposures and outcome. Then I will describe the statistical analyses and present my power calculations to support this analysis, and conclude with a discussion of the strengths and limitations of Aim 2.

i. Exposure Measurement

The approximately 195,000 SNPS genotyped on the MetaboChip were chosen to fine-map 257 loci that had been validated with cardiometabolic traits as of 2009 and capture a low coverage genome-wide backbone [233]. Because the MetaboChip contains finely mapped 21 adult BMI loci known at the time of design and an additional 13 densely-genotyped loci implicated with BMI since 2009 [14-33], Metabochip data are well suited to investigate of the genetic architecture of obesity among US Hispanic/Latinos and inform how acculturation and genetic susceptibility may jointly influence BMI. SNPs with a Gentrain score <0.6, cluster separation <0.4, more than 1 Mendelian inconsistency, discordant duplicate calls, GenoSNP score >3.3%, call rate <95%, ambiguous mapping to the genome, or that were discordant with GWAS SNP in PAGE pilot studies or had ambiguous intensity plots were excluded by the PAGE Study CC (n=16,454) [235]. This analysis was restricted to the SNPs that passed the PAGE quality control filters (n=180,271) and were within the physical bounds of the 36 fine-mapped BMI loci defined as part of the MetaboChip design. I only reported on association results for low frequency and common SNPs (≥1% minor allele frequency). Additionally I excluded the 747 SNPs that failed Hardy Weinberg Equilibrium ($p<1x10^{-6}$) overall and in at least one background group in HCHS/SOL (not including individuals reporting backgrounds inconsistent with their parents, or individuals of other or multiple backgrounds).

ii. Outcome Measurements

As described above both weight and height were measured at examination by trained staff following a standardized protocol in HCHS/SOL [43] and WHI [253]. However, in MEC these anthropometric values were self-reported [203, 254] and in EAGLE BioVU and BioME height and weight were measured in the clinic as part of a

patient's examination [30, 33, 255, 256]. BMI was then calculated as the ratio of an individual's weight (kg) to their height (m) squared.

iv. Statistical Analyses

As has been in done in previous analyses of BMI in adults [204], I excluded individuals 20 years of age or younger as their BMI values may not be comparable (e.g. due to their potential for ongoing physical maturation) to individuals older than 20 years. Two exceptions are that in CARDIA this exclusion was not applied and in HCHS/SOL we also excluded individuals 20 years of age. Individuals with extreme BMIs (<18.5 and >70 kg/m^2) were excluded from the analytic sample for Aim 2 due to the possibility of data coding errors, an underlying illness or syndrome, or rare genetic mutations as has implemented in previous PAGE Study projects [203, 204].

The distribution of BMI is commonly skewed to the right [58] and this is also the case in the studies of the PAGE Study (e.g. in WHI-SHARe women, Figure 6). As compared to other practices to adjust for skewness, such as inverse ranked normalization, the natural log transformation results in parameters that are easier to interpret (i.e. as percent increase in BMI between an individual with one risk allele versus no risk alleles) and have been used in previous PAGE Study analyses [203, 204]. Yet the natural log transformation imposes the assumption linearity on the natural log scale when a variable contains more than two categories. In other words, genetic effects would be assumed to be linear or additive on the natural log scale across the observed genotypes (e.g. AA, AG, GG).

Throughout Aim 2 I assumed an additive genetic model (e.g. a SNP with a coded allele of G and non-coded allele of A would result in three possible genotypes: AA=0, AG=1, GG=2) and as per my directed acyclic graphic analysis (Figure 7) controlling for age, gender, principal components, and study center/region (as appropriate) were

sufficient to control for all anticipated sources of confounding, regardless of the unclear directionality between acculturation and socioeconomic factors. In HCHS/SOL we also adjusted for Hispanic/Latino backgrounds. Single SNP-BMI associations were modeled to determine the SNP with the strongest evidence of association as measured by the p-value (i.e. the 'top' SNP). I anticipated that the majority of these top SNPs will be common (>5% minor allele frequency), because the previously validated BMI loci were described as part of GWAS of common variation.

Due to the relatedness in HCHS/SOL principal components were calculated using an unrelated sample of six distinct 1000 Genomes reference sample populations (CEU, YRI, MXL, PUR, CLM and CHB) using a panel of 44,883 SNPs in low linkage disequilibrium and then principal component values were projected onto the entire HCHS/SOL sample [238]. Other studies calculated their principal components among an unrelated study sample (as defined as 2nd degree relatives or beyond, ≤0.35 estimated identity-by-descent allele sharing) and projected to their full sample. Ancestral outliers of the resulting principal components were excluded from further analysis [235]. A minimum of the top three principal components were included in all models to capture the three main ancestral groups contributing to the diversity in Hispanic/Latinos: African, European, and American Indian [112-116].

In HCHS/SOL GEEs were used to adjust for the complex sampling design and relatedness (clusters defined as 1st degree relatives or individuals sampled from the same household) using an optimized Horvitz-Thompson estimator in SUGEN based on selection probabilities adjusted for household- and individual-level non-response, and trimmed marginal inclusion probabilities [238]. In other studies with family structure (GenNet, HyperGen) linear mixed models were used in GWAF [259]. Standard linear regression models for the other studies were run using PLINK [240] or R (https://cran.r-project.org). Regression modeling based on less than 100 individuals were excluded

from all analyses. All SNP results were flipped to the positive strand and risk allele prior to the reporting of the final results.

We created a Bonferroni threshold of significance for the 170 index SNPs (or if unavailable on the MetaboChip, their highest LD proxy, $r^2 \geq 0.8$ in the discovery population 1000 Genomes pilot CEU, YRI, or CHB+JPT) from previous GWAS or Metabochip-wide studies after accounting for the 4 loci with more than one racial/ethnic-specific finding in tight linkage disequilibrium (LD, $r^2 \geq 0.8$ in CEU, YRI and CHB+JPT). Replication (i.e. in the same population of discovery) or generalization (i.e. to another population) was declared if an index SNP was: 1) Bonferroni significant at this threshold and 2) had a consistent direction of effect as the previous report. This same threshold was applied to any index SNP within the 36 fine-mapped BMI loci.

For the other SNPs in the fine-mapped regions, we generated a locus-specific Bonferroni correction for multiple comparisons based on the number of independent ($r^2 \leq 0.2$) SNPs with n a 50-SNP window, which was shifted by 5 SNPs each iteration, in the ARIC Study African American sample with MetaboChip data (n=3,399). This served as a worst-case scenario of the maximum number of independent tests in our population with the smallest LD blocks.

Among the subset of the 28,573 SNPs passing quality control and located in the 36 densely-genotyped loci, we conducted inverse variance fixed-effect meta-analysis across studies (>100 observations each) in METAL (version 2011-03-25) [260] when the SNP was >0.1% MAF in the racial/ethnic group and was informed by more than half of the maximum racial/ethnic-specific sample size. Any SNP-association with evidence of heterogeneity (defined as $p<0.05$ at the top SNP) was further investigated using forest plots, or if a trans-ethnic finding, using Bayesian fine-mapping and conditional analyses of fixed-effect estimates. As described above, MEC assessed BMI through self-reported weight and height and EAGLE BioVU and BioME utilized measurements of weight and

height from the medical record. I was careful to review the evidence of heterogeneity across studies in Aim 2A, given that self-report inaccuracy (assessed in Hispanic/Latinos as part of Aim 1) or inconsistent measurement protocols across the PAGE Study could induce heterogeneity across studies.

Similarly, in Aim 2B we generated trans-ethnic meta-analyses for SNPs >0.1% MAF in each racial/ethnic group and informed by more than half of the maximum trans-ethnic sample size (n=101,979) from at least two populations. We excluded the Alaskan Natives/American Indians from WHI from our trans-ethnic fixed-effect estimates due to small sample size (n=535).

Finally the fine-mapping of causal variants was informed by estimates of population-specific allele frequencies and LD correlation (r^2, 500 Kb sliding windows) in PLINK [261] using genotypes from the ARIC (African American), HCHS/SOL (Hispanic/Latino), and WHI studies (Asian, European and American Indian/Alaskan Native). Trans-ethnic LD estimates were generated from a sample of 17,437 individuals from HCHS/SOL, and WHI, which was proportionate to the racial/ethnic groups of our trans-ethnic meta-analysis. Regional plots were generated using LocusZoom to visualize trans-ethnic association differences as well as across the LD of various racial/ethnic groups [262].

We further investigated all of the 36 fine-mapped loci for second independent signals using GCTA (version 64) [34, 188]. We included the top SNP (i.e. marker with the lowest p-value within each region) in an approximate conditional model and contrasted the conditional effect estimates and P-values of the surrounding SNPs with their unconditional estimates, to ascertain if additional SNPs arise after we adjust for the top SNPs. We repeated this approach for any additional significant lead SNPs in low LD ($r^2<0.2$) with the previous top SNPs until no additional independent significant SNPs were identified.

Lastly, I relaxed the assumption of a fixed-effect across diverse racial/ethnic groups using an empirical Bayesian trans-ethnic meta-analysis in MANTRA [53]. I estimated mean effect allele frequency differences from our trans-ethnic results. Then I allowed MANTRA to empirically calculate the number of ancestral groups, or for comparison purposes force it to *K*=1 or *K*=*N* (number of study-, acculturation- and background-specific strata) to implement either a Bayesian fixed- or random-effects meta-analysis, respectively [53]. Significance in MANTRA analyses was determined using a threshold of a log10 Bayes Factor (BF)>5. I anticipated that no significant heterogeneous effects would remain after adjustments for ancestral differences across racial/ethnic groups using MANTRA. Furthermore we also calculated the posterior probability ϕ_j that the *j*th SNP in the *k*th independent signal is causal, and then ranked all SNPs by their BF's and summed their cumulative posterior probabilities until it exceeded 99%. Assuming that each independent signal contained only one causal variant genotyped on the MetaboChip, the resulting set of SNPs constitutes the 99% credible set and defines a genomic region where there is a 99% probability of containing the causal SNP.

v. Power Calculations

Prior to analyses I had anticipated having genotype and BMI information on a maximum analytic sample of 35,606 African, 26,048 Hispanic/Latino, 22,466 Asian and 535 American Indian/Alaskan Native descent adults. Therefore, at the time I calculated my expected power to detect a range of fixed-effect genetic estimates in Aim 2 using Quanto 1.2.4 (Figure 8). I had assumed an additive genetic model with a Bonferroni correction for the number of BMI loci I anticipated to test ($p<0.05/33$). Based on preliminary observations in HCHS/SOL as a representation of the outcome distribution in the other studies, I assumed a mean BMI (standard deviation) of 28.5 (1.22) kg/m^2. I had

decreased the maximum available sample size by 15% to account for the non-independence of the HCHS/SOL observations with respect to household and/or relatedness, the potential loss in power from conducting a weighted GEE model in HCHS/SOL, or missing person-level data. Thus, an effective sample size of 19,000 would correspond to this worst-case scenario for Asian Americans, our smallest racial/ethnic group >1000 observation and of non-European descent.

As shown in Figure 8, my power to detect SNP effects under a worst-case scenario for racial/ethnic-specific estimates (Aim 2A) would be ≥80% power for common variants (≥5%) of small genetic effect sizes (>0.1kg/m^2) and rare variants of (1 to <5%) of moderate effect sizes (>0.25kg/m^2). As minor allele frequency increases so would my power to detect smaller genetic effects. Our power to detect effects in the African, Hispanic/Latino American samples, and the trans-ethnic meta-analyses (Aim 2B) would be far greater.

vi. Strengths and Limitations

Even though it is possible that many of the BMI loci discovered after 2009 may not be fine-mapped on the MetaboChip, this analysis benefits from the large sample sizes available in the PAGE Study and the ability to capture 13 additional cardiometabolic loci that have been described with BMI since 2009. I propose a Bonferroni correction for the number of independent tests performed in each analysis as a way of limiting my type I error rate to declare generalization to 5% across all independent tests. Additionally several efforts have been made in the PAGE Study to harmonize the exposure, outcome, and effect measure modifier measurements of interest. This proposed analysis builds off of the best practices of several previous PAGE Study analyses that have used centralized genotyping and quality control, as well as harmonization protocols [203, 204].

In light of the strengths of this work, there are a few notable limitations. I was cautious when presenting fixed-effect meta-analyses in the presence of effect heterogeneity by study or race/ethnicity. Although Aim 2B incorporated a trans-ethnic to account for ancestral sources of heterogeneity (as captured by mean allele frequency differences in the results), outcome ascertainment or key environmental sources of heterogeneity may still remain (such as how acculturation captures changes in the obesogenic environment by gender, linguistic or cultural groups). My current analysis was not able to interrogate these sources of heterogeneity, but I plan to consider this in future studies of how obesogenic sociocultural environments can influence obesity and genetic susceptibility. Therefore, the loci implicated from Aim 2 will be targeted for follow-up study as key components of my overarching conceptual diagram (Figure 1).

E. Human Subjects

The PAGE Study CC, participating studies, and their study-sites all received Institutional Review Board approval prior to the initiation of each study. Although the specific forms differ slightly in wording and format, written and informed consent was obtained for the above outlined research activities, as well as the collection of data and genotyping. All investigators and staff associated with this proposal have received ongoing ethics and data security training/certification. The Institutional Review Board at the University of North Carolina at Chapel Hill has deemed that the secondary-data analyses of this dissertation do not require a full Institutional Review Board approval and are in accordance with the Institutional Review Board principles for ethical human subjects research. The HCHS/SOL Publications Committee has reviewed, approved of the research activities, and verified the data analysis (Aim 1 only) of this dissertation. In addition, the PAGE Publications Committee and its sub-studies have reviewed and approved the statistical analyses for Aim 2.

F. Supporting Figures

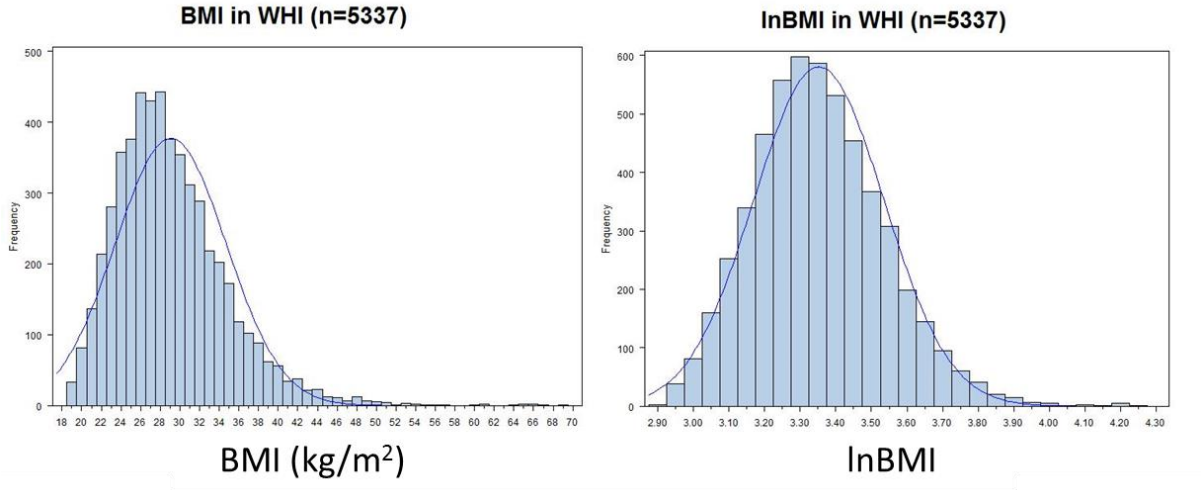


Figure 6A-B. Comparison of distributions of BMI (kg/m^2) in entire WHI sample of self-identified Hispanic/Latinas (n=5,337) before (A) and after natural log (ln) transformation (B).

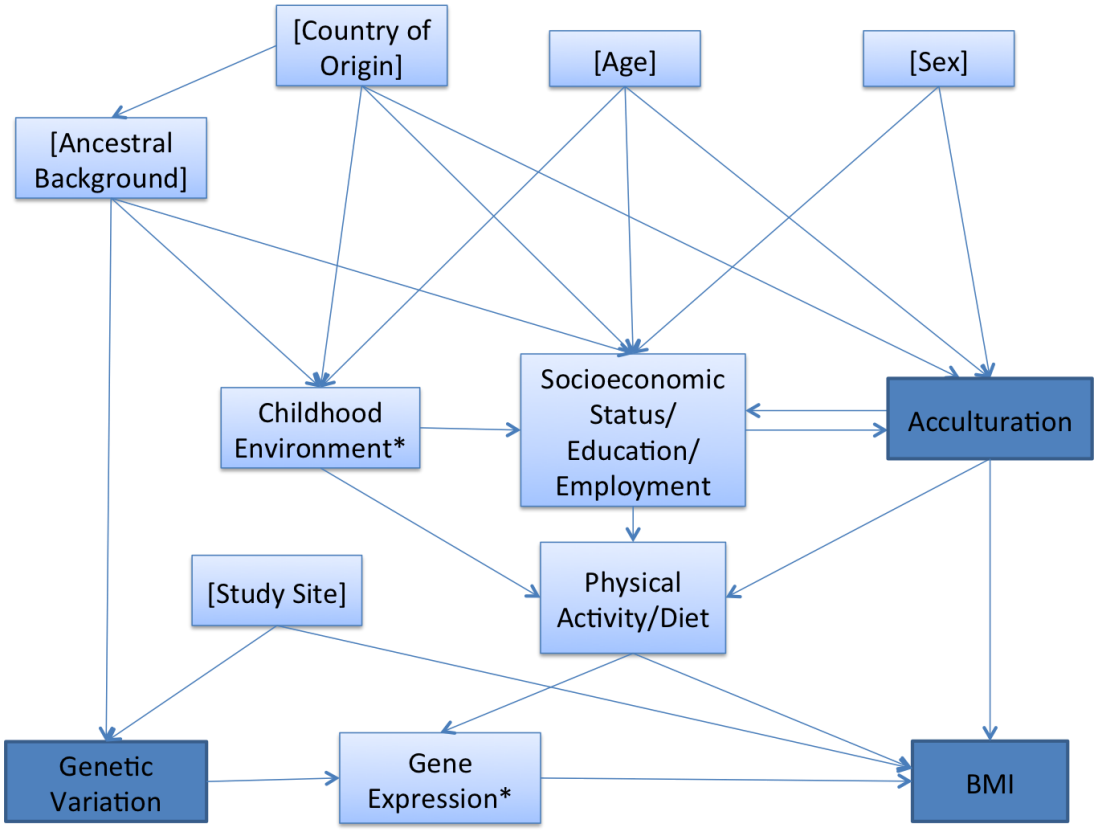


Figure 7. Directed Acyclic Graph of relationship between genetic variation and body mass index (BMI) with unclear directionality between acculturation and socioeconomic measures.

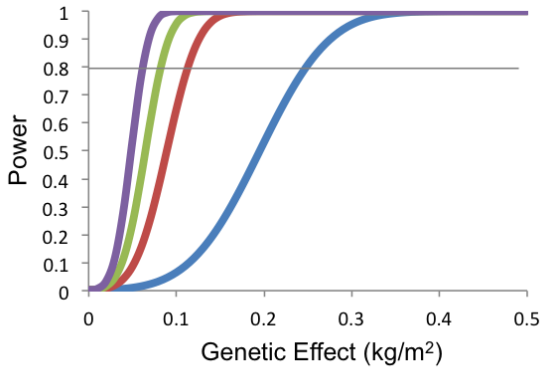


Figure 8. Power analysis of genetic effects across a range of minor allele frequencies (blue=1%, red=5%, green=10%, purple=20%) assuming an additive genetic model, a mean (SD) BMI 28.5 (1.22) kg/m^2, an alpha of 0.05/33 genetic loci, and a maximum effective sample size of n=19,000 (Quanto 1.2.4).

CHAPTER V: RESULTS

A. Manuscript 1: The Accuracy of Self-Reported Weight in a Diverse Population-Based Sample of Hispanic/Latino Adults from Four Urban United States Communities: The Hispanic Community Health Study/Study of Latinos (HCHS/SOL)

1. Overview

Background: Previous United States (U.S.) population-based studies have found that body weight may be underestimated when self-reported by individuals**.** However, this research may not apply to all U.S. Hispanics/Latinos, many of whom are immigrants with distinct cultural orientations to ideal body size. We assessed the data quality and accuracy of self-reported current weight in a population-based sample of U.S. Hispanics/Latino adults from various Hispanic/Latino backgrounds (or heritages).

Methods: Using baseline data (2008-2011) from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), we described the difference between contemporaneous self-reported and measured weight overall and across potential predictors of inaccuracy (n=16,119). Multivariate adjusted models were used to establish whether the observed trends in misreporting in a given predictor persisted after adjustment for all other potential predictors.

Results: Self-reported current body weight was well correlated with measured weight (r^2=0.95), and on average was 0.23 kg greater than measured weight. However, the following factors were associated with differential misreporting of weight: age group, gender, body mass index categories, nativity, study site by background, unit of self-report (kg or lb), and end digit preference.

Conclusions: We found slight over-reporting of weight in a diverse cohort of Hispanic/Latino adults from four U.S. urban centers, which may be due in part to characteristics of the HCHS/SOL cohort, such as its high proportion of immigrants. The direction of misreporting in self-reported weight, and thus the anticipated bias in obesity prevalence estimates based on self-reported weights, may differ in U.S. Hispanic/Latinos from that found in prior U.S. population-based studies.

2. Background

In spite of the potential for misreporting of body weight and biases associated with self-report, self-reported weight is often used in epidemiological studies when information on an individual's weight status is otherwise unavailable [151]. Previous population-based studies of the accuracy of self-reported weight indicate a tendency for participants to under-report their current weight as compared to their measured weight (with self-reported weight lower than measured weight by 0.1 kg [263] to 1.2 kg [264, 265]) with consistent differential misreporting across age, gender, and body mass index categories [151]. Yet this finding differs from a study of Mexican adults, which found that weight was over-reported by an average of 0.6 kg, indicating that the direction of misreporting may vary across cultures [47] due to the social desirability of body weight [157, 266].

Hispanic/Latinos comprise the largest United States (U.S.) minority group, representing 17% of the adult population in 2013 [96] at least half of whom were born outside of the U.S.[95] Among Hispanic/Latinos in the U.S. and abroad body mass index (BMI) categories [48, 49, 51], age [47, 49, 51, 153, 154], gender [47, 51], reproductive factors [75], household income [51], education, employment and nativity[49] have all been described as predictors of weight misreporting. Despite the recommendation to report the range of differences between self-reported and measured weight [151], we are

only aware of one study in Hispanic/Latino populations that has published this information [47]. Although U.S. Hispanic/Latinos represent diverse ancestries, cultural practices, languages and migration histories and have been documented to have variability in perceptions towards ideal body size [267], to our knowledge no previous study has compared the accuracy of self-reported weight across more than one Hispanic/Latino background (or heritage) in a predominantly foreign-born population-based cohort [47-49, 51, 75, 153, 154]. As new population-based studies collect more data on diverse U.S. Hispanic/Latinos, there is a need to reassess the data quality and accuracy of self-reports made in multiple languages or units.

Therefore using data from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), we describe the accuracy of self-reported body weight and factors associated with misreporting in a predominantly foreign-born population-based sample of U.S. Hispanic/Latino adults from various Hispanic/Latino backgrounds, including Central or South American, Cuban, Dominican, Mexican, and Puerto Rican.

3. Methods

i. Population

The HCHS/SOL is population-based cohort of 16,415 self-identified Hispanic/Latino adults (18-76 years at examination) of diverse Hispanic/Latino backgrounds (Central or South American, Cuban, Dominican, Mexican, and Puerto Rican) who were sampled using a probability study design from four U.S. urban communities (Bronx, NY; Chicago, IL; Miami, FL; San Diego, CA) between 2008-2011 [42, 43]. Centrally trained HCHS/SOL study personnel conducted the screening and baseline examinations in the participant's preferred language (English or Spanish). Women who reported being pregnant during screening were rescheduled for

examination ~3 months postpartum. Institutional Review Boards at all study sites approved the study procedures. All participants gave their informed written consent.

ii. Weight Measures

The baseline HCHS/SOL weight and height data were collected on participants who were able to stand on both feet while wearing a scrub suit or examination gown and no shoes [179] during the fasting block of the examination and immediately after a urine collection [43]. Trained and certified study personnel asked participants to self-report their body weight (to the whole kilogram, kg, or pound, lb) before measuring their weight to a tenth of a kg using a digital scale (Tanita Body Componsition Analyzer, TBF 300, Japan) and then transcribing it into an electronic data entry system. We converted self-reported weights in lb to kg and calculated the difference in kg (=self-reported weight – measured weight) between the self-reported and measured weights ("gold standard") as well as the percentage difference relative to measured weight [=(self-reported weight – measured weight)/measured weight * 100%].

iii. Data Quality Control

Quality control measures were implemented to ensure that all personnel were measuring anthropometrics with precision and included daily zero-balancing and weekly calibration of scales, centralized two day-personnel trainings that culminated with a certification that confirmed ≤0.5 kg weight agreement between a trainee and certified expert on ≥5 individuals, as well as periodic observations. Study personnel initialed each form and monthly the HCHS/SOL Coordinating Center notified the clinic managers of data points were beyond an expected error range, which resulted in refresher training(s) for the study personnel who collected those data points. With the consent of study participants, each study personnel audio recorded three baseline interviews (one per recruitment year) and were then randomly invited to share these recordings with study

personnel from others sites with the aim of determining if all sites and interviewers were implementing the study protocols and interviewing techniques consistently across the four HCHS/SOL sites.

Additionally, inter-rater reliability was assessed by randomly selecting 3-5% of participants for retest by a second certified assessor immediately after the initial anthropometric exam during the same baseline examination visit. Self-reported weights (n=565) differed between the two study personnel (=original – replicate) by 0.46 kg (95% confidence interval, CI: -0.12, 1.03 kg); whereas, measured weights (n=565) differed by 0.16 kg (95% CI: -0.18, 0.50 kg). This resulted in good reliability coefficients for self-reported and measured weight of 0.93 and 0.97 and relatively low coefficients of variation (the within-specimen variation expressed as a percentage of the mean) of 6.3 and 3.7%.

Given the large number of foreign-born Hispanic/Latino adults in the U.S. [95], we were concerned about unit confusion in the self-report (kg or lb) and sought to compare our range of difference between self-reported and measured weight with previous reports. One study from Mexico (where metric units are used) reported that among individuals ages >75 years, the differences between self-reported and measured weight ranged from -14.8 to 16.6 kg in males, and -8.6 to 14.7 kg in females [47]. As such we flagged absolute differences between self-reported and measured weight ≥15kg as possible data errors.

We then applied a data quality control protocol shown in Figure 9 and described in more detail in Appendix A to all 16,203 participants with data on both self-reported weight and measured weight (98.7% of entire sample) to: 1) address the flagged calculated differences between self-reported and measured weight as potential data errors, and 2) exclude currently pregnant women (who reported not being pregnant during the study screening but later reported being pregnant as part of their medical

history), or individuals with limb amputations (≥45 years, not otherwise affecting their ability to stand on both feet), or a body mass index <16 or >70 kg/m^2. Unless indicated otherwise, all results presented below pertain to the sample with both self-reported and measured weight that remained after applying this quality control protocol (n=16,119).

iv. Analysis

First, across and within each unit of self-report (kg or lb) we stratified self-reported and measured weights by their end digit and then compared the observed frequency with the expected frequency of 10% for each digit (i.e. under a uniform distribution of digits between 0-9). We assessed how much self-reported and measured weights deviated from homogeneity of end digit using the End Digit Preference Scores (DPS)[257] as described below:

$$DPS = 100 * \sqrt{\frac{\chi^2}{df * N_{weighted}}}$$

where df represents the degrees of freedom (= number of strata - 1) for the Rao-Scott Chi-square.

Using an *a priori* criterion for end digit preference, we interpreted DPS>20 as supportive of digit preference (i.e. heterogeneity across end digits).

Second, we described 1) the mean difference between self-reported and measured weights, 2) the mean percentage difference relative to mean measured weight, and 3) then stratified the mean differences by factors hypothesized to influence the accuracy of self-reported weight, which included demographic characteristics (age, field site, Hispanic/Latino background, gender), health behaviors, and factors relevant to weight gain/loss (body mass index categories, health insurance, smoking status, diabetes, history of cancer/malignant tumor or heart failure, menopausal status), socioeconomic and sociocultural factors (education, household income, language

preference at examination, nativity), and characteristics of self-reported weight (units of report, end digit preference). Because persons from specific Hispanic/Latino backgrounds tend to concentrate in specific geographic areas, not all Hispanic/Latino backgrounds were represented at each study center, creating confounding between background and center. In particular, participants of Cuban background were predominantly recruited in Miami, those of Dominican backgrounds were predominantly selected in the Bronx, and participants from San Diego were predominantly of Mexican background [43]. Therefore we considered the cross-classification of Hispanic/Latino background by field site to construct meaningful contrasts of the differences within either background groups or sites. All Hispanic/Latino backgrounds with <100 participants at a given study site were pooled with individuals at the same site self-identifying as being of 'Mixed' or 'Other' backgrounds.

As has been reported previously [46, 47, 50, 268], third, we used an unadjusted linear regression of measured weight on self-reported weight to estimate the overall correlation coefficient (r^2) using *a priori* criteria of good model fit of $r^2>0.9$.

Given that stratified means and correlation coefficients do not capture the complex differential sources of under- or over-reporting, we last applied multivariate linear models to assess the joint influence of potential predictors of inaccuracy on the differences between self-report and measured weights using disjoint indicator variables. Whereas stratified mean differences between self-reported and measured weight reflect the observed misreporting (kg) for all individuals in a given stratum, multivariate effect estimates represent the estimated difference (kg) in misreporting, as captured by the difference between an individual's self-report and measured weight and henceforth referred to as 'difference in misreporting', for a given stratum compared to the difference observed for the referent, after holding all other potential predictors constant. In addition to a complete case analysis, we also used multiple imputation [258] to fill in missing

predictor information (8.5% of the sample missing ≥1 predictors) and generate 25 stacked datasets for use in the multivariate analyses (20 burn in period). All statistical analyses accounted for the complex sampling design and sampling weights of HCHS/SOL in SAS 9.4 (SAS Institute, Research Triangle Park, NC).

4. Results

i. Unit and End Digit Preference

The majority of HCHS/SOL participants self-reported their body weight in pounds (96% weighted frequency). However, this varied across the four study sites. For example, 89% of individuals in San Diego and 99% of individuals in the Bronx preferred to self-report in pounds.

Over half of the participants (56%) self-reported weights ending in zero or five, which is above our expectation if interviewers or participants had no digit preference (20%, zeros and fives). Digit preference was evident for self-reported weights (DPS=23), but was the strongest for self-reports ending in zeros compared to all other digits (DPS=62). More self-reports ended in zeros and fives when the self-report was made in lb as compared to kg (58% versus 42% weighted frequencies; Appendix B) and similar trends were seen for end digit preference for even versus odd digits (Appendix C). No end digit preference was evident in the measured weights (DPS=5.8), which were transcribed by study personnel from the scale to the electronic data entry system.

ii. Raw Mean Difference

Prior to data quality control the calculated difference ranged from 74.5 kg under to 51.6 kg over-reporting, resulting in a mean difference of 0.26 kg (95% CI: 0.14, 0.37 kg; confidence limit difference, CLD: 0.23 kg) and coefficient of determination (r^2) of 0.94 (n=16,203). Although self-reported and measured weights were similar for most

individuals (i.e. most observations lay along the line of unity in Figure 11A), calculated differences were beyond four standard deviations from the mean for 129 individuals (gray lines in Figure 11C). We flagged 229 (1.4%) extreme absolute differences (≥15 kg) between self-reported and measured weight as possible instances of unit confusion.

iii. Quality Controlled Mean Difference

After data quality control, 48 of these extreme absolute differences between self-reported and measured weight were resolved reducing the number of flagged self-reported weights to 1% of the final analytic sample (181 of 16,119, Appendix A). The range of calculated differences decreased from 74.5 kg under to 51.6 kg over-reporting before quality control to 52.8 kg under to 35.4 kg over-reporting after quality control. This resulted in an attenuated mean difference (0.23 kg) and increased precision (95% CI: 0.12, 0.34 kg; CLD: 0.22 kg; Figure 11A-D). The difference between an individual's self-reported and measured weight as a percentage of their measured weights was on average 0.53% (95% CI: 0.40, 0.66%). As compared to the raw data, the quality-controlled data represent a tighter correlation between self-reported and measured weight (r^2=0.95). In a hexagonal binning plot the majority of data were placed into a few bins along the best-fit line (Appendix D). Yet differential self-reporting across measured weight was more evident after data quality control, as we observed a stronger tendency towards over-reporting among individuals with measured weight below the mean (78.8 kg) than above the mean, and a stronger tendency towards under-reporting among individuals with measured weight above the mean than below the mean (Figure 11C-D). Based on these findings all subsequent analyses were conducted using the quality-controlled data.

iv. Predictors of Misreporting

The magnitude and direction of the difference between self-reported and measured weight varied across strata of a number of potential predictors of inaccuracy including: age, categories of BMI, nativity, background by study site, unit and digit preference (Appendix E). Similar patterns were seen in both the stratified (Appendix E) and adjusted analyses (Appendix F and Table 3) of complete case and multiply imputed datasets, and therefore the results from the multiply imputed dataset are presented below. Age, categories of BMI, background by study site, digit preference, and gender/menopausal status each had at least one stratum where the multivariate adjusted difference between self-reported and measured weight and corresponding 95% CI did not contain the null, indicating that these predictors of misreporting were statistically independent of each other (Table 3).

v. Demographic Factors

There was a positive relationship between increasing categories of age and over-reporting of weight (Table 3), in which on average adults ages 18-29 years under-reported and those adults ages ≥30 years over-reported their weights (Appendix E). Although over-reporting of weight was generally observed in all strata of gender/menopausal status (Appendix E), when other predictor of misreporting were taken in account both pre-menopausal (-0.35 kg adjusted difference in misreporting) and post-menopausal (-0.43 kg) females tended to over-report their weight less than males (Table 3).

Overall individuals from all Hispanic/Latino background groups in the Bronx, and those of Puerto Rican and 'Other' backgrounds in Chicago under-reported their weight; by contrast those from all other site-background groups over-reported their weight (Appendix E). In Figure 10 we observed differential patterns of under-reporting among

the Hispanic/Latino backgrounds at the Bronx (Mexicans and Puerto Ricans under-reported their weights less than Central Americans; Appendix E), but these trends were not replicated at any other study site (Figure 10). Within the four Hispanic/Latino backgrounds that were represented at more than one study site, all adjusted contrasts contained the null except for the comparisons of Central and South Americans from the Bronx to their counterparts in Miami (Appendix G).

vi. Health-Related and Sociocultural Factors

When all other factors were held constant, underweight individuals were more likely to over-report their weight (1.58 kg adjusted difference in misreporting) as compared to normal-weight individuals, whereas over-weight and obese individuals were more likely to under-report their weights by -0.75 and -1.78 kg, respectively (Table 3). U.S.-born adults were more likely to under-report and foreign-born adults were more likely to over-report their weight (Appendix E). Although the 95% CI for the adjusted difference between self-reported and measured weight comparing U.S.-born and foreign-born individuals contained the null (-0.31 kg, 95% CI: -0.69, 0.08 kg; Table 3), the direction of misreporting was consistent with unadjusted findings [-0.41kg, difference between US-born (-0.09 kg) and Foreign-born adults (0.32 kg); Appendix E].

vii. Characteristics of Self-Reporting

On average, individuals who elected to report their weight in kg or used end digits of zeros and fives under-reported their weight, whereas those who chose to report in lb or used other end digits over-reported their weight (Appendix E). Adjusted estimates of difference in misreporting also supported the observation that individuals who used either larger units (i.e. kg, equivalent to roughly 2.2 lb), or end digits of zeros and fives to report weight were less likely to over-report their weight (Table 3).

5. Discussion

Given the diversity and recent population growth of Hispanic/Latinos residing in the U.S., there is a need to reassess the data quality and accuracy of self-reported weight, which is commonly used in obesity surveillance and epidemiologic investigations. This study is the first to our knowledge to describe the accuracy of self-reported weight in a large population-based sample of multiple Hispanic/Latino backgrounds—each with unique cultural, linguistic, and migration histories. We observed a strong correlation between self-reported and measured weight as well as slight over-reporting of weight on average after quality control (r^2=0.95, mean difference of 0.2 kg). This is consistent with good correlation described previously (r^2≥0.9) [46, 50, 268] and over-reporting in a nationally representative sample of Mexican citizens (mean difference of 0.6 kg) [47]. Yet the finding of over-reporting is contrary to the under-reporting noted in nationally representative samples of U.S. Hispanic/Latinos[156] and two convenience samples of women in the U.S. and Guatemala [75, 268]. For example, estimates from the U.S. National Health and Nutrition Examination Survey (continuous 2007-2008 NHANES) indicated that Hispanic/Latino adults under-reported their current weight on average (-0.35 kg), but the magnitude of under-reporting was less than half as much as observed among non-Hispanic/Latino Whites (-0.75 kg) [156]. To date the NHANES Hispanic/Latino samples have been primarily of Mexican heritage and US-born (foreign-born: 39% of females and 48% of males in NHANES III [49]); whereas, HCHS/SOL is a diverse sample of predominantly foreign-birth (foreign-born: 84% of females and 81% of males, Appendix A).

Although the mean difference between self-reported and measured weight did not contain the null, the magnitude of over-reporting observed in our study was considerably less than previous estimates of adult diurnal variability in weight, which have been reported to be up to 2 kg (4.4 lb) across a day [152]. We posit that the slight

over-reporting of weight may, in part, be attributable to the fact that the anthropometric assessment was conducted immediately following a urine collection within the fasting procedural block of time [43].

We are further reassured by the observation that several established predictors of inaccuracy[47-49, 51, 75] were also predicted differential self-reporting in our study both before and after applying multivariate adjustments to account for other potential predictors. However, there are some interesting differences in the magnitude and directionality of misreporting by gender in our diverse sample of Hispanic/Latino self-identified women and men. For example, in contrast to previous national estimates that described distinct effects of age by gender [153], we observed a similar tendency towards more over-reporting with increasing age in both females and males (data not shown). The magnitude of over-reporting observed in our sample of women (Appendix E) was less than the magnitude of under-reporting reported in recent national estimates of self-reported weight inaccuracy for all U.S. women (-1.38 kg) and Hispanic/Latinas (-0.59 kg) [156]. Even though the magnitude of over-reporting for males in our study was larger than these national estimates of all men and Hispanic/Latinos, the gender gap in self-report bias was narrower in our study than previous national estimates.

Given that half of the current adult U.S. Hispanic/Latino population are foreign-born,[95] previous studies from Latin America [47, 268] may inform the cultural origins of over-reporting of weight in samples like HCHS/SOL. In HCHS/SOL foreign-born adults on average over-reported their weight and their U.S.-born counterparts under-reported weight. This may be due in part to the cultural factors that influence body image and the complex process of acculturation to the dominant culture of the U.S. [157, 266, 267]. For example, qualitative studies of U.S. Hispanic/Latinas have documented a perception that their culture of origin considers a full figure to be desirable and healthy due to its connection with “wealth, affluence, and tranquility” [266], whereas in the U.S. they

perceive that it is desirable for women to be “extremely thin” [157]. Similarly in previous national estimates of Mexican Americans (NHANES III 1988-1994), foreign-born women and men (39% and 48% of sample) under-reported less and over-reported more than their U.S.-born counterparts [49].

This current work is strengthened by a transparent discussion of our data quality concerns and control procedures. In sum our efforts resulted in <1% altered observations (16 recoded and 84 excluded; Appendix A), yet interestingly we noted that the extreme observations in the raw data set obscured the differential misreporting across measured weights (Figure 11C-D). Data quality control approaches to identify data errors that are based on a single criterion (e.g. >4 standard deviations), although straightforward, may compromise the representativeness of the analytic sample and artificially inflate estimates of accuracy. As such our approach to data quality control may be useful for future accuracy studies.

In our sample 80% of participants preferred Spanish to complete the interview, but only 4% preferred to report their weight in kg, perhaps due to their participation in the U.S. medical system and its monitoring of weight in lb. This is supported by the observation that the majority of individuals who reported their weight in kg did not have current health insurance (65%), whereas less than half of individuals who reported in lb did not have current health insurance (48%). Although language preference did not appear to influence the accuracy of self-reported weight, self-reports made in kg were more accurate than those made in lb. This indicates that even though multiple units may necessitate additional quality control it may increase accuracy in future studies of U.S. Hispanic/Latinos. Additionally predictors were missing in <10% of the sample and we successfully retained these observations in our multivariate models using multiple imputation.

Our work, however, is not without limitations. First, there is an inherent confounding by study site (geography) in the design of HCHS/SOL. Therefore we assessed the cross-classification of background and site and created contrasts both within background and site for strata with ≥100 individuals to assure positivity. Although we were unable to fully decompose the effects of site and background, a predictive model including study site fit better than one with Hispanic/Latino background alone in exploratory analyses (data not shown). Our final multivariate model showed that within the Bronx there was more over-reporting of weight by Mexicans and Puerto Ricans, as compared to Central Americans (Figure 10). There was less over-reporting of weight by Central and South Americans living in the Bronx versus those living Miami and this pattern of less over-reporting in the Bronx was also consistent for Mexicans living in the Bronx versus those in San Diego (Appendix G). In light of the rigorous centralized coordination of the procedures and interviewing techniques of the HCHS/SOL baseline examination and the preference for self-reporting weights in lb at the Bronx (99%), future studies may want to examine if there is something unique about living in the Bronx that leads to Hispanic/Latino background groups to be more receptive to the U.S. cultural norms related to body size as they acculturate to the U.S.

Second, although the HCHS/SOL baseline design and data collection were extensive, we were unable to fully explore all chronic conditions that might lead to large weight fluctuations in adulthood or frequent doctor visits. However, we were able to assess prevalent diabetes, cancer, and health failure and they did not predict differential misreporting in our study.

Lastly our results are not generalizable beyond the communities sampled in HCHS/SOL. As such the slight over-reporting of weight observed in our sample of predominantly foreign-born Hispanic/Latino adults with a preference for Spanish may not

necessarily reflect the ever-increasing proportion of U.S.-born Hispanic/Latino adults [95] who may be more familiar with English and U.S. cultural biases around weight.

In summary, we observed a slight tendency towards over-reporting of weight (<0.3kg) in a community-based sample of adults from four U.S. urban centers, which was associated with demographic characteristics (age, gender, study site by background), health status (BMI categories), and self-report preferences (unit and digit preference). Etiologic analyses using self-reported weights in HCHS/SOL or similar samples of Hispanic/Latinos may need to account for the key sources of differential misreporting. Future studies of U.S. Hispanic/Latinos may increase their accuracy of self-reported weight by accommodating a participant's language or unit preferences. As such this study provides insights into the potential for a distinct pattern of social desirability towards weight among U.S. Hispanic/Latinos and serves as a model for future studies in populations of diverse backgrounds.

6. Main Tables and Figures

Table 3. Regression beta coefficients of predictors of accuracy of self-reported weight as compared to measured weight (kg) in Hispanic/Latino adults 18-76 years of age in HCHS/SOL (2008-2011) (n=16,119).

		β_{SR-M}*	SE	95% CI
Age (years)	18-22	0 (ref)		
	23-29	0.20	0.24	-0.27, 0.67
	30-44	0.48	0.20	0.08, 0.88
	45-59	0.76	0.20	0.37, 1.14
	60-76	0.82	0.24	0.36, 1.29
Background by Site	Dominicans- Bronx	0 (ref)		
	Central American- Bronx	-0.32	0.29	-0.88, 0.24
	Central American- Chicago	0.97	0.26	0.45, 1.48
	Central American- Miami	1.10	0.24	0.63, 1.57
	Cubans- Miami	1.01	0.19	0.64, 1.38
	Mexicans- Bronx	0.81	0.45	-0.07, 1.69
	Mexicans- Chicago	1.15	0.20	0.75, 1.54
	Mexican- San Diego	1.49	0.22	1.07, 1.91
	Puerto Ricans- Bronx	0.42	0.26	-0.09, 0.92
	Puerto Ricans- Chicago	0.61	0.32	-0.02, 1.24
	South American- Bronx	-0.07	0.33	-0.73, 0.58
	South American- Chicago	0.80	0.28	0.24, 1.35
	South American- Miami	1.00	0.29	0.44, 1.56
	Other- Bronx	-0.12	0.43	-0.96, 0.72
	Other- Chicago	0.33	0.52	-0.68, 1.34
	Other- Miami	1.16	0.39	0.40, 1.92
	Other- San Diego	1.58	0.41	0.79, 2.38
Body Mass Index Categories	Underweight (<18.5 kg/m^2)	1.88	0.41	1.07, 2.69
	Normal Weight (18.5-24.9 kg/m^2)	0 (ref)		
	Overweight (25.0-29.9 kg/m^2)	-0.75	0.10	-0.95, -0.56
	Obese (≥30.0 kg/m^2)	-1.78	0.13	-2.04, -1.52
Cancer History	Yes	-0.11	0.35	-0.79, 0.57
	No	0 (ref)		
Diabetic Status**	Normal Glucose Regulation	0 (ref)		
	Impaired Glucose Tolerance	-0.18	0.12	-0.42, 0.05
	Diabetes	0.17	0.16	-0.13, 0.48
Education	Less than high school or a GED	0.01	0.13	-0.23, 0.26
	At most high school or a GED	-0.10	0.12	-0.33, 0.13
	More than high school or a GED	0 (ref)		
End Digit Preference	5 or 10	-0.84	0.09	-1.02, -0.66
	1-4, 6-9	0 (ref)		
Gender	Female, pre-, peri-menopausal***	-0.35	0.13	-0.60, -0.10
	Female, post-menopausal	-0.43	0.13	-0.69, 0.17
	Male	0 (ref)		
Heart Failure History	Yes	0.12	0.35	-0.56, 0.80
	No	0 (ref)		
Health Insurance	Yes	0.09	0.12	-0.14, 0.33
	No	0 (ref)		
Language Preference	English	0.30	0.18	-0.06, 0.66
	Spanish	0 (ref)		
Nativity	Born in the United States****	-0.31	0.20	-0.69, 0.08

	Foreign Born	0 (ref)		
Physical Activity Level***	Inactive	0.03	0.12	-0.22, 0.27
	Low Activity	0.03	0.13	-0.23, 0.29
	Medium Activity	0.29	0.16	-0.02, 0.60
	High Activity	0 (ref)		
Smoking Status	Never	0 (ref)		
	Former	0.24	0.14	-0.03, 0.51
	Current	0.13	0.13	-0.13, 0.38
Socioeconomic Status	Less than $30,000 USD	0.14	0.10	-0.06, 0.35
	$30,000 or more USD	0 (ref)		
Unit of Self-Report	Kg	-1.09	0.30	-1.67, -0.51
	Lb	0 (ref)		

Abbreviations: CI=Confidence interval, GED=General Education Development Equivalent of a High School Diploma, M=Measured weight, ref=Referent, SE=Standard error, SR=Self-reported weight, USD=United States Dollars.

*Difference=self-reported minus measured weight (kg). Multivariate difference was calculated from a multivariate linear regression model of mean difference on the above possible determinants of validity (independent variables).

**As defined by the American Diabetes Association [269].

***Women reporting not reporting 'yes' to having reached menopause (change of life) were assumed to be pre- or peri-menopausal.

****As defined as being born in one of the 50 United States, not including United States Territories such as Puerto Rico.

*****As defined in the 2008 Physical Activity Guidelines for adults [270].

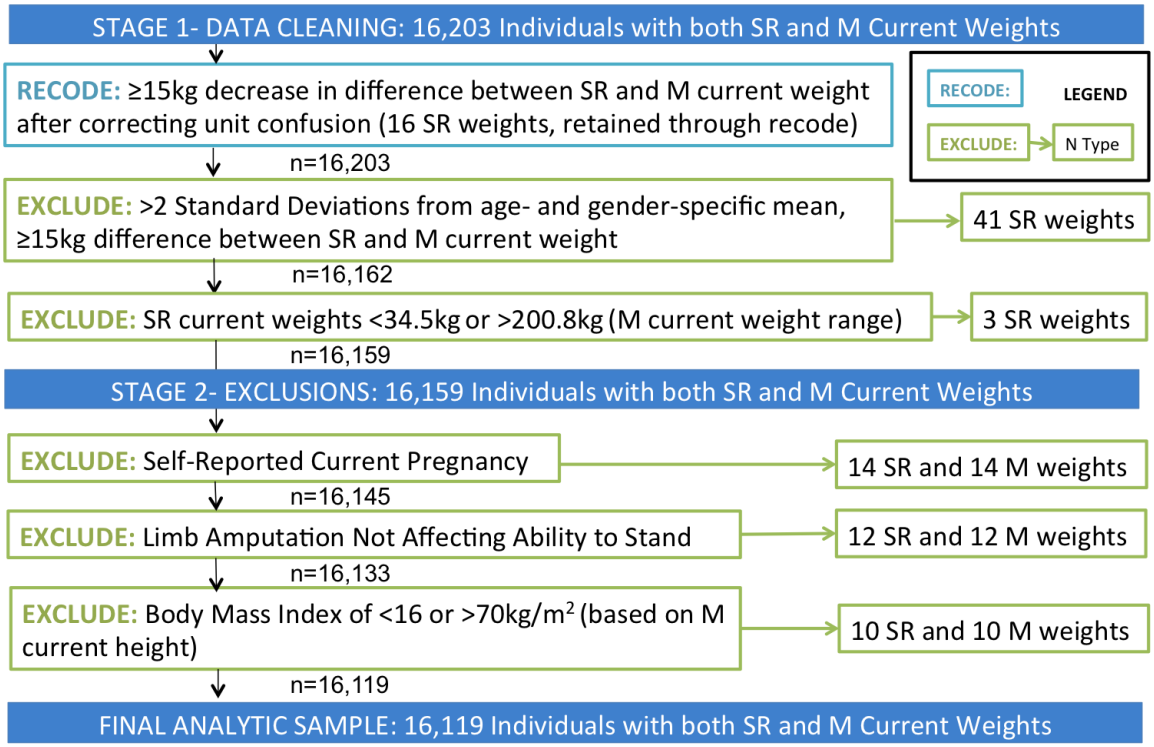


Figure 9. Flow chart of staged quality control on 16,203 adult Hispanic/Latino participants (18-76 years) with both self-reported (SR) and measured (M) weight at the baseline examination (2008-2011) of the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), resulting in 16 self-reported weights recoded due to unit confusion, 84 individuals excluded, and a final analytic sample of 16,119 participants.

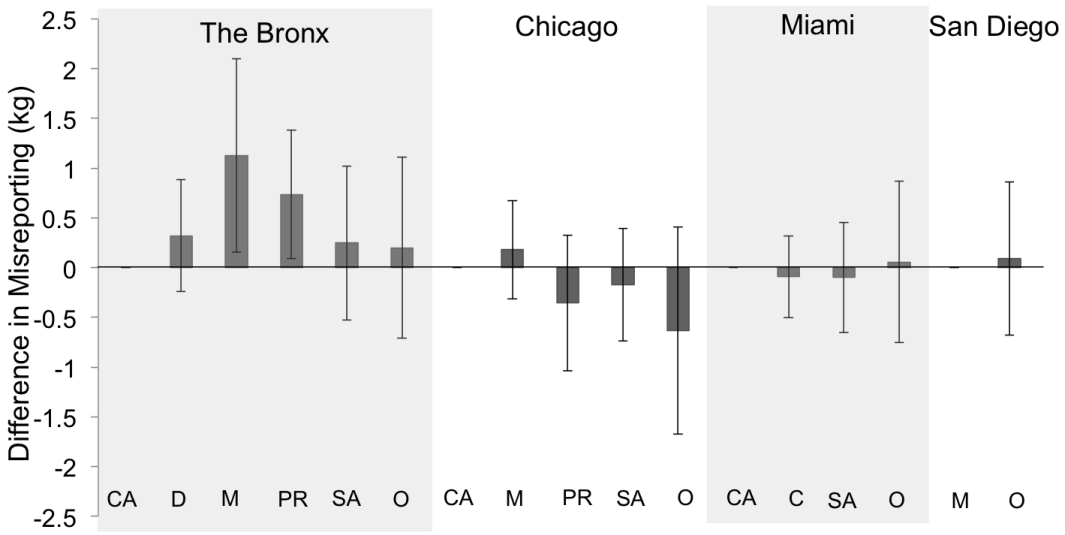


Figure 10. Multivariate estimated differences in misreporting (defined as the difference between self-reported and measured weights) and 95% confidence intervals comparing the seven Hispanic/Latino backgrounds (CA=Central Americans, referent for all sites but San Diego; C=Cubans; D=Dominicans; M=Mexicans, referent for San Diego; PR=Puerto Ricans; SA=South Americans; O=Other) within the study sites (The Bronx, NY; Chicago, IL; Miami, FL; San Diego, CA; ≥100 participants of a given background per site).

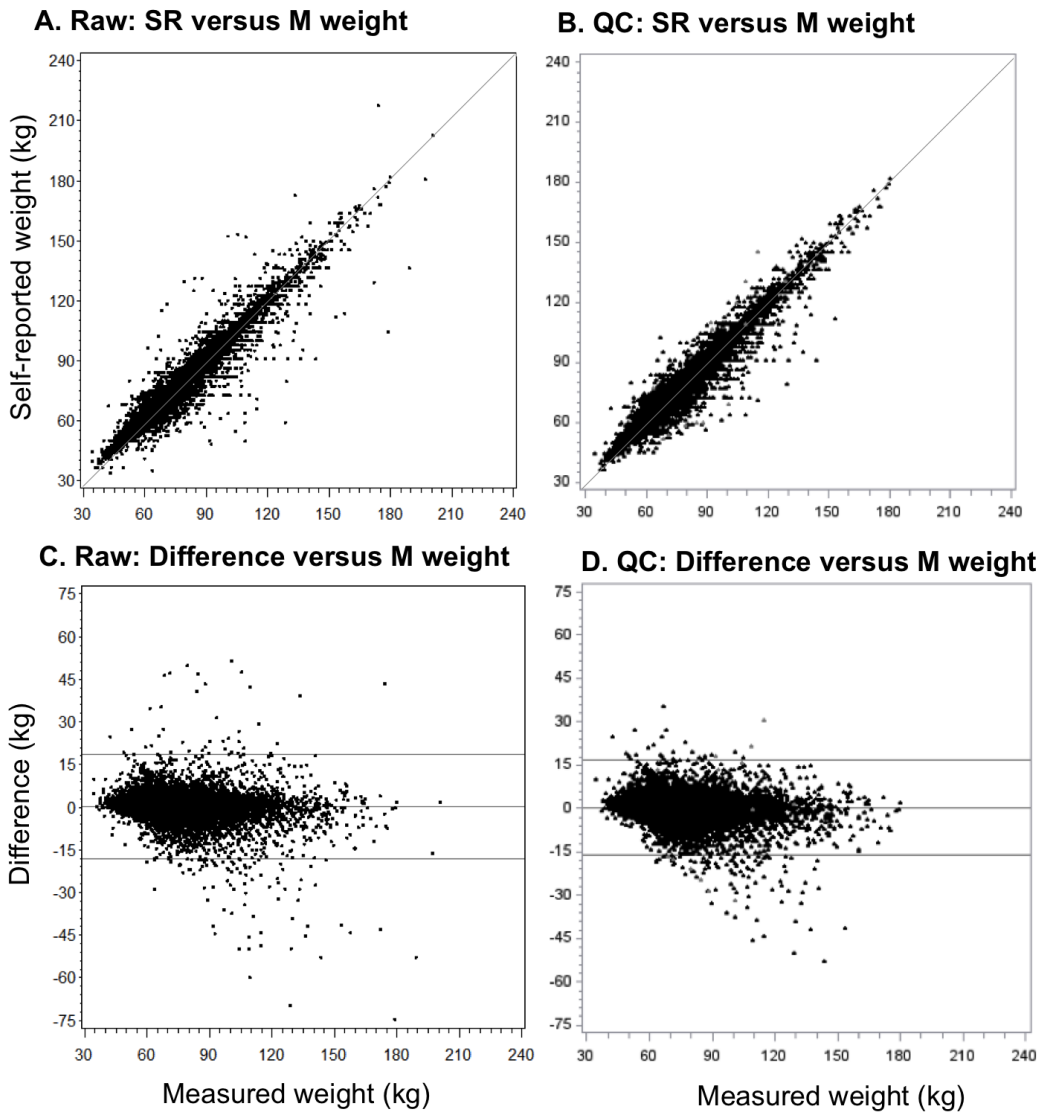


Figure 11A-D. Change in scatterplots of self-reported (SR) weight (Panel A and B, unity shown by gray line) and the difference between SR and measured (M) weight as a function of M weight (Panel C and D, mean difference and ± 4 standard deviations shown by gray lines; 129 versus 137 observations beyond 4 standard deviations from the mean, respectively) in raw (Panels A,C; n=16,203 observations in black) and quality controlled datasets (Panels B,D; n=16,119 observations in black and recoded values in gray).

B. Manuscript 2: Trans-Ethnic Fine-Mapping of Genetic Loci for Body Mass Index in the Diverse Populations of the Population Architecture using Genomics and Epidemiology (PAGE) Study

1. Overview

Most body mass index (BMI) genetic loci have been identified in studies of primarily non-Hispanic/Latino European ancestries. The effect of these loci in other racial/ethnic groups is less clear. Thus we aimed to characterize the allelic heterogeneity at 170 established BMI variants, or their proxies, to diverse US populations and trans-ethnically fine-map 36 BMI loci using a sample of >102,000 adults of African, Hispanic/Latino, Asian, European and American Indian/Alaskan Native descent from the Population Architecture using Genomics and Epidemiology Study.

We restricted our analytic sample to adults with BMI between 18.5-70kg/m^2 and performed linear regression of additive single nucleotide polymorphisms (SNPs) of the MetaboChip (Illumina, Inc.) on natural log-BMI, adjusting for age, sex, population stratification, study site or relatedness. We then performed fixed-effect meta-analyses and a Bayesian trans-ethnic meta-analysis to empirically cluster by allele frequency differences. Lastly, we approximated conditional and joint associations to test for the presence of secondary signals.

We noted directional consistency with the previously reported risk alleles beyond what would have been expected by chance (binomial $p<0.05$). Nearly a quarter of the previously described BMI index SNPs and 29 of 36 densely-genotyped BMI loci on the MetaboChip replicated/generalized in trans-ethnic analyses. We observed multiple signals at 9 loci, including the description of seven loci with novel multiple signals.

This study supports the generalization of most common genetic loci to diverse ancestral populations and emphasizes the importance of dense multi-ethnic genomic data in refining the functional variation at genetic loci of interest and describing allelic heterogeneity.

2. Introduction

Obesity is a global epidemic and has become a top public health concern given its downstream effects on cardiovascular disease, diabetes, cancer, and other diseases [1]. In the United States (US), there are marked racial/ethnic differences in obesity prevalence among adults [2]. For example, the US National Health and Nutrition Examination Survey estimated that in 2009-2010, non-Hispanic/Latino African descent (50%) and Hispanic/Latino (39%) adults had the highest burden of obesity; whereas adults of non-Hispanic/Latino European descent had the lowest (34%). Studies of Asian descent subpopulations indicate that they may have an even lower prevalence of obesity between 4-10% [271]. Given that non-European ancestries and Hispanic/Latinos collectively make up more than one third of the US population and are experiencing some of the fastest population growth [272], future public health research on the determinants of obesity in US must be relevant to these racial/ethnic minorities.

Body mass index (BMI, kg/m^2) is commonly used to classify obesity in epidemiologic studies and has been shown to be a polygenic trait with heritability estimates ranging between 40-70% [12, 13]. As numerous genome-wide association studies (GWAS) of predominantly non-Hispanic/Latino European descent populations have identified more than 100 BMI loci [21, 24-26, 29, 33, 186, 195], little is known about the effect of these loci in non-European ancestries. Therefore, the study of diverse populations can inform the generalizability and allelic heterogeneity of established loci and aid the identification of causal variants through trans-ethnic fine-mapping.

To this aim the Population Architecture using Genomics and Epidemiology (PAGE) Study was designed to extend the current body of knowledge on the genetic determinants of complex chronic diseases from studies of primarily non-Hispanic/Latino European descent populations to African, Hispanic/Latino, Asian and American Indian/Alaskan Native ancestries [41], which within the US are differentially affected by

the obesity epidemic [59, 271]. In this study of approximately 102,000 adults from diverse ancestries, we aimed to generalize a total of 170 previously described BMI index single nucleotide polymorphisms (SNPs), or their available proxies, located within 166 loci and to fine-map 36 of these BMI loci with dense genotyping on the MetaboChip (Illumina, Inc.) using trans-ethnic meta-analytic methods to narrow the putative interval for future biologic study.

