THESIS

MULTILOCUS GENETIC ASSOCIATIONS WITH OBESITY OUTCOMES IN HISPANIC AND NON-HISPANIC WHITES USING A PRINCIPAL COMPONENTS REGRESSION APPROACH: THE SAN LUIS VALLEY DIABETES STUDY

Submitted by

Andrea D. Booth-Kalajian

Department of Food Science and Human Nutrition

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2013

Master's Committee:

Advisor: Chris Melby
Co-Advisor: Tracy Nelson

John Hokanson
Jennifer Peel
Laura Bellows

ABSTRACT

MULTILOCUS GENETIC ASSOCIATIONS WITH OBESITY OUTCOMES IN HISPANIC AND NON-HISPANIC WHITES USING A PRINCIPAL COMPONENTS REGRESSION APPROACH: THE SAN LUIS VALLEY DIABETES STUDY

Introduction**:** The overweight and obesity crisis in America has reached alarming rates with little progress of reversing the trend, despite much effort. Heritability has been estimated at up to 70%, though it is still unclear how genetics respond to environmental pressures. Evaluating groups of genes that are known to influence metabolic pathways has given some insight into the variation we see in body composition and prevalence of metabolic diseases.

Methods: Data from the San Luis Valley Diabetes Study's third examination were utilized (1997-1998, n=837). One hundred seven single nucleotide polymorphisms (SNP) were selected from 22 genes that have previously been associated with obesity and type 2 diabetes (T2D) in a cohort of Hispanic and non-Hispanic white (NHW) individuals. Genetic data were reduced to a smaller set of derived factors using principal components analysis (PC). Associations were determined between factors and obesity outcomes.

Results: Hispanics were more likely to have T2D than NHW (19% vs. 11%). Sample minor allele frequencies for 100 analyzable SNPs varied between the two groups with the minor alleles of rs8059937 (*A2BP1*) and rs6822807 (*UCP1*) being significantly more prevalent in Hispanics and rs11724758 (*FABP2*) and rs2239179 (*VDR*) significantly more prevalent in NHWs. SNP variance was redistributed into orthogonal components and 32 were retained for analysis, accounting for 77% of the total variance in genetic data. The

combined genetic information increased predictive power of increases in body mass index (BMI) from the study baseline by 5.6% in Hispanics. Genetic data increased predictability of BMI and waist circumference (WC) in NHWs by 7.5% and 5.1%, respectively. Both groups had a significant increase in knowledge gained (18%) for the prevalence of T2D when genetic information was added to the base model. SNPs from *UCP1* loaded strongly onto PC4, which was associated with BMI change in Hispanics and BMI, WC, and T2D in NHWs. PC7 represented SNPs from *RBP4* and *FABP2*, which was associated with diabetes status in both groups. All obesity outcomes were associated with PC15 in Hispanics, symbolizing SNPs on the *PPARD* and *RBP4* genes. NHWs showed additional associations with components having strong loadings from SNPs on multiple genes, including *ADIPOQ, GC, VDR, PPARG, PPARGC1A, PPARD, UCP2, UCP3,* and *AlOX15*.

Conclusions: When combined together, multilocus genetic data show a larger influence on obesity outcomes than single polymorphisms alone. Base variations in *UCP1*, *RBP4*, and *FABP2* gene sequences are associated with change in BMI and diabetes status in Hispanics and BMI, WC, and diabetes status in NHWs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

The overweight and obesity crisis in America has been escalating for three decades with little hope of resolution, as only one-third of the US population remain at a healthy body weight (BMI<25). More alarming, this epidemic has gone global and is indiscriminate of age, gender, nationality, or socio-economic status (Ogden, Carroll, Kit, & Flegal, 2012; Flegal, Carroll, Ogden, & Curtin, 2010). The Centers for Disease Control and Prevention has declared obesity the number one health risk in America, leading to diseases such as type 2 diabetes mellitus (T2D), heart disease, hypertension, and the metabolic syndrome; overwhelming the health care system, both practically and financially. Despite much effort to determine the cause, prevention, and treatment of this problem, little progress has been made, leading researchers to think beyond traditional practices and theories.

By now, it is well understood that excess weight results from an imbalance of calories consumed versus calories expended. However, it is not clear why some individuals suffer from this trend of weight gain and its metabolic consequences, while others faced with the same environmental pressures do not. The answer to this challenging dilemma is multifactorial, requiring an integrated knowledge of energy balance, genetic predisposition, environmental influences, and social and cultural behaviors. It is likely that common gene-environment interactions have become vulnerable to the eat-more, exercise-less pressures in modern-day society. All things considered, a successful intervention strategy will be tailored to the risks and weaknesses of individuals' perceived genetic outcome (Wake et al., 2009; Hill, 2006; Mitchell, Catenacci, Wyatt, & Hill 2011).

Hispanics are the fasted growing subpopulation in the United States, accounting for 15% of the total US population, and currently have the greatest increase in obesity rates than any other minority group (National Hispanic Caucus of State Legislators, 2010 Policy Brief). For adults at least 20 years of age, 76.9% of Hispanic are overweight ($25 \leq BMI < 30$) with 49.3% of those being obese ($BMI \geq 30$), compared to 67.5% and 48.6% respectively for non-Hispanic whites (NHW). Hispanic women are especially vulnerable, having a 30% greater chance of being obese than NHW women (43.0% versus 33.0%). Additionally, Mexican American adults have higher rates in almost every age-group and obesity level than their Hispanic counterparts for both men and women (Flegal et al., 2010) and Mexican American children are 1.4 times more likely to be overweight than NHW children (Centers for Disease Control and Prevention, 2012). Clearly, this epidemic is growing at an accelerated pace and requires immediate research and intervention for many individuals to maintain an acceptable quality of life.

As a result of increased rates of overweight and obesity, Hispanics and particularly Mexican Americans experience more related morbidities, including T2D. A 2010 report by the Diabetes Surveillance System indicates there are 12.9 newly diagnosed cases per 1,000 Hispanics each year compared with 7.7 per 1,000 NHW, notwithstanding pre-diabetes and undiagnosed cases. In addition to higher incidence and prevalence, Hispanics are contracting the disease at a younger age and are more likely to die from diabetes in their lifetime. Therefore, a high prevalence of diabetes and other complications of obesity create stress on the quality of life and economy of the Hispanic community.

In an effort to stifle this burdening disease, researchers are determined to uncover the root of the problem through genetic studies. By learning about specific gene

contributions as well as environmental interactions, there is optimism towards finding successful prevention strategies to avoid future crises stemming from obesity and T2D. A theoretical report by Claude Bouchard advises that genetic variation cannot be dismissed as a strong determinate of an individual's predisposition to become obese (Bouchard, 2007). He and others have illustrated through twin and familial studies that body mass index is highly heritable, suggesting that one's genetic makeup is a major contributor to body weight. Similarly, genetic studies are gaining predictive power through identification of susceptible loci on the propensity towards developing T2D, which could impact the way we diagnose and treat the disease.

Although genetic contribution to obesity and T2D has been estimated at high proportions ranging from 40-70%, many studies have failed to reach this level of predictability (Lyon & Hirschhorn, 2005). One explanation could be that they have not uncovered key loci or groups of loci with significant influences. With the introduction of genome-wide association studies (GWAS), many genes have been identified that are associated with BMI, though a robust combination of genes that explain significant variation in BMI has yet to be discovered. Therefore, countless new gene combinations are obtainable for research and may potentially contain greater explanatory power.

Success was seen with the fat mass (*FTO*) and melanocortin-4 receptor (*MC4R*) genes for a genetic explanation of obesity and its metabolic consequences, such as T2D (Li et al., 2010; Cauchi et al., 2009). However, little progress has been made with other potentially relevant gene sequences. This may be partly because many genes have not been discovered and partly because individual variants have little effect by themselves.

When gene effects are added together, there is potential for much more explanation of the variability in obesity outcomes.

Peterson et al developed a genetic risk sum score with increasing appearance of 56 risk alleles identified through GWAS, which was significantly associated with BMI in European- and African-Americans (Peterson et al., 2011). Similarly, *Iwata et al* proposed a 14-allele genetic risk score and showed its association with T2D in Japanese adults (Iwata et al., 2012). As these studies are a step in the right direction, their effects have been small and incomplete, possibly due to the minimal amount of genetic information included. Therefore, a simplified approach is needed to include substantially more genetic information that can be summarized into a usable format.

Genetic data alone cannot provide a complete description of obesity risk, but must be integrated with individual lifestyle and environmental factors. With an abundance of food available and low motivation for exercise, those with even modest likelihood for obesity are succumbing to this condition. The current study aims at developing a deeper understanding of genetics, environment, and their interaction on markers for obesity and T2D in a heavily Hispanic population.

STATEMENT OF PURPOSE

The purpose of the present study is to determine the aggregate effect of appropriately selected gene variants on markers for obesity and T2D in an admixed population of Hispanics and NHW. The genes found to have the most influence on obesity and T2D are thought to interact with lifestyle and environmental factors.

The following hypotheses will be challenged:

1. Selected loci variants will be grouped into a smaller set of derived factors that explain a portion of the variance in body mass index (BMI), change in BMI, waist circumference (WC), and prevalence of T2D.

2. Influential genes will have a combined effect on BMI, change in BMI, WC, and prevalence of T2D.

3. The variance in BMI, change in BMI, WC, and prevalence of T2D explained by genetics are tested against demographic variables, including age, gender, ethnicity, and smoking status.

4. The variance in BMI, change in BMI, WC, and prevalence of T2D explained by genetics will be tested separately for Hispanics and non-Hispanic whites.

CHAPTER 2

Literature Review

The Obesity Epidemic

In the United States, obesity has become a significant concern over the last several decades and is now spreading to many regions around the world. The healthcare industry is already overwhelmed with obesity-related chronic diseases, such as T2D, cardiovascular disease, hypertension, and the metabolic syndrome; with an expected increase in future diagnoses. Two-thirds of adults in the U.S. are overweight ($25 \leq BMI < 30$), half being classified as obese ($BMI \geq 30$), with an estimated $147 billion spent each year on medical costs associated with obesity (Ogden et al., 2012; Flegal et al., 2010; Finkelstein, Trogdon, Cohen, & Dietz, 2009). Sadly, childhood obesity is on the rise with 17% of youth being obese in 2010, tripling since 1980 (Ogden et al., 2012). Studies show that BMI in adolescence is a strong indicator of BMI in adulthood where children who are obese are very likely to become obese adults (Guo & Chumlea, 1999).