3. Methods

i. Study Population

The Population Architecture using Genomics and Epidemiology (PAGE) Study is comprised of several large study sites/consortia and a coordinating center bringing together samples of diverse populations including those included in this analysis: the Atherosclerosis Risk in Communities (ARIC) Study, the Epidemiologic Architecture for Genes Linked to Environment study accessing BioVU (EAGLE BioVU), Coronary Artery Risk Development in Young Adults (CARDIA), Cardiovascular Health Study (CHS), the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), Multiethnic Cohort (MEC), the Women's Health Initiative (WHI) [41]. Additional studies collaborating in this analysis also included: the GenNet Network (GenNet), the Hypertension Genetic Epidemiology Network (HyperGEN) Study, the MEC-Slim Initiative in Genomic Medicine for the Americas Type 2 Diabetes Consortium (MEC-SIGMA), the Mount Sinai School of Medicine BioBank (BioME), and the Taiwan-MetaboChip Study for Cardiovascular Disease (TaiChi) study. A detailed description of each study can be found in our Supplemental Materials.

Racial/ethnicity was self-reported in most studies except for EAGLE BioVU where racial/ethnicity is observer-reported [273, 274]. MEC-SIGMA sample included Type 2 Diabetes cases and controls from Los Angeles, CA [275]. The TaiChi Consortium

substudies were conducted in Taiwan, the San Francisco Bay Area, and Hawaii and represent East Asian ancestry [276]. The PAGE MEC and WHI Hispanic/Latino samples predominantly represent individuals of Mexican origin [277], whereas the HCHS/SOL [6] and BioME Hispanic/Latino samples were more diverse with respect to Hispnaic/Latino backgrounds and admixture (e.g. African, European and American Indian) [278]. The majority of WHI Asian American samples were of Chinese and Japanese descent, but also included smaller samples of other backgrounds (e.g. Hawaiian, Filipino, Korean, and Vietnamese). MEC represents both Japanese and Hawaiian ancestries, which were analyzed separately based on their self-reported Asian background. Only WHI recruited American Indians/Alaskan Natives. Each study obtained approval from their Institutional Review Boards and written consent from all participants with the exception of EAGLE BioVU, which followed an opt-out program [255, 279].

ii. Genotyping and Imputation

The MetaboChip was a custom Illumina iSELECT array that contained approximately 195,000 SNPs and was designed to support large scale follow up of putative associations for cardiovascular and metabolic traits, including BMI [233]. Approximately 33% of the MetaboChip SNPs were included as replication targets and 62% were included for fine-mapping within 257 targeted densely-genotyped loci, which included 21 loci associated with BMI as of 2009 [233] and 15 additional loci (i.e. originally included on the MetaboChip for other cardiometabolic traits) associated with BMI since 2009 [21, 25, 26, 33, 186, 195]. Collectively, these 36 densely-genotyped BMI MetaboChip loci include 37,900 SNPs (Appendix I), represent 20% of all BMI loci identified as of June 2016, and contain more than a third of all BMI index SNPs, or their proxies, on the MetaboChip. We define a locus as was done as part of the design of the MetaboChip [233]. Therefore as shown in Appendix I the number of SNPs per locus,

which varied widely as a function of the base pair range of the putative region of interest (133 to 3,494 SNPs across 38 kb to 1.9Mb, respectively) and the tiered-prioritization of 11 dense-genotyping for cardiometabolic phenotypes of interest (e.g. BMI) [233].

As part of the PAGE Study, the genotyping of the MetaboChip was performed at research genomics laboratories: the Human Genetics Center of the University of Texas-Houston (Houston, TX), the Vanderbilt University Center for Human Genetics Research (CHGR) DNA Resources Core (Nashville, TN), University of Southern California Genomics Core (Los Angeles, CA), and the Translational Genomics Research Institute (Phoenix, AZ) [235]. Each genotyping center genotyped the same 90 HapMap YRI (Yoruba in Ibadan, Nigeria) samples and 2-3% study-specific blinded replicates to facilitate genotyping quality control. The study-specific SNP- and person-level quality control measures are summarized in Appendix J.

Imputation of MetaboChip SNPs was conducted in MEC-SIGMA (Hispanic/Latinos only), BioME (African and Hispanic/Latino ancestries), and WHI (representing 54% of WHI African descent women, and all of the WHI European descent women) using 1000 Genomes phase 1 reference populations, or in the case of WHI using study-specific reference samples [237], and then filtered on imputation quality (Appendix J). Less than a third of the final analytic sample genotypes were imputed.

Within each racial/ethnic group, related participants were identified within and across studies in PAGE using PLINK [261] to estimate identical-by-descent statistics. When apparent first-degree relative pairs were identified, the member from each pair with the lower call rate was excluded from further analysis with the exception of GenNet, HCHS/SOL, and HyperGen (Appendix J). In these studies the family structure was either accounted for using a linear mixed models (GenNet, HyperGen) or a generalized estimating equation incorporating clusters of 1st degree relative pairs/household members (HCHS/SOL) [202]. Any samples from studies without extensive relatedness

identified by an inbreeding coefficient (F) >0.15 were excluded [236]. Principal components of ancestry were calculated using the Eigensoft software [280, 281] and determined either among the unrelated subset, or in the 1000 Genomes reference populations, and then projected to the study sample [202]. Ancestral outliers of the resulting principal components were excluded by the PAGE Coordinating Center from further analysis [235].

iii. Ascertainment of BMI

GenNet, HCHS/SOL, HyperGen, WHI and TaiChi studies. In EAGLE BioVU, the median weight and height were calculated across the complete medical histories [282]. For BioMe, height and weight measures were obtained from participants' medical records at the time of enrollment [30, 33]. In MEC weight and height were self-reported by questionnaire with good validity [283, 284].

BMI was then calculated as the ratio of weight to height squared. Following previous PAGE study recommendations to remove extreme outliers [203, 204], BMI values <18.5 or >70 kg/m^2 are excluded from most studies due to the potential for these extremes to be coding errors, reflect underlying illnesses or rare genetic mutations. However, due to the young age of CARDIA participants, individuals <18.5kg/m^2 were retained in the analytic sample. To reduce the influence of growth and development on quantitative variation in BMI, we limited our analytic samples to adults >19 years of age in EAGLE BioVU, CARDIA, and BioME, and >20 years of age in HCHS/SOL. Across the PAGE studies we had genotype and BMI information available on a resulting analytic sample of 35,606 African, 26,048 Hispanic/Latino, 22,466 Asian and 535 American Indian/Alaskan Native descent adults (Appendices K-N).

iv. Statistical Analysis

As described previously [203, 204] the distribution of BMI was naturally log (ln) transformed to minimize the influence of outliers. All regression models were adjusted for age, sex, the top 2 to top 10 principal components, and study site, as appropriate for the racial/ethnic group and study (Appendix J). Study- and racial/ethnic-specific linear regression models were implemented in PLINK [261], R (WHI, https://cran.r-project.org), SNPTEST (BioME), GWAF (GenNet, HyperGen) [259], or a weighted version of a generalized estimating equation in SUGEN (HCHS/SOL) [202].

v. Generalization of Established SNP-Associations with BMI in Diverse Populations

We created a Bonferroni corrected threshold of significance for the 170 index SNPs (or if unavailable on the MetaboChip, their highest LD proxy, $r^2 \geq 0.8$ in the discovery population 1000 Genomes pilot CEU, YRI, or CHB+JPT) from previous GWAS or MetaboChip-wide studies after accounting for the four loci with more than one racial/ethnic specific finding in strong linkage disequilibrium (LD, $r^2 \geq 0.8$ in CEU, YRI and CHB+JPT). Replication (i.e. in the same population of discovery) or generalization (i.e. to another racial/ethnic group) was declared if an index SNP was: 1) Bonferroni significant for 166 independent tests at this threshold and 2) had a consistent direction of effect as the previous report. This same threshold was applied to any index SNP within the 36 densely-genotyped BMI loci. Using a binomial distribution, we tested if the number of observed SNPs with directional consistency between the risk allele observed in this study and prior studies was greater than would be expected by chance (50% expected allele consistency by chance, $p<0.05$ significant).

vi. Replication/Generalization of 36 Densely-Genotyped BMI Loci in Diverse Populations

To identify independent signals in the fine-mapped regions, we generated a locus-specific Bonferroni correction for multiple comparisons based on the number of independent SNPs ($r^2 \leq 0.2$, pruned in PLINK using a 50-SNP window that was shifted by five SNPs each iteration) in the African descent samples with MetaboChip data from the ARIC Study (n=3,399). This served as a worst-case scenario of the maximum number of independent tests in the present study's populations with the least LD. The resulting p-value thresholds for statistical significance ranged from 6.31×10^{-5} to 1.39×10^{-3} (Appendix I).

Among the subset of the 28,573 SNPs passing quality control and located in the 36 densely-genotyped loci (range per locus: 110 to 2,785; Appendix I), we conducted inverse variance fixed-effect meta-analysis across studies (>100 observations each) in METAL (version 2011-03-25) [260] when the SNP was >0.1% minor allele frequency (MAF) in the racial/ethnic group and was informed by more than half of the maximum racial/ethnic-specific sample size.

vii. Trans-Ethnic Meta-Analyses to Narrow the Putative Interval

Similarly, we generated trans-ethnic meta-analyses for SNPs >0.1% MAF in each racial/ethnic group and informed by at least two populations and more than half of the maximum trans-ethnic sample size (n=101,979). We excluded American Indians/Alaskan Natives from our trans-ethnic fixed-effect estimates due to their small sample size and apparent extensive allelic heterogeneity given their recruitment across all nation-wide WHI recruitment centers (n=535).

Linkage Disequilibrium: Finally the fine-mapping of causal variants was informed by estimates of population-specific allele frequencies and LD correlation (r^2, 500 Kb sliding windows) in PLINK [261] using genotypes from the ARIC (African descent), HCHS/SOL (Hispanic/Latino), and WHI studies (Asian, European, and American Indian/Alaskan Native ancestries). Trans-ethnic LD estimates were generated from a sample of 17,437 individuals from HCHS/SOL and WHI, which was proportionate to the racial/ethnic groups of our trans-ethnic meta-analysis. Regional plots were generated using LocusZoom to visualize trans-ethnic association differences as well as across the LD of various racial/ethnic groups [262].

Bayesian Trans-Ethnic Meta-Analysis: Lastly, the assumption of fixed-effects across racial/ethnic groups was relaxed in a Bayesian trans-ethnic meta-analysis in MANTRA, which allows for the empirical estimation of mean allele frequency differences between racial/ethnic groups as prior information in the clustering of the observed genetic effects across defined racial/ethnic groups [53]—in our case African, Hispanic/Latino, Asian and European ancestries. We adjusted for multiple comparisons in this Bayesian analysis by defining strong evidence in favor of association to have a Bayes Factor (BF)>5. Furthermore we also calculated the posterior probability ϕ_j that the *j*th SNP in the *k*th independent signal is causal as:

$$\varphi j \quad = \frac{BF_j}{\sum_k BF_k}$$

We then ranked all SNPs by their BFs and summed their cumulative posterior probabilities until it exceeded 99%. The resulting set of SNPs constitutes the 99% credible set and defines a genomic region where there is a 99% probability of containing the causal SNP, if the assumption holds that each region of interest contained only one causal variant.

Established and Novel Secondary Signals at Known Loci: We further investigated our trans-ethnic fixed-effect meta-analysis results at the 36 densely-genotyped loci for second independent signals using Genome-wide Complex Trait Analysis (GCTA, version 64) [34, 188]. To inform our approximations we used the same trans-ethnic genotypes of 17,437 individuals from HCHS/SOL, and WHI, which were used to calculate trans-ethnic LD above and were proportionate to the racial/ethnic groups of our trans-ethnic meta-analysis. Then we included the 'lead SNP' (i.e. the marker with the smallest p-value within each region) in an approximate conditional model and contrasted the conditional effect estimates and p-values of the surrounding SNPs with their unconditional estimates to ascertain if any additional SNPs that were at least associated unconditionally with BMI at $p<0.05$ arose as 'independent' after we adjusted for the lead SNPs. We repeated this approach including any additional significant lead conditional SNPs in the region until no additional independent significant SNPs were identified.

Then we entered these potentially independent SNP markers into an approximate joint model in GCTA, which also included all of the lead SNPs in the 36 densely-genotyped loci as well as the index SNPs outside of these regions. Joint analyses were repeated dropping out the SNPs with non-significant joint p-values, until a final joint model included only significant joint SNP associations. As a sensitivity analysis of a subset of loci with evidence of two independent signals in the approximate GCTA analyses, we performed a single round of exact conditional analyses using statistical analysis software and performed fixed-effect meta-analysis in METAL as described above. In this round we adjusted for the lead fixed-effect trans-ethnic SNP and queried the significance of the remaining SNPs within the densely-genotyped region.

The percent variance explained for a given SNP association (β) and frequency (f) was estimated for the significant joint SNP associations as previously described [33]:

$$VarExp = \beta^2(1-f)2f$$

Additionally these significantly joint SNPs were queried for functional annotation in HaploReg (version 4.1) [285]. Both GERP and SiPhy conservation, as well as GENCODE and RefSeq genetic annotations were queried on each lead SNP.

viii. Statistical Power

In Quanto version 1.2.4 [286] we calculated power to detect genetic effects on BMI. Previous PAGE meta-analyses using this transformation have estimated that genetic effects for risk variants at *FTO* could be as much as 1% change in BMI per risk allele (or 0.0119 on the natural ln scale) [204]. Using information available on the worst-case locus-specific Bonferroni correction from Appendix I (6.31×10^{-5}), the varying BMI distributions and sample sizes of the race/ethnic specific and trans-ethnic meta-analyses we calculated power to detect effects up to as large as 1% change in BMI per risk allele.

As shown in Appendix O, power was expected to be greatest in the trans-ethnic meta-analysis, which would allow for the identification of moderate genetic effects (>0.6% change per risk allele) at ≥80% power for low frequency variants (≥1%). Despite the smaller size of the Asian descent sample, we estimated that we generally would have better power in the analysis than in the African and Hispanic/Latino (>13,000 and >3,000 samples larger, respectively) descent analyses, which would allow us to describe large genetic effects at ≥80% for both low frequency and common variants (≥1%). In contrast, the African, Hispanic/Latino, and European descent analyses were expected not have sufficient power (<80%) to describe low frequency variants (e.g. ≤1%), and only had sufficient power (≥80%) to describe moderate effects (>0.6% change per risk allele) that were common (≥5%) in that specific race/ethnic group.

4. Results

Our study was comprised of 102,514 individuals from five racial/ethnic groups, with a mean age spanning from 27 years old (range: 20-37 years) in CARDIA to 73 years (65-93 years) in CHS (data not shown). The biobank studies (EAGLE BioVU, BioME), HCHS/SOL, HyperGen, and TaiChi each represented ages from more than 5 decades across the life course. Women comprised the majority (or entirety, as in the WHI) of all studies, except for the TaiChi sample, which was only 39% female. Within sex obesity prevalence varied substantially across studies (26-64% of females and 19-46% of males were obese at the time of examination/self-report). Yet obesity prevalence was higher in women and men of African, Hispanic/Latino and American Indian/Alaskan Native ancestry compared to women and men of Asian and European ancestry.

i. Generalization of SNP-Associations to BMI in Diverse Populations

Overall, 135 of 165 SNPs, or their proxies ($r^2 \geq 0.8$), were previously shown to associate with BMI, passed quality control filters in at least two racial/ethnic groups, and displayed consistent directions of effect in the trans-ethnic fixed-effect meta-analysis (Appendix P). This is more concordant than would be expected by chance (binomial p, $p_{bin}=1.63 \times 10^{-17}$). Of all 170 index SNPs, or their proxies, that passed quality control filters in at least one racial/ethnic group, 42 were significantly associated with BMI in either the trans-ethnic analyses or in at least one racial/ethnic group (Appendices K-N, P, O). For example, we replicated two African descent-specific associations at *GALNT10* (rs4569924 $p=4.79 \times 10^{-5}$ [30]*)* and *DHX34* (rs4802349, $p=3.79 \times 10^{-8}$ [204]), and demonstrated generalization of associations from previous studies of European descent populations for two SNPs at 8p12 (rs7844647, $r^2=0.96$ in CEU, $p=2.03 \times 10^{-4}$ [186]) at *AGBL4* (rs657452, $p=5.52 \times 10^{-6}$ [33]) to African and Hispanic/Latino descent individuals, respectively.

Eighteen of the 42 significant index SNP associations were only significant in the trans-ethnic sample, perhaps due to its larger sample size (Appendices K-N, P). Three SNPs exhibited significant heterogeneity across the racial/ethnic groups in the trans-ethnic fixed effect meta-analysis, yet only one of these SNPs (rs116612809 the index SNP at *BRE* and the most significant ('top') SNP in the African descent and trans-ethnic fixed-effect analyses) persisted to have evidence in favor of association after accounting for the ancestral heterogeneity in a Bayesian meta-analysis. One index SNP at *TRAF3* (rs7143963; [186]) was nominally significant and directionally consistent in both the African descent and trans-ethnic analyses, but only exhibited significant heterogeneity across the studies of African descent individuals (Appendix Q), wherein the effect estimates from two studies with <1,200 individuals were the most extreme (HyperGen n=1171, Risk Allele Frequency=66.9; MEC pilot n=433, 59.2%).

ii. Replication/Generalization of 36 Densely-Genotyped BMI Loci in Diverse Populations

In 35,606 African descent individuals, 31 of 35 index SNPs (or their proxies) that passed quality controls and were located within one of the 36 densely-genotyped BMI loci showed an association that was directionally consistent with the previously reported risk allele ($p_{bin}=1.52x10^{-6}$). We observed no significant heterogeneity within the studies contributing samples of African descent individuals at either the index or lead SNPs. Our analysis of the dense genotypes of African descent individuals led to the generalization of 14 BMI loci (Table 4), including six loci (*COBLL1*, *FLJ35779, SLC22A3, TCF7L2, MAP2K5, SH2B1)* not previously associated and eight loci that were previously generalized to African descent individuals [204]: *SEC16B, ETV5, TFAP2B, FTO* and *MC4R* with the same lead SNP, and *TMEM18, GNPDA2, and BDNF* with a different top marker (r^2 of 0.86, 0.98, 0.11, respectively). Additionally as described previously [204],

rs116612809 at *BRE* replicated as the most significant SNP for BMI in our expanded African descent sample (Table 4). Thus our findings resulted in a total of 15 BMI loci with significant evidence of association in African descent individuals, six of which were best represented by the index SNP from GWAS of European [21, 33], and non-European populations [30, 32, 204].

In a sample of 26,048 Hispanic/Latinos, 32 of 36 index SNPs in the densely-genotyped BMI loci had associations that were directionally consistent with previous reports ($p_{bin}=8.57x10^{-7}$). We also observed no significant heterogeneity within the Hispanic/Latinos studies at either the index or lead SNPs. Using the dense-genotyping at 36 BMI loci, we were able to generalize 13 BMI loci to Hispanic/Latinos (Table 5), including 8 loci that were generalized to African descent individuals (*SEC16B TMEM18, COBLL1*, *GNPDA2, TCF7L2, MAP2K5, FTO* and *MC4R*) plus an additional 5 loci (*LYPLAL1, IGF2BP2, SLC39A8, KCNQ1, MTCH2*) that only generalized to Hispanic/Latinos.

In the entire Asian descent sample (n= 22,466), 29 of 34 available index SNPs were directionally consistent ($p_{bin}=4.76x10^{-6}$). At *MAP2K5* we did observe evidence of heterogeneity across the Asian descent studies at one nominally significant SNP (rs182297248, $p=4.5x10^{-4}$, $p_{het}=2.7x10^{-4}$, Appendix R). Excluding the Hawaiian sample from the MEC (n=2,586) did diminish the effect heterogeneity ($p_{het}=2.3x10^{-3}$) and decreased the p-value, but it remained nominally significant ($p=1.7x10^{-4}$). When we included the Hawaiian samples from the MEC we were able to generalize to Asian descent adults at eight BMI loci, including loci that were previously generalized to African descent individuals (*FLJ35779, TFAP2B, BDNF),* Hispanic/Latinos (*MTCH2),* or both racial/ethnic groups (*GNPDA2, TCF7L2, FTO, MC4R)* (Table 6). The lead SNP at *MC4R* was the index SNP from GWAS of European/trans-ethnic populations [21, 32]. In addition, we replicated three loci (*CDKAL1, KCNQ1, QPCTL*) that were previously

described in only Asian populations using lead SNPs that were in strong LD (r2>0.8) with the previously reported index SNPs [25, 195], or were the Asian index SNP itself [25, 26, 195]. In summary a total of 11 BMI loci replicated or generalized to our sample of Asian Americans. We noted that *MTCH2 and MC4R* were no longer Bonferroni significant when we excluded the Hawaiian samples from the MEC in our exploratory analyses ($p<3x10^{-4}$), and therefore carried forward the full Asian descent sample in our trans-ethnic meta-analyses, below.

In the European descent sample (n=17,859), 30 of 35 available index SNPs were directionally consistent ($p_{bin}=9.45x10^{-6}$). We observed no significant heterogeneity across studies at either the index or lead SNPs. Additionally, we replicated associations at nine BMI loci, including five loci that previously had not been associated with any other racial/ethnic group (*NEGR1, LRPN6C, PRKD1, KCNJ2, KCTD15)*.

Lastly, in the small sample of 535 American Indian/Alaskan Native women 22 of 35 available BMI index SNPs were directionally consistent ($p_{bin}=4.30x10^{-2}$). We were able to generalize the lead SNP (rs73012297, 6.55% change in BMI per C allele, $p=2.2x10^{-4}$) at *SLC22A3* to American Indian/Alaskan Native women, at a different lead SNP than had generalized to African descent individuals (rs116859471, in ARIC $r^2<0.01$ with top American Indian/Alaskan Native SNP; Table 4).

iii. Trans-Ethnic Meta-Analyses to Narrow the Putative Interval

Across the ancestries carried forward to trans-ethnic analyses (African, Hispanic/Latino, Asian and European descent), we saw greater variability in risk allele frequencies than effect sizes at index BMI SNPs of the densely-genotyped BMI regions on the MetaboChip (Figure 12). Trans-ethnic fixed-effect meta-analysis in up to 101,979 individuals generalized 29 of 36 BMI loci (Table 7). Most of these loci were already replicated/generalized to at least one racial/ethnic group (Figure 13).

The Bayesian trans-ethnic meta-analysis did not reveal additional loci strongly associated with BMI, as defined as log10 Bayes Factor>5 (Table 7). However, after accounting for ancestral heterogeneity 22 loci had strong evidence in favor of association and only three of these were noted to have a different lead SNP as seen in the fixed-effect analysis. For example, at *BRE* the Bayesian approach resulted in a top/index SNP, which had significant heterogeneity across the African descent studies (Table 4) and across the racial/ethnic groups (Table 7); whereas, the fixed-effect meta-analysis resulted in a lead SNP that was located ~300kb towards *FOSL2* (Appendix S). The other two loci (*IGF2BP2, QPCTL)* with top significant SNPs that differed between the two trans-ethnic approaches appeared to be capturing the same signal across the range of LD (e.g. African to European descent) represented in our trans-ethnic meta-analysis.

Using the physical location of the top fixed-effect racial/ethnic specific results, we compared our results to the initial range of the fine-mapped region (Appendix I) and calculated a percentage reduction of our putative interval of interest (Table 8). Across the 29 loci with significant trans-ethnic fixed-effect estimates the reduction in base pairs and percentage narrowed ranged from 14,099 (37% of region) to 930,200 (72%).

Using a Bayesian approach to account for ancestral heterogeneity, we used the physical bounds of the 99% credible set to reduce the putative interval by 52,690 base pairs (bp) at *ETV5* (46% of region) to 764,979 bp at *CDKAL1* (96% of region; Table 8). Figures 12-13 illustrate the trans-ethnic fixed-effect estimates of 12 loci where the Bayesian approach narrowed the putative interval to ≤12 SNPs. The remaining 24 fine-mapped regions are plotted in the Supplement (Appendices S-V). At three of these loci (*SEC16B, TFAP2B, MC4R*) the 99% credible set reduced the interval of interest by between from 182,749-566,266 bp to a single SNP (Figure 14).

iv. Established and Novel Secondary Signals at Known Loci

First, we sought to determine if previously reported secondary signals in studies of European descent individuals (*BDNF, MC4R;* [33]) were independently associated with BMI in our trans-ethnic sample. Consistent with previous reports we observed nominally significant evidence that rs10835210 (p_c=4.22x10^{-2}) at *BDNF* was independent of rs11030104. Two *MC4R* SNPs (rs9944545 p_c=6.19x10^{-3}, rs17066842 p_c=1.48x10^{-2}) were nominally independent of the index/lead SNP, rs6567160, in our trans-ethnic sample. Additionally we noted that rs2331841, originally reported in Asian populations [26], was also nominally independent of our top finding in the region (p_c=4.10x10^{-2}). All other conditional p-values were >0.24 or were not estimable.

However, given that these index SNPs may not be the best markers of the BMI signals in our trans-ethnic sample, we then performed conditional analyses of the trans-ethnic fixed-effect estimates in the 36 densely-genotyped BMI loci after adjusting for the top trans-ethnic fixed-effect SNP. Then in an approximate joint analysis (Table 9), we confirmed the presence of Bonferroni significant secondary signals at *BDNF* and *MC4R*, which were in low LD (r^2<0.3) with our top trans-ethnic findings. Additionally, our conditional and joint analysis of *GPRC5B* (lead SNP, rs67501351; joint p, p_j=7.70x10^{-19}) and *GP2* (index SNP, rs11074446; p_j=1.69x10^{-7}) indicated their independent associations with BMI.

Similar to the previous observation of a secondary signal at *FTO* with Type 2 diabetes [287], we also noted three additional independent signals in our trans-ethnic sample with BMI in conditional (not shown) and joint analyses (Table 9), which had varying degrees of LD in our trans-ethnic sample with our lead SNP (r^2=0.05-0.41). We also observed evidence for 6 additional novel secondary signals at *LYPLAL1, COBLL1, LOC646736, SLC39A8, TFAP2B, OVCH2* (Table 9). Incidentally 2 of these 10 loci had 99% credible intervals that included 1-6 SNPs (*TFAP2B* and *FTO;* Figures 10-11).

Whereas, the 99% credible intervals for the other loci with novel conditional signals were less refined including ≥15 SNPs (Table 8).

Interestingly the top/index SNP at *BRE* was significant in the single-variant model but not statistically significant in the joint model of the most significant SNPs representing each signal, which included a variant >3 Mb upstream at *ADCY3* (rs10182181, $p_j=2.42x10^{-10}$). Conditional analyses adjusting for rs10182181 at *ADCY3* confirmed that the top fixed-effect and Bayesian SNPs in the region were no longer Bonferroni significant ($p_c=2.02x10^{-3}$ and $9.94x10^{-3}$, respectively), suggesting that this association may in part be related to long-range LD patterns.

We also conducted an exact conditional sensitivity analysis in a subset (6 of 8 loci) of the densely-genotyped BMI loci with evidence of two independent signals in the conditional and joint GCTA analyses. At these six loci we noted at four loci locus-specific Bonferroni significant conditional p-values (*COBLL1, TFAP2B, BDNF, MC4R;* $p_{c\ exact}\leq1.4x10^{-5}$) and at two loci nominally significant conditional p-values (*LYPLAL1* and *SLC39A8*; $p_{c\ exact}\leq1.6x10^{-3}$).

Collectively the lead SNPs representing multiple signals within 9 densely-genotyped regions varied dramatically in risk allele frequencies across the racial/ethnic groups (Appendix X) and explained an additional 0.025% of the variance (Table 9). This was more than a third (38%) of the variance explained by all of the 35 SNPs that remained significant in the joint model (28 lead SNPs from the densely-genotyped regions plus 7 additional index SNPs located outside of these regions).

v. Functional Annotation

Of the 39 trans-ethnic lead SNPs within 28 loci, two were annotated to be non-synonymous SNPs (*SLC39A8*, *GIPR*) and 19 were intronic SNPs (Appendix Y). Among the loci where we were able to fine-map the putative casual variant(s) there were several

interesting functional consequences. For example, the lead and index SNP 8.8kb 3' of *SEC16B* and 3.6kb 3' of RP4-798P15.2 was conserved across species, and from histone modification assessment was predicted to be an enhancer in muscle tissue. The lead SNPs 43kb 3' of *TMEM18* was predicted to change BCL and TR4 motifs, and within C10orf32-AS3MT was identified as an eQTL. The lead and index SNP within *TCF7L2* was found to be a promoter in pancreas; an enhancer in fat, muscle, and five other tissues; and changed several binding motifs. The non-synonymous lead SNP at the *QPCTL* locus was located within *GIPR,* which is conserved across species, was an enhancer, promoter, DNAse sensitive region in several tissues including fat, muscle, and pancreas, found to bind with and change the CTCF and several other binding motifs, as well as bind to CMYC.

Similarly, for the loci with multiple signals we noted varied functional consequences at several SNPs. For example, both lead SNPs upstream of the 3' of LYPLAL1 and 287kb 5' of RNU5F were predicted to be enhancers in fat and a number of other tissues, and modify motifs of a number of binding factors. Whereas the lead SNP for the primary signal at 1.7kb 3' of *COBLL1* alters the binding site for MAFK, the lead SNP for the secondary signal located intronic at *COBLL1.* Between AC068138.1 (>40kb 5') and *IRS1* (>400kb 3') lead SNPs for multiple signals were predicted to alter binding motifs and the the lead SNP for the secondary signal was also an enhancer in brain tissue. Whereas the lead and index SNP for the primary signal at *SLC39A8* was a non-synonymous mutation and conserved across species, the secondary signal was located 38kb 3' of *SLC39A8* and predicted to alter a number of binding motifs. At the primary signal lead SNP 1.6kb 5' of *TFAP2B* was predicted to modify both TATA and GAGA bind motifs, and the secondary signal was predicted to modify three other motifs. The primary and secondary signals intronic at *TRIM66* and *STK33* both were predicted to change HDAC2 sites, but the secondary site was conserved across species and was also an

enhancer in fat and skin. Both the lead SNPs of the multiple *BDNF-AS1* and *BDNF* signals were conserved across species, were predicted to be enhancer in brain and other tissues, and were DNAse sensitive regions; however, the primary signal was intronic to *BDNF* antisense RNA, which binds to GATA2 and YY1. All of the multiple signals were intronic at *FTO* andpredicted to be enhancers in at least muscle as well as either fat or brain and lead/index SNPs representing two of the four regions were DNAse sensitive in brain tissue. The primary and tertiary signals at *FTO* were conserved across species, but the tertiary signal was also predicted to be a promoter in fat and change a binding motif for FOXA1 and FOXA2 binding. Lastly, the lead and index SNP for the primary signal 209kb 3' of *MC4R* and 1.7kb 5' of U4, a small nucleor RNA, was conserved across species and DNAse sensitive in muscle, whereas the lead SNP for the secondary signal was located 44kb 5' of *MC4R* and was in high LD (r2>0.8 in 1000 Genomes AFR) with a highly conserved non-synonymous SNP (rs2229616) 44kb upstream within *MC4R*, which alters a GATA binding motif and has histone markers consistent with promoter and enhancer in brain.

5. Discussion

In this analysis we find that nearly a quarter of the previously described BMI index SNPs and >80% of the densely-genotyped BMI loci on the MetaboChip (29 of 36) met our definition for generalization in our trans-ethnic sample of 101,979 adults. The trans-ethnic meta-analyses, which are better powered than racial/ethnic specific analyses (Appendix O) for genetic loci that are shared across ancestral groups [288], demonstrate the similarity in the genetic underpinnings of obesity and transferability of common genetic loci to diverse populations.

The results show that while much of the genetic architecture underlying these adiposity-related traits is shared across ancestral groups, we also found evidence for

allelic heterogeneity and therefore unique functional variation at established loci in non-European descent populations. However, some of the BMI loci assessed in this study (7 of 36) were not significant in the trans-ethnic fixed-effect meta-analysis. Three of these replicated in European Americans only (*NEGR1, PRKD1, KCNJ2*). One locus (*SLC22A3*) generalized to African and American Indian/Alaskan Native descents, and two more were significant but were directionally inconsistent with the index report that was in low LD in European Americans from WHI with the top trans-ethnic SNP (*KCNJ11* and *TRAFD1,* r^2<0.01; Appendices K-N, P), suggesting that it may either be a spurious finding or represent a distinct haplotype at this locus. Overall these observations are consistent with the hypothesis that the majority of common genetic loci for complex traits like BMI will generalize to diverse populations given sufficient statistical power (a function of allele frequency, effect size and sample size, etc.) and that spurious findings and effect dilution can be avoided through the consideration of directional consistency as well as fine-mapping techniques [52].

The majority of the current literature on the genetic epidemiology of BMI comes from European descent populations [40]. Consistent with a recent study of BMI [33], we also observed evidence in support of multiple signals at *BDNF, MC4R and GPRC5B/GP2*. Yet due to the allelic heterogeneity of our diverse sample, we were able to describe novel independent signals at 7 BMI loci with at least two independent signals with varying risk allele frequencies across African, Hispanic/Latino, Asian and European ancestries of our study (*LYPLAL1, COBLL1, LOC646736, SLC39A8, TRAP2B, OVCH2,* and *FTO*). Additional sensitivity analysis of exact conditional analysis at a subset of these loci support the presence of multiple signals in our data. In previous work we noted a possible secondary signal in Hispanic/Latina women at *TRAP2B* [193], which was supported by this analysis. Interestingly, we noted the presence of four independent signals at *FTO* and three of the four *FTO* signals were within the physical bounds of the

putative interval of interest in African descent individuals of the PAGE Study [289]. Yet no secondary signals were observed at *FTO* in this previous study and others with less diverse samples [33, 34, 204].

Collectively the 11 secondary (or beyond) variants within 9 established BMI loci on the MetaboChip account for more than a third of the total variance explained by all of the SNPs in the joint model, which further illustrates the added value of diverse populations in mapping allelic heterogeneity. Eleven of the 20 independent primary and secondary (or beyond) SNPs at the 9 loci with multiple signals had a range of risk alleles >20% across the racial/ethnic populations included in our trans-ethnic meta-analysis (Appendix X). One such SNP at *FTO* (rs7206790) also exhibited significant evidence of heterogeneity across race/ethnicities (Table 9). However, future independent effect estimation and replication is needed to accurately describe the variance explained and the true genetic effects in similar diverse populations.

A strength of this work is that it addresses a knowledge gap on the genetic architecture of BMI [40] in populations with distinct burdens of obesity [2, 271] by expanding on previous fine-mapping efforts conducted by the PAGE Study [204], which generalized 8 of 21 BMI loci known at the time to African descent individuals. Since then the tally of BMI loci densely covered on the MetaboChip has grown to 36 and a recent large meta-analysis of >322,000 predominantly European descent samples illustrated the potential benefit for fine-mapping (at 26 of the known BMI loci at the time) [33]. Moreover, recently a handful of non-European BMI signals have been published in African descent and Asian GWAS and we were able to incorporate several of these non-European or trans-ethnic reports [25, 26, 30, 32, 195], while including almost 67,000 more Hispanic/Latino, Asian, European and American Indian/Alaskan Native descent adults than previous PAGE fine-mapping endeavors [204].

In the current study, we note the same lead SNP (e.g. *SEC16B, LOC646736, SLC39A8, FAIM2, TCF7L2, MC4R*) as previously reported in a much larger but less diverse [33] analysis using an approximate Bayesian fine-mapping approach [290]. Compared to these previous works, using a Bayesian trans-ethnic fine-mapping approach we are able to narrow the putative region of interest (in base pairs) equally well or better than these previous reports at 9 of 20 loci reported on previously that also exhibited strong evidence of association with BMI at the lead SNP in our study (*SEC16B, TMEM18, LOC646736, TFAP2B, NT5C2, TCF7L2, BDNF, MC4R, QPCTL*), and then determine if the assumption of one underlying signal held (e.g. *SEC16B, TMEM18, NT5C2, TCF7L2, QPCTL*) to interpret our 99% credible intervals as the probability of containing the underlying functional SNP at these narrowed loci.

In order to relax the assumption of fixed genetic effects in all of the racial/ethnic groups, we have also strengthened our analysis by performing a trans-ethnic Bayesian analysis, wherein we applied empirical estimates of the mean allele frequency differences to appropriately cluster the racial/ethnic groups and construct credible intervals, representing our confidence that the causal SNP lies within its bounds. We acknowledge that our findings and credible intervals are limited by the presence of multiple signals within a locus (e.g. at *TFAP2B, MC4R*). Although approximate conditional and joint analyses of fixed-effect estimates ruled out the presence of statistically significant secondary signals at 27 densely-genotyped loci, future work should focus on all aspects of genetic variation that lie within the physical bounds of the 99% credible interval and continue to test the assumption of no secondary signals at these loci or their impact on fine-mapping.

Fine-mapping resolutions depend on many factors, such as the extent of LD within the locus, allele frequencies and sample sizes of populations. Not surprisingly in this study the narrowing of the interval in trans-ethnic meta-analyses varied from one

locus to another (Table 8). Furthermore, the additional improvement in resolution offered by a Bayesian trans-ethnic meta-analysis related to the ancestral heterogeneity at a given locus, the extent to which the estimated allele frequency differences across populations captured this heterogeneity, the number of independent signals, and their allele frequencies.

Overall herein we find that nearly a quarter of the previously described BMI index SNPs and >80% of the densely-genotyped BMI loci generalize trans-ethnically. Thus study represents an important step towards prioritizing strong candidates for future epidemiologic study and targeted functional follow-up to identify causal variants for etiologic research and drug development. An improved understanding of the genetic architecture and the population diversity of complex traits such as BMI will in turn improve our ability to tailor both interventions to populations and to the individuals within them. Our systematic interrogation of the dense-genotyping at 36 BMI loci for their generalization to diverse populations further refines the regions of putative interest, and illustrates the importance of dense multi-ethnic genomic data in describing the allelic heterogeneity of diverse ancestral populations.

6. Main Tables and Figures

Table 4. Replication or generalization of 15 of the fine-mapped 36 BMI loci on the MetaboChip to 35,606 African descent adults.

Gene	SNP	rsID	Chr	Bp37	A1	A2	Freq	β (%)	SE (%)	P***	I^2	HetP	N	r^2 range**** in ARIC	r^2 range**** in WHI
SEC16B	*Index*	rs543874	1	177,889,480	g	a	0.249	1.37	0.17	6.0E-15	44.5	4.2E-02	35,604		
	Top	rs543874	1	177,889,480	g	a	0.249	1.37	0.17	6.0E-15	44.5	4.2E-02	35,604	0.32-1 (same)	0.96-1 (same)
TMEM18	*Index*	rs13021737	2	632,348	g	a	0.883	1.36	0.23	8.9E-09	26.8	1.7E-01	35,541		
	Top	rs10865549	2	631,759	a	g	0.883	1.52	0.24	6.4E-10	0	5.0E-01	33,352	0.42-1.00	1.00
BRE*,*	*Index*	rs116612809	2	28,301,171	g	a	0.097	1.39	0.25	6.4E-08	0	6.3E-01	35,583		
	Top	rs116612809	2	28,301,171	g	a	0.097	1.39	0.25	6.4E-08	0	6.3E-01	35,583	1 (same)	1 (same)
COBLL1*	*Index*	rs10184004	2	165,508,389	t	c	0.719	0.72	0.17	2.1E-05	32.2	1.2E-01	35,598		
	Top	rs10184004	2	165,508,389	t	c	0.719	0.72	0.17	2.1E-05	32.2	1.2E-01	35,598	1 (same)	1 (same)
ETV5	*Index*	rs1516725	3	185,824,004	c	t	0.817	0.64	0.20	1.2E-03	6.9	3.8E-01	35,485		
	Top	rs7647305	3	185,834,290	c	t	0.594	0.68	0.15	1.1E-05	0	5.3E-01	35,602	0.18	0.57
GNPDA2	*Index*	rs10938397	4	45,182,527	g	a	0.250	0.77	0.17	8.4E-06	51.9	1.5E-02	35,517		
	Top	rs181153926	4	45,165,656	t	c	0.249	0.87	0.18	1.6E-06	44.5	4.8E-02	32,146	0.22-0.98	-
FLJ35779	*Index*	rs2112347	5	75,015,242	t	g	0.495	0.09	0.15	5.5E-01	0	8.9E-01	35,604		
	Top	rs984976	5	74,910,870	a	g	0.150	0.88	0.22	5.4E-05	0	5.2E-01	35,595	0.09	0.36
TFAP2B	*Index*	rs2207139	6	50,845,490	g	a	0.096	0.79	0.26	2.0E-03	44.7	4.1E-02	35,605		
	Top	rs2744475	6	50,784,880	g	c	0.331	0.84	0.16	2.0E-07	7.5	3.7E-01	35,513	0.19	0.47
SLC22A3*	*Index*	rs3127574	6	160,791,370	c	g	0.587	0.03	0.15	8.3E-01	4	4.1E-01	35,597		
	Top	rs116859471	6	160,736,564	t	a	0.002	7.37	1.95	2.4E-04	53.3	1.8E-02	33,916	<0.01	<0.01
TCF7L2*	*Index*	rs7903146	10	114,758,349	c	t	0.706	0.66	0.17	6.1E-05	26.7	1.8E-01	35,604		
	Top	rs7903146	10	114,758,349	c	t	0.706	0.66	0.17	6.1E-05	26.7	1.8E-01	35,604	1 (same)	1 (same)
BDNF	*Index*	rs11030104	11	27,684,517	a	g	0.951	1.28	0.36	3.8E-04	14.6	3.0E-01	35,606		
	Top	rs7929344	11	27,743,495	a	g	0.245	0.78	0.18	1.1E-05	28.5	1.6E-01	35,586	0.02-0.05	<0.01
MAP2K5	*Index*	rs16951275	15	68,077,168	t	c	0.610	0.57	0.15	2.7E-04	0	5.7E-01	35,605		
	Top	rs3784718	15	68,098,004	c	t	0.630	0.61	0.16	1.2E-04	0	5.9E-01	34,268	0.52-0.93	0.53-0.99
SH2B1	*Index*	rs2650492	16	28,333,411	a	g	0.064	0.70	0.33	3.5E-02	0	5.3E-01	35,590		
	Top	rs8061590	16	28,895,130	g	a	0.312	0.69	0.16	2.5E-05	29.8	1.5E-01	35,592	0.82	1.00
FTO	*Index*	rs17817964	16	53,828,066	t	c	0.118	1.05	0.24	1.2E-05	33	1.2E-01	35,606		
	Top	rs62048402	16	53,803,223	a	g	0.114	1.19	0.24	1.1E-06	23	2.1E-01	35,603	0.91-0.98	0.94-1.00
MC4R	*Index*	rs6567160	18	57,829,135	c	t	0.189	1.08	0.19	2.8E-08	58.9	3.7E-03	35,599		
	Top	rs6567160	18	57,829,135	c	t	0.189	1.08	0.19	2.8E-08	58.9	3.7E-03	35,599	<0.01-1 (same)	<0.01-1 (same)

Abbreviations: A1=coded allele, A2=non-coded allele, β=Effect Size, Bp37=base pair Build 37, Chr=chromosome, Freq=coded allele frequency, HetP=heterogeneity p-value, P=p-value, SE=standard Error, SNPs=single nucleotide polymorphisms.

*Note: Starred genes represent fine-mapped loci, which were associated with BMI after the design of the MetaboChip in 2009.

**PAGE trans-ethnic discovery signal (Gong *et al.,* submitted to *Nature Communications*).

***For GWAS SNPs a Bonferroni correction for multiple tests reflected the number of independent previously-reported signals tested (=0.05/166; Appendices K-N, P). For all other SNPs in the fine-mapped BMI regions, we performed a Bonferroni correction for the number of independent SNPs per region (r^2<0.2 in ARIC African-Americans; Appendix I).

****The range of linkage disequilibrium captures any SNP within the fine-mapped loci (Appendices K-N, P) that represents the index BMI signal or secondary signal (described in European descent populations), or race/ethnic population specific marker. ARIC and WHI samples were used to represent the linkage disequilibrium for the PAGE African and European descent samples.

Table 5. Generalization of 13 of the fine-mapped 36 BMI loci on the MetaboChip to 26,048 Hispanic/Latino descent adults.

Gene	SNP	rsID	Chr	Bp37	A1	A2	Freq	β (%)	SE (%)	P***	I^2	HetP	N	r^2 range**** in HCHS/SOL	r^2 range**** in WHI
SEC16B	*Index*	rs543874	1	177,889,480	g	a	0.202	0.76	0.20	1.8E-04	0	4.4E-01	26,045		
	Top	rs543874	1	177,889,480	g	a	0.202	0.76	0.20	1.8E-04	0	4.4E-01	26,045	0.81-1 (same)	0.96-1 (same)
LYPLAL1 **	*Index*	rs2820436	1	219,640,680	a	c	0.439	0.63	0.17	1.6E-04	0	4.8E-01	26,046		
	Top	rs2820446	1	219,748,818	g	c	0.414	0.89	0.17	1.3E-07	50.8	5.8E-02	25,991	0.33	0.55
TMEM18	*Index*	rs13021737	2	632,348	g	a	0.867	1.14	0.24	3.6E-06	29.5	2.0E-01	26,016		
	Top	rs6744653	2	628,524	g	a	0.849	1.25	0.23	8.5E-08	43.9	9.8E-02	26,047	0.82-0.88	1.00
COBLL1*	*Index*	rs10184004	2	165,508,389	t	c	0.326	0.39	0.18	3.2E-02	44	9.7E-02	26,045		
	Top	rs12692738	2	165,558,252	c	t	0.252	0.77	0.20	1.1E-04	33	1.8E-01	26,045	0.67	0.45
IGF2BP2 **	*Index*	rs11927381	3	185,508,591	t	c	0.673	0.52	0.18	3.9E-03	23.6	2.5E-01	25,976		
	Top	rs6778126	3	185,405,781	g	a	0.515	0.63	0.17	1.5E-04	43.9	9.8E-02	26,043	0.14	0.05
GNPDA2	*Index*	rs10938397	4	45,182,527	g	a	0.372	0.70	0.17	4.7E-05	49.1	6.7E-02	26,020		
	Top	rs10938398	4	45,186,139	a	g	0.371	0.72	0.17	2.9E-05	44.5	9.4E-02	26,048	0.45-0.99	0.55-0.99
SLC39A8*	*Index*	rs13107325	4	103,188,709	t	c	0.046	1.03	0.39	9.5E-03	55.2	3.7E-02	26,048		
	Top	rs63519	4	103,202,914	a	c	0.142	0.85	0.24	3.4E-04	31.6	1.9E-01	26,048	0.29	0.26
TCF7L2*	*Index*	rs7903146	10	114,758,349	c	t	0.739	0.79	0.19	3.3E-05	63.8	1.1E-02	26,047		
	Top	rs7903146	10	114,758,349	c	t	0.739	0.79	0.19	3.3E-05	63.8	1.1E-02	26,047	1 (same)	1 (same)
KCNQ1*	*Index*	rs2237897	11	2,858,546	t	c	0.200	0.82	0.22	1.4E-04	0	9.1E-01	26,044		
	Top	rs60808706	11	2,857,233	a	g	0.217	0.90	0.21	1.6E-05	0	6.2E-01	26,045	0.83	0.60
MTCH2	*Index*	rs3817334	11	47,650,993	t	c	0.397	0.51	0.17	2.5E-03	0	6.1E-01	26,040		
	Top	rs11039448	11	47,918,416	t	g	0.653	0.88	0.17	4.0E-07	0	4.7E-01	26,048	0.25	0.47
MAP2K5	*Index*	rs16951275	15	68,077,168	t	c	0.531	0.37	0.17	3.4E-02	25.4	2.4E-01	26,046		
	Top	rs76616765	15	68,003,745	g	c	0.010	3.95	0.91	1.9E-05	0	6.4E-01	24,207	<0.01	<0.01
FTO	*Index*	rs17817964	16	53,828,066	t	c	0.253	1.37	0.19	2.1E-12	47.2	7.8E-02	26,046		
	Top	rs7187250	16	53,810,546	a	c	0.300	1.34	0.18	2.6E-13	47.8	7.4E-02	26,044	0.69-0.73	0.94-0.98
MC4R	*Index*	rs6567160	18	57,829,135	c	t	0.146	1.12	0.24	3.3E-06	53	4.7E-02	26,047		
	Top	rs72982988	18	57,802,714	a	g	0.151	1.22	0.24	2.8E-07	16.8	3.0E-01	26,048	<0.01-0.79	<0.01-0.75

Abbreviations: A1=coded allele, A2=non-coded allele, β=Effect Size, Bp37=base pair Build 37, Chr=chromosome, Freq=coded allele frequency, HetP=heterogeneity p-value, P=p-value, SE=standard Error, SNPs=single nucleotide polymorphisms.

**PAGE trans-ethnic discovery signal (Gong *et al.,* submitted to *Nature Communications*).

***For GWAS SNPs a Bonferroni correction for multiple tests reflected the number of independent previously-reported signals tested (=0.05/166; Appendices K-N, P). For all other SNPs in the fine-mapped BMI regions, we performed a Bonferroni correction for the number of independent SNPs per region (r^2<0.2 in ARIC African-Americans; Appendix I).

****The range of linkage disequilibrium captures any SNP within the fine-mapped loci (Appendices K-N, P) that represents the index BMI signal or secondary signal (described in European descent populations), or race/ethnic population specific marker. HCHS/SOL and WHI samples were used to represent the linkage disequilibrium for the PAGE Hispanic/Latino and European descent samples.

Table 6. Replication or generalization of 11 of the fine-mapped 36 BMI loci on the MetaboChip to 22,465 Asian descent adults.

Gene	SNP	rsID	Chr	Bp37	A1	A2	Freq	β (%)	SE (%)	P***	I^2	HetP	N	r^2 range**** in HCHS/SOL	r^2 range**** in WHI
GNPDA2	*Index*	rs10938397	4	45,182,527	g	a	0.279	0.55	0.15	2.6E-04	24	2.3E-01	22,386		
	Top	rs10938398	4	45,186,139	a	g	0.281	0.57	0.15	1.6E-04	20.5	2.6E-01	22,464	0.81-1 (same)	0.96-1 (same)
FLJ35779	*Index*	rs2112347	5	75,015,242	t	g	0.443	0.45	0.14	9.0E-04	10.3	3.5E-01	22,464		
	Top	rs56912706	5	75,037,086	a	g	0.517	0.57	0.13	2.2E-05	0.4	4.3E-01	22,464	0.33	0.55
CDKAL1*,*	*Index*	rs9356744	6	20,685,486	t	c	0.595	0.86	0.14	5.3E-10	42.6	8.4E-02	22,461		
	Top	rs9368222	6	20,686,996	c	a	0.597	0.88	0.14	2.0E-10	38.9	1.1E-01	22,393	0.82-0.88	1.00
TFAP2B	*Index*	rs2207139	6	50,845,490	g	a	0.210	0.29	0.17	8.5E-02	0	6.0E-01	22,464		
	Top	rs2076308	6	50,791,640	c	g	0.270	0.61	0.15	6.3E-05	0	8.0E-01	22,461	0.67	0.45
TCF7L2*	*Index*	rs7903146	10	114,758,349	c	t	0.934	1.49	0.32	4.6E-06	50.6	4.0E-02	22,465		
	Top	rs4506565	10	114,756,041	a	t	0.931	1.50	0.32	3.7E-06	49.5	4.5E-02	22,465	0.14	0.05
KCNQ1*	*Index*	rs2237897	11	2,858,546	t	c	0.353	0.73	0.18	3.5E-05	62.1	9.9E-03	14,181		
	Top	rs2299620	11	2,858,295	t	c	0.389	0.85	0.17	6.8E-07	42.8	9.3E-02	14,182	0.45-0.99	0.55-0.99
BDNF	*Index*	rs11030104	11	27,684,517	a	g	0.566	0.10	0.14	4.5E-01	61.9	7.2E-03	22,465		
	Top	rs11030100	11	27,677,586	g	t	0.570	0.53	0.14	1.1E-04	49.9	4.3E-02	22,465	0.29	0.26
MTCH2	*Index*	rs3817334	11	47,650,993	t	c	0.312	0.23	0.14	1.2E-01	0	8.7E-01	22,447		
	Top	rs76229852	11	47,258,369	g	a	0.958	1.48	0.34	2.0E-05	20.9	2.6E-01	22,465	1 (same)	1 (same)
FTO	*Index*	rs17817964	16	53,828,066	t	c	0.223	1.28	0.17	2.2E-14	0	4.9E-01	22,465		
	Top	rs3751812	16	53,818,460	t	g	0.185	1.56	0.17	5.5E-19	0	5.4E-01	22,463	0.83	0.60
MC4R	*Index*	rs6567160	18	57,829,135	c	t	0.197	0.67	0.17	9.0E-05	0	5.8E-01	22,461		
	Top	rs6567160	18	57,829,135	c	t	0.197	0.67	0.17	9.0E-05	0	5.8E-01	22,461	0.25	0.47
QPCTL*	*Index*	rs11671664	19	46,172,278	g	a	0.531	0.57	0.14	4.2E-05	0	8.4E-01	22,460		
	Top	rs11671664	19	46,172,278	g	a	0.531	0.57	0.14	4.2E-05	0	8.4E-01	22,460	<0.01	<0.01

Abbreviations: A1=coded allele, A2=non-coded allele, β=Effect Size, Bp37=base pair Build 37, Chr=chromosome, Freq=coded allele frequency, HetP=heterogeneity p-value, P=p-value, SE=standard Error, SNPs=single nucleotide polymorphisms.

**PAGE trans-ethnic discovery signal (Gong *et al.,* submitted to *Nature Communications*).

***For GWAS SNPs a Bonferroni correction for multiple tests reflected the number of independent previously-reported signals tested (=0.05/166; Appendices K-N, P). For all other SNPs in the fine-mapped BMI regions, we performed a Bonferroni correction for the number of independent SNPs per region (r^2<0.2 in ARIC African-Americans; Appendix I).

****The range of linkage disequilibrium captures any SNP within the fine-mapped loci (Appendices K-N, P) that represents the index BMI signal or secondary signal (described in European descent populations), or race/ethnic population specific marker. WHI samples were used to represent the linkage disequilibrium for the PAGE Asian and European descent samples.

Table 7. Trans-ethnic fixed-effect and meta-analysis of 36 BMI loci and Bayesian fine-mapping in up to 101,979 individuals.

		TOP FIXED-EFFECT														TOP MANTRA					
Gene		**rsID**	**Chr**	**Bp37**	**Ref. Risk Allele**	**A 1**	**A 2**	**Freq**	**β (%)**	**SE (%)**	**P*****	**I^2**	**HetP******	**N**	**r^2 range**** in TE Sample**	**rsID**	**Bp37**	**log 10 BF**	**Post prob. Het**	**N**	**r2 range**** in TE Sample**
NEGR1	*Ind.*	rs3101336	1	72,751,185	C	c	t	0.655	0.17	0.09	6.8E-02	69.4	2.0E-02	101,969							
	Top	rs1460939	1	72,861,567		t	a	0.868	0.47	0.13	4.3E-04	0.6	3.9E-01	101,976	0.18	rs1460939	72,861,567	2.3	0.022	101,976	0.18
TNNI3K	*Ind.*	rs12566985	1	75,002,193	G	g	a	0.719	0.42	0.11	2.8E-04	0	5.2E-01	75,627							
	Top	rs12566985	1	75,002,193		g	a	0.719	0.42	0.11	2.8E-04	0	5.2E-01	75,627	1 (s)	rs76514352	75,011,423	2.1	0.107	51,874	<0.01
SEC16B	*Ind.*	rs543874	1	177,889,480	G	g	a	0.213	0.90	0.10	3.5E-21	72.9	1.1E-02	101,972							
	Top	rs543874	1	177,889,480		g	a	0.213	0.90	0.10	3.5E-21	72.9	1.1E-02	101,972	0.62-1 (s)	rs543874	177,889,480	19.1	0.252	101,972	0.62-1 (s)
LYPLAL1**	Ind.	rs2820436	1	219,640,680	A	a	c	0.388	0.50	0.09	3.2E-08	0	4.7E-01	93,721							
	Top	rs2820436	1	219,640,680		a	c	0.388	0.50	0.09	3.2E-08	0	4.7E-01	93,721	1 (s)	rs2820436	219,640,680	6.0	0.006	93,721	1 (s)
TMEM18	*Ind.*	rs13021737	2	632,348	G	g	a	0.873	1.05	0.12	3.0E-18	42	1.6E-01	101,832							
	Top	rs6731872	2	624,205		g	t	0.877	1.09	0.12	8.3E-19	45.3	1.4E-01	101,832	0.61-0.92	rs6731872	624,205	16.6	0.037	101,832	0.61-0.92
BRE*	*Ind.*	rs116612809	2	28,301,171	G	g	a	0.088	1.05	0.23	8.8E-06	87.9	*2.6E-04*	68,016							
	Top	rs58154175	2	28,604,833		t	c	0.315	0.53	0.12	8.8E-06	12.6	3.3E-01	93,669	0.07	rs116612809	28,301,171	5.1	0.937	68,016	1 (s)
COBLL1*	*Ind.*	rs10184004	2	165,508,389	T	t	c	0.452	0.52	0.10	1.3E-07	0	5.5E-01	93,726							
	Top	rs10184004	2	165,508,389		t	c	0.452	0.52	0.10	1.3E-07	0	5.5E-01	93,726	1 (s)	rs10184004	165,508,389	5.6	0.011	93,726	1 (s)
LOC646736*	*Ind.*	rs2176040	2	227092802	A	a	g	0.2746	0.50	0.10	4.0E-07	0	5.5E-01	93,732							
	Top	rs2176040	2	227092802		a	g	0.2746	0.50	0.10	4.0E-07	0	5.5E-01	93,732	1 (s)	rs2176040	227,092,802	5.1	0.006	93,732	1 (s)
CADM2	*Ind.*	rs13078960	3	85,807,590	G	t	g	0.857	0.00	0.15	9.8E-01	0	8.2E-01	101,976							
	Top	rs115299727	3	85,843,586		g	t	0.009	2.91	0.79	2.6E-04	0	1.0E+00	51,162	<0.01	rs115299727	85,843,586	2.5	0.049	51,162	<0.01
IGF2BP2**	*Ind.*	rs11927381	3	185,508,591	T	t	c	0.563	0.49	0.09	1.3E-07	0	8.4E-01	93,626							
	Top	rs11927381	3	185,508,591		t	c	0.563	0.49	0.09	1.3E-07	0	8.4E-01	93,626	1 (s)	rs4481184	185,505,787	5.6	0.005	93,647	0.81
ETV5	Ind.	rs1516725	3	185,824,004	C	c	t	0.864	0.69	0.13	1.0E-07	0	8.7E-01	101,811							
	Top	rs7647305	3	185,834,290		c	t	0.720	0.59	0.10	3.7E-09	0	5.2E-01	101,974	0.34	rs7647305	185,834,290	7.0	0.017	101,974	0.34
GNPDA2	*Ind.*	rs10938397	4	45,182,527	G	g	a	0.325	0.60	0.08	7.3E-13	1.7	3.8E-01	101,782							
	Top	rs12507026	4	45,181,334		t	a	0.325	0.61	0.08	5.3E-13	6.7	3.6E-01	101,974	0.38-0.98	rs12507026	45,181,334	10.8	0.011	101,974	0.38-1.00
SLC39A8*	*Ind.*	rs13107325	4	103,188,709	T	t	c	0.053	1.05	0.25	3.4E-05	0	8.7E-01	79,090							
	Top	rs13107325	4	103,188,709		t	c	0.053	1.05	0.25	3.4E-05	0	8.7E-01	79,090	1 (s)	rs13107325	103,188,709	3.2	0.015	79,090	1 (s)
FLJ35779	*Ind.*	rs2112347	5	75,015,242	T	t	g	0.531	0.28	0.08	4.3E-04	3.3	3.8E-01	101,972							
	Top	rs60493905	5	75,038,426		c	t	0.630	0.48	0.09	1.9E-08	0	7.8E-01	101,968	0.17	rs60493905	75,038,426	6.4	0.008	101,968	0.17
CDKAL1*	*Ind.*	rs9356744	6	20,685,486	T	t	c	0.562	0.42	0.08	2.6E-07	81.2	1.2E-03	101,966							
	Top	rs67131976	6	20,686,878		c	t	0.729	0.60	0.10	4.0E-10	61.6	5.0E-02	101,973	0.24	rs67131976	20,686,878	8.1	0.092	101,973	0.24
TFAP2B	*Ind.*	rs2207139	6	50,845,490	G	g	a	0.211	0.41	0.10	1.0E-04	2	3.8E-01	101,973							
	Top	rs2744475	6	50,784,880		g	c	0.352	0.56	0.08	9.9E-12	39.1	1.8E-01	101,763	0.33	rs2744475	50,784,880	9.7	0.015	101,763	0.33
SLC22A3*	*Ind.*	rs3127574	6	160,791,370	C	c	g	0.497	0.13	0.08	1.2E-01	51.7	1.0E-01	93,727							
	Top	rs78739765	6	160,868,121		g	a	0.979	1.34	0.43	1.9E-03	0	6.4E-01	72,083	0.01	rs73589298	160,804,090	1.9	0.952	57,992	0.03
LRPN6C	*Ind.*	rs10968576	9	28,414,339	G	g	a	0.217	0.52	0.10	1.1E-07	0	5.9E-01	101,976							
	Top	rs17770336	9	28,414,625		t	c	0.223	0.52	0.10	9.6E-08	0	6.3E-01	101,930	0.95	rs17770336	28,414,625	5.8	0.013	101,930	0.95
NT5C2*	*Ind.*	rs11191560	10	104,869,038	C	c	t	0.204	0.54	0.11	1.8E-06	0	7.8E-01	101,966							
	Top	rs11191447	10	104,652,323		t	c	0.193	0.56	0.11	3.8E-07	0	8.4E-01	101,919	0.86	rs11191447	104,652,323	5.2	0.008	101,919	0.86
TCF7L2*	*Ind.*	rs7903146	10	114,758,349	C	c	t	0.739	0.75	0.10	2.2E-13	54	8.9E-02	101,975							
	Top	rs7903146	10	114,758,349		c	t	0.739	0.75	0.10	2.2E-13	54	8.9E-02	101,975	1 (s)	rs7903146	114,758,349	11.2	0.063	101,975	1 (s)
KCNQ1*	*Ind.*	rs2237897	11	2,858,546	T	t	c	0.237	0.66	0.12	3.3E-08	0	4.0E-01	93,516							
	Top	rs2237896	11	2,858,440		a	g	0.272	0.73	0.13	3.0E-08	0	4.6E-01	93,196	0.76	rs2237896	2,858,440	6.2	0.014	93,196	0.76
OVCH2	*Ind.*	rs4256980	11	8,673,939	G	g	c	0.509	0.22	0.08	5.2E-03	0	5.4E-01	101,492							
	Top	rs76876925	11	8,650,183		g	a	0.512	0.36	0.10	1.4E-04	0	8.7E-01	72,292	0.80	rs76876925	8,650,183	2.7	0.008	72,292	0.80
NCR3LG1/ KCNJ11*	*Ind.*	rs1557765	11	17,403,639	T	c	t	0.686	0.31	0.10	1.5E-03	0	5.0E-01	93,268							
	Top	rs7949405	11	17,085,192		a	c	0.568	0.32	0.09	4.8E-04	0	5.8E-01	86,446	0.08-	rs214933	17,194,584	2.2	0.005	93,270	0.20-

															0.09					0.21
BDNF	*Ind.*	rs11030104	11	27,684,517	A	a	g	0.694	0.40	0.10	7.3E-05	79.5	2.2E-03	101,516						
	Top	rs1519480	11	27,675,712		c	t	0.444	0.59	0.09	1.2E-11	44.3	1.5E-01	101,510	0.28-0.40	rs1519480	27,675,712	9.5	0.006	101,510	0.28-0.40
MTCH2	*Ind.*	rs3817334	11	47,650,993	T	t	c	0.338	0.28	0.08	6.3E-04	0	5.0E-01	101,940							
	Top	rs896817	11	47,394,305		c	t	0.713	0.46	0.09	4.3E-07	0	7.9E-01	101,965	0.03	rs896817	47,394,305	5.1	0.006	101,965	0.03
FAIM2	*Ind.*	rs7138803	12	50,247,468	A	a	g	0.285	0.35	0.09	9.0E-05	0	4.8E-01	101,969							
	Top	rs7138803	12	50,247,468		a	g	0.285	0.35	0.09	9.0E-05	0	4.8E-01	101,969	1 (s)	rs7138803	50,247,468	2.9	0.003	101,969	1 (s)
TRAFD1*	*Ind.*	rs11065987	12	112,072,424	A	a	g	0.716	0.21	0.12	8.8E-02	17.8	3.0E-01	93,730							
	Top	rs10774631	12	112,023,001		a	g	0.219	0.25	0.10	1.2E-02	46.6	1.3E-01	93,498	<0.01-0.06	rs10774631	112,023,001	0.9	0.024	93,498	<0.01-0.06
PRKD1	*Ind.*	rs11847697	14	30,515,112	T	t	c	0.258	0.01	0.14	9.3E-01	0	9.5E-01	93,490							
	Top	rs1957347	14	30,483,129		t	g	0.303	0.27	0.09	3.4E-03	0	5.4E-01	101,505	0.02	rs1957347	30,483,129	1.4	0.016	101,505	0.02
MAP2K5	*Ind.*	rs16951275	15	68,077,168	T	t	c	0.542	0.35	0.08	2.5E-05	27.3	2.5E-01	101,972							
	Top	rs4776970	15	68,080,886		a	t	0.422	0.38	0.08	6.1E-06	0	9.0E-01	101,972	0.56-1 (s)	rs4776970	68,080,886	4.1	0.004	101,972	0.56-1 (s)
GPRC5B/GP2	*Ind.*	rs12446632	16	19,935,389	G																
	Top	rs67501351	16	20,006,745		g	c	0.372	0.36	0.08	1.5E-05	28.8	2.4E-01	101,506	0.02	rs28461566	19,998,311	3.5	0.021	101,498	0.03
SH2B1	*Ind.*	rs2650492	16	28,333,411	A	a	g	0.146	0.54	0.14	9.5E-05	26.1	2.5E-01	99,770							
	Top	rs8061590	16	28,895,130		g	a	0.307	0.52	0.10	2.9E-07	0	3.7E-01	84,081	0.92	rs8061590	28,895,130	5.2	0.011	84,081	0.92
FTO	*Ind.*	rs17817964	16	53,828,066	T	t	c	0.256	1.23	0.10	7.1E-36	0	7.1E-01	101,976							
	Top	rs3751812	16	53,818,460		t	g	0.242	1.34	0.10	2.3E-42	15.6	3.1E-01	101,974	0.94-0.95	rs3751812	53,818,460	39.9	0.010	101,974	0.94-0.95
KCNJ2*	*Ind.*	rs312750	17	68,343,539	A	a	g	0.638	0.09	0.09	3.4E-01	0	8.9E-01	93,734							
	Top	rs72868947	17	68494065		t	a	0.0065	2.60	0.78	9.2E-04	52.5	1.2E-01	74,163	<0.01	rs72868947	68,494,065	1.7	0.105	74,163	<0.01
MC4R	*Ind.*	rs6567160	18	57,829,135	C	c	t	0.193	0.89	0.10	9.4E-19	13.8	3.2E-01	101,966							
	Top	rs6567160	18	57,829,135		c	t	0.193	0.89	0.10	9.4E-19	13.8	3.2E-01	101,966	0.01-1 (s)	rs6567160	57,829,135	16.2	0.011	101,966	0.01-1 (s)
KCTD15	*Ind.*	rs29941	19	34,309,532	G	g	a	0.562	0.26	0.09	3.9E-03	0	6.2E-01	101,951							
	Top	rs368794	19	34,320,452		a	t	0.537	0.32	0.08	1.4E-04	0	4.0E-01	99,796	0.49	rs368794	34,320,452	2.7	0.012	99,796	0.49
QPCTL*	*Ind.*	rs11671664	19	46,172,278	G	g	a	0.688	0.40	0.11	1.5E-04	79.9	1.9E-03	101,500							
	Top	rs1800437	19	46,181,392		g	c	0.817	0.64	0.11	1.6E-09	0	8.2E-01	101,488	0.09-0.88	rs11672660	46,180,184	7.0	0.007	101,501	0.09-0.88

Abbreviations: A1=coded allele, A2=non-coded allele, β=Effect Size, BF=Bayes Factor, Bp37=base pair Build 37, Chr=chromosome, Freq=coded allele frequency, HetP=heterogeneity p-value, Ind.=Index, MANTRA=Meta-ANalysis of Trans-Ethnic Association studies, P=p-value, s=same, SE=standard error, SNPs=single nucleotide polymorphisms, TE=Trans-ethnic.