The current standard treatment for a diagnosis of overweight or obesity includes education with an emphasis on healthier eating habits and exercise programs. Though these efforts are commendable, they are expensive and largely ineffective (Wake et al., 2009). To properly address the obesity challenge will require a multifactorial approach, incorporating our knowledge of energy balance, genetic predisposition, environmental influences, and social and cultural behaviors (Mitchell et al., 2011). As prevention is the best approach, an effective way to tackle and potentially reverse this epidemic is for nutrition scientists to first understand the etiology of the disease and comprehend what is

causing the shift in overall weight status of Americans. The answer may be a common gene-environment interaction that has become vulnerable to the eat-more, exercise-less pressures we are facing in today's society (Hill, 2006). Consequently, researchers are beginning to focus more resources on alternative methods for obesity treatment and prevention that incorporate individualized education with specialized programs based on a combination of factors.

Obesity and T2D in Hispanic populations

Among subpopulations, Hispanics have one of the highest rates of obesity when compared to NHW with 37.9% and 32.8%, respectively (Flegal et al., 2010). As of 2008, Hispanics made up approximately 15% of the U.S. population and are experiencing the greatest increase in prevalence of obesity than any other minority group (National Hispanic Caucus of State Legislators, 2010 Policy Brief). Further, Mexican Americans are the most overweight when compared to Cubans and Puerto Ricans within the Hispanic population. Mexican American women have the greatest disparity with 78% being overweight or obese compared to 60% of NHW women, while Mexican American children are 1.4 times more likely to be overweight than NHW children (Centers for Disease Control and Prevention, 2012). A study analyzing dietary habits of a population in Massachusetts showed that elderly Hispanics were more likely to consume a diet consistent with higher BMI and WC than elderly non-Hispanics. The study determined Hispanics who were more acculturated, or accustomed to American traditions, had a more favorable nutritional intake compared to those following a conventional Hispanic diet (Lin, Bermudez, & Tucker,

2003). Therefore, cultural influences cannot be overlooked when considering nutrition and health behaviors in subpopulations.

In 2010, the Diabetes Surveillance System reported an incidence of newly diagnosed cases of 12.9 per 1,000 Hispanics compared with only 7.7 per 1,000 NHWs. When first reported in 1997, the median age of diagnosis was almost 11 years younger for Hispanics than whites, although the gap has narrowed to less than four years difference. These numbers do not consider undiagnosed cases or pre-diabetes. The Office of Minority Health reported in 2012 that Hispanics are almost twice as likely to be diagnosed with diabetes than NHWs and they are 1.5 times more likely to die from diabetes (Centers for Disease Control and Prevention, 2012). Therefore, obesity and its health complications, especially diabetes and cardiovascular disease, pose a significant economic burden on the Hispanic community, spending a disproportionate amount of their income on healthcare costs. In addition, the higher weight status of this population may lead to discrimination and lost employment opportunities, which puts even more strain on the social and financial burden of this community.

Genetic studies in obesity

According to Claude Bouchard of Pennington Biomedical Research Center in Baton Rouge, LA, there are four main contributors to obesity: built environment (automobiles, elevators, lack of safe sidewalks), social environment (advertising, events, culture), behavior (high-fat diet, high sugar intake, TV watching, video games), and biology (individual genome). He cautions that genetic variation cannot be dismissed as a strong determinate of an individual's predisposition to become obese (Bouchard, 2007). This

genetic propensity for increased body weight has been proven in twin and familial studies on the heritability of BMI. One study published in 1990 by *Stunkard et al* included 93 pairs of identical twins who were raised apart from one another. It resulted in an intra-pair correlation coefficient for BMI of 0.70 and 0.66 for men and women, respectively, which illustrates a genetic influence cultivated in separate environmental conditions (Stunkard, Harris, Pedersen, & McClearn, 1990). Another study in which identical male twins were overfed for a prolonged period of time showed significant similarities in fat mass response within pairs, with a three-fold increase in variance between pairs compared to within pairs (Bouchard et al., 1990). These results beg the question of how genetics play a role in energy intake, energy expenditure, and the susceptibility to become fat. Likely factors would include how efficiently we utilize food for fuel and our physical activity response to over-nutrition. With childhood obesity on the rise, it is important to address the strong connection between child and parental body size. *Whitaker et al* report a 12-fold increase in risk for child obesity when both parents are obese, attributable to both genetics and lifestyle factors, independent of demographic representation (sex, age, ethnicity, socioeconomic status) (Whitaker, Jarvis, Beeken, Boniface, & Wardle, 2010). These data suggest that a person's genetics play an integral role in their likelihood of becoming obese. Thus, if clinicians could integrate DNA into a screening process, then individuals may be able to assess their risk for greater adiposity and develop more effective prevention strategies like calorie control and routine exercise.