*Note: Starred genes represent fine-mapped loci, which were associated with BMI after the design of the Metabochip in 2009.

**PAGE trans-ethnic discovery signal (Gong *et al.,* submitted to *Nature Communications*).

***For GWAS SNPs a Bonferroni correction for multiple tests reflected the number of independent previously-reported signals tested (=0.05/166; Appendices K-N, P). For all other SNPs in the fine-mapped BMI regions, we performed a Bonferroni correction for the number of independent SNPs per region (r^2<0.2 in ARIC African-Americans; Appendix I).

****The range of linkage disequilibrium captures any SNP within the fine-mapped loci (Appendices K-N, P) that represents the index BMI signal or secondary signal (described in European descent populations), or race/ethnic population specific marker. ARIC, HCHS/SOL, and WHI samples were used to represent the linkage disequilibrium for the PAGE trans-ethnic and European descent samples.

Table 8. Trans-ethnic meta-analyses to narrow the putative interval of interest at 36 BMI loci.

Gene	TOP FE SNPS			MANTRA 99% CREDIBLE INTERVAL			
	Range (bp)	Reduction	% Reduced	N SNPs	Range (bp)	Reduction	% Reduced
NEGR1	290,988	154,230	35	791	445,217	1	0
TNNI3K	101,051	16,107	14	224	117,158	0	0
SEC16B	16,575	166,174	91	1	0	182,749	100
LYPLAL1*	167,016	107,141	39	30	128,966	145,191	53
TMEM18	43,013	205,740	83	21	16,729	232,024	93
BRE*	353,982	930,200	72	26	879,946	404,236	31
COBLL1*	126,262	106,608	46	15	56,403	176,467	76
LOC646736*	54,967	128,106	70	32	87,419	95,654	52
CADM2	179,274	219,755	55	445	398,619	410	0
IGF2BP2*	204,453	53,206	21	13	38,588	219,071	85
ETV5	48,250	67,301	58	12	62,861	52,690	46
GNPDA2	22,002	66,280	75	5	10,448	77,834	88
SLC39A8*	78,859	17,861	18	141	94,766	1,954	2
FLJ35779	260,885	299,794	53	29	372,409	188,270	34
CDKAL1*	221,440	576,581	72	6	33,042	764,979	96
TFAP2B	272,303	293,963	52	1	0	566,266	100
SLC22A3*	238,014	32,672	12	720	270,605	81	0
LRPN6C	28,519	67,137	70	4	3,722	91,934	96
NT5C2*	418,829	362,996	46	22	261,330	520,495	67
TCF7L2*	52,861	23,298	31	2	4,261	71,898	94
KCNQ1*	331,607	167,414	34	7	18,885	480,136	96
OVCH2	190,703	122,255	39	368	312,450	508	0
NCR3LG1/KCNJ11*	235,612	149,042	39	295	383,901	753	0
BDNF	71,243	225,776	76	2	1,874	295,145	99
MTCH2	753,051	420,187	36	88	543,139	630,099	54
FAIM2	42,290	79,577	65	152	121,090	777	1
TRAFD1*	1,332,934	582,773	30	93	1,889,612	26,095	1
PRKD1	30,716	76,520	71	168	106,079	1,157	1
MAP2K5	408,882	156,440	28	122	493,712	71,610	13
GPRC5B/GP2	118,750	196,458	62	155	315,118	90	0
SH2B1	431,490	262,983	38	74	413,702	280,771	40
FTO	15,237	631,041	98	6	20,171	626,107	97
MC4R	73,513	293,976	80	1	0	367,489	100
KCNJ2	149,440	107,131	42	790	256,456	115	0
KCTD15	24,124	14,099	37	70	38,223	0	0
QPCTL*	197,022	73,189	27	3	21,988	248,223	92

Abbreviations: bp=base pairs, FE=Fixed-Effect, MANTRA=Meta-ANalysis of Trans-Ethnic Association studies, SNPs=single nucleotide polymorphisms.

*Note: Starred genes represent fine-mapped loci, which were associated with BMI after the design of the Metabochip in 2009.

**PAGE trans-ethnic discovery signal (Gong *et al.,* submitted to *Nature Communications*).

Table 9. Single variant and joint trans-ethnic fixed-effect estimates for the Bonferroni significant top signals at the 36 densely-genotyped BMI loci, after accounting for index SNPs (r^2<0.9 with each other, included in the trans-ethnic analyses) outside of these regions.

						SINGLE VARIANT MODEL							JOINT MODEL					
Gene	**rsID**	**Chr**	**Bp37**	**A 1**	**A 2**	**Freq**	**β (%)**	**SE (%)**	**P*****	**I^2**	**HetP******	**Actual N**	**Aprrox Freq**	**βj (%)**	**SEj (%)**	**Pj*****	**Effective N**	**% VarExp**
TNNI3K	rs12566985	1	75,002,193	g	a	0.719	0.42	0.11	2.8E-04	0	5.2E-01	75,627	0.697	0.42	0.11	1.4E-04	85,877	0.0007
SEC16B	rs543874	1	177,889,480	g	a	0.213	0.90	0.10	3.5E-21	72.9	1.1E-02	101,972	0.217	0.90	0.10	2.3E-19	125,180	0.0028
LYPLAL1**	rs2820436	1	219,640,680	a	c	0.388	0.50	0.09	3.2E-08	0	4.7E-01	93,721	0.387	0.50	0.09	3.5E-08	109,170	0.0012
LYPLAL1**	rs4445477	1	219,759,481	a	g	0.621	0.38	0.10	2.5E-04	0	6.3E-01	89,078	0.639	0.37	0.10	1.8E-04	89,156	0.0006
TMEM18	rs6731872	2	624,205	g	t	0.877	1.09	0.12	8.3E-19	45.3	1.4E-01	101,832	0.877	1.09	0.12	2.3E-19	135,002	0.0025
COBLL1*	rs10184004	2	165,508,389	t	c	0.452	0.52	0.10	1.3E-07	0	5.5E-01	93,726	0.444	0.53	0.11	4.0E-07	84,752	0.0014
COBLL1*	rs17244444	2	165,548,415	g	a	0.911	0.41	0.16	1.3E-02	58.6	6.4E-02	93,731	0.927	0.64	0.17	1.0E-04	101,259	0.0006
LOC646736*	rs2176040	2	227,092,802	a	g	0.275	0.50	0.10	4.0E-07	0	5.5E-01	93,732	0.259	0.75	0.11	5.2E-11	105,388	0.0021
LOC646736*	rs2673147	2	227,177,202	c	g	0.466	0.15	0.09	9.9E-02	0	8.6E-01	93,727	0.418	0.47	0.10	4.9E-06	104,177	0.0011
IGF2BP2**	rs11927381	3	185,508,591	t	c	0.563	0.49	0.09	1.3E-07	0	8.4E-01	93,626	0.523	0.38	0.09	4.6E-05	105,325	0.0007
ETV5	rs7647305	3	185,834,290	c	t	0.720	0.59	0.10	3.7E-09	0	5.2E-01	101,974	0.744	0.48	0.10	3.1E-06	104,120	0.0009
GNPDA2	rs12507026	4	45,181,334	t	a	0.325	0.61	0.08	5.3E-13	6.7	3.6E-01	101,974	0.317	0.61	0.08	2.5E-14	149,522	0.0016
SLC39A8*	rs28392891	4	103,134,678	a	t	0.891	0.52	0.15	4.8E-04	62.3	4.7E-02	95,585	0.913	0.51	0.15	6.9E-04	95,771	0.0004
SLC39A8*	rs13107325	4	103,188,709	t	c	0.053	1.05	0.25	3.4E-05	0	8.7E-01	79,090	0.030	1.03	0.25	4.2E-05	67,039	0.0006
FLJ35779	rs60493905	5	75,038,426	c	t	0.630	0.48	0.09	1.9E-08	0	7.8E-01	101,968	0.606	0.48	0.09	9.7E-08	111,182	0.0011
CDKAL1*	rs67131976	6	20,686,878	c	t	0.729	0.60	0.10	4.0E-10	61.6	5.0E-02	101,973	0.798	0.60	0.10	2.0E-09	106,176	0.0012
TFAP2B	rs2744475	6	50,784,880	g	c	0.352	0.56	0.08	9.9E-12	39.1	1.8E-01	101,763	0.349	0.54	0.08	2.1E-11	143,714	0.0013
TFAP2B	rs2397016	6	50,929,066	a	g	0.806	0.76	0.17	6.8E-06	39.6	1.9E-01	79,510	0.909	0.68	0.17	6.8E-05	46,410	0.0008
LRPN6C	rs17770336	9	28,414,625	t	c	0.223	0.52	0.10	9.6E-08	0	6.3E-01	101,930	0.217	0.52	0.10	2.0E-07	121,118	0.0009
NT5C2*	rs11191447	10	104,652,323	t	c	0.193	0.56	0.11	3.8E-07	0	8.4E-01	101,919	0.127	0.56	0.11	3.6E-07	111,260	0.0007
TCF7L2*	rs7903146	10	114,758,349	c	t	0.739	0.75	0.10	2.2E-13	54	8.9E-02	101,975	0.765	0.75	0.10	6.5E-14	108,781	0.0020
KCNQ1*	rs2237896	11	2,858,440	a	g	0.272	0.73	0.13	3.0E-08	0	4.6E-01	93,196	0.133	0.82	0.13	5.5E-10	62,758	0.0015
OVCH2	rs76633799	11	8,599,566	a	g	0.037	1.38	0.37	2.6E-04	0	7.1E-01	57,988	0.017	1.45	0.37	1.1E-04	42,922	0.0007
OVCH2	rs76876925	11	8,650,183	g	a	0.512	0.36	0.10	1.4E-04	0	8.7E-01	72,292	0.556	0.41	0.10	5.7E-05	84,029	0.0008
BDNF	rs1519480	11	27,675,712	c	t	0.444	0.59	0.09	1.2E-11	44.3	1.5E-01	101,510	0.513	0.64	0.09	1.6E-12	104,967	0.0020
BDNF	rs190666912	11	27,737,969	g	c	0.496	0.35	0.09	1.9E-04	16.9	3.0E-01	72,303	0.502	0.43	0.09	2.6E-06	103,685	0.0009
MTCH2	rs896817	11	47,394,305	c	t	0.713	0.46	0.09	4.3E-07	0	7.9E-01	101,965	0.735	0.46	0.09	3.2E-07	126,735	0.0008
FAIM2	rs7138803	12	50,247,468	a	g	0.285	0.35	0.09	9.0E-05	0	4.8E-01	101,969	0.254	0.35	0.09	1.0E-04	127,226	0.0005
MAP2K5	rs4776970	15	68,080,886	a	t	0.422	0.38	0.08	6.1E-06	0	9.0E-01	101,972	0.440	0.38	0.08	2.0E-06	134,469	0.0007
GPRC5B/ GP2	rs67501351	16	20,006,745	g	c	0.372	0.36	0.08	1.5E-05	28.8	2.4E-01	101,506	0.327	0.40	0.08	4.8E-07	140,434	0.0007
SH2B1	rs8061590	16	28,895,130	g	a	0.307	0.52	0.10	2.9E-07	0	3.7E-01	84,081	0.318	0.56	0.10	3.2E-08	98,598	0.0013
FTO	rs7206790	16	53,797,908	g	c	0.424	0.30	0.09	5.1E-04	94.7	*3.3E-12*	101,974	0.437	-1.14	0.14	5.6E-16	106,167	0.0064
FTO	rs73612011	16	53,809,861	t	c	0.858	0.40	0.14	3.1E-03	58.4	6.6E-02	101,978	0.879	-0.95	0.18	1.3E-07	87,717	0.0020
FTO	rs3751812	16	53,818,460	t	g	0.242	1.34	0.10	2.4E-42	15.6	3.1E-01	101,974	0.213	1.31	0.13	1.8E-24	114,260	0.0057
FTO	rs9936385	16	53,819,169	c	t	0.289	1.34	0.11	7.5E-37	6.9	3.4E-01	66,366	0.366	1.51	0.16	1.2E-20	84,239	0.0104
MC4R	rs6567160	18	57,829,135	c	t	0.193	0.89	0.10	9.4E-19	13.8	3.2E-01	101,966	0.184	0.89	0.10	7.7E-19	134,789	0.0024
MC4R	rs77901086	18	58,083,923	a	c	0.985	1.84	0.43	2.6E-05	0	4.5E-01	88,060	0.989	1.81	0.43	3.2E-05	77,866	0.0007
KCTD15	rs368794	19	34,320,452	a	t	0.537	0.32	0.08	1.4E-04	0	4.0E-01	99,796	0.581	0.32	0.08	6.3E-05	131,938	0.0005
QPCTL*	rs1800437	19	46,181,392	g	c	0.817	0.64	0.11	1.6E-09	0	8.2E-01	101,488	0.842	0.64	0.11	6.0E-09	116,182	0.0011

Abbreviations: Approx=approximate, A1=coded allele, A2=non-coded allele, β=Effect Size, βj=Joint effect Size, BF=Bayes Factor, Bp37=base pair Build 37, Chr=chromosome, Freq=coded allele frequency, HetP=heterogeneity p-value, MANTRA=Meta-ANalysis of Trans-Ethnic Association studies, P=p-value, Pj=joint p-value, SE=standard error, SEj=joint standard error, SNPs=single nucleotide polymorphisms, TE=Trans-ethnic, VarExp=Variance Explained.

*Note: Starred genes represent fine-mapped loci, which were associated with BMI after the design of the MetaboChip in 2009.

**PAGE trans-ethnic discovery signal (Gong *et al.,* submitted to *Nature Communications*).

***For GWAS SNPs a Bonferroni correction for multiple tests reflected the number of independent previously-reported signals tested (=0.05/166; Appendices K-N, P). For all other SNPs in the fine-mapped BMI regions, we performed a Bonferroni correction for the number of independent SNPs per region (r^2<0.2 in ARIC African-Americans; Appendix I).
****Locus-specific Bonferroni significant heterogeneity p-values shown in italics.
*****The range of linkage disequilibrium captures any SNP within the fine-mapped loci (Appendices K-N, P) that represents the index BMI signal or secondary signal (described in European descent populations), or race/ethnic population specific marker. ARIC, HCHS/SOL, and WHI samples were used to represent the linkage disequilibrium for the PAGE trans-ethnic and European descent samples.

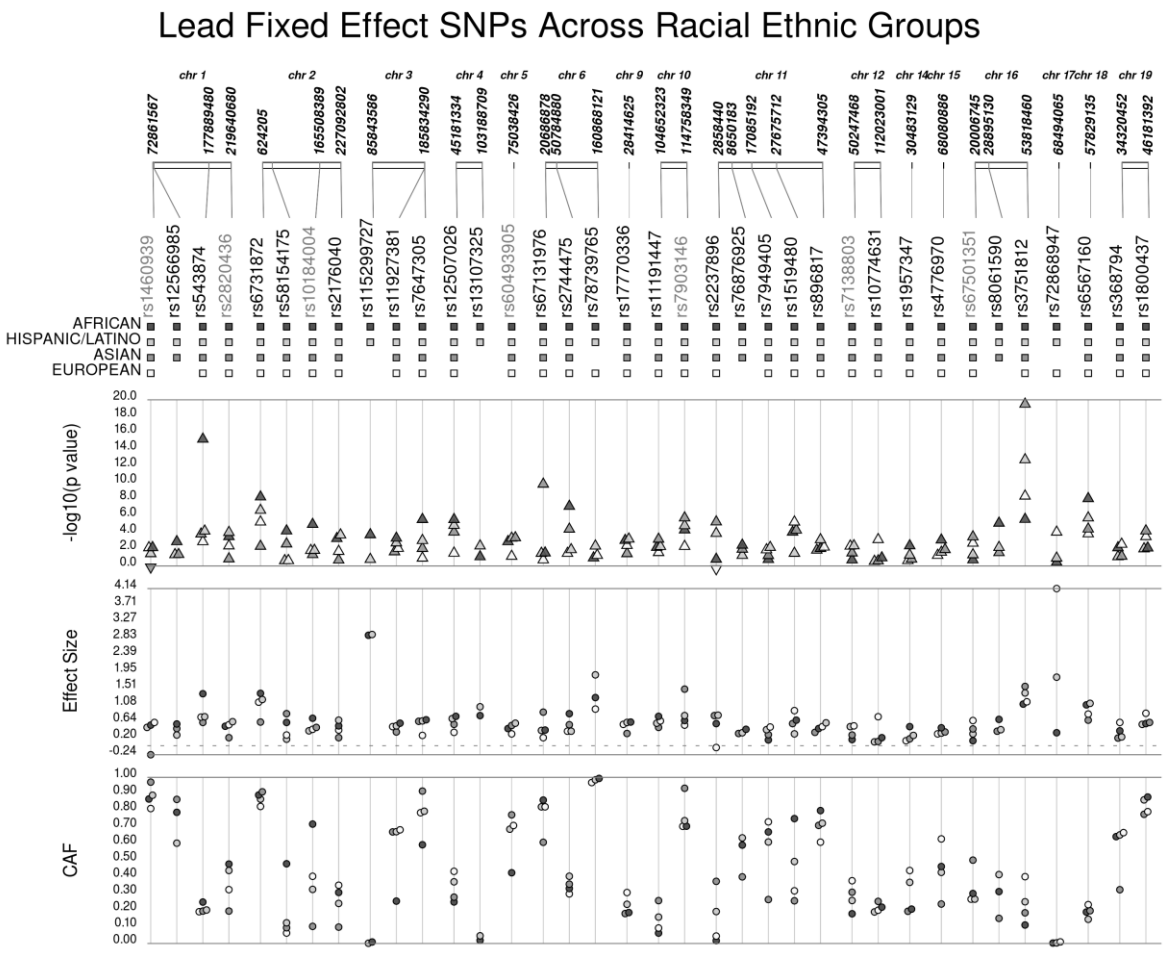


Figure 12. The comparison of the statistical significance (-log10 of the p-value), effect size (% change in BMI per risk allele) and coded allele frequencies (oriented to the risk allele in the trans-ethnic meta-analysis) across African, Hispanic/Latino, Asian and European ancestries for the lead SNPs (position noted for build 36) within the 36 densely-genotyped BMI regions on the MetaboChip with either locus-specific Bonferroni significant associations (rsid in black) or non-significant (rsid in gray).

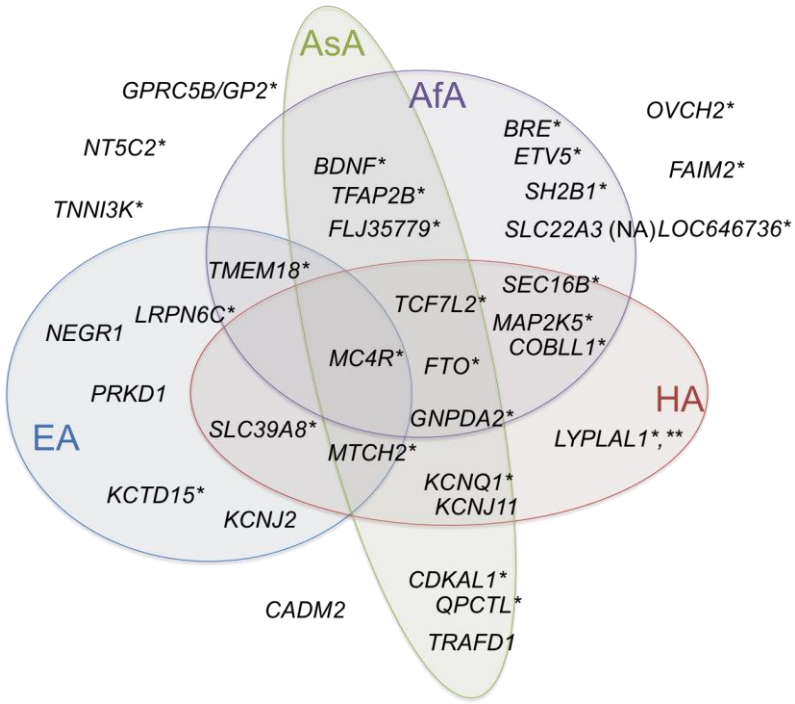


Figure 13. Venn diagram of overlap in significant lead SNP findings at each of 36 densely-genotyped BMI loci across the racial/ethnic populations [African (AfA), Hispanic/Latino (HA), Asian (AsA), European (EA), American Indian/Alaskan Native descent (NA, in parentheses)] and in the trans-ethnic fixed-effect meta-analysis of African, Hispanic/Latino, Asian and European descent adults (noted with asterisk).

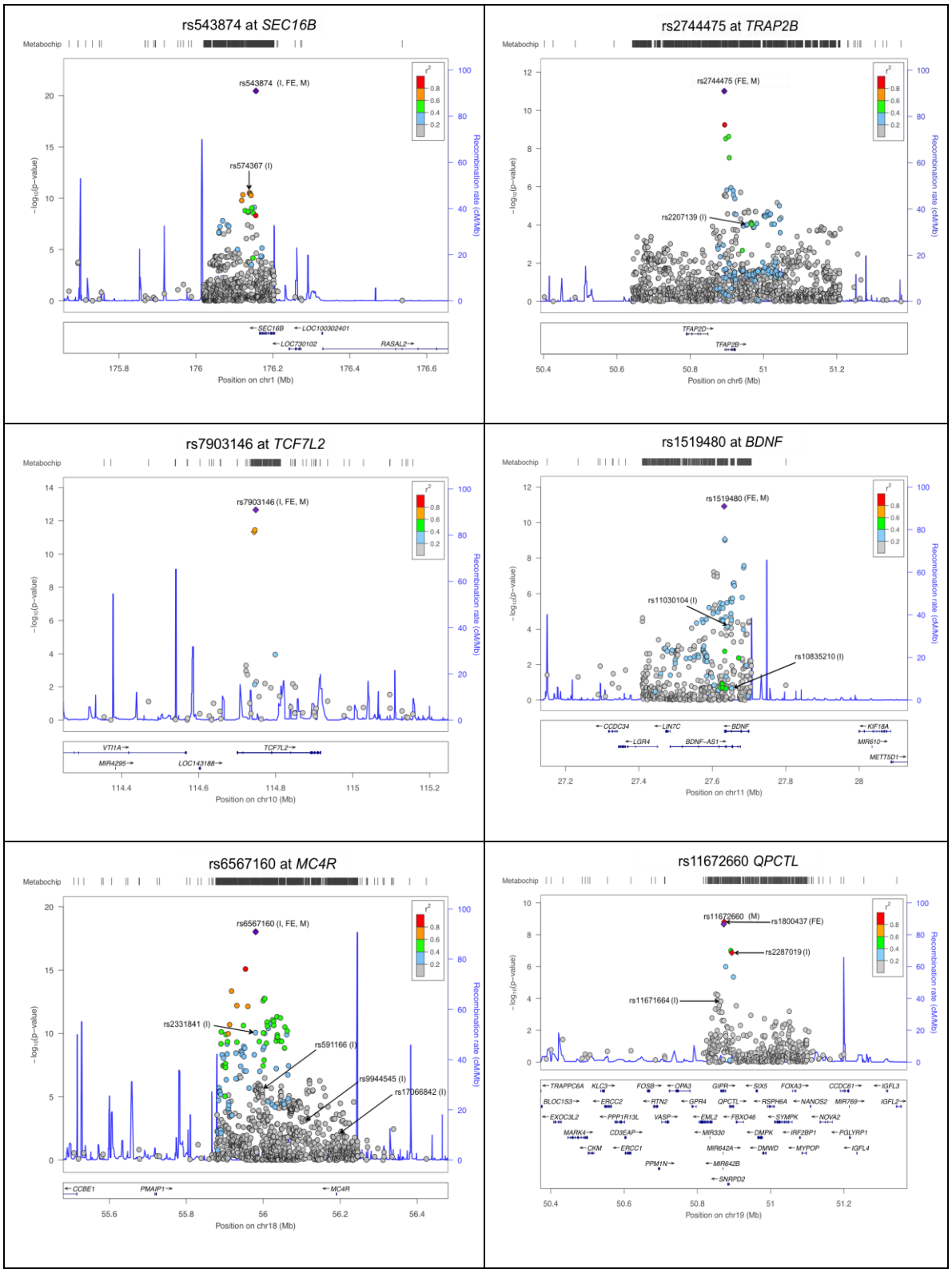


Figure 14. Regional plots of trans-ethnic fixed-effect estimates (I, index SNPs; FE, top finding) and Bayesian fine-mapping of 6 significant BMI loci to select the SNP with the highest posterior probability (M, shown in purple and reference for trans-ethnic linkage

disequilibrium) and narrow the putative interval of interest to <4 SNPs (SNPs in 99% credible interval shown in diamonds) in a sample of up to 101,979 individuals.

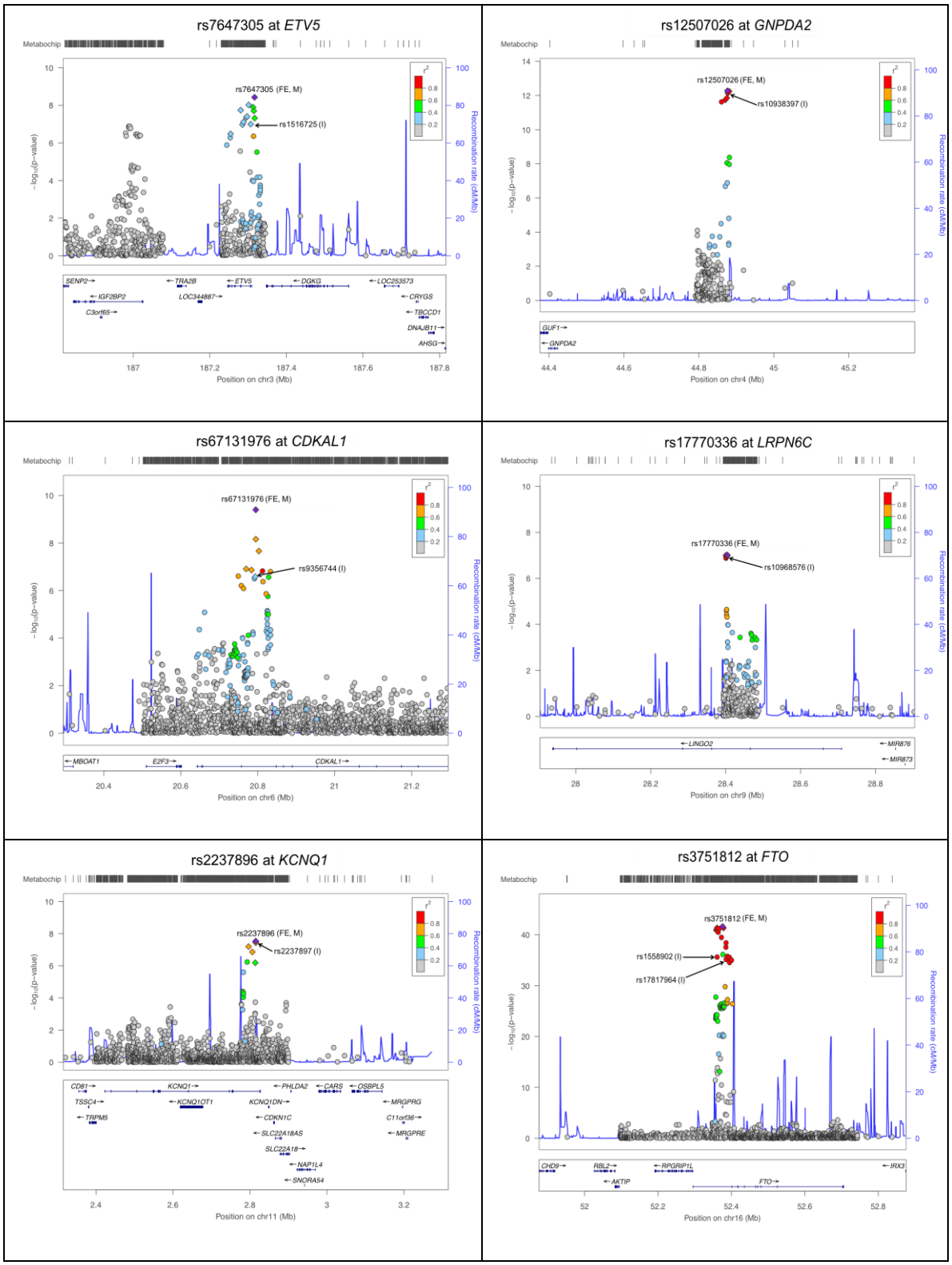


Figure 15. Regional plots of trans-ethnic fixed-effect estimates (I, index SNPs in black; FE, top finding) and Bayesian fine-mapping of 6 significant BMI loci to select the SNP with the highest posterior probability (M, shown in purple and reference for trans-ethnic

linkage disequilibrium) and narrow the putative interval of interest to 4-12 SNPs (SNPs in 99% credible interval shown in diamonds) in a sample of up to 101,979 individuals.

CHAPTER VI: CONCLUSIONS

A. Recapitulation of Aims

The disconnect between the current body of genetic epidemiology research [40] and the disparate burden of obesity and its downstream consequences in US racial/ethnic minorities, such as Hispanic/Latinos [44], has inspired me to organize my dissertation aims within a larger conceptual model (Figure 1). Specifically, as part this dissertation I performed secondary data analyses of the HCHS/SOL and the other PAGE collaborating studies to address what I determined were important gaps in the public health literature surrounding the application of alternative measures (Aim 1) and diverse populations (Aim 2) in genomics and epidemiologic research. This was done in order to inform our understanding of and future studies on the complex etiologic origins of obesity and its population-level disparities.

This conceptual model strengthened my dissertation by nesting the research aims within the broader study of genomics and epidemiology. As shown in Figure 1 my dissertation research provides two entrance points into a broader research agenda on the complex origins of racial and ethnic disparities. Specifically, Aim 1 represents an advancement of current knowledge on the accuracy of diverse measures of obesity, such as self-reported weight (Figure 1; top corner) in populations with otherwise unavailable weight data, e.g. US Hispanic/Latinos many of whom are transnational immigrants. Next, Aim 2 advanced our understanding of what are the particular genetic determinants of obesity in populations of diverse ancestries (Figure 1; right hand corner). Collectively these works provide a strong basis for the successful application of diverse

measures and populations in research on the complex mixture of genetic and non-genetic determinants of obesity, such as in the etiologic interaction between genetics and sociocultural environments (Figure 1; center set of arrows). Specifically, future gene-acculturation studies will be able to use both self-reported and measured weights and heights to calculate BMI (Aim 1), as well as decrease the burden of multiple testing by focusing on the just putative functional genetic risk factors for increased BMI (Aim 2).

B. Main Findings

In Aim 1 I established the accuracy of self-reported weight in HCHS/SOL, as a component of BMI, a useful metric for classifying obesity (Figure 1; upper corner). In Aim 1A I investigated the unadjusted difference between self-reported and measured weight at the HCHS/SOL baseline examination, and then in Aim 1B I extended this investigation to model the differences in a multivariate model including a number of potential predictors of self-reported weight inaccuracy.

Overall Findings: *I found that self-reported weight was an accurate proxy of measured weight at baseline across Hispanic/Latino backgrounds, as measured by both the correlation and the mean difference. I noted in a modified Bland-Altman plot that current self-reported weights were most imprecise at the extremes of measured weight in HCHS/SOL. Similar to previous studies I noted that the mean difference varied in the direction of effect across a number of key predictors of inaccuracy in a multiply imputed dataset, both in unadjusted and mutually-adjusted multivariate prediction models: age, gender, categories of BMI, nativity, study site by Hispanic/Latino background, self-report and end digit preference.*

In Aim 2 I utilized the fine-mapped MetaboChip data of >100,000 samples of the PAGE Study to generalize previously described SNPs and narrow the putative interval around the underlying functional SNPs at 36 established BMI loci. In Aim 2A I performed

the generalization and fine-mapping within each racial/ethnic group. Then in Aim 2B I extended this work through two trans-ethnic meta-analyses and approximate conditional and joint estimations of genetic effects.

Overall Findings: *I found that the majority of the risk alleles of the previous reported ('index') SNPs were directionally consistent with the risk alleles of previous reports, surpassing what we would expect by chance. I noted that a smaller proportion of the index SNPs generalized to a racial/ethnic group, which I defined as a SNP association being both directionally consistent and statistically significant. I observed that the additional genotypes available in the 36 fine-mapped BMI loci improved my ability to select the strongest associated SNPs ('top') and that in most cases this top SNP was not the index SNP from the discovery report. A handful of these loci had evidence of heterogeneity across racial/ethnic groups and therefore the Bayesian trans-ethnic meta-analysis was a useful way to relax the assumption of fixed-effects to bolster the evidence for the observed fixed-effect findings across multiple racial/ethnic groups. I found that the 99% credible set allowed me to fine-map the interval of putative interest for the subset of loci without independent signals. Lastly, conditional and joint analyses helped me describe allelic heterogeneity at several established BMI loci, 7 of which for the first time with BMI.*

1. Strengths

Given the ubiquitous use of proxy measures in obesity surveillance and epidemiology as well as the recent population growth of US Hispanic/Latinos, there is a clear need to reassess the data quality and accuracy of self-reported weight in this group due to the potential for differential migratory histories, or sociocultural, linguistic, etc. characteristics than in previous accuracy studies of US Hispanic/Latinos. Aim 1 represents the first study to my knowledge to describe the accuracy of self-reported

weight in a large population-based sample of multiple Hispanic/Latino backgrounds—each with unique cultural, linguistic, and migration histories. Previous accuracy studies in NHANES Hispanic/Latino samples have primarily been of Mexican heritage and US-born [49]. In contrast HCHS/SOL is more reflective of nativity status of the current adult US Hispanic/Latino population (i.e. the majority is foreign-born) [95]. Unlike most accuracy studies available in the current literature, I presented a transparent discussion of my data quality concerns and control, including reporting of the range of observed differences between self-reported and measured weight. Data quality control approaches based on a single criterion, although straightforward, could have compromised the representativeness of our analytic sample and artificially inflated estimates of accuracy. Although language preference did not appear to influence the accuracy of self-reported weight, self-reports made in kilograms were more accurate than those made in pounds. This indicates that even though multiple units may necessitate additional quality control measures, it may increase accuracy in future studies of US Hispanic/Latinos. As such the HCHS/SOL protocol for requesting self-reported weights in multiple languages or units and our approach to data quality control may be a useful model for future public health research design.

In Aim 2 of this dissertation we purposively utilized measures of BMI derived from both self-reported and measured weight, to obtain a maximum available sample of diverse populations (African, Hispanic/Latino, Asian, European and American Indian/Alaskan Native descent) with fine-mapping data from the MetaboChip as part of the PAGE Study. Given that the majority of the current literature on the genetic epidemiology of BMI comes from non-Hispanic/Latino European descent populations [40], this study was strengthened by its inclusion of ~67,000 additional diverse samples to expand on the previous fine-mapping efforts conducted by the PAGE Study [204]. Thus this study is the first to our knowledge to systematically generalize and fine-map

established BMI loci to diverse populations. Moreover, when the index SNP was not located in a fine-mapped BMI region (or was unavailable), we individually investigated this index SNP or its best proxy on the MetaboChip. Since the design of the MetaboChip, a handful of African- and Asian-descent GWAS [25, 26, 30, 195] or trans-ethnic reports [32] (Gong *et al*., submitted to *Nature Communications*) have been completed, increasing the tally of fine-mapped BMI loci to 36, which is nearly a third of all currently known BMI loci. We were able to interrogate these non-European-descent index SNPs as potential population-specific markers in our diverse study samples. Lastly, I relaxed the assumption of fixed genetic effects in all of the racial/ethnic groups by using a Bayesian analysis to appropriately cluster the racial/ethnic groups by their allele frequency differences and construct credible intervals that would reflect my confidence that the underlying functional variants were located within the intervals' bounds. Overall this study represents an important first step towards prioritizing strong candidates for future epidemiologic study and targeted functional follow-up, in the hope of identifying the exact causal variants of obesity risk.

2. Limitations

This body of work is not within limitations, however. With respect to the analyses of HCHS/SOL baseline data in Aim 1, there is an inherent confounding by geography. Even though we utilized a cross-classification of site and background, this obscured interpretability and still we could not fully decompose the complex effects of site and background. Furthermore, even if there were notable site differences across the sites of HCHS/SOL, these may have also been conflated by disparate interviewing practices despite the CC's efforts to coordinate study procedures across the sites. Lastly even though our sample of predominantly foreign-born Hispanic/Latino adults may more closely reflect the national profile of Hispanic/Latinos [95], than in reflected in previous

studies such as NHANES, our conclusions cannot generalized broadly due to its reliance on four urban US Hispanic/Latino communities (or immigrant enclaves). Yet the HCHS/SOL study is unique in its remarkable community engagement, which allows it to have stronger internal validity than previous studies like NHANES that may have poor interval validity (i.e. selection bias from immigrant non-participation) but less external validity concerns (i.e. nationally-based sampling design).

In Aim 2, we were also limited by the inability to independently contrast my results with previous fine-mapping work by the PAGE Study in African Americans (82% overlap with current analytic sample, which includes 6,455 additional individuals) [204]. It is reassuring, however, that I saw similar conclusions with respect to directional consistency, magnitude of effect, and statistical significance, even when top SNP changed in the expanded African descent and trans-ethnic meta-analyses. As well my ability to narrow the putative interval of around functional variant of interests at the 36 fine-mapped BMI loci may have violated a key assumption of the Bayesian trans-ethnic [53]. Missing SNP information on the MetaboChip, resulting from the strict SNP-level quality control measures throughout the work, may have further perturbed the Bayesian modeling. Although approximate conditional analyses of fixed-effect meta-analysis estimates ruled out the presence of secondary signals at most of the densely-genotyped loci, future work should focus on all aspects of genetic variation within the bounds of the credible interval and test the assumption of no secondary signals at these loci or their impact on fine-mapping initiatives.

C. Overall Conclusions

Future studies of US Hispanic/Latinos may increase their accuracy of self-reported weight by accommodating a participant's preference for both language and units of measurement. Even though this may an intuitive approach to studying diverse

individuals, it poses particular challenges to data quality. The data cleaning protocol and findings of Aim 1 provide support for the use of self-reported weight as a reasonably good proxy for measured weight, albeit with differential error, in diverse backgrounds of US Hispanic/Latinos. My findings indicate that there may be a distinct pattern of social desirability of weight among US Hispanic/Latinos, as compared to other racial/ethnic groups predominantly studied in the public health literature [151]. Awareness of this distinct patterning of self-report bias will be important to future etiologic obesity research in US Hispanic/Latinos.

Etiologic analyses using self-reported weights in HCHS/SOL or similar samples of Hispanic/Latinos should consider these sources of differential misreporting. For example, the bias in association results from measurement error in self-reported weight (exposure) may be corrected under certain assumptions by calibrating the regression coefficients [291]. Alternatively, in descriptive analysis of self-reported weight histories in HCHS/SOL these variables may be included as covariates in descriptive or predictive models.

The majority of the current literature on the genetic epidemiology of obesity relies on the study of BMI in non-Hispanic/Latino European-descent populations [40]. Consistent with recent reports from >300,000 European-descent samples [33], in Aim 2 we also observed evidence in support of multiple signals at *BDNF* and *MC4R*. Yet due to the allelic heterogeneity of our innovatively diverse sample, we were able to discover 7 additional BMI loci with multiples signals (*TNNI3K, LYPLAL1, COBLL1, SLC39A8, TRAP2B, OVCH2,* and *FTO*).

Targeted functional studies and drug development for obesity will benefit from our description of allelic heterogeneity in BMI. Ancestral diversity is needed to elucidate the complex biologic machinery that regulates human body mass and to determine how obesogenic environmental cues interact with these pathways. Moreover, the process by

which disparate environments become embodied to create racial/ethnic health disparities will remain unclear until we are able to document the underlying functional variants in diverse populations (i.e. minimize our multiple testing burden and increase our power to detect effects). Thus Aim 2 represents an important first step towards prioritizing strong candidates for future epidemiologic studies in GxE interactions. Overall an improved understanding of the genetic architecture and the population diversity of complex traits such as BMI, and its component parts based on either self-reported or measured anthropometrics, will in turn bolster our ability to tailor public health interventions to both diverse populations and the individuals within them.

D. Fulfillment of Doctoral Research Requirements

The Department of Epidemiology at the University of North Carolina at Chapel Hill requires that a dissertation be of appropriate scope and rigor to fulfill the goals of doctoral research. Although the committee is ultimately responsible for determining if I have met this goal through the research described herein, I can attest that since 2011 I have collaborated in the numerous methodological [201, 238] and substantive [193, 203, 204, 238, 277, 292] manuscripts that have supported the methods and analyses of this dissertation. Under the guidance of Kari E. North, I have led the concept, design, analysis, and writing of the two above scientific manuscripts.

This dissertation has benefited greatly from the input provided by the Committee Chair and its members, as well as peer-review from my coauthors and the HCHS/SOL Publications Committee. At each of the interim meetings, all members reached the consensus that the scope of the research was appropriate. Collectively the proposal defense, research preparation, and final defense fully address the specific goals of the Epidemiology Academic Policies Manual: originality, depth, scholarship, and writing skills. Thus I feel that this dissertation clearly demonstrates my ability as a public health

professional to execute epidemiologic research that integrates the methodological and substantive objectives outlined by the Department's core curriculum and requirements.

APPENDIX A: RESULTS OF STAGED DATA QUALITY CONTROL PROTOCOL ON 16,203 ADULT HISPANIC/LATINO PARTICIPANTS (18-76 YEARS) WITH BOTH SELF-REPORTED AND MEASURED WEIGHT AT THE BASELINE EXAMINATION (2008-2011) OF THE HISPANIC COMMUNITY HEALTH STUDY/STUDY OF LATINOS (HCHS/SOL).

Action	Measure(s)	Criteria	Number of Individuals (Number Affected)	Number of Remaining Individuals
Stage 1- Data Cleaning of Flagged SR Current Weights (≥15kg change)[a]				
Recoded[b]	Current SR weight	≥15kg drop in difference between SR and M current weight	16,203 (16 recoded[b])	16,203
Excluded	Current SR weight	>2 SD[c], ≥15kg difference between SR and M current weight	16,203 (41)	16,162
	Current SR weight	<34.5kg[d]	16,162 (1)	16,161
		>200.8kg[d]	16,161 (2)	16,159
Stage 2- Exclusions				
Excluded	Current SR, M weights	Current pregnancy[e]	16,159 (14)	16,145
	Current SR, M weights	Limb amputation[e]	16,145 (12)	16,133
	BMI for current SR, M weight	<16.0kg/m^2 [f]	16,133 (14)	16,119
		>70.0kg/m^2 [f]	-	-
Final Analytic Sample with both SR and M Current Weights				**16,119**

Abbreviations: BMI=Body mass index, M=Measured, SD=Standard deviation, SR=Self-reported

[a]At the beginning of the first stage, 229 current SR weights were flagged for being ≥15kg from the M weight at the same time point. After completing Stage 1, the number of flagged SR weights decreased to 183. After Stage 2, the number of flagged SR weights decreased to 181.

[b]The two possible scenarios of kg/lb SR were assessed (1- true SRs in kg were recorded as lb, 2- true SR in lb were recorded as kg) and the weight was recoded if one of the scenarios were favored according to the listed criteria.

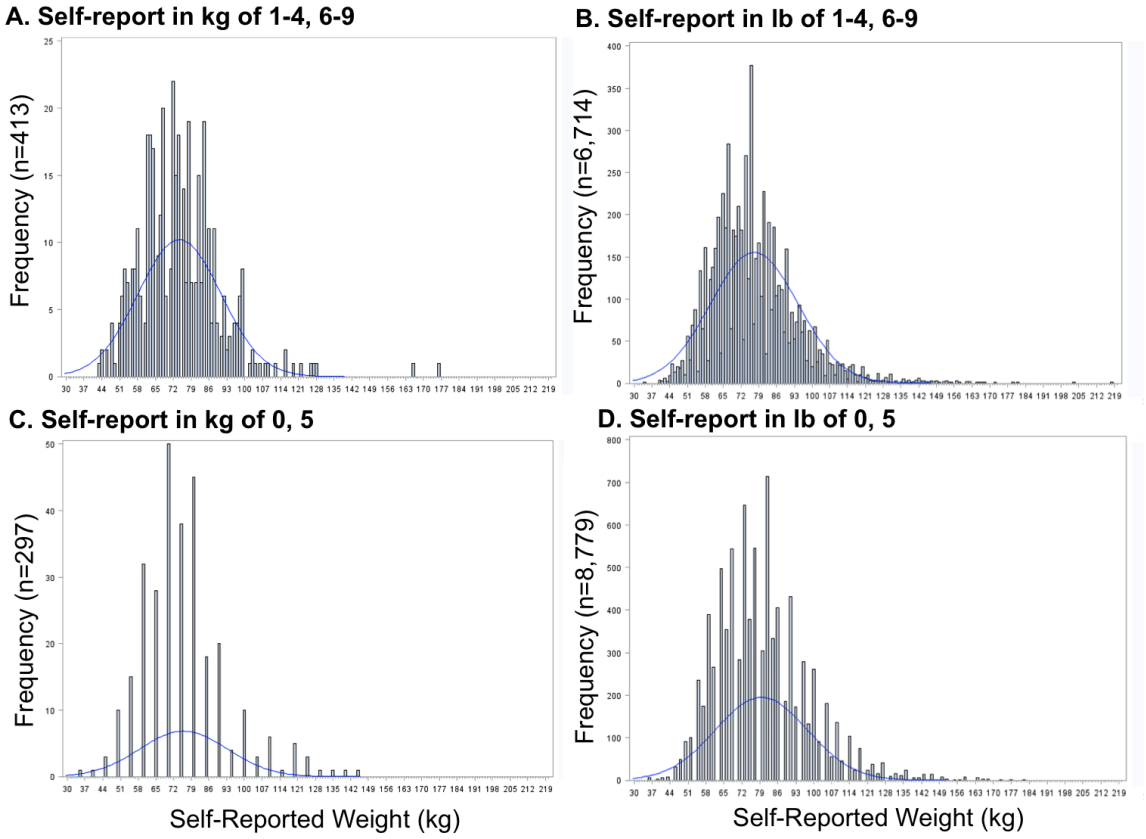
[c]Beyond 2 standard deviations from the gender and age-specific mean of any self-reported weight in HCHS/SOL (categories of age: 18-21, 22-29, 30-39, 40-49, 50-59, 60-69, 70-76 years).

[d]The criteria <34.5 and >200.8 were obtained from the range of M weights at the same time point. If social desirability were to differentially bias SR weights at the same away from extreme weights, we would expect that anything beyond the range of M weights might be a data error.

[e]Both current SR and M weights were excluded for women reporting to be currently pregnant (noted on Medical History Questionnaire Form) and for individuals with a limb amputation (noted on Ankle Arm Blood Pressure Procedure Form) who were otherwise able to stand on both feet (noted on Anthropometric Procedure Form) at the baseline examination.

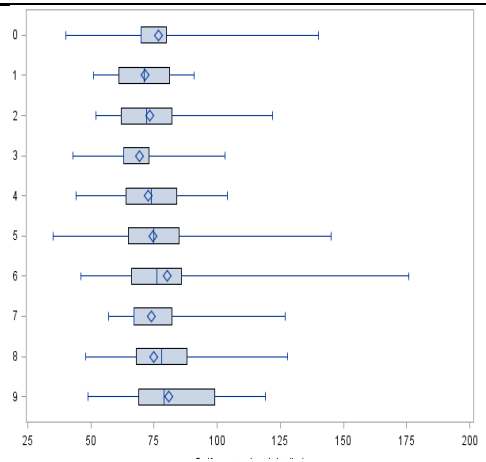
[f]BMI was calculated for all SR and M weights using an individual's M adult height at the baseline examination under the assumption that this would be static across adulthood.

APPENDIX B: HISTOGRAMS OF THE DISTRIBUTION OF SELF-REPORTED WEIGHTS STRATIFIED BY UNITS OF SELF-REPORT AND END DIGIT PREFERENCE FOR ZEROS AND FIVES (PANEL A: KG, 1-4 AND 6-9 END DIGITS, B: LB, 1-4 AND 6-9 END DIGITS, C: KG, 0 AND 5 END DIGITS, D: LB, 0 AND 5 END DIGITS) PRIOR TO DATA QUALITY CONTROL (N=16,203).

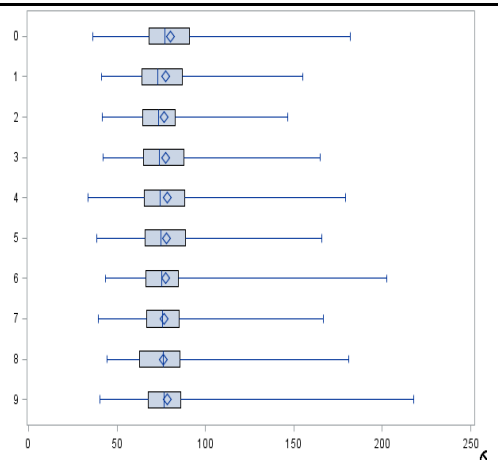


APPENDIX C: FREQUENCY PRIOR AND AFTER WEIGHTED FOR SAMPLING DESIGN AND DISTRIBUTION (BOX AND WHISKER PLOTS, WHERE THE BOX REPRESENTS THE INTERQUARTILE RANGE AND THE MEDIAN, THE DIAMOND REPRESENTS THE MEAN, THE WHISKERS REPRESENT THE MINIMUM AND MAXIMUM) OF SELF-REPORTED WEIGHTS (N=16,203) BY UNITS OF SELF-REPORT AND END DIGIT PREFERENCE.

Ending Digit	Expected Frequency (n=16,203)	Frequency (unweighted/weighted) with box and whisker plots of self-reported weights given in kg (n=710)	Frequency (unweighted/weighted) with box and whisker plots of self-reported weights given in lb (n=15,493)
0	0.10	0.25/0.24	0.37/0.38
1	0.10	0.04/0.04	0.02/0.02
2	0.10	0.10/0.12	0.07/0.07
3	0.10	0.07/0.07	0.05/0.05
4	0.10	0.09/0.09	0.06/0.06
5	0.10	0.16/0.17	0.20/0.20
6	0.10	0.08/0.08	0.06/0.05
7	0.10	0.05/0.05	0.05/0.05
8	0.10	0.10/0.10	0.08/0.08
9	0.10	0.05/0.04	0.03/0.03

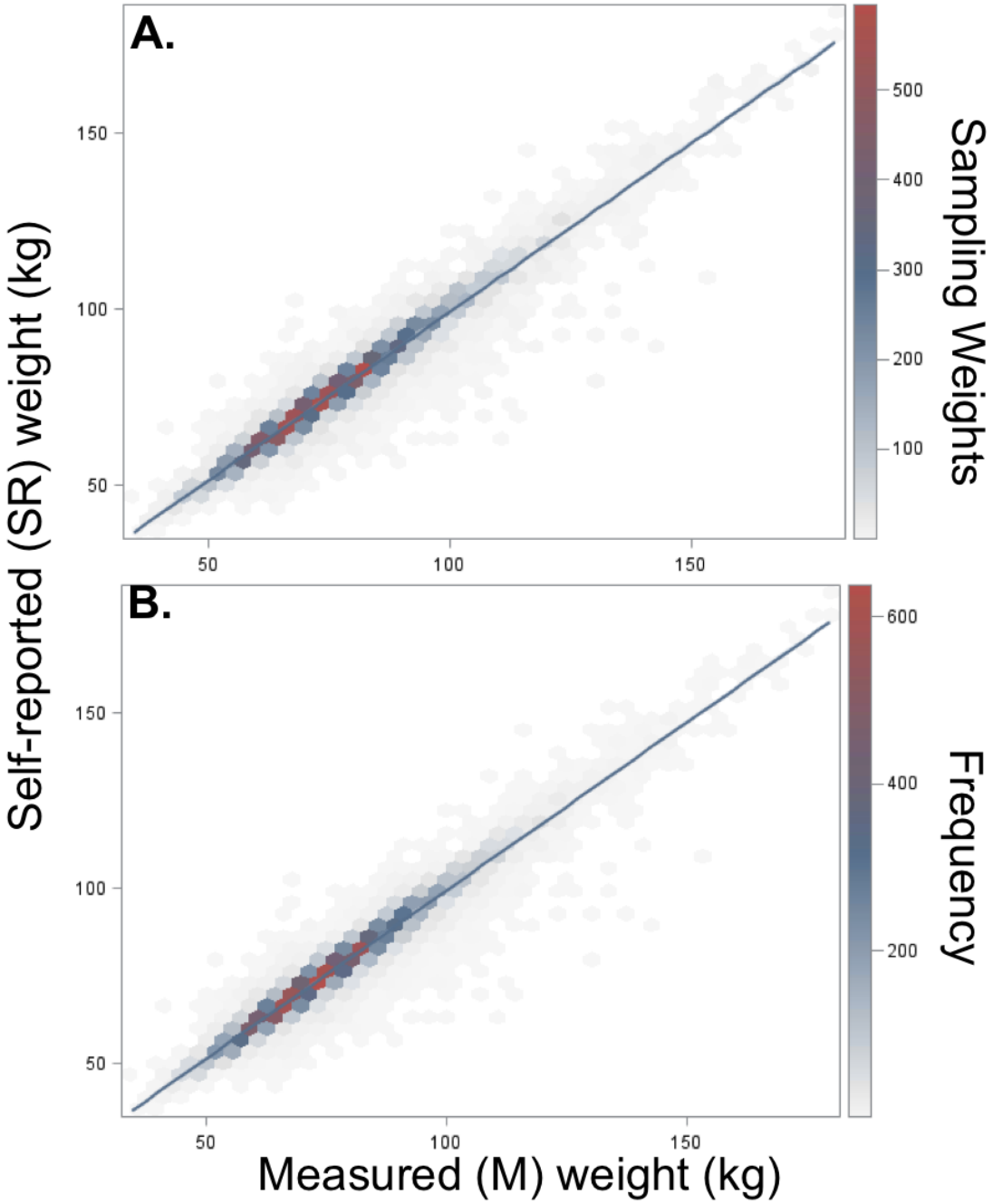


Self-Reported Weight (kg)



Self-Reported Weight (kg)

APPENDIX D: HEXAGONAL BINNING OF QUALITY CONTROLLED SELF-REPORTED WEIGHT AND MEASURED WEIGHTS AT BASELINE EXAMINATION (18-76 YEARS OF AGE) USING 60 BINS, A BEST FIT REGRESSION LINE, AND COLORED BY EITHER THE SAMPLING WEIGHT (PANEL A, ADJUSTED FOR COMPLEX SAMPLING DESIGN) OR THE OBSERVATION COUNT (PANEL B, UNWEIGHTED).