There is consensus among researchers that genetics play a critical role in understanding the variation in body weight given that some individuals are thin and some are fat when exposed to similar environmental factors. The extent of genetic contribution

has been estimated at up to 70%, yet many of these genes, or groups of genes, have not been discovered (Lyon & Hirschhorn, 2005). With the help of GWAS, it is becoming clear that obesity traits are the result of a combined effect of many small gene contributions that alone have very little explanatory power (Fall & Ingelsson, 2012). Whether the combined effect of individual polymorphic loci is additive or not is still being determined.

Over the past decade, more than 50 single nucleotide polymorphisms (SNPs) have been recognized as being associated with standard obesity measures, including BMI, WC, T2D, and cardiovascular disease. Although each locus by itself appears to have limited explanatory power on the variance in obesity measures, their cumulative effects may harbor some insight. *Li et al* found that 12 SNPs identified in a large-scale GWAS as *BCDIN3D, BDNF, ETV5, FTO, GNPDA2, KCTD15, MC4R, MTCH2, NEGR1, SEC16B, SH2B1*, and *TMEM18*, had modest predictive power for increased risk of overweight and obesity, while their individual effect sizes were trivial (Li et al., 2010). This gives hope to the possibility of uncovering a group-wise effect that incorporates clusters of the most influential gene variants. As single gene mutations provide explanation for only 1-5% of obesity cases, there is no doubt this is primarily a multi-genetic disorder with many gene groupings yet to be determined. In a 2003 review by Loos and Bouchard, they give details of possible gene grouping that target either energy intake or energy expenditure. They also propose four super-groups depending on one's genes and environment as 1) genetic obesity, 2) strong predisposition, 3) slight predisposition, and 4) genetically resistant, which highlights obesity risk as a scaled disorder rather than a dichotomous outcome (Loos & Bouchard, 2003). Ideally, knowing ones genetic position and risk of obesity will help tailor prevention strategies by focusing on ways to lessen the magnitude of excess adiposity and curtail its

manifestation. For example, someone with low satiety signals could learn to measure their food intake without having to rely on an internal sign to stop eating.

Recent research has even spread into the field of epigenetics, which encompasses a reversible alteration in gene expression without a change in DNA sequencing (Slomko, Heo, & Einstein, 2012; Drong, Lindgren, & McCarthy, 2012). In other words, genes may change the way they respond to a certain event based on a few influential factors. Epigenetics may eventually be found to contribute far more to BMI variation than genetics. After all, genes exert their influence by way of the proteins they encode, so it seems likely that variability in gene expression may be more important than genetic variability. This theory has a major impact on future hypotheses and though epigenetics is an important new area in genetic research, it is beyond the scope of this paper.

Genetic studies in T2D

As with obesity, GWAS are helping us better understand the predisposition to develop T2D. Although family history, age, and weight status, among other factors, have high predictive power, it is advantageous to know who might be at greater risk for getting the disease before it begins to manifest. Obtaining genetic information with strong predictive power can have a significant impact on how we diagnose and treat diabetes, including pre-diabetes, which could result in a higher quality of life and less money spent on complications of the disease.

The first breakthrough in identifying loci associated with T2D was made in 2006 as a variant in transcription factor 7-like 2 (*TCF7L2*). The investigation found a 1.86- and 2.15-fold increase in risk for women and men, respectively, who were homozygous carriers

of the T-allele (Zhang et al., 2006). Since then, at least 20 individual loci have been indentified as having a significant and robust association with the disease and many other loci with loose associations (McCarthy & Zeggini, 2009). Investigators are finding promising genetic data with strong connections to T2D in both obese and non-obese populations and these findings extend to multiple ethnic backgrounds. A recent genetic variation study on an admixed Mexican population found 6 new loci (*SLC30A8, HHEX, CDKN2A/2B, IGF2BP2, CDC123/CAMK1D*, and *KCNQ1*) associated with T2D that may be specific to this ethnic group (Gamboa-Melendez et al., 2012).

As great progress is being made in the identification of genetic variants associated with T2D, the question remains on what to do with this information for prevention and treatment of the disease. One cannot change their genes and, contrary to obesity genes, T2D may not be avoided simply by knowing susceptibility. Further, if someone views himself or herself as inevitably destined to develop T2D, they may be less prone to engage in prevention efforts. To complicate things even more, a study stratifying T2D cases by BMI found that significant gene variants associated with the disease differed between lean and obese subjects, with lean T2D cases being more susceptible and having more risk alleles (Perry et al., 2012). However, there is hope that knowing ones susceptibility or nature of their predisposition can help tailor new drugs and treatment programs for increased effectiveness and fewer complications.

It goes without saying that many of the genetic findings will be similar for obesity and T2D in obese populations. While biological and genetic links are being made, McCarthy points out these links are not always well translated to clinical practices and the greater discovery of loci within both polygenic diseases makes it even more difficult to formulate a

comprehensive management plan. This overlap in susceptibility should lead us to a combined solution for a healthier and better quality of life for these individuals, which McCarthy has termed 'personalized medicine' (McCarthy, 2010).

Genetic studies in Hispanic populations

Genetics do not only set individuals apart, but whole populations as well. Clinical investigators have found that using genetic markers in the form of allele differences is a much better indication of ancestry than a self-selected race or ethnicity category (Kosoy et al., 2009). However, genetic information can be a lot more revealing than just our country of origin. The Viva La Familia Study was developed between 2000-2004 in Houston, TX to genetically map childhood obesity and its metabolic consequences in Hispanics (Butte, Cai, Cole, & Comuzzie, 2006). They found a strong genetic contribution to the high prevalence of childhood obesity in Mexican-American families with significant heritability for obesity-related traits. Further, this study employed a GWAS in 815 Hispanic children that localized novel genetic loci associated with the pathophysiology of childhood obesity (Comuzzie et al., 2012).

The San Luis Valley Diabetes Study (SLVDS) was initiated in 1984 to determine the prevalence, risk factors, and complications of T2D in self-reported Hispanics and NHW. It used a case-control design with participants selected from a rural two-county region of Southern Colorado. According to the 1980 US census, when initial data collection took place, the region was 49.4% male, 43.6% Hispanic, and had median annual family income lower than the Colorado average (Hamman et al., 1989). Participants in the SLVDS control group were selected from the community using a two-step process, stratified by age, sex,

ethnic group, and county to match that of the diabetic group. Of the controls identified, 1351 attended the baseline clinic between 1984 and 1988 and were prospectively followed for the development of diseases. They were examined again from 1988 to 1992 and a third time from 1997 to 1998.

Data from the SLVDS have been published previously with many intriguing results. *Nelson et al* revealed ethnic differences in insulin sensitivity between Hispanic and NHW women related to dual-energy X-ray absorptiometry measures of abdominal fat, over and above waist circumference (Nelson, Bessesen, & Marshall, 2008). Additionally, *Rewers et al* discovered a significantly lower prevalence of coronary heart disease in Hispanic men and women than in NHW for those with T2D or impaired glucose tolerance. They speculate the reason for a lower rate of heart disease despite the higher rate of T2D is due to unknown protective factors or competing illnesses (Rewers, Shetterly, Baxter, Marshall, & Hamman, 1992). *Damcott et al* found a significant interaction between two genes in a biologic pathway whose variants influence insulin resistance and body composition in male, non-diabetic SLVDS participants (Damcott et al., 2004). These studies and others have increased our knowledge of multiple disease states and ethnic differences between Hispanic and NHW Americans.

Environment and lifestyle modification to genetics

There is no doubt that environmental changes have had a negative impact on body weight, proven by the recent spike in obesity rates corresponding to an increase in food availability and convenience factors in the U.S. (Isganaitis & Lustig, 2005). However, as the majority of people live in this obesigenic environment, we still observe a broad spectrum of

body weight, leading us to believe there are multiple interactions of environment and genetics. Strong behavioral factors like overconsumption of macronutrients and physical inactivity are clear contributors to obesity and T2D, so much that even with a genetic propensity towards these diseases, control of food intake and energy expenditure can curtail their development (Temelkova-Kurktscheiv & Stefanov, 2012). Similarly, individuals with genetic resistance to obesity are still at risk if behavior is not controlled. Other modifying factors include age, sex, race, ethnicity, and income, with greater disparities in obesity among women, Mexican-Americans, and low-socioeconomic populations (Wang & Beydoun, 2007).

As with early detection and prevention of obesity, treatment is equally susceptible to environmental influences and modifications. In a review by Choquet and Meyre, they suggest three options for treatment of obesity: lifestyle intervention, pharmacotherapy, and bariatric surgery (Choquet & Meyre, 2011). With monogenic obesity when one gene mutation is involved, extreme actions such as gastric banding and bypass techniques may be inevitable, as mutations and/or loss of function in certain genes demand a mechanical intervention. However, polygenic overweight and obesity when multiple genes are involved can benefit from lifestyle modifications and in some cases pharmacotherapy, if interventions are tailored towards an individual's obesity origin as well as their unique environmental modifiers. This has been observed with the FTO gene variant and its interaction with macronutrient composition (Grau et al., 2009) and moderate intensity exercise (Mitchell et al., 2010).

As we inch closer to an understanding of genetic contribution to body composition, nutrition researchers are gaining enthusiasm about the topic. Still, there is much to learn

before effective improvements can be made in our intervention strategies, ideally targeting obesity prevention. Early identification of obesigenic traits should not be viewed as an individual's body weight destiny, but can assist in formulating an effective prevention strategy to avoid early onset adiposity. Geneticists agree that, consistent with many diseases, prevention is the best strategy and though a genetic mapping to uncover this inherent outcome would be ideal, the best indicators to date are demographic and environmental factors (Choquet & Meyre, 2011). Additionally, both desire and ethical considerations should be considered with genetic testing to ensure information is not misinterpreted or misused.

Genetic composite score

A commonly used analytic procedure to summarize combined effects of allele frequency data involves a composite sum score incorporating a defined set of genetic variants. *Peterson et al* constructed a genetic risk sum score (GRSS) comprised of variants proven or suggested to have an association with BMI in a group of European- and African-Americans. In their study, they show a composite sum score was significantly associated with BMI with an average effect of 10 risk variants resulting in 8 extra pounds for males and 7 extra pounds for females (Peterson et al., 2011). Their composite score employs an additive model, as gene-to-gene interactions have not been suggested previously on genetic obesity data. A strong statistical association exists between BMI and GRSS, even after adjustments were made for age, sex, and race, however the GRSS only accounted for an additional 0.66% of variation in BMI. Similarly, *Iwata et al* constructed a significant genetic

sum score consisting of 14 loci known to be associated with diabetes, though only a small increase in odds of getting the disease (1.26) was detected (Iwata et al., 2012).