APPENDIX E: STRATIFIED MEAN DIFFERENCES BETWEEN SELF-REPORTED AND MEASURED WEIGHT DURING BASELINE EXAMINATION OF HISPANIC/LATINO ADULTS 18-76 YEARS OF AGE IN THE HISPANIC COMMUNITY HEALTH STUDY/STUDY OF LATINOS (HCHS/SOL) ACROSS STRATA OF KEY COVARIATES BEFORE AND AFTER ASSIGNING 1,366 INDIVIDUALS (8.5% OF SAMPLE) WITH MISSING COVARIATES TO A SPECIFIC STRATUM IN MULTIPLE IMPUTATIONS (25 STACKED DATASETS OF N=16,119).

		Before Multiple Imputation (n=15139 to 16119)			After Multiple Imputation, If Missing Covariates (n=16119)		
		N	Mean Diff. (kg)*	95% CI (kg)	Range of N	Mean Diff. (kg)*	95% CI (kg)
Age at Examination (years)	18-22	1236	-0.04	-0.40, 0.32			
	23-29	1385	-0.01	-0.34, 0.32			
	30-44	3954	0.12	-0.07, 0.31			
	45-59	6796	0.41	0.26, 0.56			
	60-76	2748	0.63	0.45, 0.81			
Back-ground by Site	Dominicans- Bronx	1362	-0.64	-0.96, -0.31	1367-1374	-0.64	-0.96, -0.32
	Central American- Bronx	217	-0.97	-1.51, -0.43	217-222	-0.97	-1.51, -0.43
	Central American- Chicago	416	0.38	-0.05, 0.82	416-419	0.38	-0.05, 0.82
	Central American- Miami	1017	0.36	0.02, 0.71	1019-1022	0.37	0.02, 0.71
	Cubans- Miami	2229	0.52	0.36, 0.68	2233-2235	0.52	0.36, 0.68
	Mexicans- Bronx	203	-0.21	-1.12, 0.71	205-207	-0.08	-0.99, 0.83
	Mexicans- Chicago	2342	0.33	0.12, 0.53	2346-2351	0.33	0.12, 0.53
	Mexican- San Diego	3771	0.66	0.40, 0.92	3790-3795	0.66	0.40, 0.93
	Puerto Ricans- Bronx	1793	-0.15	-0.48, 0.18	1797-1805	-0.15	-0.48, 0.18
	Puerto Ricans- Chicago	764	-0.14	-0.69, 0.42	764-767	-0.13	-0.68, 0.42
	South American- Bronx	186	-0.58	-1.14, -0.01	186-191	-0.58	-1.14, -0.01
	South American- Chicago	366	0.20	-0.35, -0.75	366-369	0.20	-0.34, -0.75
	South American- Miami	461	0.44	-0.05, 0.92	461-463	0.44	-0.04, 0.92
	Other- Bronx	241	-0.84	-1.69, 0.01	241-244	-0.84	-1.69, 0.01
	Other- Chicago	150	-0.49	-1.51, 0.53	150-151	-0.49	-1.50, 0.53
	Other- Miami	290	0.44	-0.23, 1.11	290-293	0.44	-0.23, 1.11
	Other- San Diego	244	0.84	0.18, 1.50	244-249	0.84	0.19, 1.50
	Missing	67			0		
Body Mass Categories	Underweight (<18.5 kg/m^2)	117	2.76	1.94, 3.58	117-118	2.76	1.94, 3.57
	Normal Weight (18.5-24.9 kg/m^2)	3153	1.09	0.94, 1.24	3153-3158	1.09	0.94, 1.24
	Overweight (25.0-29.9 kg/m^2)	6058	0.52	0.39, 0.64	6059-6063	0.52	0.39, 0.64
	Obese (≥30.0 kg/m^2)	6782	-0.59	-0.79, -0.39	6783-6788	-0.59	-0.79, -0.39

	Missing	9			0		
Cancer History	Yes	633	0.26	-0.44, 0.95	634-640	0.26	-0.43, 0.95
	No	15414	0.22	0.11, 0.33	15479-15485	0.23	0.12, 0.33
	Missing	72			0		
Diabetic Status**	Normal Glucose Regulation	6752	0.30	0.15, 0.44	6753-6759	0.30	0.15, 0.44
	Impaired Glucose Tolerance	6224	0.05	-0.12, 0.22	6226-6234	0.05	-0.12, 0.22
	Diabetes	3130	0.42	0.18, 0.66	3130-3136	0.42	0.18, 0.66
	Missing	13			0		
Education	Less than high school or a GED	6078	0.25	0.05, 0.45	6099-6119	0.26	0.06, 0.46
	At most high school or a GED	4102	0.10	-0.10, 0.29	4114-4127	0.11	-0.09, 0.30
	More than high school or a GED	5868	0.29	0.14, 0.44	5882-5895	0.29	0.14, 0.44
	Missing	71			0		
End Digit Preference	5, 10s	9029	-0.13	-0.29, 0.04			
	1-4s, 6-9s	7090	0.72	0.61, 0.82			
Gender	Female, pre-, peri-menopausal***	5784	0.03	-0.14, 0.20			
	Female, post-menopausal	3874	0.24	0.06, 0.43			
	Male	6461	0.38	0.21, 0.54			
Heart Failure History	Yes	297	0.30	-0.35, 0.95	297-305	0.31	-0.34, 0.96
	No	15743	0.22	0.11, 0.33	15814-15822	0.23	0.12, 0.34
	N Missing	79			0		
Health Insurance	Yes	8035	0.21	0.05, 0.38	8210-8233	0.20	0.03, 0.36
	No	7793	0.26	0.12, 0.40	7886-7909	0.26	0.12, 0.40
	N Missing	291			0		
Language Preference	English	3242	0.04	-0.21, 0.28			
	Spanish	12877	0.29	0.18, 0.41			
Nativity	Born in the United States****	2819	-0.09	-0.38, 0.21	2821-2830	-0.09	-0.38, 0.21
	Foreign Born	13247	0.31	0.21, 0.42	13289-13298	0.32	0.22, 0.43
	Missing	53			0		
Physical Activity Level***	Inactive	3623	0.22	0.02, 0.41	3640-3659	0.23	0.03, 0.42
	Low Activity	2152	0.16	-0.06, 0.37	2161-2172	0.17	-0.04, 0.39
	Medium Activity	1776	0.47	0.19, 0.74	1780-1794	0.47	0.20, 0.75
	High Activity	8450	0.19	0.04, 0.34	8507-8525	0.20	0.04, 0.35
	Missing	118			0		
Smoking Status	Never	9759	0.11	-0.02, 0.25	9794-9812	0.12	-0.02, 0.25
	Former	3174	0.48	0.25, 0.71	3186-3195	0.49	0.26, 0.72
	Current	3112	0.33	0.13, 0.54	3120-3131	0.34	0.13, 0.54
	Missing	74			0		

Socioeconomic Status	Less than $30,000 USD	10315	0.23	0.09, 0.36	11031-11087	0.23	0.10, 0.36
	$30,000 or more USD	4824	0.24	0.07, 0.41	5032-5088	0.23	0.05, 0.40
	Missing	980			0		
Unit of Self-Report	Kg	704	-0.03	-0.59, 0.52			
	Lb	15415	0.24	0.13, 0.35			

Abbreviations: CI=Confidence interval, Diff.=Difference, GED=General Education Development Equivalent of a High School Diploma, ref=Referent, SE=Standard error, USD=United States Dollars.
*Mean difference=self-reported minus measured weight (kg).
**As defined by the American Diabetes Association [269].
***Women reporting not reporting ‘yes’ to having reached menopause (change of life) were assumed to be pre- or peri-menopausal.
****As defined as being born in one of the 50 United States, not including United States Territories such as Puerto Rico.
*****As defined in the 2008 Physical Activity Guidelines for adults [270].

APPENDIX F: REGRESSION BETA COEFFICIENTS OF PREDICTORS OF ACCURACY OF SELF-REPORTED WEIGHT AS COMPARED TO MEASURED WEIGHT (KG) IN THE HISPANIC/LATINO ADULTS 18-76 YEARS OF AGE IN HCHS/SOL (2008-2011) INCLUDED IN A COMPLETE CASE ANALYSIS (N=14,753).

		β_{SR-M}*	SE	95% CI (kg)
Age at Examination (years)	18-22	0 (ref)		
	23-29	0.22	0.28	-0.32, 0.76
	30-44	0.50	0.23	0.05, 0.95
	45-59	0.73	0.22	0.29, 1.17
	60-76	0.76	0.27	0.23, 1.28
Background by Site	Dominicans- Bronx	0 (ref)		
	Central American- Bronx	-0.29	0.33	-0.94, 0.36
	Central American- Chicago	1.07	0.28	0.53, 1.61
	Central American- Miami	1.34	0.26	0.84, 1.85
	Cubans- Miami	1.22	0.20	0.82, 1.61
	Mexicans- Bronx	0.77	0.50	-0.21, 1.76
	Mexicans- Chicago	1.27	0.22	0.84, 1.70
	Mexican- San Diego	1.62	0.23	1.16, 2.07
	Puerto Ricans- Bronx	0.57	0.29	0.01, 1.14
	Puerto Ricans- Chicago	0.76	0.34	0.08, 1.43
	South American- Bronx	0.17	0.37	-0.56, 0.90
	South American- Chicago	0.99	0.30	0.41, 1.57
	South American- Miami	1.11	0.30	0.52, 1.70
	Other- Bronx	0.38	0.36	-0.33, 1.09
	Other- Chicago	0.65	0.51	-0.35, 1.66
	Other- Miami	1.18	0.41	0.38, 1.98
	Other- San Diego	1.70	0.42	0.88, 2.52
Body Mass Categories	Underweight (<18.5 kg/m^2)	1.81	0.49	0.85, 2.76
	Normal Weight (18.5-24.9 kg/m^2)	0 (ref)		
	Overweight (25.0-29.9 kg/m^2)	-0.77	0.10	-0.97, -0.57
	Obese (≥30.0 kg/m^2)	-1.84	0.14	-2.10, -1.57
Cancer History	Yes	-0.20	0.36	-0.91, 0.52
	No	0 (ref)		
Diabetic Status**	Normal Glucose Regulation	0 (ref)		
	Impaired Glucose Tolerance	-0.16	0.12	-0.40, 0.07
	Diabetes	0.24	0.16	-0.08, 0.56
Education	Less than high school or a GED	0.04	0.13	-0.22, 0.31
	At most high school or a GED	-0.08	0.12	-0.33, 0.16
	More than high school or a GED	0 (ref)		
End Digit Preference	5, 10s	-0.84	0.10	-1.03, -0.65
	1-4s, 6-9s	0 (ref)		
Gender	Female, pre-, peri-menopausal***	-0.36	0.13	-0.62, -0.10
	Female, post-menopausal	-0.38	0.14	-0.65, -0.11
	Male	0 (ref)		
Heart Failure History	Yes	0.33	0.32	-0.31, 0.96
	No	0 (ref)		
Health Insurance	Yes	0.13	0.13	-0.11, 0.38
	No	0 (ref)		
Language Preference	English	0.29	0.19	-0.09, 0.67
	Spanish	0 (ref)		
Nativity	Born in the United States****	-0.32	0.21	-0.73, 0.08
	Foreign Born	0 (ref)		

Physical Activity Level***	Inactive	-0.03	0.13	-0.28, 0.23
	Low Activity	0.02	0.13	-0.24, 0.29
	Medium Activity	0.38	0.16	0.07, 0.68
	High Activity	0 (ref)		
Smoking Status	Never	0 (ref)		
	Former	0.23	0.14	-0.05, 0.52
	Current	0.19	0.14	-0.08, 0.45
Socioeconomic Status	Less than $30,000 USD	0.14	0.10	-0.06, 0.34
	$30,000 or more USD	0 (ref)		
Unit of Self-Report	Kg	-0.99	0.27	-1.52, -0.46
	Lb	0 (ref)		

Abbreviations: CI=Confidence interval, GED=General Education Development Equivalent of a High School Diploma, M=Measured weight, ref=Referent, SE=Standard error, SR=Self-reported weight, USD=United States Dollars.

*Difference=self-reported minus measured weight (kg). Multivariate difference was calculated from a multivariate linear regression model of mean difference on the above possible determinants of validity (independent variables).

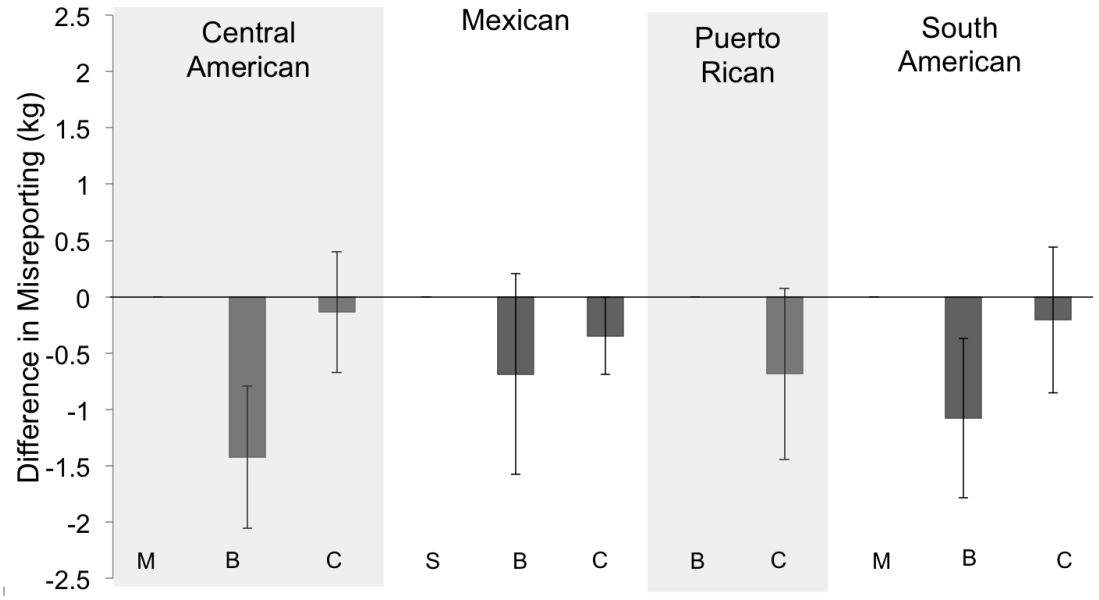
**As defined by the American Diabetes Association [269].

***Women reporting not reporting ‘yes’ to having reached menopause (change of life) were assumed to be pre- or peri-menopausal.

****As defined as being born in one of the 50 United States, not including United States Territories such as Puerto Rico.

*****As defined in the 2008 Physical Activity Guidelines for adults [270].

APPENDIX G: MULTIVARIATE ESTIMATED DIFFERENCES IN MISREPORTING (DEFINED AS THE DIFFERENCE BETWEEN SELF-REPORTED AND MEASURED WEIGHTS) AND 95% CONFIDENCE INTERVALS COMPARING THE STUDY SITES (B=THE BRONX, NY; C=CHICAGO, IL; M=MIAMI, FL; S=SAN DIEGO, CA; THE LARGEST SITE FOR A GIVEN BACKGROUND WAS USED AS REFERENT) WITHIN THE FOUR HISPANIC/LATINO BACKGROUNDS SAMPLED AT MORE THAN ONE SITE (≥100 PARTICIPANTS OF A GIVEN BACKGROUND PER SITE).



APPENDIX H: SUPPLEMENTAL DESCRIPTION OF STUDIES PARTICIPATING IN MANUSCRIPT 2.

The Population Architecture using Genomics and Epidemiology (PAGE) study is funded by the National Human Genome Research Institute to examine the epidemiologic architecture of common genetic variants that have been reproducibly associated with human diseases and traits (https://www.pagestudy.org) [41]. The PAGE study consists of a coordinating center and four sub-consortia, representing one or more US racial/ethnic groups, and includes the Atherosclerosis Risk in Communities Study (ARIC), the Vanderbilt University Medical Center's DNA biobank (EAGLE BioVU), the Coronary Artery Risk Disease in Young Adults study (CARDIA), the Cardiovascular Health Study (CHS), the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), the Multiethnic Cohort (MEC), Mount Sinai Biobank Program (BioME), and the Women's Health Initiative (WHI). Additionally, for this analysis the PAGE study also reached out to additional studies, such as the GenNet Network, the Hypertension Genetic Epidemiology Network (HyperGen), Slim Initiative in Genomic Medicine for the Americas Type 2 Diabetes Consortium (SIGMA), and the Taiwan-MetaboChip Study for Cardiovascular Disease (TaiChi) Study to expand the sample size.

The Atherosclerosis Risk in Communities Study (ARIC)

ARIC is a prospective population-based study of four U.S. communities [293]. It was designed to investigate the causes of atherosclerosis and its clinical outcomes, as well as the key components (e.g. race, gender, geographic location, time period) of variation in CVD burden, and health care utilization. The larger ARIC study includes two separate parts: The Cohort Component and the Community Surveillance Component. The Cohort Component started in 1987 when the ARIC field centers randomly selected, recruited approximately 4,000 individuals aged 45-64 years from each center, and began regular telephone follow-up of the cohort for health status updates. Weight and height were measured as part of the

ARIC cohort examinations. The Community Surveillance Component monitors the ARIC communities to determine the long-term trends in hospitalized myocardial infarction and coronary heart disease (CHD) deaths in approximately 470,000 community-dwelling men and women aged 35-84 years, but these participants were not included in the current study. All ARIC cohort study participants provided written informed consent. Only the consenting African American subjects of the ARIC cohort were genotyped on the MetaboChip.

Epidemiologic Architecture for Genes Linked to Environment study accessing BioVU (EAGLE BioVU)

BioVU is Vanderbilt University's biorepository of DNA extracted from discarded blood that was collected during routine clinical testing and then linked to de-identified health records available in the Synthetic Derivative, which contains highly detailed longitudinal clinical data for approximately one million patients, and is updated regularly to include new patients and append new data [255, 279]. Planning for BioVU began in mid-2004 under the goal of providing a resources to investigators for studies of genotype-phenotype associations and the first BioVU samples were collected in February 2007 at an accrual rate of ~500-700 samples per week. BioVU uses an "opt out" model, which was informed by an opinion from the federal Office of Human Research Protection (OHRP) that discarded biologic samples could be used and linked to de-identified clinical data for biomedical research without having to obtain prospective consenting of each individual [255, 279]. The Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study accessed all non-European descent patients as of 2011 [282]. The Vanderbilt University Center for Human Genetics Research (CHGR) DNA Resources Core genotyped these samples along with 360 HapMap samples on the MetaboChip [294]. Body mass index was calculated from the median height (centimeters) and weight (kilograms), by year and then by patient [282]. The median age of clinical visits per patient was included as a covariate. Race/ethnicity was administratively in BioVU assigned as previously described [273, 274].

The Coronary Artery Risk Development in Young Adults Study (CARDIA)

CARDIA began in 1985-1986 with a group of 5115 black and white men and women aged 18-30 years to examine the determinants of CVD and its risk factors [295]. The CARDIA participants were selected to equally represent a number of subgroups of race, gender, education (high school or less and more than high school) and age (18-24 and 25-30) across four centers: Birmingham, AL; Chicago, IL; Minneapolis, MN; and Oakland, CA. Participants were invited to participate in follow-up examinations during 1987-1988 (Year 2), 1990-1991 (Year 5), 1992-1993 (Year 7), 1995-1996 (Year 10), 2000-2001 (Year 15), and 2005-2006 (Year 20), yielding retention rates from 72-90% across all follow-ups. The CARDIA examinations have collected medical and family histories, several CVD risk factors and anthropometrics, including weight, height, and skinfold fat. The participants in the CARDIA cohort were born between 1955-1968 and provide a unique avenue to investigate the mechanisms linking obesity to derangements in CVD in individuals earlier in the life course. All CARDIA study participants provided written informed consent.

The Cardiovascular Health Study (CHS)

The CHS is an observational study of risk factors for cardiovascular disease in adults 65 years or older, which began in 1989 and continued through 1999 [296]. As part of extensive clinical examinations measurements of several CVD risk factors were taken, including traditional risk factors such as blood pressure and lipids as well as measures of subclinical disease, including echocardiography of the heart, carotid ultrasound, and cranial magnetic-resonance imaging (MRI). At six month intervals between clinic visits, and once clinic visits ended, participants were followed-up by phone (ongoing) to ascertain their hospitalizations and health status, including several outcomes: coronary heart disease, angina, heart failure (HF), stroke, transient ischemic attack (TIA), claudication, and mortality.

The GenNet Network (GenNet)

Between 1995-2003 GenNet recruited African-American (n=1101) and European-American participants (n=1497) at two field centers. Non-Hispanic/Latino European American subjects were recruited from Tecumseh, Michigan, and African-American subjects were recruited from Maywood, Illinois. First, individuals aged 18–50 years with blood pressures in the upper 20th to 25th percentile of the age/gender-specific blood pressure distribution were identified, and then second, an attempt was made to enroll all siblings and parents of the proband, irrespective of their blood pressure or hypertension treatment status [297]. All study participants provided their written informed consent. All hypertensive African-American individuals were genotyped on the MetaboChip and included in the analysis.

The Hispanic Community Health Study / Study of Latinos (HCHS/SOL)

The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) is a population-based study of four urban Hispanic/Latino communities that was designed to identify CVD risk factors playing a protective or harmful role in Hispanics/Latinos, including acculturation [43]. The target population of HCHS/SOL included 16,000 adults (18-74 years at screening) of Hispanic/Latino origin, specifically of Cuban, Puerto Rican, Mexican, and Central/South American heritage, who were living at one of four field centers affiliated with San Diego State University, Northwestern University in Chicago, Albert Einstein College of Medicine in the Bronx area of New York, and the University of Miami. Seven additional academic centers serve as scientific and logistical support centers, including the HCHS/SOL CC at the University of North Carolina at Chapel Hill. The HCHS/SOL participants underwent an extensive baseline clinic exam between 2008-2011, follow-up examination (ongoing), and annual follow-up interviews are ongoing to determine health outcomes of interest.

Hypertension Genetic Epidemiology Network (HyperGEN)

HyperGEN is part of the Family Blood Pressure Program designed to study the genetics underpinnings of hypertension and other related conditions. Participants were recruited from multiply-affected hypertensive sib-ships, which were ascertained through population-based cohorts or from the community-at-large. The study was later extended to include siblings and offspring of the original sibling pair. Hypertensive individuals were identified as those developing hypertension before age 60 and the presence of at least one additional hypertensive sibling who was willing to participate. Participants with type 1 diabetes or advanced renal disease (defined as serum creatinine level >2 mg/dL) were excluded. By 2003 two of four centers (AL, NC) recruited 1,264 African Americans, while three centers (NC, MN, and UT) recruited European Americans [298]. All study participants provided written informed consent, and all African American participants were genotyped on the MetaboChip for this analysis.

The Multiethnic Cohort Study of Diet and Cancer (MEC)

The MEC was established in 1993 to examine lifestyle risk factors and genetic susceptibility for cancer and CVD in five racial/ethnic groups at the University of Hawai‘i Cancer Center, in Honolulu, HI, and the Keck School of Medicine, University of Southern California (USC) in Los Angeles, CA [283, 299]. The MEC cohort is comprised of more than 215,000 men and women primarily of African American, Japanese, Latino, Native Hawaiian and European ancestry. Every cohort member completed a self-administered 26-page baseline questionnaire at entry to the MEC Study (1993-1996), which included an extensive diet history, demographics, medical, medication, physical activity and female reproductive histories. Incident cancer cases are identified through cancer registries that have been established by state statute in Hawai‘i and California. In addition to the baseline questionnaire, two additional questionnaires were mailed to MEC participants including a 4-page questionnaire that was sent in 1999-2001 and another 26-page questionnaire that was

sent in 2003-2008. Biological specimens were collected from selected members of the cohort, starting in 1996, but more concertedly from 2001-2006. Subjects were selected for MetaboChip genotyping based on their availability of biomarker for CVD risk factors or as described previously, Type 2 diabetes cases and controls were genotyped as part of the Slim Initiative in Genomic Medicine for the Americas Type 2 Diabetes Consortium (SIGMA) [275].

Mount Sinai Biobank Program (BioME)

The BioMe Biobank is an ongoing, prospective, hospital- and outpatient- based population research program operated by The Charles Bronfman Institute for Personalized Medicine (IPM) at Mount Sinai and has enrolled over 33,000 participants since September 2007 [278]. BioMe is an Electronic Medical Record (EMR)-linked biobank that integrates research data and clinical care information for consented patients at The Mount Sinai Medical Center, which serves diverse local communities of upper Manhattan with broad health disparities. BioMe populations include 25% of African Ancestry, 36% of Hispanic/Latino ancestry, 30% of European Ancestry, and 9% of other ancestry. The BioMe disease burden is reflective of health disparities in the local communities. BioMe operations are fully integrated in clinical care processes, including direct recruitment from clinical sites waiting areas and phlebotomy stations by dedicated recruiters independent of clinical care providers, prior to or following a clinician standard of care visit. Recruitment currently occurs at a broad spectrum of over 30 clinical care sites.

The Taiwan-MetaboChip Study for Cardiovascular Disease (TaiChi) study

The TaiChi study was formed through a collaborative effort between investigators based in the US and Taiwan, with the goal of identifying the genetic determinants of atherosclerosis and diabetes related traits in East Asians [276]. Several US academic sites participate in the TaiChi consortium: Stanford University School of Medicine in Stanford, California; Hudson-Alpha Biotechnology Institute in Huntsville, Alabama; and Cedars-Sinai

Medical Center in Los Angeles, California. The main academic sites in Taiwan include National Health Research Institutes (NHRI); National Taiwan University Hospital (NTUH); Taipei and Taichung Veteran's General Hospitals (VGH) and Tri-Service General Hospital (TSGH). To reach TaiChi's goal, a well-phenotyped East Asian sample of ~13,500 Han Chinese subjects living in Taiwan was assembled. Relevant qualitative and quantitative traits are available in either a subset or in all cohorts, and seven cohorts comprise the current TaiChi bio-resource and have information on body mass index. Each cohort is described in more detail below.

HALST (Healthy Aging Longitudinal Study in Taiwan) is a NHRI-established population-based study of older adults living in Taiwan, where more than 5,000 subjects were recruited over a four-year period from seven recruitment sites across the country.

SAPPHIRe (Stanford-Asian Pacific Program in Hypertension and Insulin Resistance) is a family-based study established in 1995 to pinpoint the major genetic loci underlying hypertension and insulin resistance through linkage analysis in East Asian populations. At the outset, SAPPHIRe involved recruitment sites in the San Francisco Bay Area, Hawaii, as well as in Taiwan. Many metabolic variables were examined in baseline and regular follow-up visits by a programmatic collaboration between the NHLBI in the US and NHRI in Taiwan.

TCAGEN (Taiwan Coronary Artery Disease GENetic) study is an ongoing cohort study that has been enrolling patients undergoing coronary angiography or other percutaneous intervention at the National Taiwan University Hospital (NTUH), when either stable angina pectoris or prior myocardial infarction has been identified. Participants are from the north of Taiwan, where the main NTU medical school/hospital is located, and from the Yulin branch of NTUH, which is located in south/central Taiwan. Peripheral blood was collected in the catheter lab specifically for buffy coat isolation and DNA extraction.

TACT (TAiwan Coronary and Transcatheter intervention) cohort study enrolled patients with angina pectoris and objective documentation of myocardial ischemia, who then underwent diagnostic coronary angiography and/or revascularization at NTUH after October 2000. Participants provided clinically relevant information including use of CVD-related medication, which is supplemented by a comprehensive electronic medical records database that includes information on drug use and surgical interventions.

Taiwan DRAGON (Taiwan Diabetes and RelAted Genetic COmplicatioN) study is a cohort study with Type 2 diabetes at the Veteran's General Hospital in Taichung, Taiwan (Taichung VGH). Participants were either newly diagnosed or known to have prevalent diabetes and sought care at an outpatient clinic. Subjects with hyperglycemia, but not diabetic, were excluded from participating in a health examination at Taichung VGH.

TUDR (Taiwan USA Diabetes Retinopathy) enrolled subjects with Type 2 diabetes receiving care at Taichung VGH or TSGH, and invited TUDR participants to complete fundoscopic examination to document the presence and extent of diabetic retinopathy, as well as a variety of other clinical phenotypes, including BMI. A total of 2,222 unrelated Type 2 diabetes subjects have consented to and undergone the MetaboChip genotyping as part of the Taiwan Dragon Study. In addition to DNA and buffy coats, fasting blood for future measurement of serum/plasma biomarkers has also been banked.

TCAD (Taichung CAD study) includes patients with a variety of CVD receiving care at the Taichung VGH. Specifically, individuals who were hospitalized for diagnostic and interventional coronary angiography examinations and treatment, or those with a history of myocardial or revascularization were included.

After acquiring appropriate IRB and Taiwan Department of Health permissions for the TaiChi Study, ~11,000 of the total 13,500 subjects included in this sample set had their buffy coat or DNA transferred to Cedars Sinai and HudsonAlpha, which was followed by careful DNA extraction, plating, and genotyping on the MetaboChip at HudsonAlpha.

The Women's Health Initiative (WHI)

The WHI is a large study of postmenopausal women's health investigating risk factors for cancer, CVD, age-related fractures and chronic disease [300]. It began in 1993 as a set of randomized controlled clinical trials (CT) and an observational study (OS). Specifically, the CT (n=68,132) included three overlapping components: The Hormone Therapy (HT) Trials (n=27,347), Dietary Modification (DM) Trial (n=48,835), and Calcium and Vitamin D (CaD) Trial (n=36,282). Eligible women could be randomized into as many as all three CTs components. Women who were ineligible or unwilling to join the CT were then invited to join the OS (n=93,676). All WHI pariticaipnt provided informed consent to submit their genotype data to dbGaP and were either directly genotyped on the MetaboChip or had previously-collected genome-wide data (Affymetrix 6.0 array) available for imputation (details, see Methods above).

APPENDIX I: CHARACTERIZATION OF 36 FINE-MAPPED REGIONS ON THE METABOCHIP WITH EVIDENCE OF GENOME-WIDE OR ARRAY-WIDE SIGNIFICANCE WITH BMI.

Gene	Chr	Bp37 start	Bp37 stop	Range	Total N SNPs	SNPs Passing QC***	Indepen-dent SNPs***	$P_{Bonferroni}$ ***	Associated Trait(s)
NEGR1	1	72,513,687	72,958,905	445,218	1377	1,076	284	1.76E-04	BMI; Weight
TNNI3K	1	74,961,817	75,078,975	117,158	368	311	95	5.26E-04	BMI
SEC16B	1	177,753,776	177,936,525	182,749	767	662	164	3.05E-04	BMI; Menarche; Weight
LYPLAL1*	1	219,533,817	219,807,974	274,157	980	767	218	2.29E-04	BMI**; Adiponectin levels; Adiposity; Fasting insulin-related traits (interaction with BMI); Height; Osteoarthritis; Visceral adipose tissue/subcutaneous adipose tissue ratio; Waist-hip ratio; Visceral to adipose tissue ratio
TMEM18	2	471,136	719,889	248,753	1126	862	257	1.95E-04	BMI; Menarche; Weight
BRE*	2	27,386,799	28,670,981	1,284,182	2702	2,157	669	7.47E-05	BMI*; Cardiovascular disease risk factors; Chronic kidney disease; Crohn's disease; Fasting glucose-related traits; Hypertriglyceridemia; Inflammatory bowel disease; LDL cholesterol; Lipoprotein-associated phospholipase A2 activity and mass; Liver enzyme levels (gamma-glutamyl transferase); Menopause (age at onset); Metabolic syndrome; Metabolic traits; Metabolite levels; Non-albumin protein levels ; Phospholipid levels (plasma); Platelet counts; Serum albumin level; Serum total protein level; Sex hormone-binding globulin levels; Triglycerides; Triglycerides-Blood Pressure (TG-BP); Two-hour glucose challenge; Type 1 diabetes; Urate levels; Uric acid levels; Waist Circumference - Triglycerides (WC-TG); Waist circumference and related phenotypes
COBLL1*	2	165,499,548	165,732,418	232,870	549	429	135	3.70E-04	BMI*; Fasting insulin-related traits (interaction with BMI); HDL cholesterol; Triglycerides; Type 2 diabetes; Waist-hip ratio
LOC646736*	2	227,007,600	227,190,673	183,073	718	592	148	3.38E-04	BMI*; Adiponectin levels; Adiposity; Coronary heart disease; Fasting insulin-related traits (interaction with BMI); HDL cholesterol; Triglycerides; Type 2 diabetes
CADM2	3	85,651,797	86,050,826	399,029	792	631	218	2.29E-04	BMI
IGF2BP2*	3	185,339,119	185,596,778	257,659	516	398	154	3.25E-04	BMI**; Diabetes (gestational); Fasting glucose-related traits (interaction with BMI); Height; Type 2 diabetes
ETV5	3	185,747,042	185,862,593	115,551	371	298	90	5.56E-04	BMI; Weight
GNPDA2	4	45,099,376	45,187,658	88,282	344	255	55	9.09E-04	BMI
SLC39A8*	4	103,121,726	103,218,446	96,720	306	254	64	7.81E-04	BMI*; Diastolic and systolic blood pressure; HDL

									cholesterol
FLJ35779	5	74,562,373	75,123,052	560,679	1129	891	244	2.05E-04	BMI; Total and LDL cholesterol; Metabolite levels
CDKAL1*	6	20,393,907	21,191,928	798,021	2372	1,957	580	8.62E-05	BMI*; Type 2 and gestational diabetes; Birth weight; Inflammatory bowel disease; Ileal carcinoids; Glycated hemoglobin levels
TFAP2B	6	50,534,485	51,100,751	566,266	1689	1,389	340	1.47E-04	BMI; Adiposity; Metabolic syndrome; Obesity; Renal function
SLC22A3*	6	160,704,943	160,975,629	270,686	1185	839	198	2.53E-04	BMI*; Colorectal cancer; Coronary heart disease; Lp (a) levels; Monocyte early outgrowth colony forming units; Prostate cancer
LRPN6C	9	28,403,443	28,499,099	95,656	344	278	66	7.58E-04	BMI
NT5C2*	10	104,217,441	104,999,266	781,825	1727	1,358	344	1.45E-04	BMI*; Blood pressure; Coronary heart disease; Intracranial aneurysm; Parkinson's disease; Schizophrenia; Systolic blood pressure
TCF7L2*	10	114,746,580	114,822,739	76,159	259	233	78	6.41E-04	BMI*: Coronary heart disease; Fasting glucose-related traits; Fasting glucose-related traits (interaction with BMI); Fasting insulin-related traits (interaction with BMI); Glycated hemoglobin levels; Metabolic syndrome; Proinsulin levels; Two-hour glucose challenge; Type 2 diabetes; Type 2 diabetes and other traits
KCNQ1*	11	2,444,094	2,943,115	499,021	2083	1,681	661	7.56E-05	BMI*; Bilirubin levels; Electrocardiographic traits; Height; Protein quantitative trait loci; QT interval; Type 2 diabetes
OVCH2	11	8,394,189	8,707,147	312,958	672	542	161	3.11E-04	BMI; Menarche
NCR3LG1/KCNJ11*	11	17,039,079	17,423,733	384,654	559	422	121	4.13E-04	BMI*: Height; Schizophrenia; Type 2 diabetes
BDNF	11	27,452,706	27,749,725	297,019	691	547	164	3.05E-04	BMI; Bone mineral density; Obesity; Smoking behavior, Weight
MTCH2	11	46,921,641	48,094,879	1,173,238	2401	1,873	566	8.83E-05	BMI; Fasting glucose; HDL cholesterol; Metabolic syndrome; Proinsulin levels; Serum albumin levels
FAIM2	12	50,168,189	50,290,056	121,867	343	260	84	5.95E-04	BMI; Waist circumference; Weight
TRAFD1*	12	111,290,599	113,206,306	1,915,707	3494	2,785	792	6.31E-05	BMI*: Alcohol consumption; Biomedical quantitative traits; Blood pressure; Celiac disease; Celiac disease and Rheumatoid arthritis; Cholesterol, total; Chronic kidney disease; Coronary heart disease; Diastolic blood pressure; Drinking behavior; Eosinophil counts; Esophageal cancer; Gamma glutamyl transpeptidase; HDL cholesterol; Hematocrit; Hematological parameters; Hemoglobin; Hypothyroidism; Intracranial aneurysm; LDL cholesterol; Mean platelet volume; Metabolite levels; Platelet counts; Red blood cell traits; Renal function-related traits (BUN); Renal function-related traits (sCR); Retinal vascular caliber; Rheumatoid arthritis; Stroke (ischemic); Systolic blood pressure; Tetralogy of Fallot;

									Triglycerides; Type 1 diabetes; Type 1 diabetes autoantibodies; Upper aerodigestive tract cancers; Urate levels; Vitiligo
PRKD1	14	30,436,558	30,543,794	107,236	251	203	76	6.58E-04	BMI
MAP2K5	15	67,649,978	68,215,300	565,322	1304	1,032	313	1.60E-04	BMI; Restless leg syndrome
GPRC5B/GP2	16	19,704,224	20,019,432	315,208	764	562	174	2.87E-04	BMI
SH2B1	16	28,306,987	29,001,460	694,473	795	501	177	2.82E-04	BMI; Inflammatory bowel disease; Type 1 diabetes; Weight
FTO	16	53,539,509	54,185,787	646,278	1817	1,501	490	1.02E-04	BMI; Waist circumference; Menarche; Adiposity; Obesity; Type 2 diabetes; Weight
KCNJ2*	17	68,259,822	68,516,393	256,571	1,096	944	229	2.18E-04	BMI*; QT interval; Thyrotoxic hypokalemic periodic paralysis
MC4R	18	57,727,147	58,094,636	367,489	1278	1,064	271	1.85E-04	BMI; Waist circumference; Height; Obesity; Weight
KCTD15	19	34,295,278	34,333,501	38,223	133	110	36	1.39E-03	BMI; Major depressive disorder; Weight
QPCTL*	19	46,136,487	46,406,698	270,211	598	445	155	3.23E-04	BMI*; 2 Hour glucose challenge; Adiposity

Abbreviations: Bp37=base pair Build 37, Chr=chromosome, SNPs=single nucleotide polymorphisms, QC=quality control.
*Note: Starred genes represent fine-mapped loci, which were associated with BMI after the design of the MetaboChip in 2009.
**PAGE trans-ethnic discovery signal (Gong *et al.,* submitted to *Nature Communications*).

APPENDIX J: GENOTYPING AND ANALYTICAL CHARACTERISTICS OF THE STUDIES COLLABORATING AS PART OF THE POPULATION ARCHITECTURE USING GENOMICS AND EPIDEMIOLOGY (PAGE) STUDY.

	Genotyping							Statistical Analysis	
	Metabochip genotype calling	**Imputation**	**HWE p-value threshold**	**SNP call rate***	**Additional SNP QC**	**Sample success rate**	**Duplicate Concord-ance rate**	**Softwar e**	**Covar-iates**
ARIC	GenomeStudio with the GenCall 2.0 algorithm	NA	African American: $p<1 \times 10^{-6}$	≥95%	GenTrain score<0.6 or cluster separation score<0.4 excluded; Mendelian errors;GenoSNP > 3.3% or PAGE consensus vs. HapMap	≥95%	≥99%	PLINK	age, sex, PCs: 1-10, center
EAGLE BioVU	GenomeStudio with the GenCall 2.0 algorithm	NA	Asian and African: $p<1 \times 10^{-6}$; Hispanic/Latino: Exclusions Identified in HCHS/SOL	≥95%	GenTrain score<0.6 or cluster separation score<0.4 excluded; Mendelian errors;GenoSNP > 3.3% or PAGE consensus vs. HapMap	≥95%	≥99%	PLINK	age, sex, PCs: 1-10
CARDIA	GenomeStudio with the GenCall 2.0 algorithm	NA	African American: $p<1 \times 10^{-6}$	≥95%	GenTrain score<0.6 or cluster separation score<0.4 excluded; Mendelian errors; GenoSNP > 3.3% or PAGE consensus vs. HapMap	≥95%	≥99%	PLINK	age, sex, PCs: 1-4, center
CHS	GenomeStudio with the GenCall 2.0 algorithm	NA	African American: $p<1 \times 10^{-6}$	≥95%	GenTrain score<0.6 or cluster separation score<0.4 excluded; Mendelian errors;GenoSNP > 3.3% or PAGE consensus vs. HapMap	≥95%	≥99%	PLINK	age, sex, PCs: 1-10, center
GenNet	GenomeStudio with the GenCall 2.0 algorithm	NA	African American: $p<1 \times 10^{-6}$	≥95%	GenTrain score<0.6 or cluster separation score<0.4 excluded; Mendelian errors; GenoSNP > 3.3% or PAGE consensus vs. HapMap	≥95%	≥99%	GWAF	age, sex, PCs: 1-10, center
HCHS/SOL	GenomeStudio with the GenCall 2.0 algorithm	NA	Hispanic/Latino: p<1 x 10-6 in HCHS/SOL both	≥95%	GenTrain score<0.6 or cluster separation score<0.4 excluded;	≥95%	≥99%	SUGEN 4.0	age, sex, PCs: 1-4, center,

			within and across self-identified background groups (n=747)		Mendelian errors; GenoSNP > 3.3% or PAGE consensus vs. HapMap				Hispanic/ Latino backgrou nd
HyperGen	GenomeStudio with the GenCall 2.0 algorithm	NA	African American: $p<1 \times 10^{-6}$	≥95%	GenTrain score<0.6 or cluster separation score<0.4 excluded; Mendelian errors; GenoSNP > 3.3% or PAGE consensus vs. HapMap	≥95%	≥99%	GWAF	age, sex, PCs: 1-10, center
MEC-Metabochip	GenomeStudio with the GenCall 2.0 algorithm	NA	Asian, African and European American: $p<1 \times 10^{-6}$	≥95%	GenTrain score<0.6 or cluster separation score<0.4 excluded; Mendelian errors;GenoSNP > 3.3% or PAGE consensus vs. HapMap	≥95%	≥99%	PLINK	age, sex, PCs: 1-10
MEC-Imputed	NA	Illumina HumanOmni 2.5 (Genome Studio) and 1000 Genome haplotype panel imputation	Hispanic/Latino: Exclusions Identified in HCHS/SOL	≥99%	Imputation (INFO) ≥ 0.6	≥95%	≥99%	PLINK	age, sex, PCs: 1-10
BioME	NA	Illumina HumanOmni ExpressExo me (Genome Studio) and 1000 Genome haplotype panel imputation	African American: p<5 x 10-5; Hispanic/Latino: Identified in HCHS/SOL, or in BioME $p<5 \times 10^{-5}$	≥95%	Proper_info ≥ 0.4	≥95%	≥99%	SNPTE ST	age, sex, PCs: 1-2 for African American s and 1-5 in Hispanic s
TaiChi	GenomeStudio with the GenCall 2.0 algorithm	NA	Asian: $p<1 \times 10^{-3}$	≥95%	Replication errors (1 or more)	≥98.5%	≥99%	PLINK	age, sex, PCs: 1-10, study cohort
WHI-Metabochip		NA	Asian, African, American	≥95%	GenTrain score<0.6 or cluster separation	≥95%	≥99%	R	age, PCs: 1-

			Indian/Alaskan Native: $p<1 \times 10^{-6}$; Hispanic/Latino: Exclusions Identified in HCHS/SOL		score<0.4 excluded; Mendelian errors; GenoSNP > 3.3% or PAGE consensus vs. HapMap				10
WHI-Imputed	NA	Affymetrix Genome-wide Human SNP Array 6.0 and 1000 Genome haplotype panel imputation	African and European American: $p<1 \times 10^{-6}$	≥95%	Imputation (Rsq)>0.3	≥95%	≥99%	R	age, sex, PCs: 1-4, center

Abbreviations: GWAS=Genome-wide association study, HWE=Hardy Weinberg Equilibrium, NA=Not applicable, PCs=Principal components, SNP=Single nucleotide polymorphism, QC=Quality Control

*An additional 2,646 SNPs across the MetaboChip were excluded due to poor call differentiation in quality control analyses.

APPENDIX K: AFRICAN DESCENT GENETIC EFFECT ESTIMATES FOR 170 BMI INDEX SNPS FROM GWAS, INCLUDING SECONDARY SIGNALS IN EUROPEAN DESCENT POPULATIONS AND/OR POPULATION-SPECIFIC MARKERS.

rsID*	Chr	Bp37	Gene	Ref. First Author	Ref. Risk Allele	A1	A2	Freq	β (%)	SE (%)	P***	I^2	HetP	N
rs2803328	1	1,874,326	*KIAA1751*	Winkler	C	c	g	0.804	0.21	0.20	2.9E-01	34.7	1.1E-01	35589
rs2271928	1	32,127,953	*COL16A1*	Winkler	A	a	g	0.485	0.08	0.15	5.9E-01	0	5.7E-01	35597
rs2275426	1	46,487,552	*MAST2*	Winkler	A	a	g	0.416	0.03	0.15	8.3E-01	59.1	3.5E-03	35598
rs977747	1	47,684,677	*TAL1*	Locke	T	t	g	0.663	0.02	0.16	9.0E-01	0	8.9E-01	35598
rs657452	1	49,589,847	*AGBL4*	Locke	A									
rs11583200	1	50,559,820	*ELAVL4*	Locke	C	c	t	0.712	0.30	0.17	7.5E-02	6.7	3.8E-01	35606
rs3101336	1	72,751,185	*NEGR1*	Speliotes	C	t	c	0.459	0.05	0.15	7.3E-01	0	4.6E-01	35603
rs12566985	1	75,002,193	*FPGT-TNNI3K*	Speliotes	G	g	a	0.790	0.57	0.19	2.6E-03	0	5.7E-01	35485
rs12401738	1	78,446,761	*FUBP1*	Locke**	A	a	g	0.117	0.48	0.24	4.3E-02	0	5.4E-01	35603
rs11165643	1	96,924,097	*PTBP2*	Speliotes	T	t	c	0.220	0.49	0.18	8.2E-03	10	3.5E-01	35598
rs17024393	1	110,154,688	*GNAT2*	Locke**	C	t	c	0.917	0.01	0.27	9.7E-01	48.3	2.6E-02	35603
rs4357530*	1	151,103,153	*SEMA6C*	Winkler	G	a	g	0.281	0.19	0.17	2.6E-01	56.6	6.3E-03	35597
rs10913118	1	175,954,755	*RFWD2*	Winkler	A	c	a	0.150	0.28	0.21	1.9E-01	0	8.2E-01	35600
rs574367* **	1	177,873,210	*SEC16B*	Wen 2014	T	t	g	0.110	0.98	0.24	4.6E-05	0	7.7E-01	35559
rs543874* **	1	177,889,480	*SEC16B*	Speliotes, Monda	G	g	a	0.249	1.37	0.17	6.0E-15	44.5	4.2E-02	35604
rs10920678	1	190,239,907	*FAM5C*	Winkler	A	a	g	0.412	0.26	0.15	9.4E-02	0	4.9E-01	35422
rs2820292	1	201,784,287	*NAV1*	Locke	C	c	a	0.364	0.08	0.16	6.0E-01	0	5.4E-01	35598
rs2820436	1	219,640,680	*LYPLAL1*	Gong (sub.)	A	a	c	0.478	0.51	0.15	8.4E-04	0	5.3E-01	35606
rs12463617* **	2	629,244	*TMEM18*	Wen 2014	C									
rs13021737* **	2	632,348	*TMEM18*	Speliotes	G	g	a	0.883	1.36	0.23	8.9E-09	26.8	1.7E-01	35541
rs11676272*,* **	2	25,141,538	*ADCY3*	Wen 2014	G	g	a	0.833	0.94	0.21	6.5E-06	0	4.5E-01	35600
rs10182181* **	2	25,150,296	*ADCY3*	Speliotes	G	g	a	0.830	0.89	0.21	1.5E-05	0	6.2E-01	35587
rs11126666	2	26,928,811	*KCNK3*	Locke	A	g	a	0.805	0.17	0.19	3.6E-01	20.3	2.4E-01	35597
rs116612809	2	28,301,171	*BRE*	Gong 2013	G	g	a	0.097	1.39	0.25	6.4E-08	0	6.3E-01	35583
rs1016287	2	59,305,625	*FLJ30838*	Speliotes	T	t	c	0.215	0.22	0.18	2.2E-01	15.8	2.9E-01	35600
rs11688816	2	63,053,048	*EHBP1*	Locke	G	a	g	0.379	0.13	0.15	4.1E-01	40.1	6.6E-02	35600
rs12622013*	2	79,501,362	*REG3A*	Winkler	G	g	a	0.200	0.62	0.19	1.1E-03	28.1	1.6E-01	35593
rs7570971	2	135,837,906	*RAB3GAP1*	Winkler	A	a	c	0.856	0.20	0.23	3.8E-01	5.3	3.9E-01	35579
rs4988235	2	136,608,646	*MCM6*	Winkler	A	g	a	0.851	0.20	0.22	3.7E-01	4.4	4.0E-01	35602
rs2121279	2	143,043,285	*LRP1B*	Speliotes	T	t	c	0.032	0.00	0.44	9.9E-01	0	7.5E-01	35577
rs1460676	2	164,567,689	*FIGN*	Locke	C	c	t	0.247	0.38	0.18	2.9E-02	41.1	6.0E-02	35603
rs10184004	2	165,508,389	*GRB14/COBLL1*	Gong (sub.)	T	t	c	0.719	0.72	0.17	2.1E-05	32.2	1.2E-01	35598
rs10930502	2	172,890,588	*METAP1D*	Gong (sub.)	A	a	g	0.700	0.40	0.17	1.6E-02	19.4	2.5E-01	35599
rs1528435	2	181,550,962	*UBE2E3*	Locke	T	t	c	0.612	0.16	0.16	3.0E-01	0	4.9E-01	35606
rs972540	2	207,244,783	*ADAM23*	Winkler	A	g	a	0.175	0.09	0.20	6.4E-01	6.7	3.8E-01	35605
rs17203016	2	208,255,518	*CREB1*	Locke	G	g	a	0.041	0.59	0.39	1.3E-01	0	8.8E-01	35590
rs7599312	2	213,413,231	*ERBB4*	Locke	G	g	a	0.623	0.09	0.16	5.7E-01	30.1	1.4E-01	35575
rs492400	2	219,349,752	*USP37*	Locke	C									
rs2176040	2	227,092,802	*LOC646736*	Speliotes	A	a	g	0.307	0.52	0.16	1.4E-03	1	4.4E-01	35602
rs9845966	3	13,433,158	*NUP210*	Winkler	T	t	g	0.232	0.03	0.20	8.6E-01	6.8	3.8E-01	29845
rs6804842	3	25,106,437	*RARB*	Locke	G	g	a	0.412	0.05	0.15	7.4E-01	0	8.3E-01	35595
rs7613875	3	49,971,514	*MON1A*	Winkler	A	a	c	0.674	0.18	0.18	3.3E-01	0	6.5E-01	27843
rs2365389	3	61,236,462	*FHIT*	Locke	C	c	t	0.203	0.22	0.19	2.5E-01	5.3	3.9E-01	35602
rs333495*	3	78,834,343	*ROBO1*	Winkler	G	t	g	0.510	0.08	0.15	5.9E-01	0	6.1E-01	35605
rs13078960	3	85,807,590	*CADM2*	Speliotes	G	t	g	0.935	0.02	0.31	9.6E-01	8.5	3.6E-01	35605
rs1720825	3	138,108,083	*MRAS*	Graff (prep.)	A	g	a	0.908	0.14	0.26	5.8E-01	11.4	3.3E-01	35606
rs2640017*	3	141,335,121	*RASA2*	Locke	G	g	a	0.018	0.34	0.57	5.5E-01	0	5.9E-01	35481
rs11927381	3	185,508,591	*IGF2BP2*	Gong (sub.)	T	t	c	0.255	0.59	0.18	7.7E-04	0	8.5E-01	35592
rs1516725	3	185,824,004	*ETV5*	Speliotes	C	c	t	0.817	0.64	0.20	1.2E-03	6.9	3.8E-01	35485
rs16992647	4	36,813,105	*KIAA1239*	Winkler	T	c	t	0.869	0.09	0.23	6.9E-01	0	5.0E-01	35478
rs16858082	4	45,175,804	*GNPDA2*	Wen 2014	T	t	c	0.599	0.39	0.15	1.1E-02	14.4	3.0E-01	35539
rs10938397	4	45,182,527	*GNPDA2*	Speliotes	G	g	a	0.250	0.77	0.17	8.4E-06	51.9	1.5E-02	35517
rs348495	4	45,184,442	*GNPDA2*	Monda	G									
rs13107325	4	103,188,709	*SLC39A8*	Speliotes	T	t	c	0.019	0.79	0.59	1.8E-01	6.6	3.8E-01	35183
rs11727676	4	145,659,064	*HHIP*	Locke	T	t	c	0.980	0.53	0.55	3.3E-01	0.4	4.4E-01	35596
rs2112347	5	75,015,242	*POC5*	Speliotes	T	t	g	0.495	0.09	0.15	5.5E-01	0	8.9E-01	35604
rs6870983	5	87,697,533	*TMEM161B−AS1*	Winkler	T	c	t	0.554	0.17	0.15	2.6E-01	39.9	6.8E-02	35601
rs11951673*	5	95,861,012	*PCSK1*	Wen 2014	C	c	t	0.593	0.48	0.15	1.6E-03	21	2.3E-01	35595
rs6864049	5	124,330,522	*ZNF608*	Winkler	A	g	a	0.803	0.31	0.19	1.1E-01	22.9	2.1E-01	35567
rs13174863	5	139,080,745	*CXXC5*	Winkler	A	g	a	0.060	0.27	0.32	4.0E-01	17.4	2.7E-01	35602
rs4569924*	5	153,540,025	*GALNT10*	Monda	T	t	c	0.355	0.65	0.16	4.8E-05	0	9.5E-01	35604
rs2228213	6	12,124,855	*HIVEP1*	Winkler	A	g	a	0.891	0.38	0.25	1.2E-01	25.6	1.9E-01	35606
rs9356744	6	20,685,486	*CDKAL1*	Wen 2014	T	t	c	0.374	0.15	0.16	3.4E-01	0	9.9E-01	35598
rs943466	6	33,731,787	*LEMD2*	Winkler	A	g	a	0.707	0.14	0.16	4.0E-01	12.9	3.2E-01	35602
rs205262	6	34,563,164	*C6orf106*	Speliotes	G	g	a	0.630	0.33	0.16	3.6E-02	0	4.9E-01	35603
rs2033529	6	40,348,653	*TDRG1*	Locke	G	a	g	0.835	0.15	0.20	4.6E-01	13.7	3.1E-01	35602
rs2207139	6	50,845,490	*TFAP2B*	Speliotes	G	g	a	0.096	0.79	0.26	2.0E-03	44.7	4.1E-02	35605
rs9400239	6	108,977,663	*FOXO3*	Locke	C	c	t	0.255	0.05	0.18	7.9E-01	0	9.8E-01	35600
rs9374842	6	120,185,665	*LOC285762*	Locke	T	c	t	0.228	0.03	0.18	8.5E-01	36.2	9.3E-02	35570
rs13201877	6	137,675,541	*IFNGR1*	Locke	G	g	a	0.033	0.32	0.43	4.6E-01	0	5.7E-01	35606
rs1281962	6	153,431,376	*RGS17*	Winkler	C	g	c	0.264	0.01	0.17	9.5E-01	38.8	7.5E-02	35599
rs3127574	6	160,791,370	*SLC22A3*	Winkler	C	c	g	0.587	0.03	0.15	8.3E-01	4	4.1E-01	35597
rs13191362	6	163,033,350	*PARK2*	Locke	A	a	g	0.947	0.97	0.34	4.1E-03	40.8	6.2E-02	35593
rs10499694*	7	50,614,173	*DDC*	Winkler	G	a	g	0.641	0.12	0.16	4.4E-01	0	8.8E-01	35598
rs1167827	7	75,163,169	*HIP1*	Locke	G	g	a	0.870	0.20	0.23	4.0E-01	0	5.1E-01	35598
rs6465468	7	95,169,514	*ASB4*	Locke	T	g	t	0.838	0.36	0.21	8.5E-02	23.8	2.0E-01	35573

rs6990042	8	14,173,974	*SGCZ*	Winkler	T	t	g	0.634	0.17	0.16	2.7E-01	0	5.5E-01	35602
rs7844647*	8	34,503,776	*Intergenic*	Winkler	T	t	c	0.285	0.63	0.17	2.0E-04	4.9	4.0E-01	35604
rs17405819	8	76,806,584	*HNF4G*	Locke**	T	t	c	0.917	0.13	0.28	6.3E-01	0	5.9E-01	35602
rs16907751	8	81,375,457	*ZBTB10*	Locke	C	c	t	0.921	0.13	0.28	6.5E-01	11.2	3.3E-01	35605
rs2033732	8	85,079,709	*RALYL*	Locke	C	t	c	0.119	0.79	0.23	7.2E-04	5.6	3.9E-01	35556
rs4740619	9	15,634,326	*C9orf93*	Locke	T	c	t	0.448	0.15	0.15	3.1E-01	48.7	2.5E-02	35575
rs10968576	9	28,414,339	*LINGO2*	Speliotes	G	g	a	0.170	0.63	0.20	1.8E-03	0	8.3E-01	35604
rs6477694	9	111,932,342	*EPB41L4B*	Locke	C	c	t	0.419	0.07	0.15	6.5E-01	0	6.9E-01	35603
rs1928295	9	120,378,483	*TLR4*	Locke	T	c	t	0.436	0.09	0.15	5.6E-01	0	4.7E-01	35594
rs10733682	9	129,460,914	*LMX1B*	Locke	A	a	g	0.288	0.39	0.17	1.9E-02	15.8	2.8E-01	35586
rs2270204	9	131,042,734	*SWI5*	Winkler	T	g	t	0.690	0.35	0.17	3.7E-02	33.3	1.2E-01	35527
rs7899106	10	87,410,904	*GRID1*	Locke	G	g	a	0.122	0.20	0.23	3.9E-01	55.3	8.2E-03	35596
rs17094222	10	102,395,440	*HIF1AN*	Locke	C	c	t	0.055	0.18	0.34	6.0E-01	0	8.7E-01	35597
rs11191560	10	104,869,038	*NT5C2*	Locke, Wen 2012	C	c	t	0.042	0.78	0.38	3.9E-02	0	7.6E-01	35601
rs7903146	10	114,758,349	*TCF7L2*	Locke	C	c	t	0.706	0.66	0.17	6.1E-05	26.7	1.8E-01	35604
rs10886017	10	118,672,531	*KIAA1598*	Winkler	A	a	c	0.462	0.28	0.15	6.3E-02	0	7.1E-01	35565
rs2237897	11	2,858,546	*KCNQ1*	Wen 2014	T	t	c	0.088	0.48	0.27	8.0E-02	3	4.2E-01	35432
rs4256980	11	8,673,939	*TRIM66*	Speliotes	G	g	c	0.515	0.26	0.15	8.8E-02	0	5.4E-01	35594
rs7928810	11	17,372,443	*NCR3LG1*	Winkler	A	a	c	0.915	0.70	0.28	1.2E-02	40.1	6.7E-02	35603
rs1557765	11	17,403,639	*KCNJ11*	Winkler	T	c	t	0.893	0.50	0.25	4.1E-02	27.6	1.7E-01	35605
rs11030104	11	27,684,517	*BDNF*	Speliotes, Wen 2014	A	a	g	0.951	1.28	0.36	3.8E-04	14.6	3.0E-01	35606
rs10835210	11	27,695,910	*BDNF*	Locke	C	c	a	0.868	0.22	0.25	3.8E-01	31	1.6E-01	29676
rs652722	11	31,905,534	*PAX6*	Wen 2012	C									
rs2176598	11	43,864,278	*HSD17B12*	Locke	T	t	c	0.363	0.02	0.16	9.0E-01	0	1.0E+00	35602
rs3817334	11	47,650,993	*MTCH2*	Speliotes	T	t	c	0.266	0.19	0.17	2.7E-01	0	7.1E-01	35594
rs1865732*	11	112,960,722	*NCAM1*	Winkler	C	t	c	0.668	0.12	0.16	4.5E-01	0	7.9E-01	35601
rs12286929	11	115,022,404	*CADM1*	Locke	G	g	a	0.570	0.05	0.15	7.4E-01	0	6.6E-01	34809
rs11611246	12	939,480	*WNK1*	Winkler	T									
rs7970953	12	24,075,508	*SOX5*	Winkler	A	a	g	0.236	0.07	0.18	7.1E-01	25.7	1.8E-01	35574
rs1405552	12	41,746,673	*PDZRN4*	Winkler	A	a	g	0.102	0.27	0.26	2.9E-01	0	4.5E-01	35606
rs11181001	12	41,948,196	*PDZRN4*	Winkler	A	g	a	0.419	0.07	0.15	6.4E-01	0	6.9E-01	35603
rs7138803	12	50,247,468	*BCDIN3D*	Speliotes	A	a	g	0.179	0.16	0.20	4.2E-01	32.9	1.2E-01	35600
rs1438994*	12	90,594,389	*Intergenic*	Winkler	T	t	c	0.112	0.43	0.24	7.4E-02	0	8.1E-01	35603
rs11065987	12	112,072,424	*BRAP*	Winkler	A	a	g	0.919	0.38	0.29	1.8E-01	0	5.7E-01	35602
rs17630235	12	112,591,686	*TRAFD1*	Winkler	A	g	a	0.920	0.43	0.29	1.4E-01	0	5.4E-01	35574
rs11057405	12	122,781,897	*CLIP1*	Locke	G									
rs1885988*	13	28,010,262	*MTIF3*	Speliotes	C	c	t	0.040	0.55	0.40	1.6E-01	40.1	6.6E-02	35605
rs12429545	13	54,102,206	*OLFM4*	Speliotes	A	a	g	0.049	0.65	0.35	6.6E-02	23.4	2.1E-01	35591
rs9540493	13	66,205,704	*MIR548X2*	Locke	A	a	g	0.610	0.37	0.15	1.5E-02	50.8	1.8E-02	35547
rs1441264	13	79,580,919	*MIR548A2*	Locke	A	a	g	0.693	0.06	0.16	7.3E-01	0	4.6E-01	35577
rs9634489	13	97,049,004	*HS6ST3*	Winkler	A	g	a	0.619	0.05	0.16	7.7E-01	0	8.9E-01	35603
rs10132280	14	25,928,179	*STXBP6*	Locke	C	c	a	0.475	0.17	0.15	2.6E-01	0	8.5E-01	35561
rs12885454	14	29,736,838	*PRKD1*	Locke	C	c	a	0.866	0.14	0.22	5.3E-01	0	1.0E+00	35585
rs11847697	14	30,515,112	*PRKD1*	Speliotes	T	t	c	0.331	0.03	0.16	8.4E-01	0	9.9E-01	35603
rs17522122	14	33,302,882	*AKAP6*	Winkler	T									
rs7141420	14	79,899,454	*NRXN3*	Speliotes	T	t	c	0.593	0.52	0.15	7.5E-04	14.4	3.0E-01	35592
rs3783890	14	93,790,276	*BTBD7*	Winkler	T	t	c	0.902	0.51	0.25	4.5E-02	0	8.9E-01	35606
rs7143963	14	103,304,425	*TRAF3*	Winkler	T	t	c	0.618	0.33	0.16	3.7E-02	69	1.2E-04	35602
rs709400	14	104,149,475	*KLC1*	Winkler	A	g	a	0.219	0.03	0.18	8.6E-01	11.8	3.3E-01	35605
rs3736485	15	51,748,610	*DMXL2*	Locke	A	a	g	0.579	0.14	0.15	3.7E-01	0	5.6E-01	35573
rs16951275	15	68,077,168	*MAP2K5*	Speliotes	T	t	c	0.610	0.57	0.15	2.7E-04	0	5.7E-01	35605
rs4776970	15	68,080,886	*MAP2K5*	Wen 2012	A	a	t	0.463	0.47	0.15	1.9E-03	24.5	2.0E-01	35602
rs7164727	15	73,093,991	*LOC100287559*	Locke	T	c	t	0.646	0.04	0.16	8.1E-01	0	5.0E-01	35600
rs7181659	15	95,267,483	*MCTP2*	Winkler	A	g	a	0.353	0.01	0.16	9.5E-01	35.3	1.0E-01	35602
rs11866815	16	387,867	*AXIN1*	Winkler	T	t	c	0.417	0.17	0.15	2.6E-01	39	7.4E-02	35540
rs12446632	16	19,935,389	*GPRC5B*	Speliotes	G									
rs11074446	16	20,255,123	*GP2*	Locke	T	t	c	0.678	0.49	0.16	2.4E-03	25.6	1.9E-01	35583
rs2650492	16	28,333,411	*SBK1*	Locke	A	a	g	0.064	0.70	0.33	3.5E-02	0	5.3E-01	35590
rs3888190	16	28,889,486	*ATP2A1*	speliotes	A	a	c	0.271	0.62	0.17	2.8E-04	25.4	1.9E-01	35587
rs4787491	16	30,015,337	*INO80E*	Locke	G	g	a	0.533	0.31	0.15	4.2E-02	2.9	4.2E-01	35602
rs9925964	16	31,129,895	*KAT8*	Locke	A	a	g	0.867	0.24	0.23	2.8E-01	22.1	2.3E-01	34839
rs2080454	16	49,062,590	*CBLN1*	Locke	C	c	a	0.654	0.27	0.16	8.4E-02	0	7.1E-01	35604
rs1558902***	16	53,803,574	*FTO*	Speliotes, Wen 2014	A									
rs17817964***	16	53,828,066	*FTO*	Monda	T	t	c	0.118	1.05	0.24	1.2E-05	33	1.2E-01	35606
rs889398	16	69,556,715	*NFAT5*	Winkler	T	c	t	0.718	0.02	0.17	8.9E-01	52.6	1.3E-02	35585
rs9914578	17	2,005,136	*SMG6*	Locke	G	c	g	0.473	0.02	0.15	8.9E-01	37.5	8.4E-02	35587
rs1000940	17	5,283,252	*RABEP1*	Locke	G	g	a	0.237	0.24	0.18	1.7E-01	0	7.8E-01	35599
rs4986044	17	21,261,560	*KCNJ12*	Winkler	T	t	c	0.654	0.07	0.16	6.8E-01	24.1	2.1E-01	34843
rs12150665	17	34,914,787	*GGNBP2*	Winkler	T	c	t	0.125	0.07	0.23	7.7E-01	42.7	5.1E-02	35601
rs11652097	17	45,316,717	*ITGB3*	Winkler	T	c	t	0.639	0.17	0.16	2.9E-01	0	7.4E-01	35559
rs6504108	17	46,292,923	*SKAP1*	Winkler	T	c	t	0.302	0.16	0.16	3.3E-01	0	6.4E-01	35599
rs8075273*	17	61,728,881	*unknown*	Winkler	C	c	a	0.635	0.02	0.16	8.9E-01	0	7.9E-01	35365
rs312750	17	68,343,539	*KCNJ2*	Winkler	A	a	g	0.810	0.06	0.19	7.6E-01	0	8.6E-01	35604
rs12940622	17	78,615,571	*RPTOR*	Locke**	G	g	a	0.446	0.25	0.15	1.0E-01	0	8.6E-01	35603
rs1808579	18	21,104,888	*C18orf8*	Speliotes	C	c	t	0.548	0.08	0.15	5.9E-01	21.1	2.3E-01	35597
rs7239883	18	40,147,671	*LOC284260*	Locke	G	a	g	0.568	0.00	0.15	9.8E-01	10.2	3.4E-01	35396
rs7243357	18	56,883,319	*GRP*	Locke	T	t	g	0.872	0.59	0.23	8.4E-03	0	9.6E-01	35606
rs2331841	18	57,828,637	*MC4R*	Okada	A	a	g	0.486	0.70	0.15	4.1E-06	0	7.5E-01	35586
rs6567160	18	57,829,135	*MC4R*	Speliotes, Pei	C	c	t	0.189	1.08	0.19	2.8E-08	58.9	3.7E-03	35599
rs591166	18	57,841,589	*MC4R*	Wen 2014	A	a	t	0.742	0.28	0.17	1.1E-01	0	7.5E-01	35594
rs9944545	18	57,958,244	*MC4R*	Locke	T	t	c	0.543	0.20	0.15	1.9E-01	0	4.5E-01	35603
rs17066842	18	58,040,624	*MC4R*	Locke	G	g	a	0.827	0.34	0.20	8.6E-02	30.2	1.4E-01	35570
rs17724992	19	18,454,825	*PGPEP1*	Locke	A	a	g	0.887	0.67	0.24	5.7E-03	15.2	2.9E-01	35599
rs17513613	19	30,286,822	*CCNE1*	Winkler	T	t	c	0.889	0.17	0.24	4.8E-01	0	9.8E-01	35605
rs29941	19	34,309,532	*KCTD15*	Speliotes	G	g	a	0.818	0.24	0.19	2.1E-01	0	7.9E-01	35593
rs2075650	19	45,395,619	*TOMM40*	Speliotes	A	g	a	0.125	0.01	0.24	9.6E-01	3.5	4.1E-01	35602
rs11671664	19	46,172,278	*GIPR, QPCTL*	Wen 2014	G	a	g	0.116	0.41	0.24	8.9E-02	0	7.5E-01	35596
rs2287019	19	46,202,172	*QPCTL*	Speliotes	C									

rs3810291	19	47,569,003	*ZC3H4*	Speliotes	A	a	g	0.210	0.47	0.25	5.8E-02	0	5.2E-01	20961
rs4802349	19	47,874,510	*DHX34*	Gong 2013	G	g	t	0.519	0.84	0.15	3.8E-08	0	5.1E-01	35507
rs8123881*	20	15,819,495	*MACROD2*	Winkler	G	g	a	0.357	0.29	0.16	6.9E-02	26.9	1.7E-01	35604
rs6091540	20	51,087,862	*ZFP64*	Locke**	C	c	t	0.775	0.22	0.18	2.2E-01	0	6.9E-01	35604
rs2836754	21	40,291,740	*ETS2*	Locke	C	c	t	0.372	0.23	0.16	1.5E-01	0	5.7E-01	35601
rs4820408	22	40,604,945	*TNRC6B*	Winkler	T	g	t	0.893	0.02	0.25	9.5E-01	10.9	3.4E-01	35606

Abbreviations: A1=coded allele, A2=non-coded allele, Bp37=base pair Build 37, Chr=chromosome, FE=Fixed-Effect, HetP=heterogeneity p-value, P=p-value, Prep=prepared reference, Ref=reference, Sub=submitted reference, SNPs=single nucleotide polymorphisms.

*When the index SNP was not genotyped on the MetaboChip, the proxy SNP in tight linkage disequilibrium ($r^2 \geq 0.8$ in 1000 Genomes pilot 1 CEU, YRI, CHB+JPT depending on the population of discovery) with the lowest p-value in the African American sample was chosen to represent the index signal. The decreasing and increasing alleles for proxies were assigned assuming that the risk index SNP would have a similar allele frequency in the 1000 Genomes population (EUR, AFR, or EAS depending on the discovery population) as the risk proxy SNP.

**These loci were also described by Berndt *et al*. for obesity (maximum sample size of 263,407) [29]. The most recent BMI references per racial/ethnic group are noted above by their first author and publication year, if applicable [21, 25, 26, 30, 32, 33, 186, 195, 204].

***For GWAS SNPs a Bonferroni correction for multiple tests in the fixed-effect analyses reflected the number of independent previously-reported signals tested (=0.05/166). The 4 noted SNP pairs above were in tight linkage disequilibrium [$r^2 \geq 0.8$ in non-European 1000 genomes pilot populations(s)] with each other, but because they were reports from distinct discovery populations we retained them in this inventory in case they were population-specific variants. Therefore, our Bonferroni correction was penalized for only 166 (=170-4) tests.

****Bonferroni significant heterogeneity p-values shown in italics.

APPENDIX L: HISPANIC/LATINO DESCENT GENETIC EFFECT ESTIMATES FOR 170 BMI INDEX SNPS FROM GWAS, INCLUDING SECONDARY SIGNALS IN EUROPEAN DESCENT POPULATIONS AND/OR POPULATION-SPECIFIC MARKERS.

rsID*	Chr	Bp37	Gene	Ref. First Author	Ref. Risk Allele	A1	A2	Freq	β (%)	SE (%)	P***	I^2	HetP	N
rs2803328	1	1,874,326	*KIAA1751*	Winkler	C	g	c	0.515	0.03	0.17	8.6E-01	19.9	2.8E-01	26044
rs2271928	1	32,127,953	*COL16A1*	Winkler	A	g	a	0.550	0.07	0.16	6.9E-01	0	6.7E-01	26040
rs2275426	1	46,487,552	*MAST2*	Winkler	A	a	g	0.495	0.26	0.17	1.2E-01	23.6	2.5E-01	26046
rs977747	1	47,684,677	*TAL1*	Locke	T	t	g	0.550	0.00	0.17	1.0E+00	0	6.4E-01	26046
rs657452	1	49,589,847	*AGBL4*	Locke	A	a	g	0.437	0.79	0.17	5.5E-06	0	9.8E-01	24479
rs11583200	1	50,559,820	*ELAVL4*	Locke	C	c	t	0.509	0.53	0.17	1.2E-03	42.7	1.1E-01	26045
rs3101336	1	72,751,185	*NEGR1*	Speliotes	C	t	c	0.299	0.04	0.18	8.3E-01	17.9	2.9E-01	26045
rs12566985	1	75,002,193	*FPGT-TNNI3K*	Speliotes	G	g	a	0.604	0.28	0.17	1.0E-01	2.4	4.1E-01	25968
rs12401738	1	78,446,761	*FUBP1*	Locke**	A	g	a	0.716	0.09	0.18	6.2E-01	11.1	3.4E-01	26045
rs11165643	1	96,924,097	*PTBP2*	Speliotes	T	t	c	0.582	0.46	0.17	6.0E-03	4.3	3.9E-01	26031
rs17024393	1	110,154,688	*GNAT2*	Locke**	C	t	c	0.976	0.04	0.64	9.5E-01	0	6.2E-01	24225
rs4357530*	1	151,103,153	*SEMA6C*	Winkler	G	g	a	0.274	0.07	0.21	7.3E-01	18.9	2.9E-01	21396
rs10913118	1	175,954,755	*RFWD2*	Winkler	A	a	c	0.762	0.27	0.20	1.7E-01	0	8.9E-01	26042
rs574367***	1	177,873,210	*SEC16B*	Wen 2014	T	t	g	0.183	0.61	0.21	3.9E-03	0	6.0E-01	26044
rs543874***	1	177,889,480	*SEC16B*	Speliotes, Monda	G	g	a	0.202	0.76	0.20	1.8E-04	0	4.4E-01	26045
rs10920678	1	190,239,907	*FAM5C*	Winkler	A	a	g	0.296	0.36	0.18	5.3E-02	38.2	1.4E-01	25997
rs2820292	1	201,784,287	*NAV1*	Locke	C	c	a	0.429	0.15	0.17	3.6E-01	0	5.6E-01	26045
rs2820436	1	219,640,680	*LYPLAL1*	Gong (sub.)	A	a	c	0.439	0.63	0.17	1.6E-04	0	4.8E-01	26046
rs12463617***	2	629,244	*TMEM18*	Wen 2014	C	c	a	0.851	1.17	0.23	6.1E-07	46.5	8.2E-02	26046
rs13021737***	2	632,348	*TMEM18*	Speliotes	G	g	a	0.867	1.14	0.24	3.6E-06	29.5	2.0E-01	26016
rs11676272*,***	2	25,141,538	*ADCY3*	Wen 2014	G	g	a	0.414	0.48	0.17	4.7E-03	0	7.3E-01	26046
rs10182181***	2	25,150,296	*ADCY3*	Speliotes	G	g	a	0.414	0.47	0.17	6.1E-03	0	8.4E-01	26045
rs11126666	2	26,928,811	*KCNK3*	Locke	A	a	g	0.194	0.21	0.21	3.0E-01	4.9	3.9E-01	26047
rs116612809	2	28,301,171	*BRE*	Gong 2013	G	a	g	0.979	1.35	0.70	5.7E-02	0	5.6E-01	22385
rs1016287	2	59,305,625	*FLJ30838*	Speliotes	T	t	c	0.272	0.10	0.19	5.7E-01	0	8.4E-01	26044
rs11688816	2	63,053,048	*EHBP1*	Locke	G	a	g	0.446	0.13	0.17	4.3E-01	69.3	3.4E-03	26044
rs12622013*	2	79,501,362	*REG3A*	Winkler	G	g	a	0.163	0.06	0.22	8.0E-01	0	5.3E-01	26015
rs7570971	2	135,837,906	*RAB3GAP1*	Winkler	A	c	a	0.221	0.21	0.20	3.0E-01	8.8	3.6E-01	26045
rs4988235	2	136,608,646	*MCM6*	Winkler	A	a	g	0.226	0.31	0.20	1.3E-01	0.5	4.2E-01	26044
rs2121279	2	143,043,285	*LRP1B*	Speliotes	T	t	c	0.062	0.01	0.34	9.7E-01	25.3	2.4E-01	26033
rs1460676	2	164,567,689	*FIGN*	Locke	C	c	t	0.134	0.24	0.25	3.3E-01	0	9.0E-01	26048
rs10184004	2	165,508,389	*GRB14/COBLL1*	Gong (sub.)	T	t	c	0.326	0.39	0.18	3.2E-02	44	9.7E-02	26045
rs10930502	2	172,890,588	*METAP1D*	Gong (sub.)	A	a	g	0.656	0.43	0.18	1.4E-02	0	5.2E-01	26043
rs1528435	2	181,550,962	*UBE2E3*	Locke	T	t	c	0.653	0.45	0.18	9.8E-03	47.2	7.8E-02	26048
rs972540	2	207,244,783	*ADAM23*	Winkler	A	g	a	0.191	0.63	0.21	2.9E-03	0	5.7E-01	26047
rs17203016	2	208,255,518	*CREB1*	Locke	G	g	a	0.130	0.12	0.25	6.2E-01	0	8.9E-01	26038
rs7599312	2	213,413,231	*ERBB4*	Locke	G	g	a	0.780	0.38	0.20	6.1E-02	41.4	1.2E-01	26046
rs492400	2	219,349,752	*USP37*	Locke	C	c	t	0.419	0.37	0.18	3.5E-02	0	5.3E-01	24446
rs2176040	2	227,092,802	*LOC646736*	Speliotes	A	a	g	0.241	0.67	0.19	6.2E-04	0	4.8E-01	26045
rs9845966	3	13,433,158	*NUP210*	Winkler	T	t	g	0.536	0.26	0.17	1.2E-01	0	5.2E-01	26044
rs6804842	3	25,106,437	*RARB*	Locke	G	g	a	0.552	0.12	0.16	4.5E-01	0	9.4E-01	26046
rs7613875	3	49,971,514	*MON1A*	Winkler	A	a	c	0.412	0.07	0.17	7.0E-01	0	7.2E-01	25991
rs2365389	3	61,236,462	*FHIT*	Locke	C	c	t	0.379	0.36	0.17	3.9E-02	0	4.7E-01	26047
rs333495*	3	78,834,343	*ROBO1*	Winkler	G	t	g	0.499	0.06	0.17	7.3E-01	0	5.5E-01	26048
rs13078960	3	85,807,590	*CADM2*	Speliotes	G	g	t	0.136	0.02	0.24	9.4E-01	0	7.5E-01	26046
rs1720825	3	138,108,083	*MRAS*	Graff (prep.)	A	g	a	0.867	0.10	0.24	6.9E-01	46.3	8.3E-02	26048
rs2640017*	3	141,335,121	*RASA2*	Locke	G	g	a	0.192	0.53	0.22	1.6E-02	38.4	1.4E-01	26046
rs11927381	3	185,508,591	*IGF2BP2*	Gong (sub.)	T	t	c	0.673	0.52	0.18	3.9E-03	23.6	2.5E-01	25976
rs1516725	3	185,824,004	*ETV5*	Speliotes	C	c	t	0.900	0.84	0.28	3.1E-03	0	1.0E+00	26021
rs16992647	4	36,813,105	*KIAA1239*	Winkler	T	t	c	0.155	0.25	0.23	2.7E-01	0	7.5E-01	26047
rs16858082	4	45,175,804	*GNPDA2*	Wen 2014	T	t	c	0.559	0.46	0.17	5.5E-03	47.5	7.6E-02	26035
rs10938397	4	45,182,527	*GNPDA2*	Speliotes	G	g	a	0.372	0.70	0.17	4.7E-05	49.1	6.7E-02	26020
rs348495	4	45,184,442	*GNPDA2*	Monda	G	g	a	0.523	0.44	0.17	8.4E-03	45.9	8.5E-02	25944
rs13107325	4	103,188,709	*SLC39A8*	Speliotes	T	t	c	0.046	1.03	0.39	9.5E-03	55.2	3.7E-02	26048
rs11727676	4	145,659,064	*HHIP*	Locke	T	c	t	0.056	0.71	0.36	4.9E-02	25.8	2.3E-01	26048
rs2112347	5	75,015,242	*POC5*	Speliotes	T	t	g	0.629	0.30	0.17	8.1E-02	0	5.6E-01	26045
rs6870983	5	87,697,533	*TMEM161B−AS1*	Winkler	T	c	t	0.778	0.37	0.21	7.1E-02	48.4	7.1E-02	26048
rs11951673*	5	95,861,012	*PCSK1*	Wen 2014	C	c	t	0.614	0.41	0.17	1.5E-02	0	5.9E-01	26042
rs6864049	5	124,330,522	*ZNF608*	Winkler	A	a	g	0.296	0.12	0.18	5.1E-01	0	6.1E-01	26044
rs13174863	5	139,080,745	*CXXC5*	Winkler	A	g	a	0.131	0.37	0.25	1.3E-01	0	9.9E-01	26047
rs4569924*	5	153,540,025	*GALNT10*	Monda	T	t	c	0.598	0.15	0.17	3.8E-01	53.7	4.4E-02	26044
rs2228213	6	12,124,855	*HIVEP1*	Winkler	A	g	a	0.741	0.01	0.19	9.6E-01	29.8	2.0E-01	26048
rs9356744	6	20,685,486	*CDKAL1*	Wen 2014	T	t	c	0.641	0.33	0.17	5.4E-02	6.2	3.8E-01	26048
rs943466	6	33,731,787	*LEMD2*	Winkler	A	g	a	0.676	0.06	0.18	7.5E-01	0	9.9E-01	26048
rs205262	6	34,563,164	*C6orf106*	Speliotes	G	g	a	0.284	0.25	0.19	1.8E-01	0	7.2E-01	26045
rs2033529	6	40,348,653	*TDRG1*	Locke	G	g	a	0.184	0.04	0.22	8.6E-01	1.5	4.1E-01	26046
rs2207139	6	50,845,490	*TFAP2B*	Speliotes	G	g	a	0.294	0.30	0.19	1.0E-01	0	9.7E-01	26045
rs9400239	6	108,977,663	*FOXO3*	Locke	C	c	t	0.586	0.03	0.17	8.4E-01	0	4.5E-01	26044
rs9374842	6	120,185,665	*LOC285762*	Locke	T	t	c	0.793	0.06	0.20	7.6E-01	0	6.6E-01	26009
rs13201877	6	137,675,541	*IFNGR1*	Locke	G	a	g	0.891	0.02	0.27	9.4E-01	0	9.0E-01	26048
rs1281962	6	153,431,376	*RGS17*	Winkler	C	c	g	0.615	0.12	0.17	4.9E-01	50.3	6.0E-02	26038
rs3127574	6	160,791,370	*SLC22A3*	Winkler	C	c	g	0.456	0.49	0.17	2.9E-03	0	7.1E-01	26048
rs13191362	6	163,033,350	*PARK2*	Locke	A	g	a	0.085	0.12	0.30	7.0E-01	0	4.9E-01	26040
rs10499694*	7	50,614,173	*DDC*	Winkler	G	g	a	0.437	0.07	0.17	6.7E-01	0	9.0E-01	26048
rs1167827	7	75,163,169	*HIP1*	Locke	G	g	a	0.452	0.45	0.18	1.1E-02	45.5	8.8E-02	26044
rs6465468	7	95,169,514	*ASB4*	Locke	T	t	g	0.229	0.04	0.20	8.3E-01	56.9	3.1E-02	26046
rs6990042	8	14,173,974	*SGCZ*	Winkler	T	g	t	0.514	0.35	0.17	3.7E-02	0	9.7E-01	26043
rs7844647*	8	34,503,776	*Intergenic*	Winkler	T	c	t	0.390	0.05	0.17	7.7E-01	16.7	3.0E-01	26047

rs17405819	8	76,806,584	*HNF4G*	Locke**	T	t	c	0.739	0.06	0.19	7.5E-01	0	7.9E-01	26043
rs16907751	8	81,375,457	*ZBTB10*	Locke	C	c	t	0.925	0.13	0.32	6.8E-01	0	7.0E-01	26047
rs2033732	8	85,079,709	*RALYL*	Locke	C	c	t	0.784	0.23	0.20	2.5E-01	18	2.9E-01	26033
rs4740619	9	15,634,326	*C9orf93*	Locke	T	t	c	0.398	0.07	0.17	6.8E-01	30.8	1.9E-01	26044
rs10968576	9	28,414,339	*LINGO2*	Speliotes	G	g	a	0.231	0.64	0.20	1.1E-03	19.8	2.8E-01	26048
rs6477694	9	111,932,342	*EPB41L4B*	Locke	C	c	t	0.402	0.17	0.17	3.2E-01	43.7	9.9E-02	26047
rs1928295	9	120,378,483	*TLR4*	Locke	T	t	c	0.560	0.24	0.16	1.5E-01	0	8.7E-01	26045
rs10733682	9	129,460,914	*LMX1B*	Locke	A	a	g	0.571	0.34	0.17	4.3E-02	53.5	4.5E-02	26048
rs2270204	9	131,042,734	*SWI5*	Winkler	T	g	t	0.391	0.12	0.17	4.9E-01	10.4	3.5E-01	26027
rs7899106	10	87,410,904	*GRID1*	Locke	G	a	g	0.947	0.02	0.38	9.5E-01	0	5.5E-01	26048
rs17094222	10	102,395,440	*HIF1AN*	Locke	C	c	t	0.219	0.30	0.20	1.3E-01	20.1	2.8E-01	26047
rs11191560	10	104,869,038	*NT5C2*	Locke, Wen 2012	C	c	t	0.157	0.58	0.23	1.2E-02	0	7.1E-01	26048
rs7903146	10	114,758,349	*TCF7L2*	Locke	C	c	t	0.739	0.79	0.19	3.3E-05	63.8	1.1E-02	26047
rs10886017	10	118,672,531	*KIAA1598*	Winkler	A	c	a	0.683	0.19	0.18	2.8E-01	0	4.4E-01	26047
rs2237897	11	2,858,546	*KCNQ1*	Wen 2014	T	t	c	0.200	0.82	0.22	1.4E-04	0	9.1E-01	26044
rs4256980	11	8,673,939	*TRIM66*	Speliotes	G	g	c	0.570	0.04	0.17	8.1E-01	13.4	3.3E-01	26039
rs7928810	11	17,372,443	*NCR3LG1*	Winkler	A	a	c	0.654	0.10	0.18	5.8E-01	45.1	9.1E-02	26045
rs1557765	11	17,403,639	*KCNJ11*	Winkler	T	c	t	0.653	0.12	0.18	4.8E-01	38.6	1.3E-01	26043
rs11030104	11	27,684,517	*BDNF*	Speliotes, Wen 2014	A	a	g	0.831	0.42	0.22	5.4E-02	49.6	6.4E-02	26047
rs10835210	11	27,695,910	*BDNF*	Locke	C	c	a	0.709	0.17	0.18	3.4E-01	0	8.0E-01	25824
rs652722	11	31,905,534	*PAX6*	Wen 2012	C	c	t	0.685	0.18	0.18	3.2E-01	0	7.3E-01	26022
rs2176598	11	43,864,278	*HSD17B12*	Locke	T	t	c	0.405	0.05	0.17	7.8E-01	14.8	3.2E-01	26045
rs3817334	11	47,650,993	*MTCH2*	Speliotes	T	t	c	0.397	0.51	0.17	2.5E-03	0	6.1E-01	26040
rs1865732*	11	112,960,722	*NCAM1*	Winkler	C	t	c	0.579	0.05	0.17	7.7E-01	0	8.3E-01	26044
rs12286929	11	115,022,404	*CADM1*	Locke	G	g	a	0.526	0.28	0.17	9.2E-02	0	9.1E-01	26044
rs11611246	12	939,480	*WNK1*	Winkler	T	t	g	0.266	0.55	0.19	3.7E-03	28	2.1E-01	25991
rs7970953	12	24,075,508	*SOX5*	Winkler	A	a	g	0.441	0.13	0.17	4.5E-01	63.9	1.1E-02	26041
rs1405552	12	41,746,673	*PDZRN4*	Winkler	A	g	a	0.545	0.14	0.17	4.0E-01	21.4	2.7E-01	26047
rs11181001	12	41,948,196	*PDZRN4*	Winkler	A	a	g	0.450	0.24	0.17	1.4E-01	0	5.4E-01	26045
rs7138803	12	50,247,468	*BCDIN3D*	Speliotes	A	a	g	0.259	0.50	0.19	7.9E-03	12.4	3.3E-01	26047
rs1438994*	12	90,594,389	*Intergenic*	Winkler	T	t	c	0.235	0.04	0.20	8.3E-01	15.8	3.1E-01	26046
rs11065987	12	112,072,424	*BRAP*	Winkler	A	a	g	0.743	0.03	0.19	8.9E-01	10.1	3.5E-01	26045
rs17630235	12	112,591,686	*TRAFD1*	Winkler	A	g	a	0.745	0.02	0.20	9.1E-01	0	4.8E-01	26045
rs11057405	12	122,781,897	*CLIP1*	Locke	G	g	a	0.926	0.25	0.32	4.4E-01	38.3	1.4E-01	25991
rs1885988*	13	28,010,262	*MTIF3*	Speliotes	C	c	t	0.106	0.31	0.27	2.5E-01	40.1	1.2E-01	26048
rs12429545	13	54,102,206	*OLFM4*	Speliotes	A	a	g	0.273	0.84	0.19	1.1E-05	0	8.2E-01	26048
rs9540493	13	66,205,704	*MIR548X2*	Locke	A	a	g	0.485	0.46	0.17	5.5E-03	32.5	1.8E-01	26040
rs1441264	13	79,580,919	*MIR548A2*	Locke	A	a	g	0.678	0.26	0.18	1.4E-01	64.2	1.0E-02	26039
rs9634489	13	97,049,004	*HS6ST3*	Winkler	A	g	a	0.399	0.12	0.17	5.1E-01	0	9.7E-01	26048
rs10132280	14	25,928,179	*STXBP6*	Locke	C	c	a	0.682	0.62	0.18	4.7E-04	10.7	3.5E-01	26043
rs12885454	14	29,736,838	*PRKD1*	Locke	C	c	a	0.736	0.19	0.19	3.2E-01	0	4.7E-01	26047
rs11847697	14	30,515,112	*PRKD1*	Speliotes	T	c	t	0.906	0.05	0.31	8.6E-01	16.1	3.1E-01	26045
rs17522122	14	33,302,882	*AKAP6*	Winkler	T	t	g	0.396	0.25	0.17	1.4E-01	0	6.5E-01	25973
rs7141420	14	79,899,454	*NRXN3*	Speliotes	T	t	c	0.626	0.30	0.17	7.5E-02	37.7	1.4E-01	26047
rs3783890	14	93,790,276	*BTBD7*	Winkler	T	c	t	0.207	0.13	0.20	5.3E-01	19.9	2.8E-01	26048
rs7143963	14	103,304,425	*TRAF3*	Winkler	T	t	c	0.349	0.07	0.17	6.8E-01	0	7.5E-01	26048
rs709400	14	104,149,475	*KLC1*	Winkler	A	a	g	0.751	0.12	0.19	5.5E-01	0	4.4E-01	26044
rs3736485	15	51,748,610	*DMXL2*	Locke	A	a	g	0.484	0.10	0.17	5.3E-01	43.7	1.0E-01	26037
rs16951275	15	68,077,168	*MAP2K5*	Speliotes	T	t	c	0.531	0.37	0.17	3.4E-02	25.4	2.4E-01	26046
rs4776970	15	68,080,886	*MAP2K5*	Wen 2012	A	a	t	0.428	0.33	0.17	5.9E-02	4.8	3.9E-01	26047
rs7164727	15	73,093,991	*LOC100287559*	Locke	T	t	c	0.551	0.24	0.17	1.5E-01	0	7.8E-01	26046
rs7181659	15	95,267,483	*MCTP2*	Winkler	A	a	g	0.618	0.12	0.17	5.0E-01	0	6.0E-01	26045
rs11866815	16	387,867	*AXIN1*	Winkler	T	t	c	0.220	0.11	0.20	6.0E-01	0	5.8E-01	26048
rs12446632	16	19,935,389	*GPRC5B*	Speliotes	G	g	a	0.918	0.16	0.30	6.0E-01	0	6.3E-01	26020
rs11074446	16	20,255,123	*GP2*	Locke	T	t	c	0.805	0.32	0.21	1.4E-01	6	3.8E-01	26038
rs2650492	16	28,333,411	*SBK1*	Locke	A	a	g	0.140	0.41	0.24	9.1E-02	0	8.4E-01	26047
rs3888190	16	28,889,486	*ATP2A1*	speliotes	A	a	c	0.409	0.44	0.17	8.7E-03	34.4	1.7E-01	26041
rs4787491	16	30,015,337	*INO80E*	Locke	G	g	a	0.428	0.36	0.17	3.2E-02	8.6	3.6E-01	26047
rs9925964	16	31,129,895	*KAT8*	Locke	A	a	g	0.596	0.47	0.17	6.5E-03	38.5	1.4E-01	26047
rs2080454	16	49,062,590	*CBLN1*	Locke	C	a	c	0.637	0.01	0.17	9.7E-01	60.7	1.8E-02	26045
rs1558902***	16	53,803,574	*FTO*	Speliotes, Wen 2014	A	a	t	0.260	1.39	0.19	5.9E-13	45.6	8.8E-02	26004
rs17817964***	16	53,828,066	*FTO*	Monda	T	t	c	0.253	1.37	0.19	2.1E-12	47.2	7.8E-02	26046
rs889398	16	69,556,715	*NFAT5*	Winkler	T	t	c	0.343	0.07	0.17	6.8E-01	0	7.1E-01	26048
rs9914578	17	2,005,136	*SMG6*	Locke	G	g	c	0.275	0.11	0.19	5.5E-01	0	5.6E-01	25996
rs1000940	17	5,283,252	*RABEP1*	Locke	G	a	g	0.652	0.23	0.17	1.8E-01	0	8.0E-01	26043
rs4986044	17	21,261,560	*KCNJ12*	Winkler	T	t	c	0.469	0.00	0.17	9.9E-01	0	7.9E-01	26046
rs12150665	17	34,914,787	*GGNBP2*	Winkler	T	t	c	0.697	0.45	0.18	1.3E-02	33.5	1.7E-01	26045
rs11652097	17	45,316,717	*ITGB3*	Winkler	T	c	t	0.658	0.25	0.17	1.5E-01	0	9.5E-01	26046
rs6504108	17	46,292,923	*SKAP1*	Winkler	T	t	c	0.760	0.14	0.19	4.7E-01	3.5	4.0E-01	26046
rs8075273*	17	61,728,881	*unknown*	Winkler	C	c	a	0.771	0.21	0.19	2.8E-01	27.8	2.2E-01	26040
rs312750	17	68,343,539	*KCNJ2*	Winkler	A	a	g	0.599	0.15	0.17	3.8E-01	0	7.7E-01	26047
rs12940622	17	78,615,571	*RPTOR*	Locke**	G	g	a	0.659	0.25	0.18	1.6E-01	38.4	1.4E-01	26036
rs1808579	18	21,104,888	*C18orf8*	Speliotes	C	c	t	0.420	0.38	0.17	2.5E-02	0	6.4E-01	26031
rs7239883	18	40,147,671	*LOC284260*	Locke	G	g	a	0.319	0.05	0.18	7.7E-01	0	8.8E-01	26024
rs7243357	18	56,883,319	*GRP*	Locke	T	g	t	0.240	0.06	0.19	7.6E-01	0	7.1E-01	26048
rs2331841	18	57,828,637	*MC4R*	Okada	A	a	g	0.344	0.39	0.17	2.6E-02	0	5.3E-01	26041
rs6567160	18	57,829,135	*MC4R*	Speliotes, Pei	C	c	t	0.146	1.12	0.24	3.3E-06	53	4.7E-02	26047
rs591166	18	57,841,589	*MC4R*	Wen 2014	A	a	t	0.389	0.41	0.17	1.6E-02	0	5.2E-01	26045
rs9944545	18	57,958,244	*MC4R*	Locke	T	t	c	0.229	0.44	0.20	2.9E-02	0	9.5E-01	26046
rs17066842	18	58,040,624	*MC4R*	Locke	G	g	a	0.953	0.58	0.42	1.7E-01	14.7	3.2E-01	26035
rs17724992	19	18,454,825	*PGPEP1*	Locke	A	a	g	0.664	0.63	0.18	4.3E-04	0	6.0E-01	26045
rs17513613	19	30,286,822	*CCNE1*	Winkler	T	c	t	0.193	0.21	0.21	3.2E-01	30.4	2.0E-01	26048
rs29941	19	34,309,532	*KCTD15*	Speliotes	G	g	a	0.646	0.15	0.17	4.0E-01	0	7.0E-01	26043
rs2075650	19	45,395,619	*TOMM40*	Speliotes	A	a	g	0.898	0.35	0.27	2.1E-01	0	9.8E-01	26047
rs11671664	19	46,172,278	*GIPR, QPCTL*	Wen 2014	G	g	a	0.906	0.80	0.28	4.5E-03	46.5	8.2E-02	26046

rs2287019	19	46,202,172	*QPCTL*	Speliotes	C	c	t	0.871	0.64	0.25	9.1E-03	0	5.5E-01	26048
rs3810291	19	47,569,003	*ZC3H4*	Speliotes	A	a	g	0.529	0.56	0.17	8.2E-04	59.7	2.1E-02	26014
rs4802349	19	47,874,510	*DHX34*	Gong 2013	G	t	g	0.241	0.33	0.20	9.1E-02	0	8.1E-01	26023
rs8123881*	20	15,819,495	*MACROD 2*	Winkler	G	g	a	0.150	0.72	0.24	2.5E-03	42.8	1.1E-01	26045
rs6091540	20	51,087,862	*ZFP64*	Locke**	C	t	c	0.315	0.07	0.18	7.1E-01	0	5.6E-01	26045
rs2836754	21	40,291,740	*ETS2*	Locke	C	c	t	0.434	0.26	0.17	1.3E-01	0	9.1E-01	26038
rs4820408	22	40,604,945	*TNRC6B*	Winkler	T	t	g	0.338	0.04	0.18	8.1E-01	54.6	4.0E-02	26048

Abbreviations: A1=coded allele, A2=non-coded allele, Bp37=base pair Build 37, Chr=chromosome, FE=Fixed-Effect, HetP=heterogeneity p-value, P=p-value, Prep=prepared reference, Ref=reference, Sub=submitted reference, SNPs=single nucleotide polymorphisms.

*When the index SNP was not genotyped on the MetaboChip, the proxy SNP in tight linkage disequilibrium ($r^2 \geq 0.8$ in 1000 Genomes pilot 1 CEU, YRI, CHB+JPT depending on the population of discovery) with the lowest p-value in the African American sample was chosen to represent the index signal. The decreasing and increasing alleles for proxies were assigned assuming that the risk index SNP would have a similar allele frequency in the 1000 Genomes population (EUR, AFR, or EAS depending on the discovery population) as the risk proxy SNP.

**These loci were also described by Berndt *et al*. for obesity (maximum sample size of 263,407) [29]. The most recent BMI references per racial/ethnic group are noted above by their first author and publication year, if applicable [21, 25, 26, 30, 32, 33, 186, 195, 204].

***For GWAS SNPs a Bonferroni correction for multiple tests in the fixed-effect analyses reflected the number of independent previously-reported signals tested (=0.05/166). The 4 noted SNP pairs above were in tight linkage disequilibrium [$r^2 \geq 0.8$ in non-European 1000 genomes pilot populations(s)] with each other, but because they were reports from distinct discovery populations we retained them in this inventory in case they were population-specific variants. Therefore, our Bonferroni correction was penalized for only 166 (=170-4) tests.

APPENDIX M: ASIAN DESCENT GENETIC EFFECT ESTIMATES FOR 170 BMI INDEX SNPS FROM GWAS, INCLUDING SECONDARY SIGNALS IN EUROPEAN DESCENT POPULATIONS AND/OR POPULATION-SPECIFIC MARKERS.

rsID*	Chr	Bp37	Gene	Ref. First Author	Ref. Risk Allele	A 1	A 2	Freq	β (%)	SE (%)	P***	I^2	HetP	N
rs2803328	1	1,874,326	*KIAA1751*	Winkler	C	g	c	0.497	0.15	0.16	3.7E-01	21.5	2.6E-01	14209
rs2271928	1	32,127,953	*COL16A1*	Winkler	A	g	a	0.326	0.02	0.18	9.2E-01	0	7.1E-01	14225
rs2275426	1	46,487,552	*MAST2*	Winkler	A	a	g	0.581	0.17	0.17	3.3E-01	0	5.5E-01	14222
rs977747	1	47,684,677	*TAL1*	Locke	T	t	g	0.896	0.68	0.37	6.9E-02	0	6.3E-01	14223
rs657452	1	49,589,847	*AGBL4*	Locke	A	a	g	0.646	0.33	0.17	5.9E-02	22.1	2.6E-01	13928
rs11583200	1	50,559,820	*ELAVL4*	Locke	C	c	t	0.815	0.41	0.22	5.8E-02	0	8.8E-01	14226
rs3101336	1	72,751,185	*NEGR1*	Speliotes	C	c	t	0.910	0.46	0.24	5.4E-02	0	4.7E-01	22462
rs12566985	1	75,002,193	*FPGT-TNNI3K*	Speliotes	G	g	a	0.867	0.45	0.27	9.1E-02	0	8.5E-01	14174
rs12401738	1	78,446,761	*FUBP1*	Locke**	A	a	g	0.064	0.70	0.70	3.2E-01	0	6.2E-01	14225
rs11165643	1	96,924,097	*PTBP2*	Speliotes	T	c	t	0.263	0.14	0.19	4.5E-01	39.7	1.1E-01	14216
rs17024393	1	110,154,688	*GNAT2*	Locke**	C									
rs4357530*	1	151,103,153	*SEMA6C*	Winkler	G	g	a	0.119	0.55	0.26	3.5E-02	16.8	3.0E-01	14225
rs10913118	1	175,954,755	*RFWD2*	Winkler	A	a	c	0.558	0.10	0.17	5.3E-01	0	9.1E-01	14214
rs574367***	1	177,873,210	*SEC16B*	Wen 2014	T	t	g	0.195	0.61	0.17	4.3E-04	0	6.7E-01	22452
rs543874***	1	177,889,480	*SEC16B*	Speliotes, Monda	G	g	a	0.195	0.61	0.17	4.2E-04	0	6.5E-01	22464
rs10920678	1	190,239,907	*FAM5C*	Winkler	A	a	g	0.351	0.60	0.18	6.1E-04	40.5	1.1E-01	13789
rs2820292	1	201,784,287	*NAV1*	Locke	C	c	a	0.269	0.25	0.19	1.9E-01	0	4.7E-01	14224
rs2820436	1	219,640,680	*LYPLAL1*	Gong (sub.)	A	a	c	0.195	0.21	0.21	3.2E-01	24.2	2.4E-01	14210
rs12463617***	2	629,244	*TMEM18*	Wen 2014	C	c	a	0.914	0.61	0.24	1.0E-02	0	7.0E-01	22465
rs13021737***	2	632,348	*TMEM18*	Speliotes	G	g	a	0.913	0.61	0.24	1.0E-02	0	6.3E-01	22417
rs11676272*,***	2	25,141,538	*ADCY3*	Wen 2014	G	g	a	0.474	0.50	0.16	2.2E-03	18.8	2.8E-01	14224
rs10182181***	2	25,150,296	*ADCY3*	Speliotes	G	g	a	0.477	0.51	0.16	1.9E-03	10.3	3.5E-01	14218
rs11126666	2	26,928,811	*KCNK3*	Locke	A	g	a	0.349	0.07	0.18	7.0E-01	0	8.6E-01	14225
rs116612809	2	28,301,171	*BRE*	Gong 2013	G									
rs1016287	2	59,305,625	*FLJ30838*	Speliotes	T	t	c	0.189	0.06	0.21	7.7E-01	50.9	4.7E-02	14202
rs11688816	2	63,053,048	*EHBP1*	Locke	G	g	a	0.740	0.30	0.19	1.1E-01	0	9.9E-01	14226
rs12622013*	2	79,501,362	*REG3A*	Winkler	G	g	a	0.181	0.44	0.22	4.7E-02	0	7.0E-01	14226
rs7570971	2	135,837,906	*RAB3GAP1*	Winkler	A	c	a	0.109	0.84	0.61	1.7E-01	28.9	2.0E-01	14224
rs4988235	2	136,608,646	*MCM6*	Winkler	A	a	g	0.110	0.88	0.60	1.5E-01	39.8	1.1E-01	14225
rs2121279	2	143,043,285	*LRP1B*	Speliotes	T									
rs1460676	2	164,567,689	*FIGN*	Locke	C	c	t	0.368	0.23	0.17	1.9E-01	0	7.1E-01	14224
rs10184004	2	165,508,389	*GRB14/COBLL1*	Gong (sub.)	T	t	c	0.103	0.48	0.29	9.2E-02	0	8.9E-01	14224
rs10930502	2	172,890,588	*METAP1D*	Gong (sub.)	A	a	g	0.333	0.57	0.18	1.3E-03	46.9	6.8E-02	14220
rs1528435	2	181,550,962	*UBE2E3*	Locke	T	t	c	0.699	0.05	0.18	7.8E-01	15.2	3.1E-01	14226
rs972540	2	207,244,783	*ADAM23*	Winkler	A	g	a	0.182	0.11	0.21	5.9E-01	0	7.6E-01	14226
rs17203016	2	208,255,518	*CREB1*	Locke	G	g	a	0.204	0.18	0.21	3.9E-01	17.1	2.9E-01	14214
rs7599312	2	213,413,231	*ERBB4*	Locke	G	g	a	0.934	0.11	0.38	7.8E-01	0	5.3E-01	14223
rs492400	2	219,349,752	*USP37*	Locke	C	c	t	0.239	0.37	0.20	6.2E-02	0	9.8E-01	14035
rs2176040	2	227,092,802	*LOC646736*	Speliotes	A	a	g	0.099	0.21	0.28	4.6E-01	0	8.2E-01	14226
rs9845966	3	13,433,158	*NUP210*	Winkler	T	t	g	0.577	0.31	0.17	6.0E-02	0	4.9E-01	14220
rs6804842	3	25,106,437	*RARB*	Locke	G	g	a	0.612	0.36	0.17	3.5E-02	0	6.7E-01	14223
rs7613875	3	49,971,514	*MON1A*	Winkler	A	a	c	0.202	0.15	0.22	4.8E-01	0	7.9E-01	14192
rs2365389	3	61,236,462	*FHIT*	Locke	C	c	t	0.128	0.31	0.27	2.5E-01	53.1	3.7E-02	14224
rs333495*	3	78,834,343	*ROBO1*	Winkler	G	g	t	0.280	0.60	0.18	1.2E-03	0	9.6E-01	14223
rs13078960	3	85,807,590	*CADM2*	Speliotes	G	g	t	0.035	0.66	0.72	3.6E-01	0	7.3E-01	22466
rs1720825	3	138,108,083	*MRAS*	Graff (prep.)	A	a	g	0.037	0.13	0.59	8.3E-01	0	4.9E-01	14226
rs2640017*	3	141,335,121	*RASA2*	Locke	G	g	a	0.275	0.35	0.19	6.8E-02	0	5.1E-01	14223
rs11927381	3	185,508,591	*IGF2BP2*	Gong (sub.)	T	t	c	0.672	0.36	0.18	4.2E-02	39.6	1.2E-01	14201
rs1516725	3	185,824,004	*ETV5*	Speliotes	C	c	t	0.930	0.52	0.31	9.3E-02	0	5.9E-01	22446
rs16992647	4	36,813,105	*KIAA1239*	Winkler	T	t	c	0.460	0.00	0.17	9.9E-01	67	3.5E-03	14225
rs16858082	4	45,175,804	*GNPDA2*	Wen 2014	T	t	c	0.345	0.42	0.14	3.0E-03	12.5	3.3E-01	22441
rs10938397	4	45,182,527	*GNPDA2*	Speliotes	G	g	a	0.279	0.55	0.15	2.6E-04	24	2.3E-01	22386
rs348495	4	45,184,442	*GNPDA2*	Monda	G	g	a	0.364	0.42	0.17	1.3E-02	0	6.1E-01	15429
rs13107325	4	103,188,709	*SLC39A8*	Speliotes	T									
rs11727676	4	145,659,064	*HHIP*	Locke	T	t	c	0.984	0.30	1.24	8.1E-01	11.2	3.4E-01	13682
rs2112347	5	75,015,242	*POC5*	Speliotes	T	t	g	0.443	0.45	0.14	9.0E-04	10.3	3.5E-01	22464
rs6870983	5	87,697,533	*TMEM161B-AS1*	Winkler	T	t	c	0.057	0.44	0.73	5.5E-01	16.6	3.0E-01	13682
rs11951673*	5	95,861,012	*PCSK1*	Wen 2014	C	c	t	0.423	0.73	0.17	1.7E-05	34	1.6E-01	14221
rs6864049	5	124,330,522	*ZNF608*	Winkler	A	g	a	0.677	0.24	0.18	1.8E-01	21.1	2.6E-01	14206
rs13174863	5	139,080,745	*CXXC5*	Winkler	A	g	a	0.066	0.07	0.34	8.3E-01	0	9.6E-01	14226
rs4569924*	5	153,540,025	*GALNT10*	Monda	T	c	t	0.174	0.16	0.35	6.6E-01	0	6.4E-01	14223
rs2228213	6	12,124,855	*HIVEP1*	Winkler	A	g	a	0.734	0.37	0.19	5.2E-02	0	5.6E-01	14226
rs9356744	6	20,685,486	*CDKAL1*	Wen 2014	T	t	c	0.595	0.86	0.14	5.3E-10	42.6	8.4E-02	22461
rs943466	6	33,731,787	*LEMD2*	Winkler	A	g	a	0.866	0.11	0.25	6.6E-01	42	9.9E-02	14226
rs205262	6	34,563,164	*C6orf106*	Speliotes	G	a	g	0.857	0.06	0.27	8.4E-01	37.4	1.3E-01	14223
rs2033529	6	40,348,653	*TDRG1*	Locke	G	g	a	0.284	0.10	0.19	6.0E-01	0	4.7E-01	14223
rs2207139	6	50,845,490	*TFAP2B*	Speliotes	G	g	a	0.210	0.29	0.17	8.5E-02	0	6.0E-01	22464
rs9400239	6	108,977,663	*FOXO3*	Locke	C	c	t	0.708	0.19	0.19	3.2E-01	18.2	2.9E-01	14224
rs9374842	6	120,185,665	*LOC285762*	Locke	T	c	t	0.176	0.27	0.22	2.3E-01	25.2	2.3E-01	14217
rs13201877	6	137,675,541	*IFNGR1*	Locke	G	a	g	0.944	0.50	0.36	1.7E-01	0	6.9E-01	14226
rs1281962	6	153,431,376	*RGS17*	Winkler	C	c	g	0.820	0.18	0.22	4.1E-01	0	4.8E-01	14222
rs3127574	6	160,791,370	*SLC22A3*	Winkler	C	c	g	0.433	0.06	0.17	7.2E-01	0	7.2E-01	14223
rs13191362	6	163,033,350	*PARK2*	Locke	A									

rs10499694*	7	50,614,173	*DDC*	Winkler	G	g	a	0.399	0.06	0.17	7.3E-01	0	8.2E-01	14226
rs1167827	7	75,163,169	*HIP1*	Locke	G	g	a	0.104	0.05	0.36	8.9E-01	0	4.7E-01	14224
rs6465468	7	95,169,514	*ASB4*	Locke	T	t	g	0.068	0.65	0.34	5.5E-02	36.2	1.4E-01	14222
rs6990042	8	14,173,974	*SGCZ*	Winkler	T	g	t	0.264	0.28	0.19	1.3E-01	0	9.2E-01	14225
rs7844647*	8	34,503,776	*Intergenic*	Winkler	T	t	c	0.506	0.53	0.17	1.2E-03	0	5.0E-01	14225
rs17405819	8	76,806,584	*HNF4G*	Locke**	T	t	c	0.609	0.11	0.17	5.3E-01	0	8.5E-01	14222
rs16907751	8	81,375,457	*ZBTB10*	Locke	C	c	t	0.853	0.10	0.24	6.7E-01	58.8	1.7E-02	14224
rs2033732	8	85,079,709	*RALYL*	Locke	C	c	t	0.723	0.29	0.19	1.3E-01	0	5.0E-01	14085
rs4740619	9	15,634,326	*C9orf93*	Locke	T	t	c	0.248	0.06	0.19	7.7E-01	0	9.9E-01	14223
rs10968576	9	28,414,339	*LINGO2*	Speliotes	G	g	a	0.180	0.32	0.18	7.2E-02	45.2	6.8E-02	22465
rs6477694	9	111,932,342	*EPB41L4B*	Locke	C	c	t	0.592	0.17	0.17	3.0E-01	46.8	6.9E-02	14226
rs1928295	9	120,378,483	*TLR4*	Locke	T	t	c	0.581	0.43	0.17	1.2E-02	46.6	7.0E-02	14219
rs10733682	9	129,460,914	*LMX1B*	Locke	A	g	a	0.273	0.21	0.19	2.7E-01	7.9	3.7E-01	14225
rs2270204	9	131,042,734	*SWI5*	Winkler	T	g	t	0.566	0.16	0.17	3.4E-01	0	8.7E-01	14020
rs7899106	10	87,410,904	*GRID1*	Locke	G									
rs17094222	10	102,395,440	*HIF1AN*	Locke	C	c	t	0.324	0.13	0.18	4.8E-01	2.4	4.1E-01	14224
rs11191560	10	104,869,038	*NT5C2*	Locke, Wen 2012	C	c	t	0.270	0.45	0.15	4.0E-03	25	2.2E-01	22458
rs7903146	10	114,758,349	*TCF7L2*	Locke	C	c	t	0.934	1.49	0.32	4.6E-06	50.6	4.0E-02	22465
rs10886017	10	118,672,531	*KIAA1598*	Winkler	A	a	c	0.364	0.23	0.17	1.7E-01	35.9	1.4E-01	14223
rs2237897	11	2,858,546	*KCNQ1*	Wen 2014	T	t	c	0.353	0.73	0.18	3.5E-05	62.1	9.9E-03	14181
rs4256980	11	8,673,939	*TRIM66*	Speliotes	G	g	c	0.398	0.35	0.14	1.0E-02	0	7.4E-01	22462
rs7928810	11	17,372,443	*NCR3LG1*	Winkler	A	a	c	0.682	0.22	0.18	2.2E-01	0	5.0E-01	14222
rs1557765	11	17,403,639	*KCNJ11*	Winkler	T	c	t	0.665	0.28	0.18	1.1E-01	17.2	2.9E-01	14222
rs11030104	11	27,684,517	*BDNF*	Speliotes, Wen 2014	A	a	g	0.566	0.10	0.14	4.5E-01	61.9	7.2E-03	22465
rs10835210	11	27,695,910	*BDNF*	Locke	C	c	a	0.664	0.02	0.17	9.0E-01	46.4	6.1E-02	16161
rs652722	11	31,905,534	*PAX6*	Wen 2012	C									
rs2176598	11	43,864,278	*HSD17B12*	Locke	T	c	t	0.863	0.09	0.25	7.1E-01	1.5	4.2E-01	14225
rs3817334	11	47,650,993	*MTCH2*	Speliotes	T	t	c	0.312	0.23	0.14	1.2E-01	0	8.7E-01	22447
rs1865732*	11	112,960,722	*NCAM1*	Winkler	C	c	t	0.161	0.03	0.24	9.1E-01	0	5.3E-01	14225
rs12286929	11	115,022,404	*CADM1*	Locke	G	g	a	0.243	0.27	0.20	1.7E-01	0	9.4E-01	14214
rs11611246	12	939,480	*WNK1*	Winkler	T	t	g	0.288	0.08	0.19	6.9E-01	5	3.9E-01	14034
rs7970953	12	24,075,508	*SOX5*	Winkler	A	a	g	0.692	0.23	0.18	2.1E-01	0	6.0E-01	14202
rs1405552	12	41,746,673	*PDZRN4*	Winkler	A	g	a	0.435	0.30	0.17	7.6E-02	0	4.3E-01	14223
rs11181001	12	41,948,196	*PDZRN4*	Winkler	A	a	g	0.390	0.36	0.17	3.6E-02	0	7.1E-01	14218
rs7138803	12	50,247,468	*BCDIN3D*	Speliotes	A	a	g	0.309	0.28	0.15	5.4E-02	35.7	1.3E-01	22463
rs1438994*	12	90,594,389	*Intergenic*	Winkler	T	t	c	0.372	0.10	0.17	5.6E-01	0	6.2E-01	14226
rs11065987	12	112,072,424	*BRAP*	Winkler	A	g	a	0.078	0.60	0.66	3.7E-01	67.9	2.7E-03	14224
rs17630235	12	112,591,686	*TRAFD1*	Winkler	A	a	g	0.076	0.65	0.66	3.3E-01	71.4	9.3E-04	14140
rs11057405	12	122,781,897	*CLIP1*	Locke	G	g	a	0.984	0.49	1.21	6.8E-01	17.9	2.9E-01	14137
rs1885988*	13	28,010,262	*MTIF3*	Speliotes	C	t	c	0.825	0.29	0.22	2.0E-01	0	7.4E-01	14226
rs12429545	13	54,102,206	*OLFM4*	Speliotes	A	a	g	0.208	0.44	0.20	3.1E-02	0	5.8E-01	14221
rs9540493	13	66,205,704	*MIR548X2*	Locke	A	g	a	0.280	0.08	0.19	6.8E-01	50.9	4.7E-02	14223
rs1441264	13	79,580,919	*MIR548A2*	Locke	A	a	g	0.543	0.29	0.17	7.9E-02	0	6.2E-01	14226
rs9634489	13	97,049,004	*HS6ST3*	Winkler	A	g	a	0.494	0.15	0.17	3.5E-01	0	4.7E-01	14225
rs10132280	14	25,928,179	*STXBP6*	Locke	C	a	c	0.088	0.08	0.30	8.0E-01	0	6.6E-01	14225
rs12885454	14	29,736,838	*PRKD1*	Locke	C	c	a	0.502	0.21	0.17	2.1E-01	57.1	2.2E-02	14225
rs11847697	14	30,515,112	*PRKD1*	Speliotes	T	c	t	0.962	0.46	0.93	6.2E-01	0	7.4E-01	14444
rs17522122	14	33,302,882	*AKAP6*	Winkler	T	t	g	0.379	0.26	0.17	1.3E-01	29.2	2.0E-01	14121
rs7141420	14	79,899,454	*NRXN3*	Speliotes	T	c	t	0.621	0.17	0.17	3.1E-01	0	7.6E-01	14223
rs3783890	14	93,790,276	*BTBD7*	Winkler	T	t	c	0.635	0.28	0.17	1.1E-01	0	4.4E-01	14223
rs7143963	14	103,304,425	*TRAF3*	Winkler	T	t	c	0.415	0.10	0.17	5.8E-01	0	5.3E-01	14222
rs709400	14	104,149,475	*KLC1*	Winkler	A	a	g	0.870	0.08	0.25	7.6E-01	0	6.8E-01	14222
rs3736485	15	51,748,610	*DMXL2*	Locke	A	a	g	0.753	0.16	0.19	4.1E-01	0	5.3E-01	14225
rs16951275	15	68,077,168	*MAP2K5*	Speliotes	T	t	c	0.407	0.16	0.14	2.5E-01	2	4.2E-01	22462
rs4776970	15	68,080,886	*MAP2K5*	Wen 2012	A	a	t	0.238	0.36	0.16	2.3E-02	30.6	1.7E-01	22464
rs7164727	15	73,093,991	*LOC100287559*	Locke	T	c	t	0.766	0.05	0.20	7.9E-01	13.5	3.2E-01	14224
rs7181659	15	95,267,483	*MCTP2*	Winkler	A	a	g	0.450	0.23	0.17	1.8E-01	29.7	1.9E-01	14194
rs11866815	16	387,867	*AXIN1*	Winkler	T	c	t	0.804	0.01	0.21	9.7E-01	15.1	3.1E-01	14188
rs12446632	16	19,935,389	*GPRC5B*	Speliotes	G									
rs11074446	16	20,255,123	*GP2*	Locke	T	t	c	0.828	0.70	0.22	1.7E-03	16.1	3.0E-01	14212
rs2650492	16	28,333,411	*SBK1*	Locke	A	a	g	0.082	0.94	0.28	7.0E-04	33.8	1.5E-01	22457
rs3888190	16	28,889,486	*ATP2A1*	speliotes	A	a	c	0.151	0.37	0.20	5.8E-02	57.2	1.7E-02	22463
rs4787491	16	30,015,337	*INO80E*	Locke	G	g	a	0.348	0.19	0.18	2.8E-01	0	6.9E-01	14220
rs9925964	16	31,129,895	*KAT8*	Locke	A	g	a	0.838	0.16	0.25	5.3E-01	0	9.1E-01	14226
rs2080454	16	49,062,590	*CBLN1*	Locke	C	c	a	0.506	0.36	0.17	2.8E-02	0	6.9E-01	14223
rs1558902***	16	53,803,574	*FTO*	Speliotes, Wen 2014	A	a	t	0.201	1.73	0.20	1.2E-17	0	8.0E-01	16185
rs17817964***	16	53,828,066	*FTO*	Monda	T	t	c	0.223	1.28	0.17	2.2E-14	0	4.9E-01	22465
rs889398	16	69,556,715	*NFAT5*	Winkler	T	c	t	0.800	0.09	0.23	6.8E-01	0	9.2E-01	14084
rs9914578	17	2,005,136	*SMG6*	Locke	G	c	g	0.802	0.16	0.21	4.4E-01	22.3	2.5E-01	14224
rs1000940	17	5,283,252	*RABEP1*	Locke	G	g	a	0.600	0.07	0.17	6.9E-01	0	5.0E-01	14225
rs4986044	17	21,261,560	*KCNJ12*	Winkler	T	c	t	0.369	0.25	0.17	1.4E-01	0	8.8E-01	14221
rs12150665	17	34,914,787	*GGNBP2*	Winkler	T	t	c	0.625	0.36	0.17	3.8E-02	0	7.0E-01	14221
rs11652097	17	45,316,717	*ITGB3*	Winkler	T	c	t	0.619	0.29	0.17	8.6E-02	0	4.5E-01	14182
rs6504108	17	46,292,923	*SKAP1*	Winkler	T	c	t	0.204	0.35	0.21	8.6E-02	60.9	1.2E-02	14222
rs8075273*	17	61,728,881	*unknown*	Winkler	C	a	c	0.087	0.10	0.33	7.6E-01	25.4	2.3E-01	13835
rs312750	17	68,343,539	*KCNJ2*	Winkler	A	g	a	0.340	0.02	0.18	9.3E-01	0	7.0E-01	14224
rs12940622	17	78,615,571	*RPTOR*	Locke**	G	g	a	0.665	0.49	0.18	5.7E-03	0	5.1E-01	14224
rs1808579	18	21,104,888	*C18orf8*	Speliotes	C	t	c	0.662	0.61	0.18	5.7E-04	0	5.7E-01	14224
rs7239883	18	40,147,671	*LOC284260*	Locke	G	g	a	0.277	0.08	0.19	6.9E-01	33.8	1.6E-01	14198
rs7243357	18	56,883,319	*GRP*	Locke	T	t	g	0.865	0.44	0.25	7.8E-02	0	4.4E-01	14224
rs2331841	18	57,828,637	*MC4R*	Okada	A	a	g	0.229	0.46	0.16	4.1E-03	0	6.0E-01	22447
rs6567160	18	57,829,135	*MC4R*	Speliotes, Pei	C	c	t	0.197	0.67	0.17	9.0E-05	0	5.8E-01	22461
rs591166	18	57,841,589	*MC4R*	Wen 2014	A	a	t	0.231	0.43	0.16	6.8E-03	0	7.7E-01	22455

rs9944545	18	57,958,244	*MC4R*	Locke	T	t	c	0.133	0.03	0.21	8.9E-01	47.4	5.5E-02	22458
rs17066842	18	58,040,624	*MC4R*	Locke	G	g	a	0.980	0.12	0.49	8.1E-01	0	6.3E-01	22148
rs17724992	19	18,454,825	*PGPEP1*	Locke	A	g	a	0.472	0.06	0.17	7.3E-01	24.5	2.3E-01	14219
rs17513613	19	30,286,822	*CCNE1*	Winkler	T	c	t	0.109	0.11	0.28	6.9E-01	0	9.4E-01	14226
rs29941	19	34,309,532	*KCTD15*	Speliotes	G	g	a	0.238	0.23	0.16	1.5E-01	7.8	3.7E-01	22456
rs2075650	19	45,395,619	*TOMM40*	Speliotes	A	g	a	0.169	0.09	0.23	6.8E-01	0	5.2E-01	14224
rs11671664	19	46,172,278	*GIPR, QPCTL*	Wen 2014	G	g	a	0.531	0.57	0.14	4.2E-05	0	8.4E-01	22460
rs2287019	19	46,202,172	*QPCTL*	Speliotes	C	c	t	0.789	0.56	0.17	6.7E-04	0	5.7E-01	22465
rs3810291	19	47,569,003	*ZC3H4*	Speliotes	A									
rs4802349	19	47,874,510	*DHX34*	Gong 2013	G	g	t	0.613	0.12	0.17	4.8E-01	0	6.9E-01	14193
rs8123881*	20	15,819,495	*MACROD2*	Winkler	G	a	g	0.906	0.15	0.28	6.0E-01	34.7	1.5E-01	14218
rs6091540	20	51,087,862	*ZFP64*	Locke**	C	c	t	0.643	0.33	0.17	5.5E-02	0	5.3E-01	14223
rs2836754	21	40,291,740	*ETS2*	Locke	C	c	t	0.272	0.06	0.19	7.7E-01	0	7.9E-01	14219
rs4820408	22	40,604,945	*TNRC6B*	Winkler	T	t	g	0.534	0.40	0.17	1.6E-02	36.4	1.4E-01	14220

Abbreviations: A1=coded allele, A2=non-coded allele, Bp37=base pair Build 37, Chr=chromosome, FE=Fixed-Effect, HetP=heterogeneity p-value, P=p-value, Prep=prepared reference, Ref=reference, Sub=submitted reference, SNPs=single nucleotide polymorphisms.