Regardless of their limited predictive power, genetic composite scores have potential for future research in finding groups of alleles that are jointly correlated with obesity outcomes. Also, they provide further evidence of polygenetic diseases such as obesity and T2D. A genetic risk score comprised of 29 SNPs was significantly correlated with BMI in a 4-decade longitudinal study conducted in New Zealand evaluating a birth cohort through 38 years of age. The study reveals a polygenic risk for obesity, manifesting early in life and resulting in rapid childhood growth, particularly following adiposity rebound when a child's BMI begins to increase. The correlation between risk score and BMI over 4 decades of life was independent of family history captured by parental BMI (Belsky, Moffitt, & Houts, 2012). This proves genetic composite scores could be beneficial in practice if incorporated into child and adult wellness protocols as a form of obesity risk assessment.

Principal components analysis

Principal components analysis (PCA) is an analytic technique used to find patterns in data sets with a very large number of explanatory variables. It creates subsets of variables by grouping them based on correlative properties and then obtains a linear combination of each group such that the groups are perpendicular, or uncorrelated, with each other. PCA is useful in exploratory data analysis when wanting to group variables that have similar descriptive power in order to compress the data set into a few explanatory groups called 'factors' that explain the relevant information in the sample data. Factors are

chosen based on a hierarchy according to the variance they retain from the full data set (Jackson, 2004).

Another situation in which PCA is a valuable tool arises when a data set is in jeopardy of multicollinearity, or when two or more explanatory variables are correlated with each other. This occurs when variables have overlapping descriptive power and may result in an unstable model and inaccurate conclusions about the outcome variable. While multicollinearity may not be an issue when the objective is to predict an outcome, it becomes highly problematic when explanation is the ultimate goal (Berry & Feldman, 1985). PCA is a remedy for the problems that occur with multicollinearity by creating a new set of orthogonal variables. However, it is no longer possible to determine the significance of individual predictors on the outcome variable, as predictors are then considered as a group (Dohoo, Ducrot, Fourichon, Donald, & Hurnik, 1997). Then again, with genetic data, in particular SNPs, the individual contribution of each input variable is trivial and grouped data is more informative. PCA is a powerful filter and reduction tool, especially in exploratory analysis, by allowing the data to determine its own descriptive factors.

A practical application of PCA in genetic studies is its use on ancestral data where individuals of an admixture population are traced to their continent of origin using differences in allele frequencies to create ancestry informative markers (AIMs). While previously having to rely on self-identified ethnic affiliation, AIMs have been effective in population and disease association studies where ethnic background is likely to create an issue with confounding (Kosoy et al., 2009).

Single Nucleotide Polymorphism Selections

The ataxin-2 binding protein 1 gene (*A2BP1*) is a splicing regulator that controls neuronal excitation in the brain and has been associated with percent body fat in Pima Indians. *Ma et al* attempted to replicate this by searching for the same association in French adults and children, Caucasians, and Native Americans. Though no single variant in *A2BP1* was found to be significantly associated with obesity, their analysis suggest that variations in the gene may influence obesity and adiposity through the hypothalamic *MC4R* pathway (Ma et al., 2010).

Obesity has been described as an inflammatory disease based on the fact that obese individuals have increased levels of biochemical markers of inflammation. An abundance of adipose tissue can confuse the endocrine system with over-secretion of adipokines, causing the release of inflammatory mediators and affecting systemic processes. One cytokine that plays a major role in stimulating the immune system is interleukin-6 (*IL-6*). It is clear that white adipose tissue-derived *IL-6* is over-expressed with obesity, resulting in chronic low-grade inflammation. If the inflammation persists for a long period of time, the risk for cardiovascular disease and T2D increases (Emanuela et al., 2012). It seems this process may be viewed as a consequence of obesity and not a cause; however it is possible that *IL-6* is affected by one's diet, physical activity, and environment prior to development of the obesity phenotype. Further, the interleukin-6 receptor (*IL6R*) is expressed in the hypothalamus of the brain and helps regulate appetite and energy intake (Wallenius et al., 2002). Variants in the *IL6R* gene were studied in Pima Indians, a population known for increased adiposity, and shown to be significantly associated with BMI (Wolford, Colligan, Gruber, & Bogardus, 2003).