*When the index SNP was not genotyped on the MetaboChip, the proxy SNP in tight linkage disequilibrium ($r^2 \geq 0.8$ in 1000 Genomes pilot 1 CEU, YRI, CHB+JPT depending on the population of discovery) with the lowest p-value in the African American sample was chosen to represent the index signal. The decreasing and increasing alleles for proxies were assigned assuming that the risk index SNP would have a similar allele frequency in the 1000 Genomes population (EUR, AFR, or EAS depending on the discovery population) as the risk proxy SNP.

**These loci were also described by Berndt *et al*. for obesity (maximum sample size of 263,407) [29]. The most recent BMI references per racial/ethnic group are noted above by their first author and publication year, if applicable [21, 25, 26, 30, 32, 33, 186, 195, 204].

***For GWAS SNPs a Bonferroni correction for multiple tests in the fixed-effect analyses reflected the number of independent previously-reported signals tested (=0.05/166). The 4 noted SNP pairs above were in tight linkage disequilibrium [$r^2 \geq 0.8$ in non-European 1000 genomes pilot populations(s)] with each other, but because they were reports from distinct discovery populations we retained them in this inventory in case they were population-specific variants. Therefore, our Bonferroni correction was penalized for only 166 (=170-4) tests.

APPENDIX N: EUROPEAN DESCENT GENETIC EFFECT ESTIMATES FOR 170 BMI INDEX SNPS FROM GWAS, INCLUDING SECONDARY SIGNALS IN EUROPEAN DESCENT POPULATIONS AND/OR POPULATION-SPECIFIC MARKERS.

rsID*	Chr	Bp37	Gene	Ref. First Author	Ref. Risk Allele	A 1	A 2	Freq	β (%)	SE (%)	P***	I^2	HetP	N
rs2803328	1	1,874,326	*KIAA1751*	Winkler	C	c	g	0.527	0.00	0.23	1.0E+00	0	3.7E-01	17859
rs2271928	1	32,127,953	*COL16A1*	Winkler	A	g	a	0.599	0.36	0.20	8.0E-02	0	8.9E-01	17859
rs2275426	1	46,487,552	*MAST2*	Winkler	A	a	g	0.439	0.11	0.20	5.6E-01	0	7.5E-01	17859
rs977747	1	47,684,677	*TAL1*	Locke	T	t	g	0.403	0.13	0.20	5.2E-01	23.7	2.5E-01	17859
rs657452	1	49,589,847	*AGBL4*	Locke	A									
rs11583200	1	50,559,820	*ELAVL4*	Locke	C	c	t	0.387	0.14	0.20	4.9E-01	5.4	3.0E-01	17859
rs3101336	1	72,751,185	*NEGR1*	Speliotes	C	c	t	0.625	0.61	0.20	3.0E-03	0	4.4E-01	17859
rs12566985	1	75,002,193	*FPGT-TNNI3K*	Speliotes	G									
rs12401738	1	78,446,761	*FUBP1*	Locke**	A	g	a	0.651	0.31	0.21	1.4E-01	0	5.9E-01	17859
rs11165643	1	96,924,097	*PTBP2*	Speliotes	T	t	c	0.584	0.01	0.20	9.8E-01	0	9.3E-01	17859
rs17024393	1	110,154,688	*GNAT2*	Locke**	C	c	t	0.029	1.48	0.61	1.5E-02	83.6	1.3E-02	17859
rs4357530*	1	151,103,153	*SEMA6C*	Winkler	G	g	a	0.318	0.26	0.21	2.3E-01	0	7.6E-01	17859
rs10913118	1	175,954,755	*RFWD2*	Winkler	A	c	a	0.333	0.06	0.21	7.8E-01	0	4.7E-01	17859
rs574367* **	1	177,873,210	*SEC16B*	Wen 2014	T	t	g	0.195	0.71	0.25	4.8E-03	0	5.5E-01	17858
rs543874* **	1	177,889,480	*SEC16B*	Speliotes, Monda	G	g	a	0.191	0.75	0.25	2.8E-03	0	5.6E-01	17859
rs10920678	1	190,239,907	*FAM5C*	Winkler	A	a	g	0.424	0.42	0.19	2.9E-02	63.7	9.7E-02	17858
rs2820292	1	201,784,287	*NAV1*	Locke	C	c	a	0.545	0.20	0.19	3.1E-01	0	9.7E-01	17858
rs2820436	1	219,640,680	*LYPLAL1*	Gong (sub.)	A	a	c	0.324	0.55	0.21	9.8E-03	0	3.6E-01	17859
rs12463617* **	2	629,244	*TMEM18*	Wen 2014	C	c	a	0.825	1.07	0.26	5.2E-05	33.4	2.2E-01	17859
rs13021737* **	2	632,348	*TMEM18*	Speliotes	G	g	a	0.825	1.07	0.25	3.0E-05	34.4	2.2E-01	17858
rs11676272*,* **	2	25,141,538	*ADCY3*	Wen 2014	G	g	a	0.472	0.51	0.19	8.4E-03	0	3.4E-01	17859
rs10182181* **	2	25,150,296	*ADCY3*	Speliotes	G	g	a	0.473	0.50	0.19	1.1E-02	0	4.1E-01	17859
rs11126666	2	26,928,811	*KCNK3*	Locke	A	g	a	0.734	0.14	0.23	5.4E-01	0	4.2E-01	17859
rs116612809	2	28,301,171	*BRE*	Gong 2013	G	a	g	0.998	5.18	3.51	1.4E-01	0	4.1E-01	10048
rs1016287	2	59,305,625	*FLJ30838*	Speliotes	T	t	c	0.296	0.22	0.21	3.1E-01	0	9.3E-01	17859
rs11688816	2	63,053,048	*EHBP1*	Locke	G	g	a	0.512	0.14	0.19	4.7E-01	0	3.9E-01	17859
rs12622013*	2	79,501,362	*REG3A*	Winkler	G	a	g	0.890	0.12	0.31	7.1E-01	0	5.7E-01	17859
rs7570971	2	135,837,906	*RAB3GAP1*	Winkler	A									
rs4988235	2	136,608,646	*MCM6*	Winkler	A									
rs2121279	2	143,043,285	*LRP1B*	Speliotes	T	t	c	0.129	0.47	0.29	1.1E-01	60.2	1.1E-01	17858
rs1460676	2	164,567,689	*FIGN*	Locke	C	c	t	0.165	0.42	0.26	1.1E-01	0	8.9E-01	17859
rs10184004	2	165,508,389	*GRB14/C OBLL1*	Gong (sub.)	T	t	c	0.404	0.43	0.20	3.7E-02	0	9.0E-01	17859
rs10930502	2	172,890,588	*METAP1D*	Gong (sub.)	A	a	g	0.697	0.56	0.21	8.9E-03	0	9.9E-01	17859
rs1528435	2	181,550,962	*UBE2E3*	Locke	T	t	c	0.623	0.44	0.20	3.0E-02	51.7	1.5E-01	17859
rs972540	2	207,244,783	*ADAM23*	Winkler	A	a	g	0.728	0.04	0.22	8.7E-01	28.6	2.4E-01	17859
rs17203016	2	208,255,518	*CREB1*	Locke	G	g	a	0.186	0.31	0.25	2.2E-01	64.4	9.4E-02	17858
rs7599312	2	213,413,231	*ERBB4*	Locke	G	g	a	0.736	0.63	0.22	4.6E-03	0	8.9E-01	17859
rs492400	2	219,349,752	*USP37*	Locke	C	c	t	0.427	0.34	0.21	1.1E-01	68.7	7.4E-02	17841
rs2176040	2	227,092,802	*LOC646736*	Speliotes	A	a	g	0.351	0.41	0.20	4.5E-02	0	5.6E-01	17859
rs9845966	3	13,433,158	*NUP210*	Winkler	T	t	g	0.449	0.16	0.19	4.2E-01	45.9	1.7E-01	17859
rs6804842	3	25,106,437	*RARB*	Locke	G	g	a	0.584	0.05	0.19	8.0E-01	0	8.2E-01	17858
rs7613875	3	49,971,514	*MON1A*	Winkler	A									
rs2365389	3	61,236,462	*FHIT*	Locke	C	c	t	0.592	0.28	0.20	1.8E-01	0	7.8E-01	17859
rs333495*	3	78,834,343	*ROBO1*	Winkler	G	g	t	0.415	0.29	0.20	1.6E-01	0	9.5E-01	17859
rs13078960	3	85,807,590	*CADM2*	Speliotes	G	t	g	0.792	0.07	0.24	7.8E-01	30.3	2.3E-01	17859
rs1720825	3	138,108,083	*MRAS*	Graff (prep.)	A	a	g	0.197	0.18	0.25	4.8E-01	0	4.2E-01	17859
rs2640017*	3	141,335,121	*RASA2*	Locke	G	g	a	0.064	0.84	0.42	4.4E-02	0	8.4E-01	17859
rs11927381	3	185,508,591	*IGF2BP2*	Gong (sub.)	T	t	c	0.684	0.50	0.21	2.0E-02	0	9.5E-01	17857
rs1516725	3	185,824,004	*ETV5*	Speliotes	C	c	t	0.866	0.77	0.29	8.3E-03	63.8	9.6E-02	17859
rs16992647	4	36,813,105	*KIAA1239*	Winkler	T	c	t	0.840	0.25	0.27	3.5E-01	82.8	1.6E-02	17859
rs16858082	4	45,175,804	*GNPDA2*	Wen 2014	T	t	c	0.587	0.40	0.20	4.8E-02	0	3.4E-01	17858
rs10938397	4	45,182,527	*GNPDA2*	Speliotes	G	g	a	0.434	0.36	0.19	6.8E-02	0	5.9E-01	17859
rs348495	4	45,184,442	*GNPDA2*	Monda	G									
rs13107325	4	103,188,709	*SLC39A8*	Speliotes	T	t	c	0.075	1.17	0.39	3.0E-03	20.8	2.6E-01	17859
rs11727676	4	145,659,064	*HHIP*	Locke	T	t	c	0.905	0.58	0.37	1.2E-01	0	6.7E-01	17859
rs2112347	5	75,015,242	*POC5*	Speliotes	T	t	g	0.640	0.26	0.20	2.1E-01	86.3	6.9E-03	17859
rs6870983	5	87,697,533	*TMEM161B−AS1*	Winkler	T	c	t	0.770	0.09	0.23	7.1E-01	0	5.5E-01	17859
rs11951673*	5	95,861,012	*PCSK1*	Wen 2014	C	c	t	0.614	0.35	0.20	8.5E-02	0	8.2E-01	17859
rs6864049	5	124,330,522	*ZNF608*	Winkler	A	g	a	0.527	0.21	0.19	2.7E-01	62	1.0E-01	17859
rs13174863	5	139,080,745	*CXXC5*	Winkler	A	g	a	0.153	0.03	0.30	9.3E-01	0	4.1E-01	17859
rs4569924*	5	153,540,025	*GALNT10*	Monda	T	c	t	0.570	0.01	0.19	9.7E-01	0	7.7E-01	17859
rs2228213	6	12,124,855	*HIVEP1*	Winkler	A	g	a	0.655	0.12	0.20	5.6E-01	0	5.2E-01	17859
rs9356744	6	20,685,486	*CDKAL1*	Wen 2014	T	t	c	0.690	0.06	0.21	7.9E-01	0	5.8E-01	17859
rs943466	6	33,731,787	*LEMD2*	Winkler	A	g	a	0.770	0.08	0.23	7.4E-01	0	6.2E-01	17859
rs205262	6	34,563,164	*C6orf106*	Speliotes	G	g	a	0.284	0.27	0.22	2.2E-01	0	7.0E-01	17859
rs2033529	6	40,348,653	*TDRG1*	Locke	G	g	a	0.282	0.20	0.21	3.5E-01	0	7.0E-01	17859
rs2207139	6	50,845,490	*TFAP2B*	Speliotes	G	g	a	0.172	0.49	0.26	6.1E-02	0	9.7E-01	17859
rs9400239	6	108,977,663	*FOXO3*	Locke	C	c	t	0.699	0.35	0.21	1.0E-01	0	8.8E-01	17858
rs9374842	6	120,185,665	*LOC285762*	Locke	T	t	c	0.765	0.19	0.23	4.1E-01	85.2	9.3E-03	17858
rs13201877	6	137,675,541	*IFNGR1*	Locke	G	g	a	0.136	0.18	0.29	5.4E-01	0	7.8E-01	17859
rs1281962	6	153,431,376	*RGS17*	Winkler	C	c	g	0.529	0.32	0.19	9.8E-02	72.7	5.6E-02	17859
rs3127574	6	160,791,370	*SLC22A3*	Winkler	C	g	c	0.518	0.07	0.19	7.3E-01	0	9.7E-01	17859
rs13191362	6	163,033,350	*PARK2*	Locke	A	a	g	0.883	0.71	0.30	1.9E-02	76.2	4.0E-02	17859
rs10499694*	7	50,614,173	*DDC*	Winkler	G	a	g	0.482	0.53	0.19	6.2E-03	0	8.7E-01	17859
rs1167827	7	75,163,169	*HIP1*	Locke	G	g	a	0.561	0.36	0.20	7.8E-02	0	4.2E-01	17859
rs6465468	7	95,169,514	*ASB4*	Locke	T									
rs6990042	8	14,173,974	*SGCZ*	Winkler	T	g	t	0.477	0.20	0.20	3.4E-01	42.7	1.9E-01	17859
rs7844647*	8	34,503,776	*Intergenic*	Winkler	T	t	c	0.738	0.07	0.23	7.7E-01	90.4	1.3E-03	17859

rs17405819	8	76,806,584	*HNF4G*	Locke**	T	t	c	0.699	0.56	0.21	8.4E-03	0	6.6E-01	17859
rs16907751	8	81,375,457	*ZBTB10*	Locke	C	t	c	0.105	0.08	0.33	8.1E-01	0	3.4E-01	17859
rs2033732	8	85,079,709	*RALYL*	Locke	C	t	c	0.251	0.42	0.22	6.3E-02	0	5.3E-01	17859
rs4740619	9	15,634,326	*C9orf93*	Locke	T	t	c	0.547	0.66	0.20	1.1E-03	16.8	2.7E-01	17851
rs10968576	9	28,414,339	*LINGO2*	Speliotes	G	g	a	0.306	0.55	0.21	1.0E-02	0	8.5E-01	17859
rs6477694	9	111,932,342	*EPB41L4B*	Locke	C	c	t	0.344	0.47	0.21	2.6E-02	0	5.0E-01	17859
rs1928295	9	120,378,483	*TLR4*	Locke	T	t	c	0.552	0.08	0.19	6.8E-01	66.4	8.4E-02	17859
rs10733682	9	129,460,914	*LMX1B*	Locke	A	a	g	0.482	0.58	0.20	4.3E-03	0	7.1E-01	17859
rs2270204	9	131,042,734	*SWI5*	Winkler	T	g	t	0.255	0.42	0.23	7.2E-02	7.7	3.0E-01	17857
rs7899106	10	87,410,904	*GRID1*	Locke	G	g	a	0.049	0.32	0.47	4.9E-01	0	8.1E-01	17859
rs17094222	10	102,395,440	*HIF1AN*	Locke	C	c	t	0.215	0.12	0.25	6.4E-01	0	3.2E-01	17859
rs11191560	10	104,869,038	*NT5C2*	Locke, Wen 2012	C	c	t	0.091	0.73	0.35	3.6E-02	0	9.6E-01	17859
rs7903146	10	114,758,349	*TCF7L2*	Locke	C	c	t	0.705	0.54	0.21	1.1E-02	0	4.7E-01	17859
rs10886017	10	118,672,531	*KIAA1598*	Winkler	A	a	c	0.250	0.29	0.23	2.2E-01	0	8.0E-01	17859
rs2237897	11	2,858,546	*KCNQ1*	Wen 2014	T	c	t	0.952	0.02	0.50	9.7E-01	64	9.6E-02	17859
rs4256980	11	8,673,939	*TRIM66*	Speliotes	G	g	c	0.639	0.16	0.20	4.4E-01	0	7.8E-01	17397
rs7928810	11	17,372,443	*NCR3LG1*	Winkler	A	a	c	0.621	0.52	0.21	1.5E-02	0	7.2E-01	17398
rs1557765	11	17,403,639	*KCNJ11*	Winkler	T	c	t	0.621	0.47	0.20	2.1E-02	0	5.7E-01	17398
rs11030104	11	27,684,517	*BDNF*	Speliotes, Wen 2014	A	a	g	0.791	0.89	0.24	2.5E-04	91.5	6.1E-04	17398
rs10835210	11	27,695,910	*BDNF*	Locke	C	c	a	0.570	0.13	0.20	5.2E-01	0	5.5E-01	17396
rs652722	11	31,905,534	*PAX6*	Wen 2012	C	t	c	0.259	0.48	0.22	3.2E-02	0	7.2E-01	17859
rs2176598	11	43,864,278	*HSD17B12*	Locke	T	t	c	0.253	0.48	0.22	3.2E-02	41.4	1.9E-01	17859
rs3817334	11	47,650,993	*MTCH2*	Speliotes	T	t	c	0.407	0.21	0.20	3.0E-01	0	5.9E-01	17859
rs1865732*	11	112,960,722	*NCAM1*	Winkler	C	c	t	0.556	0.18	0.20	3.7E-01	7.4	3.0E-01	17398
rs12286929	11	115,022,404	*CADM1*	Locke	G	g	a	0.534	0.59	0.20	3.9E-03	0	4.3E-01	17859
rs11611246	12	939,480	*WNK1*	Winkler	T	t	g	0.205	0.03	0.25	9.2E-01	77.1	3.7E-02	17836
rs7970953	12	24,075,508	*SOX5*	Winkler	A	a	g	0.298	0.01	0.21	9.7E-01	72	5.9E-02	17859
rs1405552	12	41,746,673	*PDZRN4*	Winkler	A	g	a	0.545	0.11	0.20	6.0E-01	0	7.8E-01	17398
rs11181001	12	41,948,196	*PDZRN4*	Winkler	A	a	g	0.466	0.04	0.20	8.3E-01	21.5	2.6E-01	17398
rs7138803	12	50,247,468	*BCDIN3D*	Speliotes	A	a	g	0.378	0.52	0.20	1.1E-02	0	6.8E-01	17859
rs1438994*	12	90,594,389	*Intergenic*	Winkler	T	c	t	0.747	0.17	0.23	4.7E-01	0	9.9E-01	17398
rs11065987	12	112,072,424	*BRAP*	Winkler	A	a	g	0.569	0.40	0.20	5.0E-02	76.1	4.1E-02	17859
rs17630235	12	112,591,686	*TRAFD1*	Winkler	A	g	a	0.575	0.39	0.21	7.0E-02	73.1	5.4E-02	17858
rs11057405	12	122,781,897	*CLIP1*	Locke	G	g	a	0.904	0.99	0.35	4.6E-03	65.9	8.7E-02	17859
rs1885988*	13	28,010,262	*MTIF3*	Speliotes	C	c	t	0.180	0.28	0.26	2.8E-01	55	1.4E-01	17398
rs12429545	13	54,102,206	*OLFM4*	Speliotes	A	a	g	0.127	0.53	0.30	8.0E-02	43.8	1.8E-01	17859
rs9540493	13	66,205,704	*MIR548X2*	Locke	A	g	a	0.565	0.18	0.20	3.8E-01	0	5.5E-01	17859
rs1441264	13	79,580,919	*MIR548A2*	Locke	A	a	g	0.602	0.50	0.21	1.8E-02	0	6.3E-01	17398
rs9634489	13	97,049,004	*HS6ST3*	Winkler	A	g	a	0.507	0.36	0.20	8.0E-02	0	6.5E-01	17859
rs10132280	14	25,928,179	*STXBP6*	Locke	C	c	a	0.689	0.40	0.21	6.3E-02	43.6	1.8E-01	17857
rs12885454	14	29,736,838	*PRKD1*	Locke	C	c	a	0.665	0.57	0.21	8.0E-03	0	6.4E-01	17398
rs11847697	14	30,515,112	*PRKD1*	Speliotes	T	t	c	0.046	0.12	0.49	8.1E-01	0	7.7E-01	17398
rs17522122	14	33,302,882	*AKAP6*	Winkler	T	t	g	0.481	0.67	0.20	9.8E-04	5.4	3.0E-01	17842
rs7141420	14	79,899,454	*NRXN3*	Speliotes	T	t	c	0.526	0.38	0.19	5.3E-02	0	9.0E-01	17859
rs3783890	14	93,790,276	*BTBD7*	Winkler	T	c	t	0.184	0.28	0.25	2.6E-01	0	7.8E-01	17859
rs7143963	14	103,304,425	*TRAF3*	Winkler	T	t	c	0.177	0.28	0.26	2.8E-01	0	5.5E-01	17859
rs709400	14	104,149,475	*KLC1*	Winkler	A	a	g	0.625	0.16	0.20	4.4E-01	0	5.8E-01	17859
rs3736485	15	51,748,610	*DMXL2*	Locke	A	a	g	0.470	0.35	0.20	8.4E-02	0	5.9E-01	17858
rs16951275	15	68,077,168	*MAP2K5*	Speliotes	T	t	c	0.765	0.27	0.23	2.4E-01	0	7.9E-01	17859
rs4776970	15	68,080,886	*MAP2K5*	Wen 2012	A	a	t	0.629	0.31	0.20	1.3E-01	0	7.1E-01	17859
rs7164727	15	73,093,991	*LOC100287559*	Locke	T	t	c	0.674	0.26	0.20	2.1E-01	57.3	1.3E-01	17859
rs7181659	15	95,267,483	*MCTP2*	Winkler	A	a	g	0.493	0.02	0.19	9.2E-01	0	5.4E-01	17858
rs11866815	16	387,867	*AXIN1*	Winkler	T	c	t	0.751	0.04	0.23	8.6E-01	0	7.6E-01	17859
rs12446632	16	19,935,389	*GPRC5B*	Speliotes	G	g	a	0.857	0.83	0.28	3.2E-03	0	9.9E-01	17398
rs11074446	16	20,255,123	*GP2*	Locke	T	t	c	0.868	0.71	0.29	1.6E-02	0	8.2E-01	17396
rs2650492	16	28,333,411	*SBK1*	Locke	A	a	g	0.267	0.21	0.27	4.4E-01	0	9.1E-01	15676
rs3888190	16	28,889,486	*ATP2A1*	speliotes	A	a	c	0.382	0.26	0.21	2.3E-01	81.9	1.9E-02	17858
rs4787491	16	30,015,337	*INO80E*	Locke	G	g	a	0.536	0.37	0.20	6.8E-02	0	7.3E-01	17859
rs9925964	16	31,129,895	*KAT8*	Locke	A	a	g	0.626	0.54	0.20	8.4E-03	72.1	5.9E-02	17398
rs2080454	16	49,062,590	*CBLN1*	Locke	C	c	a	0.378	0.02	0.20	9.2E-01	74.4	4.8E-02	17398
rs1558902***	16	53,803,574	*FTO*	Speliotes, Wen 2014	A	a	t	0.414	1.20	0.20	5.1E-09	0	8.2E-01	17832
rs17817964***	16	53,828,066	*FTO*	Monda	T	t	c	0.404	1.14	0.20	2.7E-08	0	6.7E-01	17859
rs889398	16	69,556,715	*NFAT5*	Winkler	T	c	t	0.576	0.57	0.20	5.0E-03	78.1	3.3E-02	17398
rs9914578	17	2,005,136	*SMG6*	Locke	G	c	g	0.801	0.19	0.25	4.5E-01	0	5.1E-01	17859
rs1000940	17	5,283,252	*RABEP1*	Locke	G	g	a	0.308	0.45	0.21	3.5E-02	0	1.0E+00	17859
rs4986044	17	21,261,560	*KCNJ12*	Winkler	T	c	t	0.528	0.15	0.21	4.7E-01	0	6.4E-01	15215
rs12150665	17	34,914,787	*GGNBP2*	Winkler	T	t	c	0.590	0.33	0.20	1.1E-01	0	4.8E-01	17859
rs11652097	17	45,316,717	*ITGB3*	Winkler	T	t	c	0.388	0.40	0.20	5.0E-02	0	3.2E-01	17859
rs6504108	17	46,292,923	*SKAP1*	Winkler	T	c	t	0.287	0.22	0.21	3.0E-01	0	6.0E-01	17859
rs8075273*	17	61,728,881	*unknown*	Winkler	C	c	a	0.723	0.27	0.22	2.3E-01	0	4.4E-01	17859
rs312750	17	68,343,539	*KCNJ2*	Winkler	A	a	g	0.489	0.15	0.19	4.3E-01	0	7.7E-01	17859
rs12940622	17	78,615,571	*RPTOR*	Locke**	G	g	a	0.569	0.15	0.19	4.5E-01	0	5.6E-01	17859
rs1808579	18	21,104,888	*C18orf8*	Speliotes	C	c	t	0.524	0.63	0.19	1.3E-03	0	7.1E-01	17859
rs7239883	18	40,147,671	*LOC284260*	Locke	G	g	a	0.391	0.43	0.20	3.6E-02	0	5.2E-01	17859
rs7243357	18	56,883,319	*GRP*	Locke	T	t	g	0.827	0.08	0.26	7.6E-01	0	3.5E-01	17859
rs2331841	18	57,828,637	*MC4R*	Okada	A	a	g	0.430	0.57	0.19	3.2E-03	0	4.5E-01	17859
rs6567160	18	57,829,135	*MC4R*	Speliotes, Pei	C	c	t	0.234	0.83	0.23	4.0E-04	24.3	2.5E-01	17859
rs591166	18	57,841,589	*MC4R*	Wen 2014	A	a	t	0.435	0.56	0.19	3.8E-03	0	5.9E-01	17859
rs9944545	18	57,958,244	*MC4R*	Locke	T	t	c	0.295	0.74	0.22	9.4E-04	41.4	1.9E-01	17859
rs17066842	18	58,040,624	*MC4R*	Locke	G	g	a	0.964	1.06	0.55	5.3E-02	10.2	2.9E-01	17852
rs17724992	19	18,454,825	*PGPEP1*	Locke	A	a	g	0.732	0.08	0.22	7.1E-01	0	4.9E-01	17859
rs17513613	19	30,286,822	*CCNE1*	Winkler	T	c	t	0.328	0.15	0.21	5.0E-01	75.2	4.5E-02	17398
rs29941	19	34,309,532	*KCTD15*	Speliotes	G	g	a	0.677	0.50	0.21	2.1E-02	0	9.3E-01	17859
rs2075650	19	45,395,619	*TOMM40*	Speliotes	A	a	g	0.870	0.91	0.29	1.9E-03	43.1	1.8E-01	17398
rs11671664	19	46,172,278	*GIPR, QPCTL*	Wen 2014	G	g	a	0.891	0.42	0.33	2.0E-01	69.9	6.8E-02	17398

rs2287019	19	46,202,172	*QPCTL*	Speliotes	C	c	t	0.806	0.88	0.26	8.1E-04	59.5	1.2E-01	17398
rs3810291	19	47,569,003	*ZC3H4*	Speliotes	A									
rs4802349	19	47,874,510	*DHX34*	Gong 2013	G	t	g	0.104	0.16	0.33	6.3E-01	0	9.1E-01	17858
rs8123881*	20	15,819,495	*MACROD 2*	Winkler	G	a	g	0.875	0.13	0.30	6.7E-01	0	9.3E-01	17398
rs6091540	20	51,087,862	*ZFP64*	Locke**	C	c	t	0.717	0.23	0.22	3.0E-01	0	6.0E-01	17398
rs2836754	21	40,291,740	*ETS2*	Locke	C	c	t	0.625	0.31	0.20	1.2E-01	0	7.8E-01	17859
rs4820408	22	40,604,945	*TNRC6B*	Winkler	T	t	g	0.411	0.46	0.19	1.8E-02	67.4	8.0E-02	17859

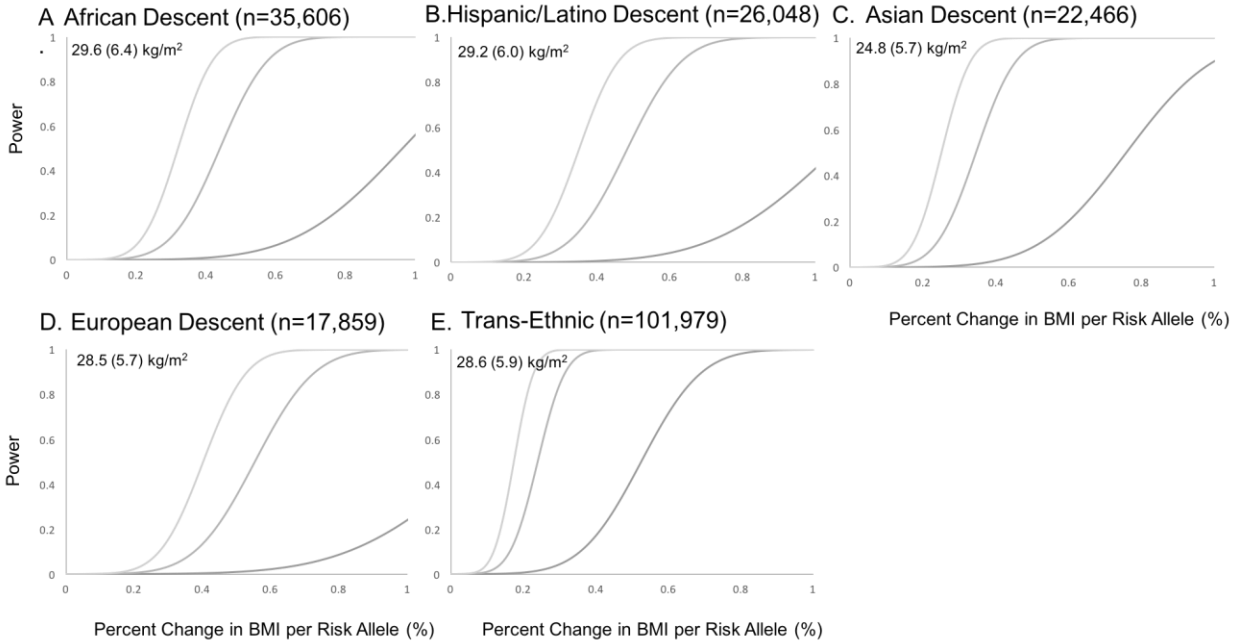
Abbreviations: A1=coded allele, A2=non-coded allele, Bp37=base pair Build 37, Chr=chromosome, FE=Fixed-Effect, HetP=heterogeneity p-value, P=p-value, Prep=prepared reference, Ref=reference, Sub=submitted reference, SNPs=single nucleotide polymorphisms.

*When the index SNP was not genotyped on the MetaboChip, the proxy SNP in tight linkage disequilibrium ($r^2 \geq 0.8$ in 1000 Genomes pilot 1 CEU, YRI, CHB+JPT depending on the population of discovery) with the lowest p-value in the African American sample was chosen to represent the index signal. The decreasing and increasing alleles for proxies were assigned assuming that the risk index SNP would have a similar allele frequency in the 1000 Genomes population (EUR, AFR, or EAS depending on the discovery population) as the risk proxy SNP.

**These loci were also described by Berndt *et al*. for obesity (maximum sample size of 263,407) [29]. The most recent BMI references per racial/ethnic group are noted above by their first author and publication year, if applicable [21, 25, 26, 30, 32, 33, 186, 195, 204].

***For GWAS SNPs a Bonferroni correction for multiple tests in the fixed-effect analyses reflected the number of independent previously-reported signals tested (=0.05/166). The 4 noted SNP pairs above were in tight linkage disequilibrium [$r^2 \geq 0.8$ in non-European 1000 genomes pilot populations(s)] with each other, but because they were reports from distinct discovery populations we retained them in this inventory in case they were population-specific variants. Therefore, our Bonferroni correction was penalized for only 166 (=170-4) tests.

APPENDIX O: STATISTICAL POWER TO DETECT LOW FREQUENCY TO COMMON VARIANTS (1% LIGHT GRAY, 5% MEDIUM GRAY, 10% DARK GRAY) ACROSS A RANGE OF BMI GENETIC EFFECTS (% CHANGE PER ALLELE) WHILE VARYING SAMPLE SIZES AND BMI MEAN (STANDARD DEVIATIONS) OF THE VARIOUS RACIAL/ETHNIC STRATIFIED (PANELS A-D) AND THE TRANS-ETHNIC META-ANALYSES (PANEL E), AND ASSUMING A WORST-CASE SCENARIO LOCUS-SPECIFIC ALPHA OF 6.31X10-5 (E.G. 792 INDEPENDENT TESTS AT *TRAFD1*).



APPENDIX P: TRANS-ETHNIC DESCENT GENETIC EFFECT ESTIMATES FOR 170 BMI INDEX SNPS FROM PREVIOUS REPORTS, INCLUDING SECONDARY SIGNALS IN EUROPEAN DESCENT POPULATIONS AND/OR POPULATION-SPECIFIC MARKERS.

					FIXED EFFECTS											MANTRA		
rsID*	**Chr**	**Bp37**	**Gene**	**Ref. Risk Allele**	**A1**	**A2**	**Freq**	**Min Freq**	**Max Freq**	**β (%)**	**SE (%)**	**P****	**I^2**	**HetP*****	**N**	**log 10 BF**	**Post prob het**	**N**
rs2803328	1	1,874,326	*KIAA1751*	C	g	c	0.435	0.4855	0.8039	0.01	0.09	8.8E-01	0	5.7E-01	93701			
rs2271928	1	32,127,953	*COL16A1*	A	g	a	0.498	0.4011	0.6739	0.06	0.08	4.6E-01	5.5	3.7E-01	93721	-0.4	1.6E-02	93721
rs2275426	1	46,487,552	*MAST2*	A	a	g	0.481	0.4155	0.5807	0.14	0.08	1.1E-01	0	7.8E-01	93725			
rs977747	1	47,684,677	*TAL1*	T	t	g	0.581	0.403	0.8956	0.08	0.10	3.8E-01	1.7	3.8E-01	93726			
rs657452	1	49,589,847	*AGBL4*	A														
rs11583200	1	50,559,820	*ELAVL4*	C	c	t	0.598	0.1852	0.6135	0.35	0.09	1.5E-04	0	5.0E-01	93736	2.7	1.3E-02	93736
rs3101336	1	72,751,185	*NEGR1*	C	c	t	0.655	0.0901	0.4591	0.17	0.09	6.8E-02	69.4	2.0E-02	101969	-0.1	4.1E-02	101969
rs12566985	1	75,002,193	*FPGT-TNNI3K*	G	g	a	0.719	0.1328	0.3962	0.42	0.11	2.8E-04	0	5.2E-01	75627	2.1	8.0E-03	75627
rs12401738	1	78,446,761	*FUBP1*	A	a	g	0.258	0.0636	0.3487	0.00	0.12	1.0E+00	59.6	5.9E-02	93732	-1.1	5.1E-02	93732
rs11165643	1	96,924,097	*PTBP2*	T	t	c	0.524	0.2202	0.7368	0.23	0.09	1.2E-02	66.4	3.0E-02	93704	1.0	4.8E-02	93704
rs17024393	1	110,154,688	*GNAT2*	C	c	t	0.067	0.9174	0.9763	0.20	0.23	3.9E-01	60.6	7.9E-02	77687	-0.4	7.3E-02	77687
rs4357530*	1	151,103,153	*SEMA6C*	G	g	a	0.423	0.2812	0.881	0.10	0.10	3.5E-01	53.9	8.9E-02	89077			
rs10913118	1	175,954,755	*RFWD2*	A	a	c	0.694	0.5584	0.8501	0.02	0.10	8.1E-01	24.4	2.6E-01	93715	-0.4	1.0E-02	93715
rs574367***	1	177,873,210	*SEC16B*	T	t	g	0.176	0.1099	0.1954	0.70	0.11	2.9E-11	0	6.1E-01	101913	9.1	1.4E-02	101913
rs543874***	1	177,889,480	*SEC16B*	G	g	a	0.213	0.7507	0.8087	0.90	0.10	3.5E-21	72.9	1.1E-02	101972	19.1	2.5E-01	101972
rs10920678	1	190,239,907	*FAM5C*	A	a	g	0.374	0.2964	0.4242	0.39	0.09	5.0E-06	0	5.4E-01	93066			
rs2820292	1	201,784,287	*NAV1*	C	c	a	0.400	0.4554	0.7315	0.16	0.09	6.7E-02	0	9.1E-01	93725	0.3	6.0E-03	93725
rs2820436	1	219,640,680	*LYPLAL1*	A	a	c	0.388	0.1952	0.4781	0.50	0.09	3.2E-08	0	4.7E-01	93721	6.0	6.0E-03	93721
rs12463617***	2	629,244	*TMEM18*	C	c	a	0.865	0.0864	0.1751	0.94	0.14	1.5E-11	33.7	2.2E-01	66370	9.4	2.3E-02	66370
rs13021737***	2	632,348	*TMEM18*	G	g	a	0.873	0.087	0.1748	1.05	0.12	3.0E-18	42	1.6E-01	101832	16.0	3.8E-02	101832
rs11676272*,***	2	25,141,538	*ADCY3*	G	g	a	0.523	0.1669	0.5856	0.58	0.09	1.3E-10	18.2	3.0E-01	93729	8.4	2.2E-02	93729
rs10182181***	2	25,150,296	*ADCY3*	G	g	a	0.523	0.1699	0.5858	0.57	0.09	2.9E-10	0	4.0E-01	93709	8.2	9.0E-03	93709
rs11126666	2	26,928,811	*KCNK3*	A	g	a	0.651	0.1942	0.6513	0.05	0.10	6.3E-01	0	5.5E-01	93728	-0.6	5.0E-03	93728
rs116612809	2	28,301,171	*BRE*	G	g	a	0.088	0.903	0.9982	1.05	0.23	8.8E-06	87.9	*2.6E-04*	68016	5.1	9.4E-01	68016
rs1016287	2	59,305,625	*FLJ30838*	T	t	c	0.242	0.1892	0.2959	0.15	0.10	1.2E-01	0	9.2E-01	93705	0.1	4.0E-03	93705
rs11688816	2	63,053,048	*EHBP1*	G	g	a	0.606	0.2596	0.488	0.01	0.09	8.7E-01	31	2.3E-01	93729	-0.4	1.3E-02	93729
rs12622013*	2	79,501,362	*REG3A*	G	g	a	0.174	0.7996	0.8899	0.33	0.11	3.2E-03	51.6	1.0E-01	93693			
rs7570971	2	135,837,906	*RAB3GAP1*	A	c	a	0.183	0.7794	0.8915	0.08	0.15	5.9E-01	42.1	1.8E-01	75848			
rs4988235	2	136,608,646	*MCM6*	A	a	g	0.186	0.1104	0.2257	0.13	0.14	3.8E-01	56.7	9.9E-02	75871			
rs2121279	2	143,043,285	*LRP1B*	T	t	c	0.087	0.0316	0.1285	0.22	0.20	2.6E-01	0	5.0E-01	79468	-0.2	1.3E-02	79468
rs1460676	2	164,567,689	*FIGN*	C	c	t	0.259	0.6324	0.8663	0.31	0.10	2.4E-03	0	8.9E-01	93734	1.6	4.0E-03	93734
rs10184004	2	165,508,389	*GRB14/COBLL1*	T	t	c	0.452	0.103	0.7186	0.52	0.10	1.3E-07	0	5.5E-01	93726	5.6	1.1E-02	93726
rs10930502	2	172,890,588	*METAP1D*	A	a	g	0.593	0.3328	0.7	0.48	0.09	1.4E-07	0	8.7E-01	93721			
rs1528435	2	181,550,962	*UBE2E3*	T	t	c	0.646	0.6115	0.6993	0.26	0.09	3.6E-03	18.4	3.0E-01	93739	1.4	5.0E-03	93739
rs972540	2	207,244,783	*ADAM23*	A	g	a	0.203	0.7281	0.8253	0.20	0.10	5.6E-02	48.9	1.2E-01	93737	0.4	9.0E-03	93737
rs17203016	2	208,255,518	*CREB1*	G	g	a	0.163	0.7962	0.9595	0.24	0.13	5.8E-02	0	7.5E-01	93700	-0.2	8.0E-03	93700
rs7599312	2	213,413,231	*ERBB4*	G	g	a	0.715	0.0658	0.3767	0.29	0.10	5.1E-03	31.5	2.2E-01	93703			
rs492400	2	219,349,752	*USP37*	C	c	t	0.364	0.5734	0.7614	0.36	0.11	1.4E-03	0	9.9E-01	56322			
rs2176040	2	227,092,802	*LOC646736*	A	a	g	0.275	0.0986	0.3506	0.50	0.10	4.0E-07	0	5.5E-01	93732	5.1	6.0E-03	93732
rs9845966	3	13,433,158	*NUP210*	T	t	g	0.466	0.2319	0.5767	0.20	0.09	2.4E-02	0	7.3E-01	87968	0.7	5.0E-03	87968
rs6804842	3	25,106,437	*RARB*	G	g	a	0.530	0.388	0.5883	0.14	0.08	8.6E-02	0	5.2E-01	93722	0.3	9.0E-03	93722
rs7613875	3	49,971,514	*MON1A*	A	a	c	0.455	0.2015	0.6741	0.13	0.11	2.3E-01	0	9.0E-01	68026			
rs2365389	3	61,236,462	*FHIT*	C	c	t	0.349	0.4078	0.8724	0.30	0.10	3.0E-03	0	9.6E-01	93732	1.2	7.0E-03	93732
rs333495*	3	78,834,343	*ROBO1*	G	g	t	0.431	0.4993	0.7202	0.15	0.09	8.3E-02	71.7	1.4E-02	93735			
rs13078960	3	85,807,590	*CADM2*	G	t	g	0.857	0.792	0.9648	0.00	0.15	9.8E-01	0	8.2E-01	101976	-0.4	1.7E-02	101976
rs1720825	3	138,108,083	*MRAS*	A	g	a	0.865	0.0372	0.1966	0.01	0.14	9.4E-01	0	8.0E-01	93739	-0.4	1.0E-02	93739
rs2640017*	3	141,335,121	*RASA2*	G	g	a	0.210	0.725	0.9819	0.46	0.13	4.6E-04	0	7.3E-01	93609	2.2	1.1E-02	93609
rs11927381	3	185,508,591	*IGF2BP2*	T	t	c	0.563	0.2551	0.684	0.49	0.09	1.3E-07	0	8.4E-01	93626	5.5	3.0E-03	93626
rs1516725	3	185,824,004	*ETV5*	C	c	t	0.864	0.0698	0.1832	0.69	0.13	1.0E-07	0	8.7E-01	101811	5.6	9.0E-03	101811
rs16992647	4	36,813,105	*KIAA1239*	T	t	c	0.273	0.1306	0.4603	0.00	0.11	9.7E-01	0	5.3E-01	93609			
rs16858082	4	45,175,804	*GNPDA2*	T	t	c	0.505	0.345	0.5987	0.42	0.08	2.1E-07	0	9.9E-01	101873	5.4	8.0E-03	101873
rs10938397	4	45,182,527	*GNPDA2*	G	g	a	0.325	0.5657	0.7499	0.60	0.08	7.3E-13	1.7	3.8E-01	101782	10.5	1.0E-02	101782
rs348495	4	45,184,442	*GNPDA2*	G														
rs13107325	4	103,188,709	*SLC39A8*	T	t	c	0.053	0.0188	0.0745	1.05	0.25	3.4E-05	0	8.7E-01	79090	3.2	1.5E-02	79090

rs11727676	4	145,659,064	*HHIP*	T	t	c	0.936	0.9046	0.9842	0.04	0.23	8.7E-01	59	6.3E-02	93185	-0.5	4.2E-02	93185
rs2112347	5	75,015,242	*POC5*	T	t	g	0.531	0.4432	0.6402	0.28	0.08	4.3E-04	3.3	3.8E-01	101972	2.2	1.1E-02	101972
rs6870983	5	87,697,533	*TMEM161B−AS1*	T	c	t	0.666	0.0573	0.4458	0.19	0.11	7.3E-02	0	6.4E-01	93190			
rs11951673*	5	95,861,012	*PCSK1*	C	c	t	0.560	0.3858	0.5768	0.50	0.08	3.5E-09	0	4.4E-01	93717			
rs6864049	5	124,330,522	*ZNF608*	A	g	a	0.679	0.1969	0.473	0.15	0.09	9.4E-02	9.3	3.5E-01	93676			
rs13174863	5	139,080,745	*CXXC5*	A	g	a	0.109	0.8467	0.94	0.21	0.15	1.6E-01	0	8.1E-01	93734			
rs4569924*	5	153,540,025	*GALNT10*	T	t	c	0.486	0.3547	0.8263	0.26	0.10	5.6E-03	69.4	2.0E-02	93730			
rs2228213	6	12,124,855	*HIVEP1*	A	g	a	0.742	0.109	0.345	0.20	0.10	4.6E-02	0	4.8E-01	93739			
rs9356744	6	20,685,486	*CDKAL1*	T	t	c	0.562	0.3735	0.6899	0.42	0.08	2.6E-07	81.2	1.2E-03	101966	5.9	7.4E-01	101966
rs943466	6	33,731,787	*LEMD2*	A	g	a	0.733	0.1343	0.3242	0.10	0.10	3.0E-01	0	9.9E-01	93735			
rs205262	6	34,563,164	*C6orf106*	G	g	a	0.399	0.3697	0.8571	0.24	0.10	1.5E-02	0	6.7E-01	93730			
rs2033529	6	40,348,653	*TDRG1*	G	g	a	0.231	0.7164	0.8354	0.05	0.10	6.5E-01	0	6.6E-01	93730	-0.5	8.0E-03	93730
rs2207139	6	50,845,490	*TFAP2B*	G	g	a	0.211	0.7062	0.9036	0.41	0.10	1.0E-04	2	3.8E-01	101973	2.9	1.2E-02	101973
rs9400239	6	108,977,663	*FOXO3*	C	c	t	0.549	0.292	0.7448	0.14	0.09	1.4E-01	0	6.3E-01	93726	-0.1	8.0E-03	93726
rs9374842	6	120,185,665	*LOC285762*	T	c	t	0.213	0.765	0.8243	0.01	0.10	8.9E-01	0	5.2E-01	93654	-0.7	9.0E-03	93654
rs13201877	6	137,675,541	*IFNGR1*	G	a	g	0.904	0.864	0.9674	0.01	0.16	9.7E-01	0	4.1E-01	93739			
rs1281962	6	153,431,376	*RGS17*	C	c	g	0.667	0.5291	0.8202	0.14	0.09	1.3E-01	0	6.3E-01	93718	0.1	6.0E-03	93718
rs3127574	6	160,791,370	*SLC22A3*	C	c	g	0.497	0.4325	0.5868	0.13	0.08	1.2E-01	51.7	1.0E-01	93727	0.0	5.0E-03	93727
rs13191362	6	163,033,350	*PARK2*	A	a	g	0.912	0.8825	0.9465	0.48	0.18	7.1E-03	70	3.6E-02	79492	0.9	4.4E-02	79492
rs10499694*	7	50,614,173	*DDC*	G	a	g	0.578	0.4816	0.6405	0.11	0.09	2.0E-01	57.2	7.2E-02	93731	-0.1	2.1E-02	93731
rs1167827	7	75,163,169	*HIP1*	G	g	a	0.548	0.1298	0.896	0.33	0.11	2.9E-03	0	7.1E-01	93725			
rs6465468	7	95,169,514	*ASB4*	T	g	t	0.823	0.0681	0.2286	0.03	0.13	8.4E-01	69.7	3.7E-02	75841	-0.7	5.6E-02	75841
rs6990042	8	14,173,974	*SGCZ*	T	g	t	0.406	0.4865	0.7356	0.14	0.09	1.0E-01	49.2	1.2E-01	93729			
rs7844647*	8	34,503,776	*Intergenic*	T	t	c	0.509	0.2848	0.7379	0.32	0.09	3.4E-04	72.1	1.3E-02	93735			
rs17405819	8	76,806,584	*HNF4G*	T	t	c	0.707	0.6087	0.9166	0.20	0.10	4.5E-02	21.5	2.8E-01	93726	0.4	1.0E-02	93726
rs16907751	8	81,375,457	*ZBTB10*	C	c	t	0.893	0.0747	0.1472	0.08	0.14	5.7E-01	0	9.6E-01	93735			
rs2033732	8	85,079,709	*RALYL*	C	t	c	0.222	0.1194	0.2766	0.11	0.10	3.1E-01	83.4	4.3E-04	93533			
rs4740619	9	15,634,326	*C9orf93*	T	t	c	0.447	0.2481	0.5519	0.11	0.09	2.3E-01	71.9	1.4E-02	93693	-0.4	9.7E-02	93693
rs10968576	9	28,414,339	*LINGO2*	G	g	a	0.217	0.6942	0.8301	0.52	0.10	1.1E-07	0	5.9E-01	101976	5.6	1.1E-02	101976
rs6477694	9	111,932,342	*EPB41L4B*	C	c	t	0.446	0.4083	0.6563	0.19	0.09	2.9E-02	0	4.9E-01	93735			
rs1928295	9	120,378,483	*TLR4*	T	t	c	0.565	0.5517	0.581	0.15	0.08	6.3E-02	47.6	1.3E-01	93717	0.3	2.2E-02	93717
rs10733682	9	129,460,914	*LMX1B*	A	a	g	0.508	0.2881	0.727	0.28	0.09	2.1E-03	68.2	2.4E-02	93718	1.7	1.4E-01	93718
rs2270204	9	131,042,734	*SWI5*	T	g	t	0.504	0.3096	0.7451	0.24	0.09	7.3E-03	0	6.3E-01	93431	1.1	5.0E-03	93431
rs7899106	10	87,410,904	*GRID1*	G	g	a	0.095	0.8783	0.9511	0.17	0.18	3.6E-01	0	8.3E-01	79503	-0.3	9.0E-03	79503
rs17094222	10	102,395,440	*HIF1AN*	C	c	t	0.241	0.6756	0.9455	0.19	0.11	9.5E-02	0	9.2E-01	93727	-0.4	9.0E-03	93727
rs11191560	10	104,869,038	*NT5C2*	C	c	t	0.204	0.7305	0.9583	0.54	0.11	1.8E-06	0	7.8E-01	101966	4.4	8.0E-03	101966
rs7903146	10	114,758,349	*TCF7L2*	C	c	t	0.739	0.0658	0.2949	0.75	0.10	2.2E-13	54	8.9E-02	101975	11.2	6.3E-02	101975
rs10886017	10	118,672,531	*KIAA1598*	A	a	c	0.370	0.2499	0.4623	0.16	0.09	7.8E-02	38.9	1.8E-01	93694	0.2	6.0E-03	93694
rs2237897	11	2,858,546	*KCNQ1*	T	t	c	0.237	0.0476	0.3527	0.66	0.12	3.3E-08	0	4.0E-01	93516	6.2	9.0E-03	93516
rs4256980	11	8,673,939	*TRIM66*	G	g	c	0.509	0.3614	0.6023	0.22	0.08	5.2E-03	0	5.4E-01	101492	1.1	9.0E-03	101492
rs7928810	11	17,372,443	*NCR3LG1*	A	a	c	0.690	0.6206	0.9151	0.31	0.10	1.9E-03	34	2.1E-01	93268	1.6	1.7E-02	93268
rs1557765	11	17,403,639	*KCNJ11*	T	c	t	0.686	0.1071	0.3794	0.31	0.10	1.5E-03	0	5.0E-01	93268	1.8	1.0E-03	93268
rs11030104	11	27,684,517	*BDNF*	A	a	g	0.694	0.5664	0.9513	0.40	0.10	7.3E-05	79.5	2.2E-03	101516	3.2	3.1E-01	101516
rs10835210	11	27,695,910	*BDNF*	C	c	a	0.686	0.1318	0.4298	0.12	0.10	2.2E-01	0	9.0E-01	89057	-0.3	8.0E-03	89057
rs652722	11	31,905,534	*PAX6*	C														
rs2176598	11	43,864,278	*HSD17B12*	T	t	c	0.323	0.1371	0.4054	0.10	0.10	3.0E-01	23	2.7E-01	93731	-0.3	1.6E-02	93731
rs3817334	11	47,650,993	*MTCH2*	T	t	c	0.338	0.2663	0.4073	0.28	0.08	6.3E-04	0	5.0E-01	101940	2.0	6.0E-03	101940
rs1865732*	11	112,960,722	*NCAM1*	C	t	c	0.619	0.4436	0.8391	0.01	0.09	9.0E-01	0	6.9E-01	93268	-0.5	4.0E-03	93268
rs12286929	11	115,022,404	*CADM1*	G	g	a	0.488	0.4298	0.7569	0.26	0.09	3.3E-03	36.2	2.0E-01	92926			
rs11611246	12	939,480	*WNK1*	T	t	g	0.261	0.2049	0.2881	0.25	0.12	3.4E-02	50.8	1.3E-01	57861			
rs7970953	12	24,075,508	*SOX5*	A	a	g	0.426	0.2359	0.692	0.12	0.09	2.0E-01	0	8.7E-01	93676			
rs1405552	12	41,746,673	*PDZRN4*	A	g	a	0.558	0.1015	0.5654	0.13	0.10	1.8E-01	11.2	3.4E-01	93274	0.0	8.0E-03	93274
rs11181001	12	41,948,196	*PDZRN4*	A	a	g	0.480	0.39	0.5806	0.13	0.08	1.1E-01	29	2.4E-01	93264			
rs7138803	12	50,247,468	*BCDIN3D*	A	a	g	0.285	0.179	0.3777	0.35	0.09	9.0E-05	0	4.8E-01	101969	2.9	3.0E-03	101969
rs1438994*	12	90,594,389	*Intergenic*	T	t	c	0.266	0.1124	0.3721	0.09	0.10	3.7E-01	10.5	3.4E-01	93273	-0.2	7.0E-03	93273
rs11065987	12	112,072,424	*BRAP*	A	a	g	0.716	0.5693	0.9221	0.21	0.12	8.8E-02	17.8	3.0E-01	93730	0.2	1.5E-02	93730
rs17630235	12	112,591,686	*TRAFD1*	A	g	a	0.723	0.0762	0.4249	0.21	0.13	9.9E-02	23.3	2.7E-01	93617	0.2	1.9E-02	93617
rs11057405	12	122,781,897	*CLIP1*	G	g	a	0.919	0.0156	0.0956	0.58	0.23	1.2E-02	18.1	3.0E-01	57987			
rs1885988*	13	28,010,262	*MTIF3*	C	c	t	0.144	0.8201	0.9602	0.11	0.13	4.3E-01	45.4	1.4E-01	93277	-0.4	4.8E-02	93277
rs12429545	13	54,102,206	*OLFM4*	A	a	g	0.203	0.0491	0.2728	0.63	0.12	8.4E-08	0	5.2E-01	93719			
rs9540493	13	66,205,704	*MIR548X2*	A	a	g	0.567	0.4347	0.7201	0.19	0.09	2.5E-02	68.1	2.4E-02	93669			
rs1441264	13	79,580,919	*MIR548A2*	A	a	g	0.632	0.5428	0.6933	0.25	0.09	4.9E-03	0	4.1E-01	93240	1.2	1.0E-02	93240
rs9634489	13	97,049,004	*HS6ST3*	A	g	a	0.508	0.381	0.6009	0.15	0.09	7.9E-02	0	6.8E-01	93735			
rs10132280	14	25,928,179	*STXBP6*	C	c	a	0.622	0.0877	0.5253	0.32	0.10	8.4E-04	47.4	1.3E-01	93686	1.9	2.3E-02	93686
rs12885454	14	29,736,838	*PRKD1*	C	c	a	0.669	0.1338	0.4983	0.27	0.10	5.8E-03	0	4.4E-01	93255	0.9	8.0E-03	93255
rs11847697	14	30,515,112	*PRKD1*	T	t	c	0.258	0.0384	0.3313	0.01	0.14	9.3E-01	0	9.5E-01	93490	-0.5	1.1E-02	93490