Adiponectin, C1Q and collagen domain containing (*ADIPOQ*) is a gene expressed exclusively in adipose tissue and is involved with metabolic and hormonal processes. Adiponectin is negatively correlated with amount of adipose tissue and sends hunger related signals to the brain allowing modification of a persons eating behavior. The gene was first reported in 1996 to be involved with a signaling function from adipocytes and has reduced expression in obese humans (Hu, Liang, & Spiegelman, 1996; Emanuela et al., 2012). *ADIPOQ* variant rs1501299 was previously studied in a young Mexican-American population, but failed to show significance with obesity traits (Duran-Gonzalez et al., 2011), possibly because it was not combined with appropriate co-variants. However, a group from the Finnish Diabetes Prevention Study was found to have variants rs266729, rs16861205, rs1501299, rs3821799 and rs6773957 of the *ADIPOQ* gene that were significantly associated with body weight (Siitonen et al., 2011). Adiponectin receptor genes, *ADIPOR1* and *ADIPOR2*, also affect fatty acid catabolism and glucose levels by activating an AMP-activated kinase-signaling pathway. All three genes and numerous variants within the genes have been studied extensively for their suspected involvement in obesity, cardiovascular disease, and T2D.

A gene encoding for an enzyme responsible for lipid peroxidation, arachidonate 15-lipoxigenase (*ALOX15*), has been implicated in the pathogenesis of inflammatory disorders, including obesity. The rs916055 variant of this gene was found to be significantly associated with fat mass percent in a study of Chinese men and women (Ke et al., 2012).

PPARs are genes encoding for proteins known as peroxisome proliferator-activated receptors that regulate target genes by increasing or decreasing their transcription. The receptor has three forms, α, δ, and γ, which together are expressed in almost all tissues

throughout the body. In addition, PPARs can change their function based on whether they bind with a co-activator. This group of receptors is typically activated by fatty acids, either from diet or adipose tissue, or their metabolites and plays an important role in lipid homeostasis and adipocyte differentiation. Additionally, transcription of adiponectin is regulated by *PPARγ*, affecting a variety of metabolic reactions. Many variants of these receptor genes have been studies for their interaction with obesity measures. Specifically, the Pro12Ala variation of *PPARγ* (rs1801282) is associated with obesity in a population of Mexican Americans where those with at least one risk allele had significantly higher BMI and WC (Cole et al., 2000). There is countless research in the area of PPARs and their potential role in development of obesity and thus all will be included in the present analysis.

Intracellular fatty acid binding proteins (*FABP*) participate in the transport of long-chain fatty acids. The polymorphism Ala54Thr (rs1799883) in its intestinal version has been studied considerably and found to be associated with obesity measures. *Martinez-Lopez et al* noticed differences in response to a moderate-fat diet for those with one or two mutant alleles of the *FABP2* gene in a Mexican population (Martinez-Lopez et al., 2013). The expression of this protein in adipose tissue is called *FABP4* and plays an important role in maintaining glucose and lipid homeostasis. Circulating levels of *FABP4* are higher in obese children and correlates with BMI, HOMA (an estimate of insulin resistance), and hsCRP (a marker for inflammation), according to a study performed in Louisville, KY (Khalyfa et al., 2010). Interestingly in the SLVDS, *FABP4* was found to interact with *PPARγ* and have a significant effect on insulin sensitivity and body composition in Hispanic and NHW males (Damcott et al., 2004). Not only does it utilize the same data set as this one, but

the SLVDS also highlights the importance of considering multiple genes and their interactions with each another when exploring the etiology of complex diseases such as obesity.

The fat mass and obesity associated gene (*FTO*) was first discovered in 2007 and has since been the subject of much research, though the exact physiological function of this gene is still not clear. During this discovery, *Frayling et al* concluded that the 16% of adults in his study homozygous for the risk allele were 6.6 pounds heavier and 1.67 times more likely to be obese than those without the risk allele or heterozygous (Frayling et al., 2007). *FTO* variant rs8050136 has been consistently associated with obesity risk in both children and adults, however differences in ethnicity have not yet been determined (Mei et al., 2012).

There is speculation that vitamin D deficiency is a main cause of common obesity, although it is controversial as to which came first when explaining their relationship. In a study of 108 obese subjects undergoing bariatric surgery, over 70% were deficient in vitamin D, which could lead to poor calcium metabolism and decreased parathyroid function (Hultin, Edfeldt, Sundbom, & Hellman, 2010). A relationship in vitamin D binding protein polymorphisms and obesity was reported for Caucasians, revealing a strong association between SNP rs17467825 with percent body fat (Jiang et al., 2007). Similarly, the vitamin D receptor gene has been studied for its contribution to obesity risk, being highly interconnected with the vitamin D endocrine system. Specifically, the variant rs3782905 was associated with both BMI and WC in a random sample of NHW women (Ochs-Balcom et al., 2011). Though individual significances were not found, a study comparing vitamin D-related genes and BMI in Chinese women included potential variants

rs17467825, rs222020, rs222029, and rs2298849 in the vitamin D binding protein (*GC*) and rs10783219, rs2239179, rs4334089, rs4760648 in the vitamin D receptor (*VDR*) with a per-allele p-value for BMI and weight (kg) less than 0.50 and will be replicated in this study (Dorjgochoo et al., 2012).

The melanocortin 4-receptor gene (*MC4R*) is expressed in the hypothalamus and codes for a protein that binds alpha-malanocyte stimulating hormone, which produces satiety factors signaling an individual to reduce food intake. Mutations in this gene can have a profound effect on energy intake and be a risk for early onset obesity. The *MC4R*-obesity linkage was first discovered in the late 1990s and has been replicated many times; including a recent study in Mexican children that highlights the effect of SNP rs17782313 near the *MC4R* gene (Mejia-Benitez et al., 2013).