rs17522122	14	33,302,882	*AKAP6*	T	t	g	0.412	0.3787	0.481	0.37	0.10	3.9E-04	36.8	2.1E-01	57936	2.3	1.2E-02	57936
rs7141420	14	79,899,454	*NRXN3*	T	t	c	0.535	0.3788	0.6255	0.27	0.08	1.3E-03	69.5	2.0E-02	93721	1.8	1.6E-01	93721
rs3783890	14	93,790,276	*BTBD7*	T	t	c	0.756	0.6348	0.9017	0.11	0.10	2.9E-01	59.6	5.9E-02	93736			
rs7143963	14	103,304,425	*TRAF3*	T	t	c	0.432	0.1767	0.6178	0.19	0.09	3.9E-02	0	6.5E-01	93731			
rs709400	14	104,149,475	*KLC1*	A	a	g	0.748	0.6251	0.8703	0.08	0.10	4.4E-01	0	9.0E-01	93730	-0.3	4.0E-03	93730
rs3736485	15	51,748,610	*DMXL2*	A	a	g	0.570	0.47	0.7527	0.17	0.09	4.6E-02	0	8.0E-01	93693			
rs16951275	15	68,077,168	*MAP2K5*	T	t	c	0.542	0.4071	0.765	0.35	0.08	2.5E-05	27.3	2.5E-01	101972	2.4	1.2E-02	101972
rs4776970	15	68,080,886	*MAP2K5*	A	a	t	0.422	0.2375	0.6294	0.38	0.08	6.1E-06	0	9.0E-01	101972	4.1	4.0E-03	101972
rs7164727	15	73,093,991	*LOC100287559*	T	t	c	0.450	0.234	0.6741	0.10	0.09	2.8E-01	0	4.5E-01	93729	-0.2	1.5E-02	93729
rs7181659	15	95,267,483	*MCTP2*	A	a	g	0.558	0.4499	0.6466	0.09	0.09	2.9E-01	0	7.4E-01	93699			
rs11866815	16	387,867	*AXIN1*	T	t	c	0.299	0.1958	0.4172	0.08	0.09	3.8E-01	0	8.4E-01	93635	-0.3	1.3E-02	93635
rs12446632	16	19,935,389	*GPRC5B*	G														
rs11074446	16	20,255,123	*GP2*	T	t	c	0.765	0.678	0.8677	0.52	0.10	3.8E-07	0	5.6E-01	93229			
rs2650492	16	28,333,411	*SBK1*	A	a	g	0.146	0.0644	0.2669	0.54	0.14	9.5E-05	26.1	2.5E-01	99770	2.8	2.0E-02	99770
rs3888190	16	28,889,486	*ATP2A1*	A	a	c	0.307	0.1509	0.4086	0.44	0.09	1.6E-06	0	5.8E-01	101949	4.6	1.3E-02	101949
rs4787491	16	30,015,337	*INO80E*	G	g	a	0.464	0.4636	0.6518	0.31	0.09	3.7E-04	0	8.9E-01	93728	2.2	6.0E-03	93728
rs9925964	16	31,129,895	*KAT8*	A	a	g	0.585	0.1625	0.8666	0.34	0.10	1.1E-03	48	1.2E-01	92510			
rs2080454	16	49,062,590	*CBLN1*	C	c	a	0.488	0.346	0.6373	0.17	0.09	4.5E-02	9.7	3.4E-01	93270	-0.5	1.6E-02	93270
rs1558902* **	16	53,803,574	*FTO*	A	a	t	0.290	0.2006	0.4135	1.44	0.11	2.4E-36	44.6	1.6E-01	60021	34.0	1.7E-02	60021
rs17817964* **	16	53,828,066	*FTO*	T	t	c	0.256	0.1177	0.4037	1.23	0.10	7.1E-36	0	7.1E-01	101976	33.5	5.0E-03	101976
rs889398	16	69,556,715	*NFAT5*	T	c	t	0.682	0.1996	0.4244	0.13	0.09	1.8E-01	55	8.3E-02	93115	-0.4	2.9E-02	93115
rs9914578	17	2,005,136	*SMG6*	G	c	g	0.651	0.4728	0.8023	0.04	0.09	6.7E-01	0	7.2E-01	93666	-1.0	8.0E-03	93666
rs1000940	17	5,283,252	*RABEP1*	G	g	a	0.384	0.3996	0.7629	0.10	0.09	2.8E-01	58.2	6.6E-02	93726	-0.2	1.1E-02	93726
rs4986044	17	21,261,560	*KCNJ12*	T	c	t	0.432	0.4694	0.6543	0.07	0.09	4.2E-01	0	5.3E-01	90325	-0.4	1.1E-02	90325
rs12150665	17	34,914,787	*GGNBP2*	T	t	c	0.681	0.5904	0.8752	0.30	0.10	1.5E-03	12.5	3.3E-01	93726			
rs11652097	17	45,316,717	*ITGB3*	T	c	t	0.634	0.3421	0.3875	0.12	0.09	1.8E-01	64.5	3.8E-02	93646	-0.4	4.4E-02	93646
rs6504108	17	46,292,923	*SKAP1*	T	c	t	0.264	0.6981	0.7958	0.14	0.09	1.5E-01	9.9	3.4E-01	93726			
rs8075273*	17	61,728,881	*unknown*	C	c	a	0.719	0.0872	0.3655	0.12	0.10	2.5E-01	0	6.8E-01	93099	-0.1	1.5E-02	93099
rs312750	17	68,343,539	*KCNJ2*	A	a	g	0.638	0.4891	0.8103	0.09	0.09	3.4E-01	0	8.9E-01	93734	-0.2	8.0E-03	93734
rs12940622	17	78,615,571	*RPTOR*	G	g	a	0.571	0.3349	0.5544	0.28	0.09	9.9E-04	0	5.9E-01	93722			
rs1808579	18	21,104,888	*C18orf8*	C	c	t	0.464	0.4518	0.6622	0.11	0.09	1.9E-01	88.5	*9.3E-06*	93711	2.2	9.9E-01	93711
rs7239883	18	40,147,671	*LOC284260*	G	g	a	0.363	0.5684	0.7229	0.11	0.09	2.0E-01	7.2	3.6E-01	93477			
rs7243357	18	56,883,319	*GRP*	T	t	g	0.822	0.7605	0.8721	0.23	0.11	4.4E-02	48.4	1.2E-01	93737			
rs2331841	18	57,828,637	*MC4R*	A	a	g	0.373	0.2293	0.4857	0.54	0.08	8.6E-11	0	5.3E-01	101933	8.8	8.0E-03	101933
rs6567160	18	57,829,135	*MC4R*	C	c	t	0.193	0.7662	0.8544	0.89	0.10	9.4E-19	13.8	3.2E-01	101966	16.2	1.1E-02	101966
rs591166	18	57,841,589	*MC4R*	A	a	t	0.443	0.231	0.7417	0.41	0.09	1.5E-06	0	7.5E-01	101953	4.3	5.0E-03	101953
rs9944545	18	57,958,244	*MC4R*	T	t	c	0.346	0.1334	0.5432	0.32	0.09	7.3E-04	54.2	8.8E-02	101966	2.1	2.0E-02	101966
rs17066842	18	58,040,624	*MC4R*	G	g	a	0.874	0.0204	0.1731	0.41	0.16	1.1E-02	0	5.7E-01	101605	0.9	1.4E-02	101605
rs17724992	19	18,454,825	*PGPEP1*	A	a	g	0.670	0.5282	0.8874	0.30	0.10	2.6E-03	73.3	1.1E-02	93722	1.5	5.7E-02	93722
rs17513613	19	30,286,822	*CCNE1*	T	c	t	0.200	0.6719	0.8911	0.09	0.12	4.5E-01	0	6.6E-01	93277	-0.4	1.3E-02	93277
rs29941	19	34,309,532	*KCTD15*	G	g	a	0.562	0.1824	0.7617	0.26	0.09	3.9E-03	0	6.2E-01	101951	1.1	4.0E-03	101951
rs2075650	19	45,395,619	*TOMM40*	A	a	g	0.866	0.8308	0.8978	0.22	0.13	8.1E-02	65.2	3.5E-02	93271	0.2	5.1E-02	93271
rs11671664	19	46,172,278	*GIPR, QPCTL*	G	g	a	0.688	0.0936	0.4691	0.40	0.11	1.5E-04	79.9	1.9E-03	101500	3.0	6.6E-01	101500
rs2287019	19	46,202,172	*QPCTL*	C	c	t	0.813	0.1294	0.211	0.65	0.12	1.4E-07	0	5.9E-01	65911	5.6	1.3E-02	65911
rs3810291	19	47,569,003	*ZC3H4*	A														
rs4802349	19	47,874,510	*DHX34*	G	g	t	0.631	0.1035	0.4808	0.28	0.09	2.8E-03	88.4	*1.0E-05*	93581			
rs8123881*	20	15,819,495	*MACROD2*	G	g	a	0.238	0.6429	0.9057	0.25	0.11	2.2E-02	60.2	5.7E-02	93265			
rs6091540	20	51,087,862	*ZFP64*	C	c	t	0.702	0.2247	0.3572	0.18	0.09	5.4E-02	0	4.2E-01	93270			
rs2836754	21	40,291,740	*ETS2*	C	c	t	0.417	0.3746	0.7276	0.22	0.09	1.5E-02	0	8.1E-01	93717	0.9	6.0E-03	93717
rs4820408	22	40,604,945	*TNRC6B*	T	t	g	0.385	0.1071	0.5338	0.25	0.10	8.5E-03	33.7	2.1E-01	93733			

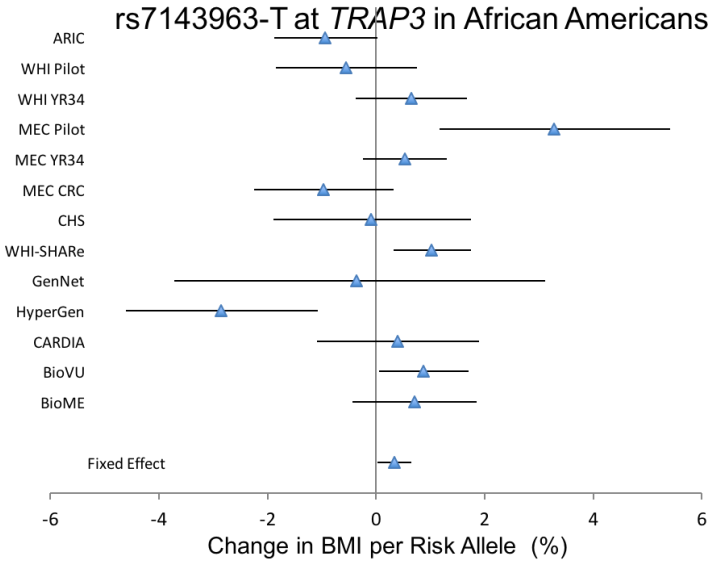
Abbreviations: A1=coded allele, A2=non-coded allele, BF=Bayes Factor, Bp37=base pair Build 37, Chr=chromosome, FE=Fixed-Effect, HetP=heterogeneity p-value, Max=maximum, Min=minimum, P=p-value, Post Prob Het=posterior probability of heterogeneity, Prep=prepared reference, Ref=reference, Sub=submitted reference, SNPs=single nucleotide polymorphisms.

*When the index SNP was not genotyped on the MetaboChip, the proxy SNP in tight linkage disequilibrium ($r^2 \geq 0.8$ in 1000 Genomes pilot 1 CEU, YRI, CHB+JPT depending on the population of discovery) with the lowest p-value in the African American sample was chosen to represent the index signal. The decreasing and increasing alleles for proxies were assigned assuming that the risk index SNP would have a similar allele frequency in the 1000 Genomes population (EUR, AFR, or EAS depending on the discovery population) as the risk proxy SNP.

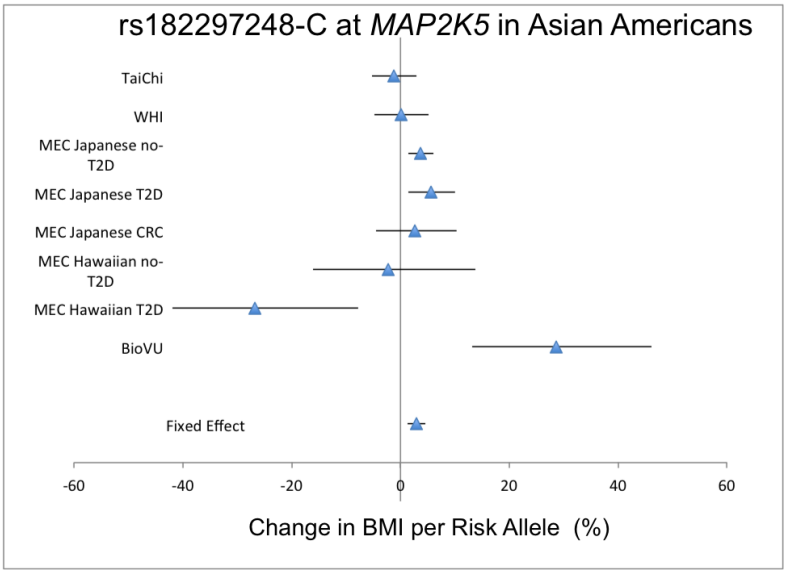
**For GWAS SNPs a Bonferroni correction for multiple tests in the fixed-effect analyses reflected the number of independent previously-reported signals tested (=0.05/166). The 4 noted SNP pairs above were in tight linkage disequilibrium [$r^2 \geq 0.8$ in non-European 1000 genomes pilot

populations(s)] with each other, but because they were reports from distinct discovery populations we retained them in this inventory in case they were population-specific variants. Therefore, our Bonferroni correction was penalized for only 166 (=170-4) tests.
***Bonferroni significant heterogeneity p-values in italics.

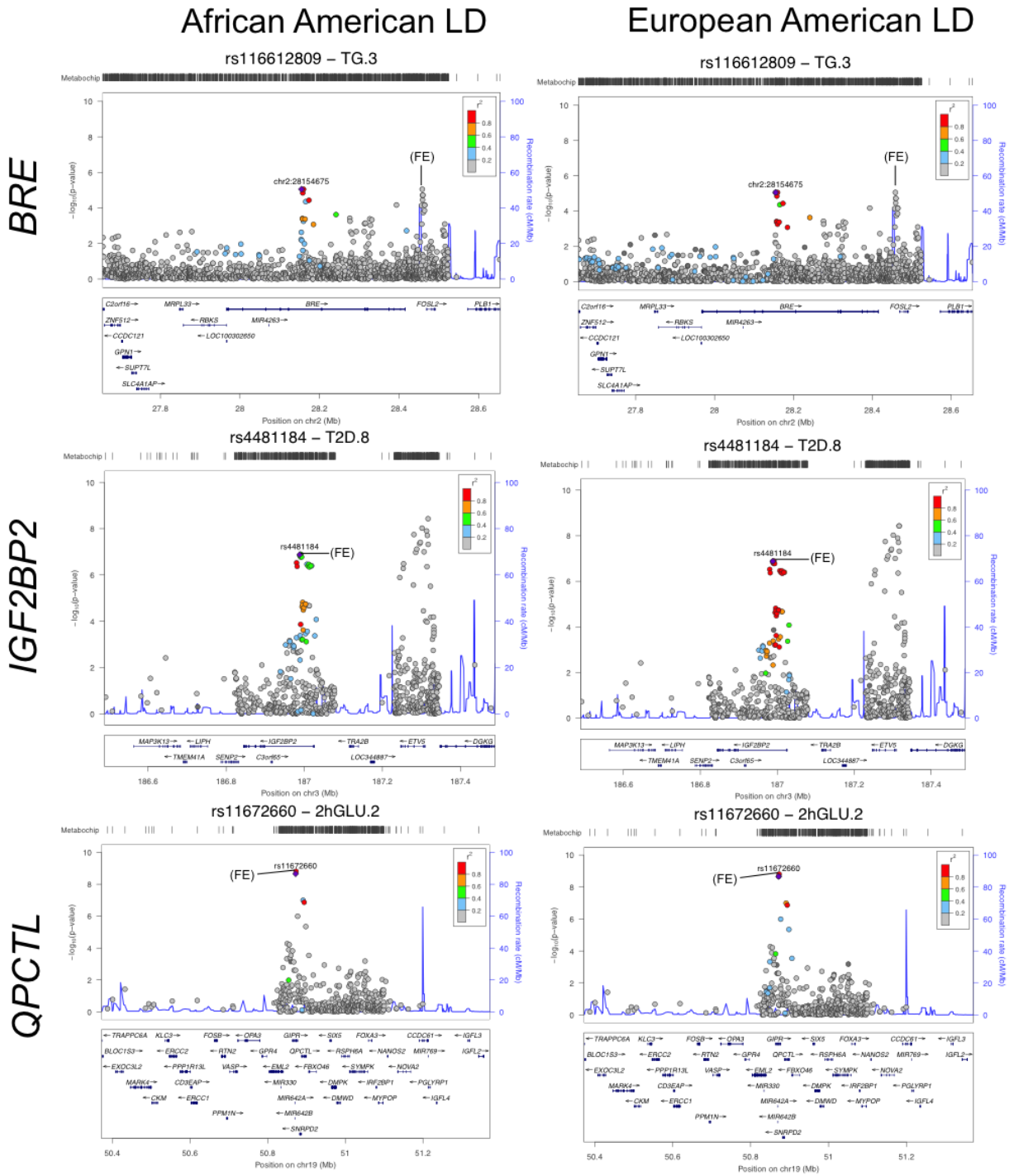
APPENDIX Q: FOREST PLOT OF EFFECT HETEROGENEITY (P_{HET}=1.16X10^{-4}, I^2=69.0) AT *TRAF3* (RS7143963-T, 62% RISK ALLELE FREQUENCY, RANGE 59-67%) IN 35,602 AFRICAN DESCENT ADULTS IN PAGE.



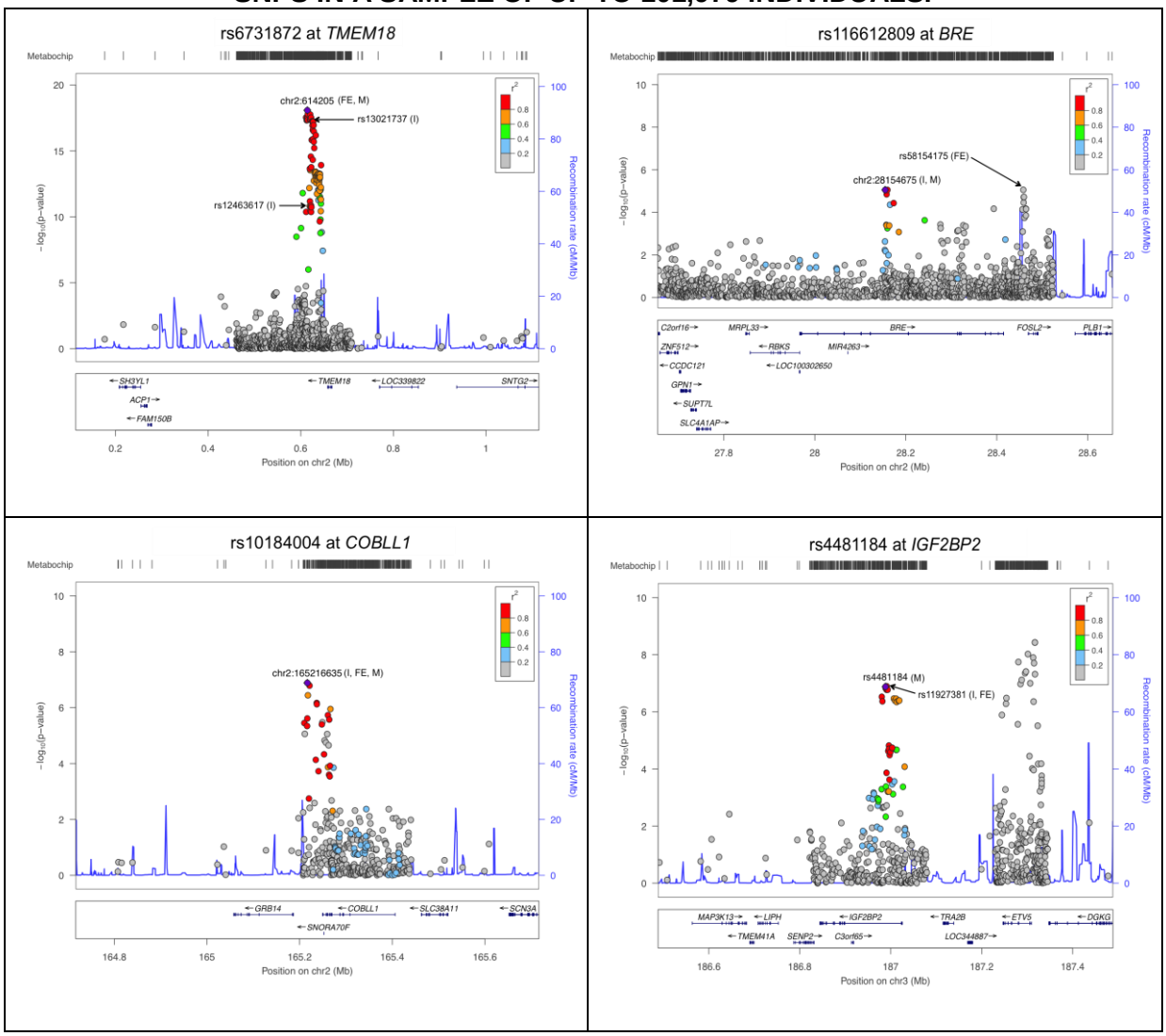
APPENDIX R: FOREST PLOT OF EFFECT HETEROGENEITY (P_{HET}=2.71X10^{-4}, I^2=74.5) AT *MAP2K5* (RS182297248-C, 0.9% RISK ALLELE FREQUENCY, RANGE 0.2-1.2%) IN 21,974 ASIAN DESCENT ADULTS IN PAGE.

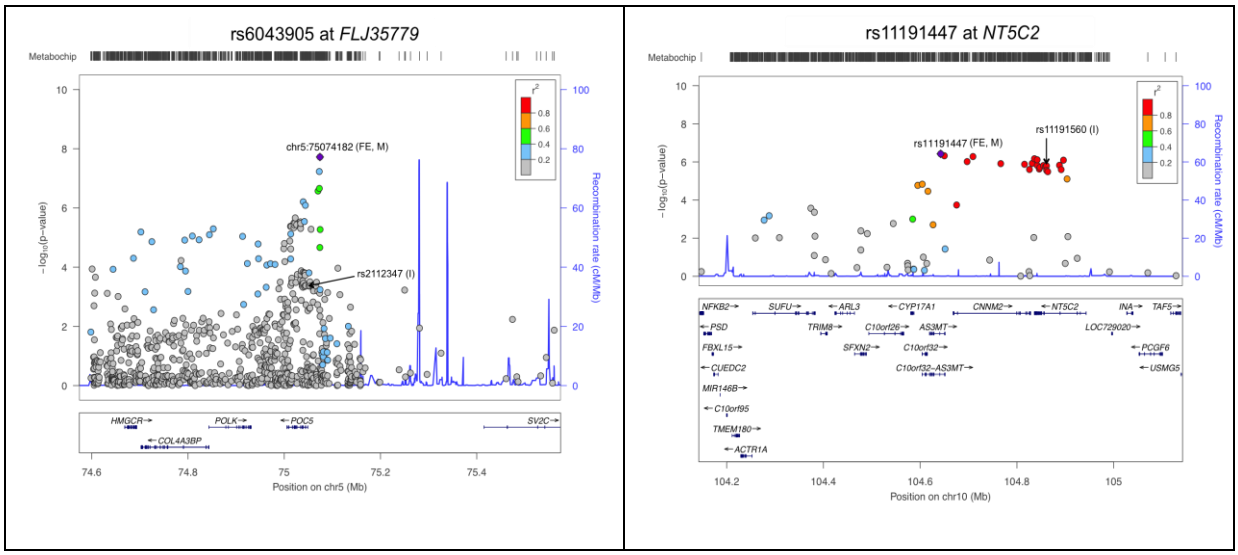


APPENDIX S: COMPARISON OF AFRICAN AND EUROPEAN DESCENT LINKAGE DISEQUILIBRIUM (LD) PATTERNS AT THREE LOCI WITH SIGNIFICANT TRANS-ETHNIC FIXED-EFFECT (FE, -LOG10 P-VALUES SHOWN HERE) ESTIMATES AND DIFFERING TOP SNPS IN THE FIXED-EFFECT AND BAYESIAN TRANS-ETHNIC META-ANALYSES (LABELED AND SHOWN IN PURPLE, REFERENCE FOR LD).

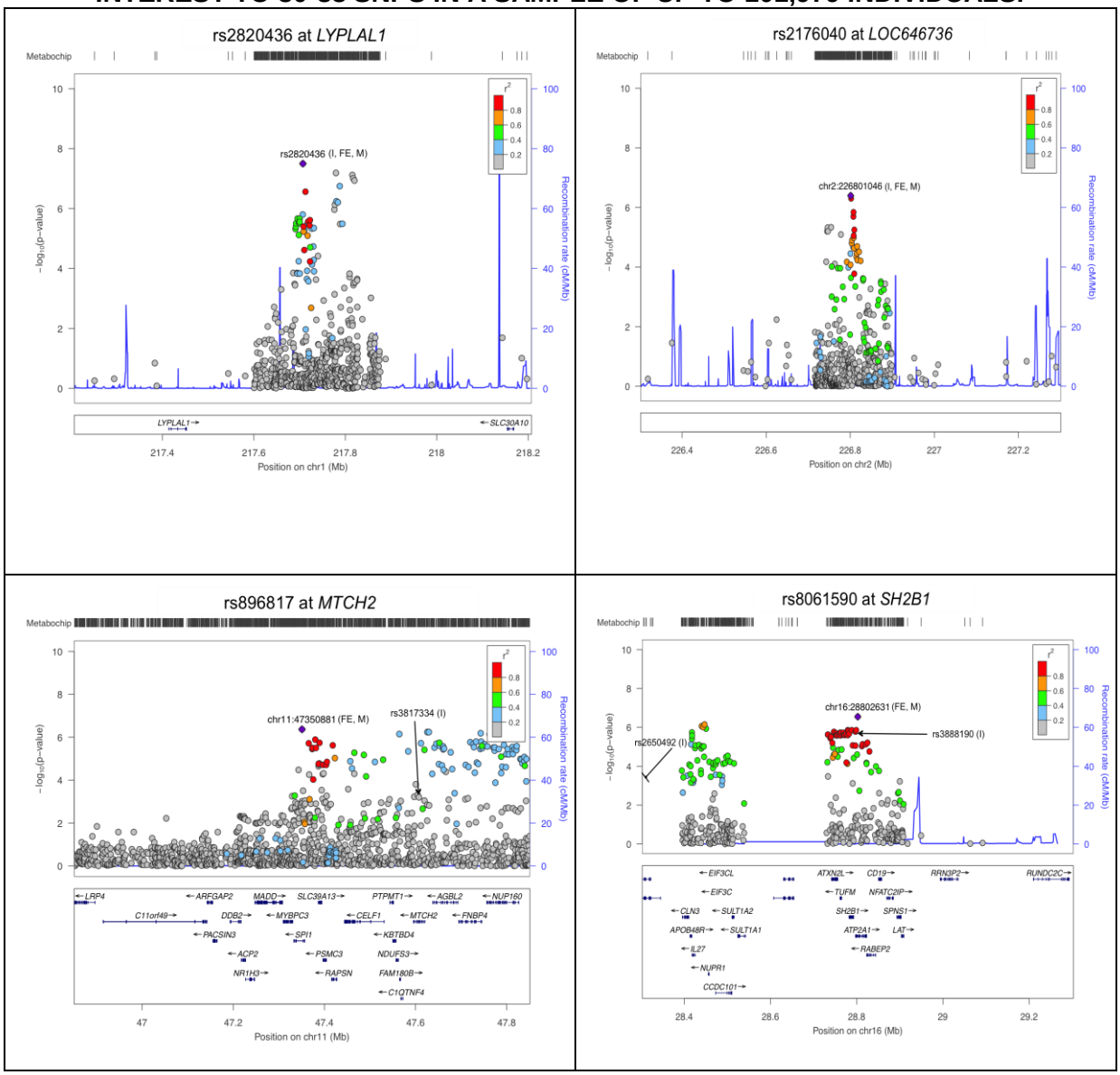


APPENDIX T: REGIONAL PLOTS OF TRANS-ETHNIC FIXED-EFFECT ESTIMATES (I, INDEX SNPS; FE, TOP FINDING) AND BAYESIAN FINE-MAPPING OF 6 SIGNIFICANT BMI LOCI TO SELECT THE SNP WITH THE HIGHEST POSTERIOR PROBABILITY (M, SHOWN IN PURPLE AND REFERENCE FOR TRANS-ETHNIC LINKAGE DISEQUILIBRIUM) AND NARROW THE PUTATIVE INTERVAL OF INTEREST TO 13-29 SNPS IN A SAMPLE OF UP TO 101,979 INDIVIDUALS.

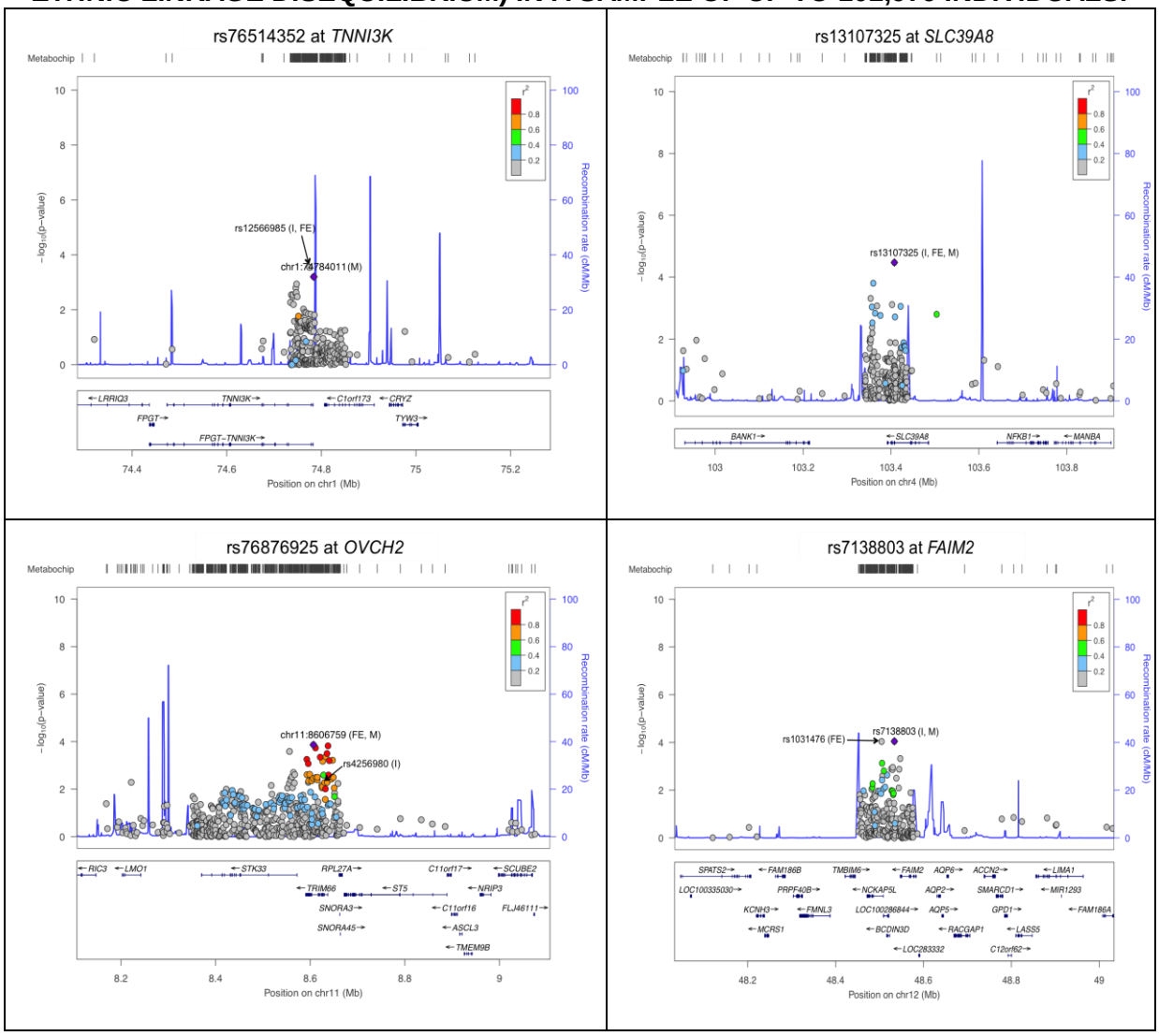


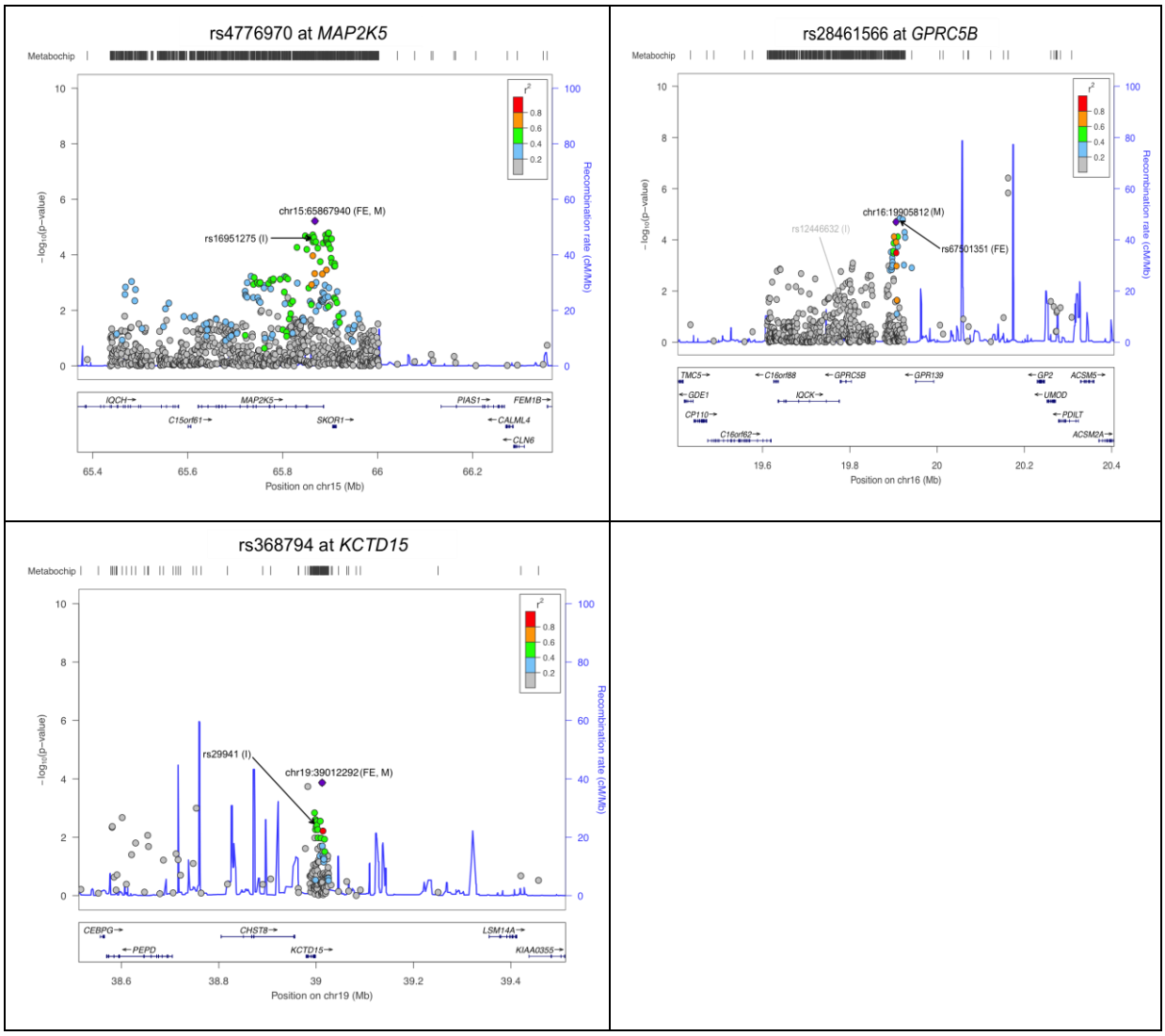

rs6043905 at FLJ35779
Metabochip
chr5:75074182 (FE, M)
rs2112347 (I)
−log10(p-value)
Recombination rate (cM/Mb)
HMGCR
POLK
POC5
SV2C
COL4A3BP
Position on chr5 (Mb)
rs11191447 at NT5C2
rs11191560 (I)
NFKB2
SUFU
ARL3
CYP17A1
CNNM2
NT5C2
INA
TAF5
PSD
TRIM8
C10orf26
AS3MT
LOC729020
FBXL15
SFXN2
C10orf32
PCGF6
CUEDC2
C10orf32-AS3MT
USMG5
MIR146B
C10orf95
TMEM180
ACTR1A
Position on chr10 (Mb)

APPENDIX U: REGIONAL PLOTS OF TRANS-ETHNIC FIXED-EFFECT ESTIMATES (I, INDEX SNPS; FE, TOP FINDING) AND BAYESIAN FINE-MAPPING OF 4 BMI SIGNIFICANT LOCI TO SELECT THE SNP WITH THE HIGHEST POSTERIOR PROBABILITY (M, SHOWN IN PURPLE AND REFERENCE FOR TRANS-ETHNIC LINKAGE DISEQUILIBRIUM) AND TO NARROW THE PUTATIVE INTERVAL OF INTEREST TO 30-88 SNPS IN A SAMPLE OF UP TO 101,979 INDIVIDUALS.

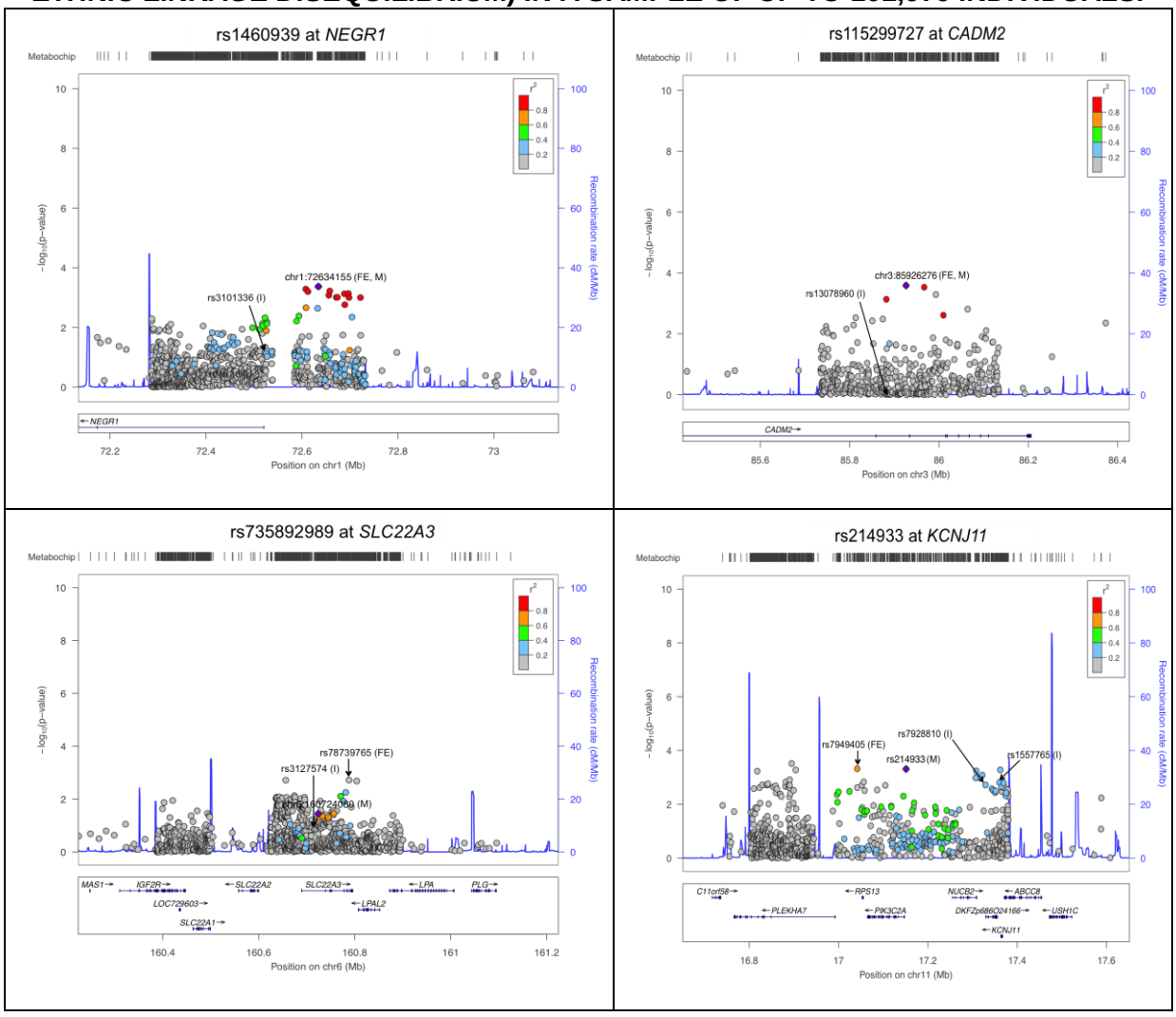


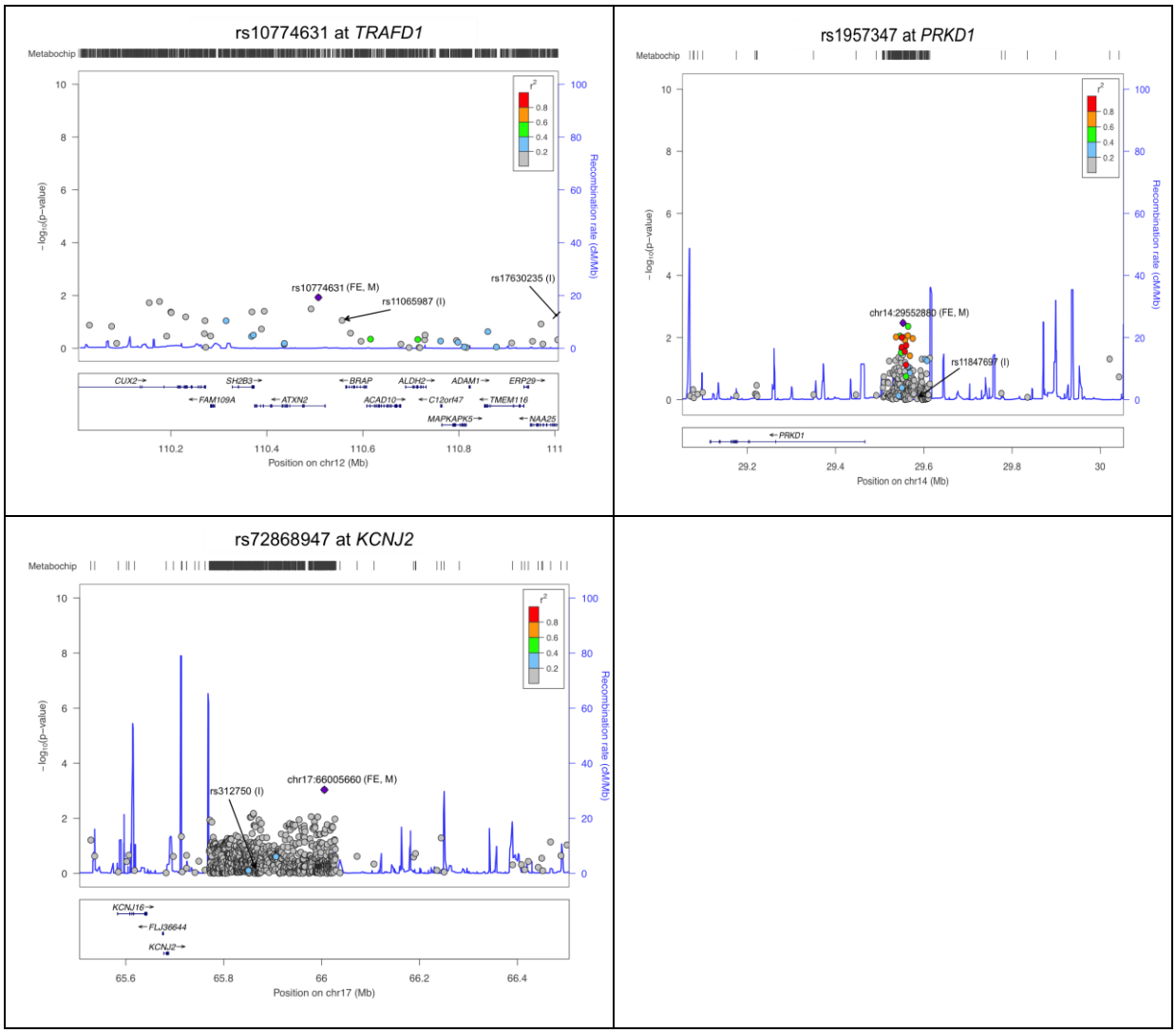
APPENDIX V: REGIONAL PLOTS OF 7 BMI LOCI WITH SIGNIFICANT TRANS-ETHNIC FIXED-EFFECT ESTIMATES (I, INDEX SNPS; FE, TOP FINDING; M, HIGHEST POSTERIOR PROBABILITY, SHOWN IN PURPLE AND REFERENCE FOR TRANS-ETHNIC LINKAGE DISEQUILIBRIUM) IN A SAMPLE OF UP TO 101,979 INDIVIDUALS.



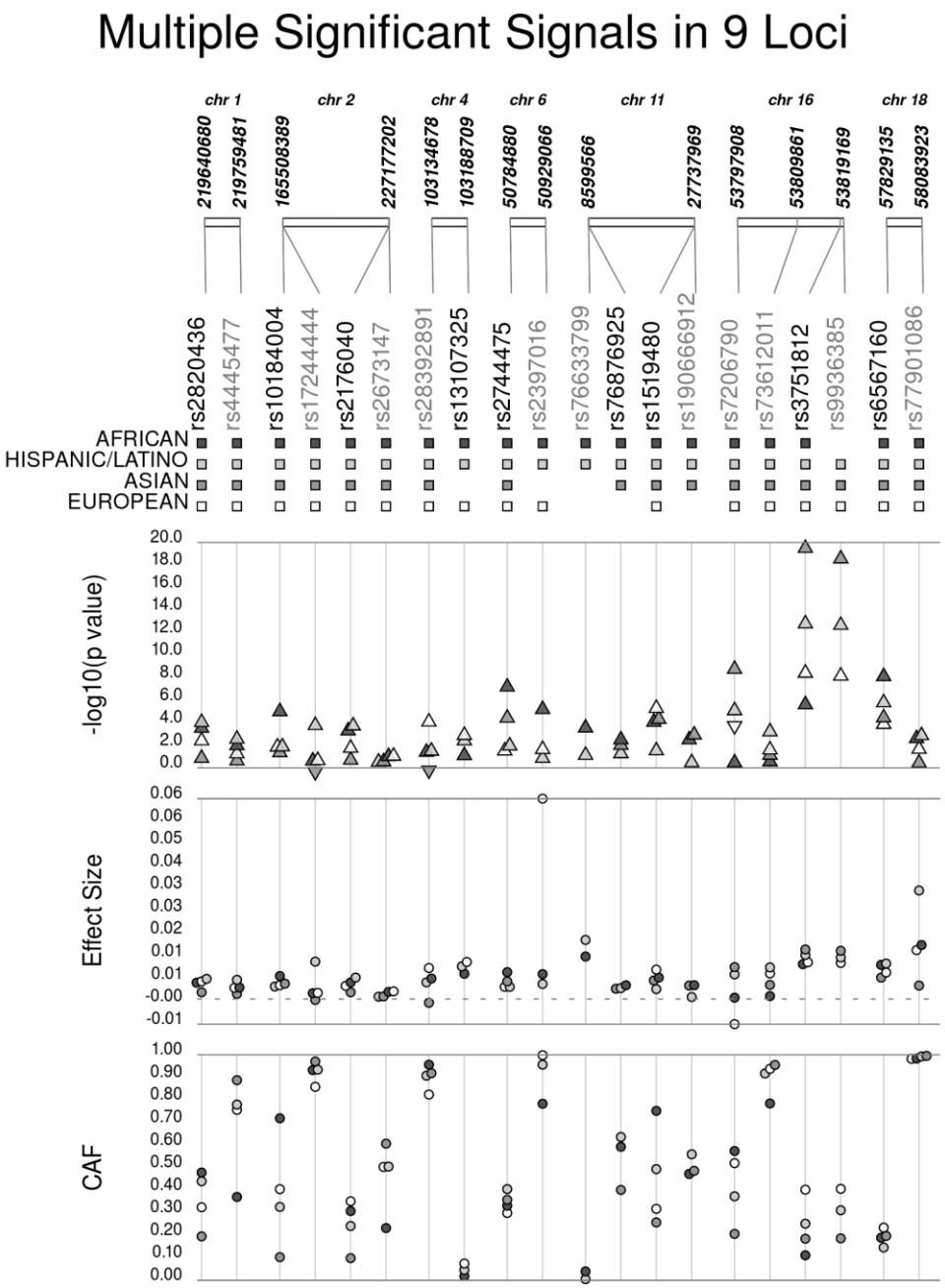

rs4776970 at MAP2K5
Metabochip
chr15:65867940 (FE, M)
rs16951275 (I)
−log10(p−value)
Recombination rate (cM/Mb)
IQCH
MAP2K5
PIAS1
FEM1B
C15orf61
SKOR1
CALML4
CLN6
Position on chr15 (Mb)
rs28461566 at GPRC5B
chr16:19905812 (M)
rs12446632 (I)
rs67501351 (FE)
TMC5
C16orf88
GPRC5B
GPR139
GP2
ACSM5
GDE1
IQCK
UMOD
CP110
PDILT
C16orf62
ACSM2A
Position on chr16 (Mb)
rs368794 at KCTD15
rs29941 (I)
chr19:39012292 (FE, M)
CEBPG
CHST8
LSM14A
PEPD
KCTD15
KIAA0355
Position on chr19 (Mb)

APPENDIX W: REGIONAL PLOTS OF 7 BMI LOCI WITH NON-SIGNIFICANT TRANS-ETHNIC FIXED-EFFECT ESTIMATES (I, INDEX SNPS; FE, TOP FINDING; M, HIGHEST POSTERIOR PROBABILITY, SHOWN IN PURPLE AND REFERENCE FOR TRANS-ETHNIC LINKAGE DISEQUILIBRIUM) IN A SAMPLE OF UP TO 101,979 INDIVIDUALS.




rs10774631 at TRAFD1
Metabochip
rs10774631 (FE, M)
rs11065987 (I)
rs17630235 (I)
CUX2
SH2B3
BRAP
ALDH2
ADAM1
ERP29
FAM109A
ATXN2
ACAD10
C12orf47
TMEM116
MAPKAPK5
NAA25
Position on chr12 (Mb)
rs1957347 at PRKD1
chr14:29552880 (FE, M)
rs11847697 (I)
PRKD1
Position on chr14 (Mb)
rs72868947 at KCNJ2
chr17:66005660 (FE, M)
rs312750 (I)
KCNJ16
FLJ36644
KCNJ2
Position on chr17 (Mb)
Recombination rate (cM/Mb)

APPENDIX X: THE COMPARISON OF THE STATISTICAL SIGNIFICANCE (-LOG10 OF THE P-VALUE), EFFECT SIZE (% CHANGE IN BMI PER RISK ALLELE) AND CODED ALLELE FREQUENCIES (ORIENTED TO THE RISK ALLELE IN THE TRANS-ETHNIC META-ANALYSIS) ACROSS AFRICAN, HISPANIC/LATINO, ASIAN AND EUROPEAN ANCESTRIES OF 9 DENSELY-GENOTYPED REGIONS WITH EVIDENCE OF MULTIPLE SIGNALS (RSID OF PRIMARY SIGNALS IN BLACK; RSID OF OTHER SIGNALS IN GRAY).



APPENDIX Y: FUNCTIONAL ANNOTATION IN HAPLOREG (VERSION 4.1) OF THE LEAD SNPS REPRESENTING MULTIPLE LOCUS-SPECIFIC BONFERRONI SIGNALS IN A JOINT FIXED-EFFECTS MODEL AT THE 28 OF THE 36 DENSELY-GENOTYPED BMI LOCI.

Gene	rsID	Chr	Bp37	Enhancer	DNAse	Proteins Bound	Motifs Changed	GWAS catalog report ($P<5x10^{-8}$)	Selected eQTL Hits (Correleated gene [tissue, Pvalue])	GENCODE Genes	RefSeq Genes	In-tron	NS	r^2>0.8 with NS
TNNI3K	rs12566985	1	75,002,193				AP-1, TATA	BMI		FPGT-TNNI3K	FPGT-TNNI3K	x		
SEC16B	rs543874	1	177,889,480	MUS			Pbx-1, TCF4	BMI, age at menarche		3.6kb 3' of RP4-798P15.2	8.8kb 3' of SEC16B			
LYPLAL1*	rs2820436	1	219,640,680	FAT, STRM, SKIN, BONE			MZF1::1-4,PLAG1,TFIIA	Waist to hip ratio	RP11-392O17.1 [Skin-Leg,P=6.1E-9]	25kb 5' of RP11-95P13.1	254kb 3' of LYPLAL1			
LYPLAL1*	rs4445477	1	219,759,481	ESDR, BRST, BLD, FAT, GI, ADRL, PLCNT, HRT, SPLN	SKIN,ADRL,PLCN		Ets,HEN1,LBP-1,RFX5	Waist to hip ratio		28kb 3' of RP11-95P13.2+W6:AJ7	287kb 5' of RNU5F			
TMEM18	rs6731872	2	624,205				BCL,TR4	BMI		43kb 3' of TMEM18	44kb 3' of TMEM18			
COBLL1*	rs10184004	2	165,508,389		SKIN	MAFK	ATF4,Maf,NF-E2,Nrf-2,PLZF,RREB-1	Waist to hip ratio	SLC38A11 [Muscle_Skeletal, P=5.19E-6]	1.7kb 3' of COBLL1	30kb 5' of GRB14			
COBLL1*	rs17244444	2	165,548,415	ESDR, LNG, LIV			ELF1,NERF1a, Nrf-2,p300	Waist to hip ratio		COBLL1	COBLL1	x		
LOC646736*	rs2176040	2	227,092,802	ESDR, LIV			Nr2e3,Pax-2		RP11-395N3.2 [Adipose_Subcutaneous, P=5.35E-8], IRS1 [Adipose_Subcutaneous, P=3.78E-6]	43kb 5' of AC068138.1	503kb 3' of IRS1			
LOC646736*	rs2673147	2	227,177,202	BRN, LNG	LNG		Foxp3,RXRA,SIX5		RP11-395N3.2 [Adipose_Subcutaneous, P=1.46E-5], IRS1 [Adipose_Subcutaneous, P=1.42E-5]	127kb 5' of AC068138.1	419kb 3' of IRS1			
IGF2BP2*	rs11927381	3	185,508,591	ESDR, IPSC, ESC, FAT, BRST, STRM, BLD, MUS, SKIN, ADRL, LIV, VAS, BRN, BONE				Type 2 diabetes	IGF2BP2 [Thyroid, P=23.4E06]	IGF2BP2	IGF2BP2	x		

ETV5	rs7647305	3	185,834,290				Cdx2,Foxj1,Foxl1,Foxp1,Gfi1,Pou1f1,RREB-1,TATA	BMI, Weight		DGKG	7.4kb 5' of ETV5		
GNPDA2	rs12507026	4	45,181,334		BRST,SKIN	JUND	HNF1,Hic1,RXRA,TCF12	BMI		128kb 3' of RP11-362I1.1	453kb 5' of GNPDA2		
SLC39A8*	rs28392891	4	103,134,678				Nkx3,Pou5f1,Sox,p300	BMI		38kb 3' of SLC39A8	38kb 3' of SLC39A8		
SLC39A8*	rs13107325	4	103,188,709				Arid5a,PRDM1,Pax-1	HDL cholesterol,BMI, diastolic & systolic blood pressure, hypertension		SLC39A8	SLC39A8		x
FLJ35779	rs60493905	5	75,038,426	THYM			Duxl,GATA,Lhx4,Lhx8,Myf	BMI	POC5 [Lymphoblastoid,P=2.33E-8],ANKDD1B [Thyroid,P=1.46E-6]	25kb 5' of POC5	25kb 5' of POC5		
CDKAL1*	rs67131976	6	20,686,878		ESC, IPSC, BLD		AP-1	BMI		CDKAL1	CDKAL1	x	
TFAP2B	rs2744475	6	50,784,880	ESDR, ESC	ESDR,GI		EWSR1-FLI1,GATA,HDAC2,Irf,PRDM1,TATA	BMI		1.6kb 5' of TFAP2B	1.6kb 5' of TFAP2B		
TFAP2B	rs2397016	6	50,929,066		BLD		Cdx2,PLZF,STAT			114kb 3' of TFAP2B	114kb 3' of TFAP2B		
LRPN6C	rs17770336	9	28,414,625				GCM,Mef2,TEF	BMI		LINGO2	LINGO2	x	
NT5C2*	rs11191447	10	104,652,323					BMI	31 hits - only relevant ones listed here: MARCKSL1P1 [Adipose_Subcutaneous, P=6.57E-7], MARCKSL1P1 [Adrenal_Gland, P=8.0E-6], MIR1307 [Muscle_Skeletal,P=4.06E-10]; MIR1307 [Thyroid,P=1.08E-7]	AS3MT	C10orf32-AS3MT	x	
TCF7L2*	rs7903146	10	114,758,349	FAT, BRST, MUS, BRN, GI, LNG, OVRY			Dbx1,Dobox4,E4BP4,HLF,PLZF,Pou3f2,TATA	Type 2 diabetes, Metabolic Syndrome, Fasting glucose, HbA1c, Proinsulin		TCF7L2	TCF7L2	x	

								levels,				
KCNQ1*	rs2237896	11	2,858,440	ESC, ESDR, LNG, IPSC, FAT, BRST, BLD, SKIN, GI, ADRL, HRT, MUS, PLCNT, THYM, PANC, SPLN	ESC,ESDR,ESDR,ESDR,ESC, LNG,IPSC,IPSC,BLD,SKIN,SKIN,ADRL,HRT, GI,GI,KID,LNG, MUS,MUS,PLCNT,GI,THYM,GI,OVRY,PANC, MUS,GI,LNG,CRVX,LIV,MUS, MUS,BLD,BLD, SKIN,LNG	CTCF,AP2ALPHA,AP2GAMMA,POL2,RAD21,SMC3	AP-2,MAZ,Pbx3,SRF,YY1	BMI		KCNQ1	KCNQ1	x
OVCH2	rs76633799	11	8,599,566	FAT, SKIN	LNG,BRST,SKIN,SKIN,SKIN,ADRL,CRVX,BRST,MUS,SKIN, LNG		Foxa,HDAC2,TCF12	BMI		STK33	STK33	x
OVCH2	rs76876925	11	8,650,183				Foxp1,HDAC2	BMI		TRIM66	TRIM66	x
BDNF	rs1519480	11	27,675,712	BRN	ESDR,LNG,BRN,LNG	GATA2, YY1	Gfi1,Hoxb8,SP1	BMI	LIN7C[Thyroid, P=7.88E-6],BDNF-AS1[Lung,P=1.82E-8]	BDNF Antisense RNA	BDNF-AS1	x
BDNF	rs190666912	11	27,737,969	LNG, STRM, BRN	ESDR,SKIN,BRN,BRN		Fox,Foxp1	BMI		BDNF	BDNF	x
MTCH2	rs896817	11	47,394,305	BLD, SKIN	ESDR,SKIN,BRN,BRN			BMI	16 hits, relevants ones listed here: C1QTNF4[Adipose_Subcutaneous,P=7.1E-6], RP11-750H9.5[Brain_Cerebellar_hemisphere,P=2.95E-7], PSMC3[Cerebellum,P=3.9E-6]	SPI1	SPI1	x
FAIM2	rs7138803	12	50,247,468	ESC			ERalpha-a,NR4A,RAR,SF1	BMI, Weight, Waist circumference, Obesity, Age at menarche	GAR1[Blood,P=6.3E-9], HOXD13[Blood, P=4.38E-6]	7.5kb 5' of RP11-70F11.7	11kb 5' of BCDIN3D	
MAP2K5	rs4776970	15	68,080,886	IPSC	IPSC		DMRT5,Foxa,Irf,Mef2,PPAR,Pax-5,RXRA,STAT, ZEB1,p300	BMI	7 hits, relevant ones listed here: SKOR1[Adipose_Subcutaneous, P=8.96E-8], SKOR1[Muscle_Skeletal,P=8.73E-6], MAP2K5[Whole_Blood,P=9	MAP2K5	MAP2K5	x

								.38E-5]				
GPRC5B/ GP2	rs67501351	16	20,006,745			Foxm1,Obox6, Pax-5,Pax-8,Pbx3			36kb 3' of GPR139	36kb 3' of GPR139		
SH2B1	rs8061590	16	28,895,130	ESC, ESDR, IPSC	IPSC,IPSC,MUS,THYM		Ets,Zfp410	4 hits	191 hits, only relevant ones listed here, all results listed have Pvalues<5E-6:CDC37P1,EIP3C,EPI3CL,RP11-1348G14.4,SH2B1,SULT1A2,TUFM[Adipose_Subcutaneous], CDC37P1,RP11-1348G14.4,SH2B1,SULT1A2 [Adipose_Visceral_Omentum], RP11-1348G14.4 [Brain_cerebellar_hemisphere], TUFM[Brain_Hippocampus, Brain_Nucleus_basal_ganglia], CCDC101,LAT,RP11-1348G14.4,SULT1A1,SULT1A2[Muscle_Skeletal]	ATP2A1	ATP2A1	x
FTO	rs7206790	16	53,797,908	ESDR, BRST, STRM, BRN, BLD, MUS, LNG, LIV	IPSC,BLD		Irf,Nkx3,PRDM1,SETDB1,STAT			FTO	FTO	x
FTO	rs73612011	16	53,809,861	FAT, BRST, BLD, STRM, MUS, BRN, SKIN, LIV, GI, HRT, ADRL, ESC, ESDR, LNG, IPSC, PANC, PLCNT, OVRY, BONE	ESDR,LNG,CRVX	FOXA1,FOXA2	Arid5a,DMRT2, Foxa,Foxc1,Fox k1,HDAC2,Pax-4,Pou2f2,Pou3f 2,Sox,p300			FTO	FTO	x
FTO	rs3751812	16	53,818,460	ESDR, LNG, STRM, BRST, BLD, SKIN, BRN, PANC, CRVX, LIV, MUS, BONE	ESDR,BRST,SKIN,BRST,BRN,SKIN,LNG		Mrg,TBX5,Tgif1	BMI		FTO	FTO	x
FTO	rs9936385	16	53,819,169	FAT, STRM, BRST, MUS,	LNG,BRST,SKIN,HRT,GI,THY		HDAC2,Pax-5	Type 2 diabetes		FTO	FTO	x

				SKIN, PANC, LNG, BONE	M,BRST,MUS, BRN,LNG								
MC4R	rs6567160	18	57,829,135	BLD	MUS		Hoxb13,Hoxb9, Hoxd10,Mef2,P ou5f1	BMI, Fat body mass		1.7kb 5' of U4	209kb 3' of MC4R		
MC4R	rs77901086	18	58,083,923				Foxp1,HDAC2, Sin3Ak-20	BMI		44kb 5' of MC4R	44kb 5' of MC4R		x
KCTD15	rs368794	19	34,320,452				HMG-IY,HP1-site-factor,Pax-4,Pax-6,Zfp105	BMI		14kb 3' of KCTD15	14kb 3' of KCTD15		
QPCTL*	rs1800437	19	46,181,392	ESDR, BRST, BLD, STRM, BRN, FAT, LIV, GI, HRT, MUS, THYM, LNG, PLCNT, SPLN, VAS	ESC,ESDR,ESC,IPSC,IPSC,BLD,BLD,BLD,BLD,BLD,BLD,BLD,SKIN,SKIN,SKIN,SKIN,HRT,GI,KID,LNG,PLCNT,GI,THYM,GI,PANC,GI,LNG,LIV,MUS,BLD,SKIN,LNG	CTCF,CMYC	BDP1,CTCF,Lmo2-complex,Myf,Rad21,SMC3,TAL1,TCF12	BMI	FBXO46[Whole_blood, P=5.44E-6],VASP[Whole_blood,P=2.75E-10]	GIPR	GIPR	x	

General Abbreviations: Bp37=base pair Build 37, Chr=chromosome, GWAS=Genome-wide association study, NS=Non-synonymous mutation, SNPs=single nucleotide polymorphisms.
Tissue Abbreviations: BLD [blood], BRN [brain], MUS [skeletal muscle], FAT [adipose], GI [digestive], HRT [heart], LIV [liver], LNG [lung], STRM [Mesenchymal Stem Cell], BRST [breast], SKIN [skin, epithelial], THYM [thymus], ADRL [adrenal gland], ESC [embryonic stem cells], ESDR [embryonic stem cells, derived from iPSC cells], IPSC [induced pluripotent stem cells], PANC [pancreas], PLCNT [placenta], OVRY [ovary], BONE [osteoblast].
*Note: Starred genes represent fine-mapped loci, which were associated with BMI after the design of the MetaboChip in 2009.
**PAGE trans-ethnic discovery signal (Gong *et al.,* submitted to *Nature Communications*).
***Total SNPs in signal (r^2>0.8 in 1000 Genomes AFR) including the queried SNP.

APPENDIX Z: SUPPLEMENTAL ACKNOWLEDGEMENTS OF MANUSCRIPT 2

LFR was supported by the Cardiovascular Disease Epidemiology Training Grant from the National Heart, Lung, and Blood Institute (T32HL007055) and the American Heart Association (AHA) predoctoral grant (13PRE16100015). RHM was supported by the Women's Health Initiative Regional Field Center Program HHSN268201100002C. KEN was supported by R01-DK089256; 2R01HD057194; U01HG007416; R01DK101855, and AHA grant 13GRNT16490017.

The Population Architecture Using Genomics and Epidemiology (PAGE) program is funded by the National Human Genome Research Institute (NHGRI), supported by U01HG004803 (CALiCo), U01HG004798 (EAGLE), U01HG004802 (MEC), U01HG004790 (WHI), and U01HG004801 (Coordinating Center), and their respective NHGRI ARRA supplements. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The complete list of PAGE members can be found at PAGE website (http://www.pagestudy.org). The data and materials included in this report result from a collaboration between the following studies:

The "Epidemiologic Architecture for Genes Linked to Environment (EAGLE)" at Vanderbilt University is funded through the NHGRI PAGE program (U01HG004798 and its NHGRI ARRA supplement). The dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center's EAGLE BioVU, which is supported by institutional funding and by the Vanderbilt CTSA grant UL1 TR000445 from NCATS/NIH. The Vanderbilt University Center for Human Genetics Research, Computational Genomics Core provided computational and/or analytical support for this work.

The Multiethnic Cohort study (MEC) characterization of epidemiological architecture is funded through the NHGRI PAGE program (U01HG004802 and its NHGRI ARRA supplement). The MEC study is funded through the National Cancer Institute (R37CA54281, R01 CA63, P01CA33619, U01CA136792, and U01CA98758).

Funding support for the “Epidemiology of putative genetic variants: The Women’s Health Initiative” study is provided through the NHGRI PAGE program (U01HG004790 and its NHGRI ARRA supplement). The WHI program is funded by the National Heart, Lung, and Blood Institute; NIH; and U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf.

Funding support for the Genetic Epidemiology of Causal Variants Across the Life Course (CALiCo) program was provided through the NHGRI PAGE program (U01HG004803 and its NHGRI ARRA supplement). The following studies contributed to this manuscript and are funded by the following agencies: The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC- 55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022. The Coronary Artery Risk Development in Young Adults (CARDIA) study is supported by the following National Institutes of Health, National Heart, Lung and Blood Institute contracts: N01-HC-95095; N01-HC-48047; N01-HC-48048; N01-HC-48049; N01-HC-48050; N01-HC-45134; N01-HC-05187; and N01-HC-45205. The Cardiovascular Health Study (CHS) is supported by contracts HHSN268201200036C, N01-HC-85239, N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01--HC-55222, N01-HC-75150, N01-HC-45133, and grant HL080295 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG-023629, AG-15928, AG-20098, and AG-027058 from the National Institute on Aging (NIA).

Assistance with phenotype harmonization, SNP selection and annotation, data cleaning, data management, integration and dissemination, and general study coordination was provided by the PAGE Coordinating Center (U01HG004801-01 and its NHGRI ARRA supplement). The National Institutes of Mental Health also contributes to the support for the Coordinating Center.

The PAGE Study thanks the staff and participants of all studies for their important contributions. The authors also gratefully acknowledge Dr. Ben Voight for sharing the MetaboChip linkage disequilibrium information.

REFERENCES

1. Popkin, B.M., *Recent dynamics suggest selected countries catching up to US obesity.* Am J Clin Nutr, 2009. **91**(1): p. 284S-288S.

2. Flegal, K.M., et al., *Prevalence of Obesity and Trends in the Distribution of Body Mass Index Among US Adults, 1999-2010.* JAMA, 2012.

3. Ogden, C.L., et al., *Prevalence of Obesity and Trends in Body Mass Index Among US Children and Adolescents, 1999-2010.* JAMA, 2012.

4. Ennis, S.R., M. Ríos-Vargas, and N.G. Albert, *The Hispanic Population: 2010*, U.S.C. Bureau, Editor 2011, U.S. Government Printing Office: Washington, DC.

5. Singh, G.K., et al., *Dramatic Increases in Obesity and Overweight Prevalence and Body Mass Index Among Ethnic-Immigrant and Social Class Groups in the United States, 1976-2008.* Journal of Community Health, 2011. **36**(1): p. 94-110.

6. Daviglus, 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, 2012. **308**(17): p. 1775-84.

7. Delavari, M., et al., *Acculturation and obesity among migrant populations in high income countries - a systematic review.* BMC Public Health, 2013. **13**.

8. Oza-Frank, R. and S.A. Cunningham, *The weight of US residence among immigrants: a systematic review.* Obes Rev, 2010. **11**(4): p. 271-80.

9. Thomson, M.D. and L. Hoffman-Goetz, *Defining and measuring acculturation: a systematic review of public health studies with Hispanic populations in the United States.* Soc Sci Med, 2009. **69**(7): p. 983-91.

10. Antecol, H. and K. Bedard, *Unhealthy assimilation: why do immigrants converge to American health status levels?* Demography, 2006. **43**(2): p. 337-60.

11. Van Hook, J., et al., *Canaries in a coalmine: Immigration and overweight among Mexican-origin children in the US and Mexico.* Soc Sci Med, 2012. **74**(2): p. 125-34.

12. Hjelmborg, J., et al., *Genetic influences on growth traits of BMI: a longitudinal study of adult twins.* Obesity (Silver Spring), 2008. **16**(4): p. 847-52.

13. Maes, H.H., M.C. Neale, and L.J. Eaves, *Genetic and environmental factors in relative body weight and human adiposity.* Behav Genet, 1997. **27**(4): p. 325-51.

14. Croteau-Chonka, D.C., et al., *Genome-wide association study of anthropometric traits and evidence of interactions with age and study year in Filipino women.* Obesity (Silver Spring), 2011. **19**(5): p. 1019-27.

15. Fox, C.S., et al., *Genome-wide association to body mass index and waist circumference: the Framingham Heart Study 100K project.* BMC Med Genet, 2007. **8 Suppl 1**: p. S18.

16. Frayling, T.M., et al., *A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity.* Science, 2007. **316**(5826): p. 889-94.

17. Johansson, A., et al., *Linkage and genome-wide association analysis of obesity-related phenotypes: association of weight with the MGAT1 gene.* Obesity (Silver Spring), 2010. **18**(4): p. 803-8.

18. Liu, J.Z., et al., *Genome-wide association study of height and body mass index in Australian twin families.* Twin Res Hum Genet, 2010. **13**(2): p. 179-93.

19. Loos, R.J., et al., *Common variants near MC4R are associated with fat mass, weight and risk of obesity.* Nat Genet, 2008. **40**(6): p. 768-75.

20. Ng, M.C., et al., *Genome-Wide Association of BMI in African Americans.* Obesity (Silver Spring), 2011.

21. Speliotes, E.K., et al., *Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index.* Nat Genet, 2010. **42**(11): p. 937-48.

22. Thorleifsson, G., et al., *Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity.* Nat Genet, 2009. **41**(1): p. 18-24.

23. Tonjes, A., et al., *Association of FTO variants with BMI and fat mass in the self-contained population of Sorbs in Germany.* Eur J Hum Genet, 2010. **18**(1): p. 104-10.

24. Willer, C.J., et al., *Six new loci associated with body mass index highlight a neuronal influence on body weight regulation.* Nat Genet, 2009. **41**(1): p. 25-34.

25. Wen, W., et al., *Meta-analysis identifies common variants associated with body mass index in east Asians.* Nat Genet, 2012. **44**(3): p. 307-11.

26. Okada, Y., et al., *Common variants at CDKAL1 and KLF9 are associated with body mass index in east Asian populations.* Nat Genet, 2012. **44**(3): p. 302-6.

27. Wang, K.S., et al., *A novel locus for body mass index on 5p15.2: a meta-analysis of two genome-wide association studies.* Gene, 2012. **500**(1): p. 80-4.

28. Yang, J., et al., *FTO genotype is associated with phenotypic variability of body mass index.* Nature, 2012. **490**(7419): p. 267-72.

29. Berndt, S.I., et al., *Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture.* Nat Genet, 2013. **45**(5): p. 501-12.

30. Monda, K.L., et al., *A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry.* Nat Genet, 2013. **45**(6): p. 690-6.

31. Graff, M., et al., *Genome-wide analysis of BMI in adolescents and young adults reveals additional insight into the effects of genetic loci over the life course.* Hum Mol Genet, 2013. **22**(17): p. 3597-3607.

32. Pei, Y.F., et al., *Meta-analysis of genome-wide association data identifies novel susceptibility loci for obesity.* Hum Mol Genet, 2014. **23**(3): p. 820-830.

33. Locke, A.E., et al., *Genetic studies of body mass index yield new insights for obesity biology.* Nature, 2015. **518**: p. 197-206.

34. Yang, J., et al., *Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits.* Nat Genet, 2012. **44**(4): p. 369-75, S1-3.

35. Temelkova-Kurktschiev, T. and T. Stefanov, *Lifestyle and Genetics in Obesity and type 2 Diabetes.* Exp Clin Endocrinol Diabetes, 2012. **120**(1): p. 1-6.

36. Flowers, E., E.S. Froelicher, and B.E. Aouizerat, *Gene-environment interactions in cardiovascular disease.* Eur J Cardiovasc Nurs, 2011.

37. Joseph, P.G., G. Pare, and S.S. Anand, *Exploring gene-environment relationships in cardiovascular disease.* Can J Cardiol, 2013. **29**(1): p. 37-45.

38. 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.* Lancet, 2011. **377**(9765): p. 557-67.

39. Pijl, H., *Obesity: evolution of a symptom of affluence.* Neth J Med, 2011. **69**(4): p. 159-66.

40. Bustamante, C.D., E.G. Burchard, and F.M. De la Vega, *Genomics for the world.* Nature, 2011. **475**(7355): p. 163-5.

41. Matise, T.C., et al., *The Next PAGE in understanding complex traits: design for the analysis of Population Architecture Using Genetics and Epidemiology (PAGE) Study.* Am J Epidemiol, 2011. **174**(7): p. 849-59.

42. Lavange, L.M., et al., *Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos.* Ann Epidemiol, 2010. **20**(8): p. 642-9.

43. Sorlie, P.D., et al., *Design and implementation of the Hispanic Community Health Study/Study of Latinos.* Ann Epidemiol, 2010. **20**(8): p. 629-41.

44. Davidson, J.A., et al., *Avoiding the looming Latino/Hispanic cardiovascular health crisis: a call to action.* J Cardiometab Syndr, 2007. **2**(4): p. 238-43.

45. WHO, *Global Health Observatory Data Repository: Overweight/Obesity Data by country*, 2008.

46. Alvarez-Torices, J.C., et al., *Self-reported height and weight and prevalence of obesity. Study in a Spanish population.* Int J Obes Relat Metab Disord, 1993. **17**(11): p. 663-7.

47. Avila-Funes, J.A., L.M. Gutierrez-Robledo, and S. Ponce De Leon Rosales, *Validity of height and weight self-report in Mexican adults: results from the national health and aging study.* J Nutr Health Aging, 2004. **8**(5): p. 355-61.

48. Field, A.E., P. Aneja, and B. Rosner, *The validity of self-reported weight change among adolescents and young adults.* Obesity, 2007. **15**(9): p. 2357-2364.

49. Lee, S.K., *Validity of self-reported weight and height: comparison between immigrant and non-immigrant Mexican Americans in NHANES III.* J Immigr Health, 2005. **7**(2): p. 127-31.

50. McAdams, M.A., R.M. Van Dam, and F.B. Hu, *Comparison of self-reported and measured BMI as correlates of disease markers in US adults.* Obesity (Silver Spring), 2007. **15**(1): p. 188-96.

51. Merrill, R.M. and J.S. Richardson, *Validity of self-reported height, weight, and body mass index: findings from the National Health and Nutrition Examination Survey, 2001-2006.* Prev Chronic Dis, 2009. **6**(4): p. A121.

52. 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, 2013. **11**(9): p. e1001661.

53. Morris, A.P., *Transethnic meta-analysis of genomewide association studies.* Genet Epidemiol, 2011. **35**(8): p. 809-22.

54. Stevens, J., J.E. McClain, and K.P. Truesdale, *Selection of measures in epidemiologic studies of the consequences of obesity.* Int J Obes (Lond), 2008. **32 Suppl 3**: p. S60-6.

55. Hu, F.B., *Measurements of Adiposity and Body Composition*, in *Obesity Epidemiology*. 2008, Oxford University Press: Oxford, New York. p. 53-83.

56. Kuczmarski, R.J. and K.M. Flegal, *Criteria for definition of overweight in transition: background and recommendations for the United States.* Am J Clin Nutr, 2000. **72**(5): p. 1074-81.

57. CDC. *Body Mass Index*. 2012 June 11, 2014]; Available from: http://www.cdc.gov/healthyweight/assessing/bmi/index.html.

58. Ogden, C.L., et al., *The epidemiology of obesity.* Gastroenterology, 2007. **132**(6): p. 2087-102.

59. Flegal, K.M. and R.P. Troiano, *Changes in the distribution of body mass index of adults and children in the US population.* Int J Obes Relat Metab Disord, 2000. **24**(7): p. 807-18.

60. Skinner, A.C., E.M. Perrin, and J.A. Skelton, *Prevalence of obesity and severe obesity in US children, 1999-2014.* Obesity (Silver Spring), 2016. **24**(5): p. 1116-23.

61. Skinner, A.C. and J.A. Skelton, *Prevalence and trends in obesity and severe obesity among children in the United States, 1999-2012.* JAMA Pediatr, 2014. **168**(6): p. 561-6.

62. WHO. *Obesity and overweight*. 2014 June 11, 2014]; Available from: http://www.who.int/mediacentre/factsheets/fs311/en/.

63. Popkin, B.M., et al., *Is there a lag globally in overweight trends for children compared with adults?* Obesity, 2006. **14**(10): p. 1846-1853.

64. Guo, S.S. and W.C. Chumlea, *Tracking of body mass index in children in relation to overweight in adulthood.* Am J Clin Nutr, 1999. **70**(1 Part 2): p. 145S-148S.

65. Freedman, D.S., et al., *The relation of childhood BMI to adult adiposity: the Bogalusa Heart Study.* Pediatrics, 2005. **115**(1): p. 22-7.

66. Freedman, D.S., et al., *Relationship of childhood obesity to coronary heart disease risk factors in adulthood: the Bogalusa Heart Study.* Pediatrics, 2001. **108**(3): p. 712-8.

67. The, N.S., et al., *Association of Adolescent Obesity With Risk of Severe Obesity in Adulthood.* Jama-Journal of the American Medical Association, 2010. **304**(18): p. 2042-2047.

68. Rainwater, D.L., et al., *Association among 5-year changes in weight, physical activity, and cardiovascular disease risk factors in Mexican Americans.* Am J Epidemiol, 2000. **152**(10): p. 974-82.

69. Andrade, F.C., et al., *One-year follow-up changes in weight are associated with changes in blood pressure in young Mexican adults.* Public Health, 2012. **126**(6): p. 535-40.

70. Baer, H.J., et al., *Risk factors for mortality in the nurses' health study: a competing risks analysis.* Am J Epidemiol, 2011. **173**(3): p. 319-29.

71. Wing, R.R., *Obesity and weight gain during adulthood: a health problem for United States women.* Womens Health Issues, 1992. **2**(2): p. 114-20; discussion 120-2.

72. Robinson, W.R., et al., *Birth cohort effects among US-born adults born in the 1980s: foreshadowing future trends in US obesity prevalence.* Int J Obes (Lond), 2013. **37**(3): p. 448-54.

73. Elliott-Sale, K.J., C.T. Barnett, and C. Sale, *Exercise interventions for weight management during pregnancy and up to 1 year postpartum among normal weight, overweight and obese women: a systematic review and meta-analysis.* Br J Sports Med, 2014.

74. Lovejoy, J.C., *Weight Gain in Women at Midlife:The Influence of Menopause.* Obesity Management, 2009. **5**(2): p. 52-56.

75. Griebeler, M.L., et al., *Self-reported versus measured height and weight in Hispanic and non-Hispanic menopausal women.* J Womens Health (Larchmt), 2011. **20**(4): p. 599-604.

76. Mackay, D.F., L. Gray, and J.P. Pell, *Impact of smoking and smoking cessation on overweight and obesity: Scotland-wide, cross-sectional study on 40,036 participants.* BMC Public Health, 2013. **13**: p. 348.

77. Chiolero, A., et al., *Association of cigarettes smoked daily with obesity in a general adult population.* Obesity (Silver Spring), 2007. **15**(5): p. 1311-8.

78. Clair, C., et al., *Dose-dependent positive association between cigarette smoking, abdominal obesity and body fat: cross-sectional data from a population-based survey.* BMC Public Health, 2011. **11**.

79. Wardle, J., et al., *Stress and adiposity: a meta-analysis of longitudinal studies.* Obesity (Silver Spring), 2011. **19**(4): p. 771-8.

80. Luppino, F.S., et al., *Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies.* Arch Gen Psychiatry, 2010. **67**(3): p. 220-9.

81. Solovieva, S., et al., *Psychosocial factors at work, long work hours, and obesity: a systematic review.* Scandinavian Journal of Work Environment & Health, 2013. **39**(3): p. 241-258.

82. Wang, Y. and M.A. Beydoun, *The obesity epidemic in the United States--gender, age, socioeconomic, racial/ethnic, and geographic characteristics: a systematic review and meta-regression analysis.* Epidemiol Rev, 2007. **29**: p. 6-28.

83. Wallace, P.M., et al., *A Review of Acculturation Measures and Their Utility in Studies Promoting Latino Health.* Hispanic Journal of Behavioral Sciences, 2010. **32**(1): p. 37-54.

84. Berry, J.W., *Psychology of group relations: Cultural and social dimensions.* Aviation Space and Environmental Medicine, 2004. **75**(7): p. C52-C57.

85. Carter-Pokras, O. and L. Bethune, *Defining and measuring acculturation: a systematic review of public health studies with Hispanic populations in the United States. A commentary on Thomson and Hoffman-Goetz.* Soc Sci Med, 2009. **69**(7): p. 992-5; discussion 999-1001.

86. Abraido-Lanza, A.F., et al., *Toward a theory-driven model of acculturation in public health research.* Am J Public Health, 2006. **96**(8): p. 1342-6.

87. Hunt, L.M., S. Schneider, and B. Comer, *Should "acculturation" be a variable in health research? A critical review of research on US Hispanics.* Soc Sci Med, 2004. **59**(5): p. 973-86.

88. Schwartz, S.J., et al., *Rethinking the Concept of Acculturation Implications for Theory and Research.* American Psychologist, 2010. **65**(4): p. 237-251.

89. Lopez-Class, M., F.G. Castro, and A.G. Ramirez, *Conceptions of acculturation: A review and statement of critical issues.* Social Science & Medicine, 2011. **72**(9): p. 1555-1562.

90. Viruell-Fuentes, E.A., P.Y. Miranda, and S. Abdulrahim, *More than culture: Structural racism, intersectionality theory, and immigrant health.* Social Science & Medicine, 2012. **75**(12): p. 2099-2106.

91. Ranadive, S.A. and C. Vaisse, *Lessons from extreme human obesity: monogenic disorders.* Endocrinol Metab Clin North Am, 2008. **37**(3): p. 733-51, x.

92. Waalen, J., *The genetics of human obesity.* Transl Res, 2014. **164**(4): p. 293-301.

93. Jaimes, N., V. Londono, and A.C. Halpern, *The term Hispanic/Latino: a note of caution.* JAMA Dermatol, 2013. **149**(3): p. 274-5.

94. *When Labels Don't Fit: Hispanics and Their Views of Identity*, 2012, The Pew Research Center: Washington, DC.

95. *Hispanic Nativity Shift*, 2014, The Pew Research Center: Washington, DC.

96. U.S. Census Bureau, P.D., *Annual Estimates of the Resident Population by Sex, Race, and Hispanic Origin for the United States, States, and Counties: April 1, 2010 to July 1, 2012* 2013.

97. Kershaw, K.N. and S.S. Albrecht, *Metropolitan-level ethnic residential segregation, racial identity, and body mass index among U.S. Hispanic adults: a multilevel cross-sectional study.* BMC Public Health, 2014. **14**: p. 283.

98. Miyares, I.M. and K.J. Gowen, *Recreating borders? The geography of Latin Americans in New York City.* Yearbook 1998, 1998. **24**: p. 31-43.

99. Hall, M. and J. Stringfield, *Undocumented migration and the residential segregation of Mexicans in new destinations.* Social Science Research, 2014. **47**: p. 61-78.

100. Hernández-Coss, R., *The U.S.–Mexico Remittance Corridor: Lessons on Shifting from Informal to Formal Transfer Systems*, 2005, The World Bank: Washington, DC.

101. Sorlie, P.D., et al., *Prevalence of hypertension, awareness, treatment, and control in the Hispanic community health study/study of Latinos.* Am J Hypertens, 2014. **27**(6): p. 793-800.

102. *A Portrait of Unauthorized Immigrants in the United States*, 2009, The Pew Research Center: Washington, D.C.

103. *Population Decline of Unauthorized Immigrants Stalls, May Have Reversed*, 2013, The Pew Research Center: Washington, DC.

104. Holmes, S.M., *"Is it worth risking your life?": Ethnography, risk and death on the US-Mexico border.* Social Science & Medicine, 2013. **99**: p. 153-161.

105. Martinez, O., et al., *Evaluating the Impact of Immigration Policies on Health Status Among Undocumented Immigrants: A Systematic Review.* J Immigr Minor Health, 2013.

106. Sullivan, M.M. and R. Rehm, *Mental health of undocumented Mexican immigrants - A review of the literature.* Advances in Nursing Science, 2005. **28**(3): p. 240-251.

107. Aptekar, S., *Citizenship Status and Patterns of Inequality in the United States and Canada.* Social Science Quarterly, 2014. **95**(2): p. 343-359.

108. Hall, M., E. Greenman, and G. Farkas, *Legal Status and Wage Disparities for Mexican Immigrants.* Social Forces, 2010. **89**(2): p. 491-513.

109. Massey, D.S. and K. Gentsch, *Undocumented Migration to the United States and the Wages of Mexican Immigrants.* International Migration Review, 2014. **48**(2): p. 482-499.

110. México, I.N.E.s.y.G.a.-.-. *Perfil sociodemográfico: Estados Unidos Mexicanos: Censo de Población y Vivienda 2010*, 2013, The Mexican National Institute of Statistics and Geography: Aguascalientes, Mexico.

111. Gonzalez Burchard, E., et al., *Latino populations: a unique opportunity for the study of race, genetics, and social environment in epidemiological research.* Am J Public Health, 2005. **95**(12): p. 2161-8.

112. Bryc, K., et al., *Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations.* Proc Natl Acad Sci U S A, 2010. **107 Suppl 2**: p. 8954-61.

113. Halder, I., et al., *Measurement of admixture proportions and description of admixture structure in different U.S. populations.* Hum Mutat, 2009. **30**(9): p. 1299-309.

114. Shtir, C.J., et al., *Variation in genetic admixture and population structure among Latinos: the Los Angeles Latino eye study (LALES).* BMC Genet, 2009. **10**: p. 71.

115. Wang, S., et al., *Geographic patterns of genome admixture in Latin American Mestizos.* PLoS Genet, 2008. **4**(3): p. e1000037.

116. Manichaikul, A., et al., *Population structure of Hispanics in the United States: the multi-ethnic study of atherosclerosis.* PLoS Genet, 2012. **8**(4): p. e1002640.

117. Moreno-Estrada, A., et al., *Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits.* Science, 2014. **344**(6189): p. 1280-5.

118. Gurdasani, D., et al., *The African Genome Variation Project shapes medical genetics in Africa.* Nature, 2014.

119. Ríos, M., F. Romero, and R. Ramírez, *Race Reporting Among Hispanics: 2010*, 2014, Population Division, US Census Bureau: Washington, DC.

120. Divers, J., et al., *Comparing self-reported ethnicity to genetic background measures in the context of the Multi-Ethnic Study of Atherosclerosis (MESA).* BMC Genet, 2011. **12**: p. 28.

121. Aponte, J., *Diabetes-related risk factors across Hispanic subgroups in the Hispanic health and nutritional examination survey (1982-1984).* Public Health Nurs, 2009. **26**(1): p. 23-38.

122. Khan, L.K., J. Sobal, and R. Martorell, *Acculturation, socioeconomic status, and obesity in Mexican Americans, Cuban Americans, and Puerto Ricans.* Int J Obes Relat Metab Disord, 1997. **21**(2): p. 91-6.

123. Albrecht, S.S. and P. Gordon-Larsen, *Ethnic differences in body mass index trajectories from adolescence to adulthood: a focus on Hispanic and Asian subgroups in the United States.* PLoS One, 2013. **8**(9): p. e72983.

124. DeNavas-Walt, C., B.D. Proctor, and J.C. Smith, *Income, Poverty, and Health Insurance Coverage in the United States: 2009*, U.S.C. Bureau, Editor 2010, U.S. Government Printing Office: Washington, DC.

125. Vega, W.A., M.A. Rodriguez, and E. Gruskin, *Health disparities in the Latino population.* Epidemiol Rev, 2009. **31**: p. 99-112.

126. Adams, P.F., et al., *Summary health statistics for the U.S. population: National Health Interview Survey, 2010.*, in *Vital Health Stat*2011, National Center for Health Statistics.

127. Dockertman, D. *Country of Origin Profiles*. 2011 [cited January 24, 2012; Available from: http://www.pewhispanic.org/2011/05/26/country-of-origin-profiles/.

128. Marmot, M.G., *Status syndrome: a challenge to medicine.* JAMA, 2006. **295**(11): p. 1304-7.

129. Guendelman, S., et al., *Birthplace, language use, and body size among Mexican American women and men: findings from the National Health and Nutrition Examination Survey (NHANES) 2001-2006.* J Health Care Poor Underserved, 2011. **22**(2): p. 590-605.

130. Barcenas, C.H., et al., *Birthplace, years of residence in the United States, and obesity among Mexican-American adults.* Obesity (Silver Spring), 2007. **15**(4): p. 1043-52.

131. Duffey, K.J., et al., *Birthplace is associated with more adverse dietary profiles for US-born than for foreign-born Latino adults.* J Nutr, 2008. **138**(12): p. 2428-35.

132. Gordon-Larsen, P., et al., *Acculturation and overweight-related behaviors among Hispanic immigrants to the US: the National Longitudinal Study of Adolescent Health.* Soc Sci Med, 2003. **57**(11): p. 2023-34.

133. Sundquist, J. and M. Winkleby, *Country of birth, acculturation status and abdominal obesity in a national sample of Mexican-American women and men.* Int J Epidemiol, 2000. **29**(3): p. 470-7.

134. Van Wieren, A.J., et al., *Acculturation and Cardiovascular Behaviors Among Latinos in California by Country/Region of Origin.* Journal of Immigrant and Minority Health, 2011. **13**(6): p. 975-981.

135. Liu, J.H., et al., *Generation and acculturation status are associated with dietary intake and body weight in Mexican American adolescents.* J Nutr, 2012. **142**(2): p. 298-305.

136. Roshania, R., K.M.V. Narayan, and R. Oza-Frank, *Age at Arrival and Risk of Obesity Among US Immigrants.* Obesity, 2008. **16**(12): p. 2669-2675.

137. Perez-Escamilla, R. and P. Putnik, *The role of acculturation in nutrition, lifestyle, and incidence of type 2 diabetes among Latinos.* Journal of Nutrition, 2007. **137**(4): p. 860-870.

138. Isasi, C.R., et al., *Is acculturation related to obesity in Hispanic/Latino adults? Results from the Hispanic community health study/study of Latinos.* J Obes, 2015. **2015**: p. 186276.

139. Akresh, I.R., *Overweight and obesity among foreign-born and U.S.-born Hispanics.* Biodemography Soc Biol, 2008. **54**(2): p. 183-99.

140. Albrecht, S.S., et al., *Secular trends in the association between nativity/length of US residence with body mass index and waist circumference among Mexican-Americans, 1988-2008.* International Journal of Public Health, 2013. **58**(4): p. 573-581.

141. Park, J., et al., *Immigrant obesity and unhealthy assimilation: Alternative estimates of convergence or divergence, 1995-2005.* Social Science & Medicine, 2009. **69**(11): p. 1625-1633.

142. Jackson, M.I., *Foreign-born health integration during the transition to adulthood: The case of weight.* Social Science Research, 2011. **40**(5): p. 1419-1433.

143. Balistreri, K.S. and J. Van Hook, *Socioeconomic Status and Body Mass Index Among Hispanic Children of Immigrants and Children of Natives.* American Journal of Public Health, 2009. **99**(12): p. 2238-2246.

144. Harris, K.M., K.M. Perreira, and D. Lee, *Obesity in the Transition to Adulthood Predictions Across Race/Ethnicity, Immigrant Generation, and Sex.* Archives of Pediatrics & Adolescent Medicine, 2009. **163**(11): p. 1022-1028.

145. Albrecht, S.S. and P. Gordon-Larsen, *Socioeconomic gradients in body mass index (BMI) in US immigrants during the transition to adulthood: examining the roles of parental education and intergenerational educational mobility.* J Epidemiol Community Health, 2014.

146. Chu, Y.H., et al., *Higher generational and acculturation status are associated with poorer diet and greater body weight among Mexican American adolescents.* Faseb Journal, 2011. **25**.

147. Creighton, M.J., et al., *Durational and generational differences in Mexican immigrant obesity: is acculturation the explanation?* Soc Sci Med, 2012. **75**(2): p. 300-10.

148. Albrecht, S.S., et al., *Immigrant Assimilation and BMI and Waist Size: A Longitudinal Examination Among Hispanic and Chinese Participants in the Multi-ethnic Study of Atherosclerosis.* Obesity, 2013. **21**(8): p. 1695-1703.

149. Berry, J.W., *Stress perspectives on acculturation*, in *The Cambridge Handbook of Acculturation Psychology*, D.L. Sam and J.W. Berry, Editors. 2006, Cambridge University Press: Cambridge, UK. p. 43-57.

150. Portes, A., P. Fernandez-Kelly, and W. Haller, *Segmented assimilation on the ground: The new second generation in early adulthood.* Ethnic and Racial Studies, 2005. **28**(6): p. 1000-1040.

151. Connor Gorber, S., et al., *A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review.* Obes Rev, 2007. **8**(4): p. 307-26.

152. Stevens, J., et al., *The definition of weight maintenance.* Int J Obes (Lond), 2006. **30**(3): p. 391-9.

153. Kuczmarski, M.F., R.J. Kuczmarski, and M. Najjar, *Effects of age on validity of self-reported height, weight, and body mass index: findings from the Third National Health and Nutrition Examination Survey, 1988-1994.* J Am Diet Assoc, 2001. **101**(1): p. 28-34; quiz 35-6.

154. Gillum, R.F. and C.T. Sempos, *Ethnic variation in validity of classification of overweight and obesity using self-reported weight and height in American women and men: the Third National Health and Nutrition Examination Survey.* Nutr J, 2005. **4**: p. 27.

155. Santillan, A.A. and C.A. Camargo, *Body mass index and asthma among Mexican adults: the effect of using self-reported vs measured weight and height.* Int J Obes Relat Metab Disord, 2003. **27**(11): p. 1430-3.

156. Wen, M. and L. Kowaleski-Jones, *Sex and ethnic differences in validity of self-reported adult height, weight and body mass index.* Ethn Dis, 2012. **22**(1): p. 72-8.

157. Lindberg, N.M. and V.J. Stevens, *Immigration and Weight Gain: Mexican-American Women's Perspectives.* Journal of Immigrant and Minority Health, 2011. **13**(1): p. 155-160.

158. Kyulo, N.L., et al., *Validation of recall of body weight over a 26-year period in cohort members of the Adventist Health Study 2.* Ann Epidemiol, 2012. **22**(10): p. 744-6.

159. Kovalchik, S., *Validity of adult lifetime self-reported body weight.* Public Health Nutr, 2009. **12**(8): p. 1072-7.

160. Stevens, J., et al., *Accuracy of current, 4-year, and 28-year self-reported body weight in an elderly population.* Am J Epidemiol, 1990. **132**(6): p. 1156-63.

161. Troy, L.M., et al., *The Validity of Recalled Weight among Younger Women.* International Journal of Obesity, 1995. **19**(8): p. 570-572.

162. Clarke, P., et al., *Accuracy of self-reported versus measured weight over adolescence and young adulthood: findings from the national longitudinal study of adolescent health, 1996-2008.* Am J Epidemiol, 2014. **180**(2): p. 153-9.

163. Gallagher, D., et al., *How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups?* Am J Epidemiol, 1996. **143**(3): p. 228-39.

164. Gallagher, D., et al., *Healthy percentage body fat ranges: an approach for developing guidelines based on body mass index.* Am J Clin Nutr, 2000. **72**(3): p. 694-701.

165. Deurenberg, P., M. Yap, and W.A. van Staveren, *Body mass index and percent body fat: a metaanalysis among different ethnic groups.* Int J Obes, 1998. **22**(12): p. 1164-1171.

166. Deurenberg, P., M. Deurenberg-Yap, and S. Guricci, *Asians are different from Caucasians and from each other in their body mass index/body fat per cent relationship.* Obes Rev, 2002. **3**(3): p. 141-6.

167. Casas, Y.G., et al., *Total and regional body composition across age in healthy Hispanic and white women of similar socioeconomic status.* Am J Clin Nutr, 2001. **73**(1): p. 13-8.

168. Fernandez, J.R., et al., *Is percentage body fat differentially related to body mass index in Hispanic Americans, African Americans, and European Americans?* Am J Clin Nutr, 2003. **77**(1): p. 71-5.

169. Olivares, J., et al., *Comparisons of body volumes and dimensions using three-dimensional photonic scanning in adult Hispanic-Americans and Caucasian-Americans.* J Diabetes Sci Technol, 2007. **1**(6): p. 921-8.

170. Bastien, M., et al., *Overview of epidemiology and contribution of obesity to cardiovascular disease.* Prog Cardiovasc Dis, 2014. **56**(4): p. 369-81.

171. Carmienke, S., et al., *General and abdominal obesity parameters and their combination in relation to mortality: a systematic review and meta-regression analysis.* Eur J Clin Nutr, 2013. **67**(6): p. 573-85.

172. Monteverde, M., et al., *Obesity and Excess Mortality among the Elderly in the United States and Mexico.* Demography, 2010. **47**(1): p. 79-96.

173. Stern, M.P., et al., *Overweight and mortality in Mexican Americans.* Int J Obes, 1990. **14**(7): p. 623-9.

174. Markides, K.S. and J. Coreil, *The Health of Hispanics in the Southwestern United-States - an Epidemiologic Paradox.* Public Health Reports, 1986. **101**(3): p. 253-265.

175. Fontaine, K.R., et al., *Body mass index and mortality rate among Hispanic adults: a pooled analysis of multiple epidemiologic data sets.* Int J Obes (Lond), 2011.

176. Goran, M.I., *Ethnic-specific pathways to obesity-related disease: the Hispanic vs. African-American paradox.* Obesity (Silver Spring), 2008. **16**(12): p. 2561-5.

177. Abdullah, A., et al., *The magnitude of association between overweight and obesity and the risk of diabetes: a meta-analysis of prospective cohort studies.* Diabetes Res Clin Pract, 2010. **89**(3): p. 309-19.

178. Taylor, A.E., et al., *Comparison of the associations of body mass index and measures of central adiposity and fat mass with coronary heart disease, diabetes, and all-cause mortality: a study using data from 4 UK cohorts.* Am J Clin Nutr, 2010. **91**(3): p. 547-56.

179. Kaplan, R.C., et al., *Body Mass Index, Sex, and Cardiovascular Disease Risk Factors Among Hispanic/Latino Adults: Hispanic Community Health Study/Study of Latinos.* Journal of the American Heart Association, 2014. **3**(4).

180. Albrecht, S.S. and B. Popkin. *The Association between BMI and Diabetes in US Adults: Examination of Secular and Race/Ethnic Trends*. in *The Obesity Society Annual Meeting*. 2014. Boston, MA.

181. Klipstein-Grobusch, K., A. Kroke, and H. Boeing, *Reproducibility of self-reported past body weight.* Eur J Clin Nutr, 1998. **52**(7): p. 525-8.

182. Perez-Cueto, F.J. and W. Verbeke, *Reliability and validity of self-reported weight and height in Belgium.* Nutr Hosp, 2009. **24**(3): p. 366-7.

183. Lin, C.J., et al., *Accuracy and reliability of self-reported weight and height in the Sister Study.* Public Health Nutr, 2012. **15**(6): p. 989-99.

184. Ulijaszek, S.J. and D.A. Kerr, *Anthropometric measurement error and the assessment of nutritional status.* Br J Nutr, 1999. **82**(3): p. 165-77.

185. Marks, G.C., J.P. Habicht, and W.H. Mueller, *Reliability, dependability, and precision of anthropometric measurements. The Second National Health and Nutrition Examination Survey 1976-1980.* Am J Epidemiol, 1989. **130**(3): p. 578-87.

186. Winkler, T.W., et al., *The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study.* PLoS Genet, 2015. **11**(10): p. e1005378.

187. Kraft, P., *Curses--winner's and otherwise--in genetic epidemiology.* Epidemiology, 2008. **19**(5): p. 649-51; discussion 657-8.

188. Yang, J., et al., *GCTA: a tool for genome-wide complex trait analysis.* Am J Hum Genet, 2011. **88**(1): p. 76-82.

189. Evangelou, E. and J.P.A. Ioannidis, *Meta-analysis methods for genome-wide association studies and beyond.* Nature Reviews Genetics, 2013. **14**(6): p. 379-389.

190. Han, B. and E. Eskin, *Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-wide Association Studies.* Am J Hum Genet, 2011. **88**(5): p. 586-598.

191. Kim, J., Y. Bai, and W. Pan, *An Adaptive Association Test for Multiple Phenotypes with GWAS Summary Statistics.* Genet Epidemiol, 2015. **39**(8): p. 651-63.

192. Malhotra, A., et al., *A genome-wide association study of BMI in American Indians.* Obesity (Silver Spring), 2011. **19**(10): p. 2102-6.

193. Graff, M., et al., *Generalization of adiposity genetic loci to US Hispanic women.* Nutr Diabetes, 2013. **3**: p. e85.

194. Yang, F., et al., *Genome wide association study: searching for genes underlying body mass index in the Chinese.* Biomedical and Environmental Sciences, 2014. **27**(5): p. 360-70.

195. Wen, W., et al., *Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index.* Hum Mol Genet, 2014. **23**(20): p. 5492-504.

196. Ahmad, S., T.V. Varga, and P.W. Franks, *Gene x environment interactions in obesity: the state of the evidence.* Hum Hered, 2013. **75**(2-4): p. 106-15.

197. 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, 2013. **11**(9): p. e1001661.

198. Boardman, J.D., et al., *Is the Gene-Environment Interaction Paradigm Relevant to Genome-Wide Studies? The Case of Education and Body Mass Index.* Demography, 2014. **51**(1): p. 119-139.

199. Gravel, S., et al., *Reconstructing Native American migrations from whole-genome and whole-exome data.* PLoS Genet, 2013. **9**(12): p. e1004023.

200. Moreno-Estrada, A., et al., *Reconstructing the population genetic history of the Caribbean.* PLoS Genet, 2013. **9**(11): p. e1003925.

201. 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.* Am J Hum Genet, 2016. **98**(1): p. 165-84.

202. Lin, D.Y., et al., *Genetic Association Analysis under Complex Survey Sampling: The Hispanic Community Health Study/Study of Latinos.* Am J Hum Genet, 2014. **95**(6): p. 675-688.

203. Fesinmeyer, M.D., et al., *Genetic Risk Factors for BMI and Obesity in an Ethnically Diverse Population: Results From the Population Architecture Using Genomics and Epidemiology (PAGE) Study.* Obesity (Silver Spring), 2012.

204. Gong, J., et al., *Fine Mapping and Identification of BMI Loci in African Americans.* Am J Hum Genet, 2013. **93**(4): p. 661-71.

205. Demerath, E.W., et al., *The positive association of obesity variants with adulthood adiposity strengthens over an 80-year period: a gene-by-birth year interaction.* Hum Hered, 2013. **75**(2-4): p. 175-85.

206. Cooper, R.S., *Gene-environment interactions and the etiology of common complex disease.* Ann Intern Med, 2003. **139**(5 Pt 2): p. 437-40.

207. Manning, A.K., et al., *Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP x environment regression coefficients.* Genet Epidemiol, 2011. **35**(1): p. 11-8.

208. Aschard, H., et al., *Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects.* Hum Hered, 2010. **70**(4): p. 292-300.

209. 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, 2013. **9**(6): p. e1003500.

210. Graff, M., et al., *Screen time behaviours may interact with obesity genes, independent of physical activity, to influence adolescent BMI in an ethnically diverse cohort.* Pediatr Obes, 2013. **8**(6): p. e74-9.

211. Demerath, E.W., et al., *Interaction of FTO and physical activity level on adiposity in African-American and European-American adults: the ARIC Study.* Obesity (Silver Spring), 2011. **19**(9): p. 1866-72.

212. Rukh, G., et al., *Genetic susceptibility to obesity and diet intakes: association and interaction analyses in the Malmo Diet and Cancer Study.* Genes Nutr, 2013. **8**(6): p. 535-47.

213. Wickrama, K.K., C.W. O'Neal, and T.K. Lee, *Early community context, genes, and youth body mass index trajectories: an investigation of gene-community interplay over early life course.* J Adolesc Health, 2013. **53**(3): p. 328-34.

214. Riedel, C., et al., *Interactions of genetic and environmental risk factors with respect to body fat mass in children: results from the ALSPAC study.* Obesity (Silver Spring), 2013. **21**(6): p. 1238-42.

215. Edwards, T.L., et al., *HTR1B, ADIPOR1, PPARGC1A, and CYP19A1 and obesity in a cohort of Caucasians and African Americans: an evaluation of gene-environment interactions and candidate genes.* Am J Epidemiol, 2012. **175**(1): p. 11-21.

216. Jourdan, C., et al., *Gene-PUFA interactions and obesity risk.* Br J Nutr, 2011. **106**(8): p. 1263-72.

217. Wang, J.W., et al., *The Impact of Lipid-metabolizing Genetic Polymorphisms on Body Mass Index and Their Interactions with Soybean Food Intake: A Study in a Chinese Population.* Biomedical and Environmental Sciences, 2014. **27**(3): p. 176-185.

218. Velez Edwards, D.R., et al., *Gene-environment interactions and obesity traits among postmenopausal African-American and Hispanic women in the Women's Health Initiative SHARe Study.* Human Genetics, 2013. **132**(3): p. 323-36.

219. Rosenquist, J.N., et al., *Cohort of birth modifies the association between FTO genotype and BMI.* Proc Natl Acad Sci U S A, 2015. **112**(2): p. 354-359.

220. Singh, A., et al., *Gene by stress genome-wide interaction analysis and path analysis identify EBF1 as a cardiovascular and metabolic risk gene.* Eur J Hum Genet, 2014.

221. Xie, B., et al., *Gender difference in interactions between MAOA promoter uVNTR polymorphism and negative familial stressors on body mass index among Chinese adolescents.* Pediatric Obesity, 2014. **9**(5): p. E80-E90.

222. Barth, S.W., et al., *Moderate effects of apple juice consumption on obesity-related markers in obese men: impact of diet-gene interaction on body fat content.* European Journal of Nutrition, 2012. **51**(7): p. 841-850.

223. Cuypers, K., et al., *Leisure time activities in adolescence in the presence of susceptibility genes for obesity: Risk or Resilience against overweight in adulthood? The HUNT study.* European Journal of Public Health, 2012. **22**: p. 33-33.

224. Qi, Q.B., et al., *Fried food consumption, genetic risk, and body mass index: gene-diet interaction analysis in three US cohort studies.* Bmj-British Medical Journal, 2014. **348**.

225. Underwood, P.C., et al., *Replication and meta-analysis of the gene-environment interaction between body mass index and the interleukin-6 promoter polymorphism with higher insulin resistance.* Metabolism-Clinical and Experimental, 2012. **61**(5): p. 667-671.

226. Aschard, H., et al., *Challenges and opportunities in genome-wide environmental interaction (GWEI) studies.* Human Genetics, 2012. **131**(10): p. 1591-1613.

227. Abecasis, G.R., et al., *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.

228. Ullmann, S.H., N. Goldman, and A.R. Pebley, *Contextual factors and weight change over time: A comparison between US Hispanics and other population sub-groups.* Social Science & Medicine, 2013. **90**: p. 40-48.

229. Bryan, H., T. Mehlman, and P. Gildner. *A study recruitment system using ultra-mobile computers with handwriting recognition*. in *Society for Clinical Trials at 30th Annual Meeting*. 2009. Atlanta, GA.

230. Liang, K.Y. and S.L. Zeger, *Longitudinal Data-Analysis Using Generalized Linear-Models.* Biometrika, 1986. **73**(1): p. 13-22.

231. Binder, D.A., *On the Variances of Asymptotically Normal Estimators from Complex Surveys.* International Statistical Review, 1983. **51**(3): p. 279-292.

232. Lin, D.Y., *On fitting Cox's proportional hazards models to survey data.* Biometrika, 2000. **87**(1): p. 37-47.

233. Voight, B.F., et al., *The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits.* PLoS Genet, 2012. **8**(8): p. e1002793.

234. Sub-Committee, H.S.S. *Public Manuals and Documents*. 2012 February 22, 2012]; Available from: http://www.cscc.unc.edu/hchs/public/docfilter.php?study=hchs&filter_type=public.

235. Buyske, S., et al., *Evaluation of the metabochip genotyping array in African Americans and implications for fine mapping of GWAS-identified loci: the PAGE study.* PLoS One, 2012. **7**(4): p. e35651.

236. Weale, M.E., *Quality control for genome-wide association studies.* Methods Mol Biol, 2010. **628**: p. 341-72.

237. Liu, E.Y., et al., *Genotype imputation of Metabochip SNPs using a study-specific reference panel of ~4,000 haplotypes in African Americans from the Women's Health Initiative.* Genet Epidemiol, 2012. **36**(2): p. 107-17.

238. Lin, D.Y., et al., *Genetic Association Analysis under Complex Survey Sampling: The Hispanic Community Health Study/Study of Latinos.* Am J Hum Genet, 2014. **95**(6): p. 675-88.

239. Hartl, D.L. and A.G. Clark, *Principles of Population Genetics*. 3rd ed. ed. 1997, Sunderland, MA.: Sinauer Associates, Inc.

240. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.

241. Amos, C.I., *Robust variance-components approach for assessing genetic linkage in pedigrees.* Am J Hum Genet, 1994. **54**(3): p. 535-43.

242. Amos, C.I., D.K. Zhu, and E. Boerwinkle, *Assessing genetic linkage and association with robust components of variance approaches.* Ann Hum Genet, 1996. **60**(Pt 2): p. 143-60.

243. Almasy, L. and J. Blangero, *Multipoint quantitative-trait linkage analysis in general pedigrees.* Am J Hum Genet, 1998. **62**(5): p. 1198-211.

244. Abecasis, G.R., L.R. Cardon, and W.O. Cookson, *A general test of association for quantitative traits in nuclear families.* Am J Hum Genet, 2000. **66**(1): p. 279-92.

245. Diao, G. and D.Y. Lin, *A powerful and robust method for mapping quantitative trait loci in general pedigrees.* Am J Hum Genet, 2005. **77**(1): p. 97-111.

246. Burton, P.R., et al., *Genetic variance components analysis for binary phenotypes using generalized linear mixed models (GLMMs) and Gibbs sampling.* Genet Epidemiol, 1999. **17**(2): p. 118-140.

247. Diao, G. and D.Y. Lin, *Semiparametric variance-component models for linkage and association analyses of censored trait data.* Genet Epidemiol, 2006. **30**(7): p. 570-81.

248. Diao, G. and D.Y. Lin, *Variance-components methods for linkage and association analysis of ordinal traits in general pedigrees.* Genet Epidemiol, 2010. **34**(3): p. 232-7.

249. Lee, E.W., L.J. Wei, and D.A. Amato, *Cox-type regression analysis for large numbers of small groups of correlated failure time observations.*, in *Survival Analysis: State of the Art*, J.P. Klein and P.K. Goel, Editors. 1992, Kluwer Academic Publishers: Dordrecht. p. 237-247.

250. Price, A.L., et al., *Principal components analysis corrects for stratification in genome-wide association studies.* Nat Genet, 2006. **38**(8): p. 904-9.

251. Sankararaman, S., et al., *Estimating local ancestry in admixed populations.* Am J Hum Genet, 2008. **82**(2): p. 290-303.

252. Patterson, N., A.L. Price, and D. Reich, *Population structure and eigenanalysis.* PLoS Genet, 2006. **2**(12): p. e190.

253. Anderson, G., et al., *Design of the Women's Health Initiative Clinical Trial and Observational Study.* Controlled Clinical Trials, 1998. **19**(1): p. 61-109.

254. Kolonel, L.N., et al., *A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics.* Am J Epidemiol, 2000. **151**(4): p. 346-57.

255. Pulley, J., et al., *Principles of human subjects protections applied in an opt-out, de-identified biobank.* Clin Transl Sci, 2010. **3**(1): p. 42-8.

256. Cronin, R.M., et al., *Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index.* Front Genet, 2014. **5**: p. 250.

257. Kuulasmaa, K., H.W. Hense, and H. Tolonen *Quality Assessment of Data on Blood Pressure in the WHO MONICA Project*. WHO MONICA Project e-publications, 1998.

258. Yuan, Y., *Multiple Imputation Using SAS Software.* Journal of Statistical Software, 2011. **45**(6): p. 1-25.

259. Chen, M.H. and Q. Yang, *GWAF: an R package for genome-wide association analyses with family data.* Bioinformatics, 2010. **26**(4): p. 580-1.

260. Willer, C.J., Y. Li, and G.R. Abecasis, *METAL: fast and efficient meta-analysis of genomewide association scans.* Bioinformatics, 2010. **26**(17): p. 2190-1.

261. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* American journal of human genetics, 2007. **81**(3): p. 559-75.

262. Pruim, R.J., et al., *LocusZoom: regional visualization of genome-wide association scan results.* Bioinformatics, 2010. **26**(18): p. 2336-7.

263. Schmidt, M.I., et al., *Validity of self-reported weight--a study of urban Brazilian adults.* Rev Saude Publica, 1993. **27**(4): p. 271-6.

264. Stunkard, A.J. and J.M. Albaum, *The accuracy of self-reported weights.* Am J Clin Nutr, 1981. **34**(8): p. 1593-9.

265. Bostrom, G. and F. Diderichsen, *Socioeconomic differentials in misclassification of height, weight and body mass index based on questionnaire data.* Int J Epidemiol, 1997. **26**(4): p. 860-6.

266. Weitzman, P.F., et al., *Bodily Aesthetic Ideals Among Latinas With Type 2 Diabetes Implications for Treatment Adherence, Access, and Outcomes.* Diabetes Educator, 2013. **39**(6): p. 856-863.

267. MossavarRahmani, Y., et al., *Determinants of body size perceptions and dieting behavior in a multiethnic group of hospital staff women.* J Am Diet Assoc, 1996. **96**(3): p. 252-256.

268. Maupin, J.N. and D.J. Hruschka, *Assessing the accuracy of two proxy measures for BMI in a semi-rural, low-resource setting in Guatemala.* BMC Public Health, 2014. **14**.

269. American Diabetes, A., *Diagnosis and classification of diabetes mellitus.* Diabetes Care, 2010. **33 Suppl 1**: p. S62-9.

270. *Physical activity guidelines for Americans.* Okla Nurse, 2008. **53**(4): p. 25.

271. Oza-Frank, R., et al., *Asian Americans: diabetes prevalence across U.S. and World Health Organization weight classifications.* Diabetes Care, 2009. **32**(9): p. 1644-6.

272. Humes, K.R., N.A. Jones, and R.R. Ramirez *Overview of Race and Hispanic Origin: 2010*. 2010 Census Briefs, 2011.

273. Dumitrescu, L., et al., *Assessing the accuracy of observer-reported ancestry in a biorepository linked to electronic medical records.* Genet Med, 2010. **12**(10): p. 648-50.

274. Hall, J.B., et al., *Accuracy of administratively-assigned ancestry for diverse populations in an electronic medical record-linked biobank.* PLoS One, 2014. **9**(6): p. e99161.

275. Williams, A.L., et al., *Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico.* Nature, 2014. **506**(7486): p. 97-101.

276. Assimes, T.L., et al., *Genetics of Coronary Artery Disease in Taiwan: A Cardiometabochip Study by the Taichi Consortium.* PLoS One, 2016. **11**(3): p. e0138014.

277. Carty, C.L., et al., *Replication of genetic loci for ages at menarche and menopause in the multi-ethnic Population Architecture using Genomics and Epidemiology (PAGE) study.* Human Reproduction, 2013. **28**(6): p. 1695-1706.

278. Tayo, B.O., et al., *Genetic background of patients from a university medical center in Manhattan: implications for personalized medicine.* PLoS One, 2011. **6**(5): p. e19166.

279. Roden, D.M., et al., *Development of a large-scale de-identified DNA biobank to enable personalized medicine.* Clin Pharmacol Ther, 2008. **84**(3): p. 362-9.

280. Patterson, N., A.L. Price, and D. Reich, *Population structure and eigenanalysis.* PLoS genetics, 2006. **2**(12): p. e190.

281. Price, A.L., et al., *Principal components analysis corrects for stratification in genome-wide association studies.* Nature genetics, 2006. **38**(8): p. 904-9.

282. Crawford, D.C., et al., *Leveraging Epidemiologic and Clinical Collections for Genomic Studies of Complex Traits.* Hum Hered, 2015. **79**(3-4): p. 137-46.

283. Lim, U., et al., *Asian Women Have Greater Abdominal and Visceral Adiposity Than Caucasian Women With Similar Body Mass Index.* Obesity, 2011. **19**: p. S224-S224.

284. Lim, U., et al., *Susceptibility Variants for Waist Size in Relation to Abdominal, Visceral, and Hepatic Adiposity in Postmenopausal Women.* J Acad Nutr Diet, 2012. **112**(7): p. 1048-1055.

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

286. Gauderman, W.J. and J.M. Morrison, *QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies*, 2006.

287. Maller, J.B., et al., *Bayesian refinement of association signals for 14 loci in 3 common diseases.* Nat Genet, 2012. **44**(12): p. 1294-301.

288. Wang, X., et al., *Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies.* Hum Mol Genet, 2013. **22**(11): p. 2303-11.

289. Peters, U., et al., *A Systematic Mapping Approach of 16q12.2/FTO and BMI in More Than 20,000 African Americans Narrows in on the Underlying Functional Variation: Results from the Population Architecture using Genomics and Epidemiology (PAGE) Study.* PLoS Genet, 2013. **9**(1).

290. Wakefield, J., *A Bayesian measure of the probability of false discovery in genetic epidemiology studies.* Am J Hum Genet, 2007. **81**(2): p. 208-27.

291. Fahey, M.T., A.B. Forbes, and A.M. Hodge, *Correcting for the bias caused by exposure measurement error in epidemiological studies.* Respirology, 2014. **19**(7): p. 979-84.

292. Chen, C.T., et al., *Replication of loci influencing ages at menarche and menopause in Hispanic women: the Women's Health Initiative SHARe Study.* Hum Mol Genet, 2011.

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

294. Crawford, D.C., et al., *Characterization of the Metabochip in diverse populations from the International HapMap Project in the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) project.* Pac Symp Biocomput, 2013: p. 188-99.

295. Friedman, G.D., et al., *CARDIA: study design, recruitment, and some characteristics of the examined subjects.* J Clin Epidemiol, 1988. **41**(11): p. 1105-16.

296. Fried, L.P., et al., *The Cardiovascular Health Study: design and rationale.* Ann Epidemiol, 1991. **1**(3): p. 263-76.

297. *Multi-center genetic study of hypertension: The Family Blood Pressure Program (FBPP).* Hypertension, 2002. **39**(1): p. 3-9.

298. Williams, R.R., et al., *NHLBI family blood pressure program: methodology and recruitment in the HyperGEN network. Hypertension genetic epidemiology network.* Ann Epidemiol, 2000. **10**(6): p. 389-400.

299. Kolonel, L.N., D. Altshuler, and B.E. Henderson, *The multiethnic cohort study: exploring genes, lifestyle and cancer risk.* Nat Rev Cancer, 2004. **4**(7): p. 519-27.

300. *Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group.* Controlled Clinical Trials, 1998. **19**(1): p. 61-109.