Uncoupling proteins (*UCP*) are distributed in the mitochondria of cells and cause an inefficient use of energy by separating oxidative phosphorylation from ATP synthesis. They weaken the membrane potential by creating a channel that transports protons inside of the inner mitochondrial membrane without producing ATP, also referred to as the proton leak. The energy potential is dissipated as heat, leading to more nutrients required for a given energy output. Consequently, low expression of UCPs lead to a more efficient energy conversion process, thus excess nutrient inputs are stored as fat rather than lost as heat. We now know these proteins are expressed in various tissues throughout the body, mainly white and brown adipose tissue and skeletal muscle, and may contribute to a protective effect against obesity. Polymorphism rs659366 in the *UCP2* gene has been shown to correlate with obesity and fat distribution in Spanish men and women (Martinez-Hervas et al., 2012). Further, when the SNP mentioned above was combined with SNP

rs1800849 in *UCP3*, the association with obesity was even stronger in a group of Spanish children and adolescents (Ochoa et al., 2007). Another study confirmed a strong association in variants of *UCP3* and BMI, particularly variant rs2075577, which may decrease uncoupling activity (van Abeelen et al., 2008). There is evidence that all three genes encoding for uncoupling proteins, UCP1-3, can potentially explain variations in body weight among individuals.

Retinol binding protein 4 (*RBP4*) is found in the plasma of humans and functions to carry vitamin A from the liver to peripheral tissues. The protein is highly expressed and secreted by adipocytes with circulating levels being a useful marker for obesity. SNP rs3758539 is involved in the regulation of *RBP4* and has been associated with obesity, perhaps through its adipogenesis properties (Munkhtulga et al., 2010).

In summary, obesity and T2D are both very complex diseases that encompass many dimensions, including heredity, environment, culture, lifestyle, and genetics. As we have pointed out, none of these dimensions can be considered alone, as they are interconnected and influence one another with a synergistic effect. Therefore, scientists can only hope to uncover a portion of the variability seen in body composition and metabolic diseases in a multiethnic population.

CHAPTER 3

Materials and Methods

Study Participants

The present research is part of the San Luis Valley Diabetes Study (SLVDS), which was initiated in 1984 to determine the prevalence, risk factors, and complications of T2D in self-reported Hispanics and NHWs. The SLVDS used a case-control design with participants selected from a rural two-county region of Southern Colorado. According to the 1980 US census, when initial data collection took place, the region was 49.4% male, 43.6% Hispanic, and had median annual family income lower than the Colorado average. Methods for data collection and results of original analyses are described elsewhere (Hamman et al., 1989). Participants in the SLVDS control group were selected from the community using a two-step process, stratified by age, sex, ethnic group, and county to match that of the diabetic group. Of the controls identified, 1351 attended the baseline clinic between 1984 and 1988 and were followed for an additional decade for prospective analysis of disease development. They were examined again from 1988 to 1992 and a third time from 1997 to 1998 (n=837). The current study utilizes data from the third and final examination in a cross-sectional analysis of multiple gene variants, environmental factors, and disease status. Informed consent was obtained from all study participants and the University of Colorado Institutional Review Board approved each study protocol.

Clinical Measurements

Comparative measurements were collected from study participants using calibrated equipment and trained personnel. Body mass index was calculated as weight in kilograms divided by height in meters squared (kg/m^2). Participants' weight was measured with a balance beam scale calibrated weekly with standard weights. Height was measured without shoes using a stadiometer to the nearest 0.1 cm. Waist circumference was assessed at the 10^{th} rib with a steel tape to the nearest 0.1 cm. Diabetic status was determined from blood glucose levels following an overnight fast of at least 8 hours or at the 2-h point following a 75 g oral glucose challenge. Samples were collected during an oral glucose tolerance test (OGTT), pre and 2 hours post a 75 g oral load and analyzed by the glucose oxidase method. Diagnosis was based on the American Diabetes Association Standards of Medical Care–2011 as fasting venous plasma glucose ≥ 126 mg/dl or 2-h post-load glucose ≥ 200 mg/dl. Individuals who were taking insulin or other glucose lowering medications at the time of the examination were classified as diabetic regardless of their blood glucose levels. Age was calculated as the participants' birth date subtracted from the clinic visit date. Ethnicity and smoking status was assessed through a questionnaire given at the visit.

Genetic variants

Study participants underwent genotyping to determine the presence of polymorphisms, which are variations that occur in human DNA found in more than one percent of the general population. These variants reside at a single base-pair site within the genome, involving substitutions of A, T, C, or G. The present study uses 384 candidate

allele variations from 170 gene sequences. SNPs were genotyped by Illumina, Inc (San Diego, CA) with 1678 DNA samples (1398 genomic DNA samples and 280 whole genomic amplification samples) in 2009-2010. The genotyping success rate for the whole genomic samples and whole genomic amplification samples were 98.4% and 93.9%, respectfully.

Single Nucleotide Polymorphism Selections

Selection of SNPs was based on a careful review of current obesity literature related to each candidate gene. SNP data were recorded as either heterozygous or homozygous for one of the two alleles. Candidate genes were initially filtered using the HuGE Navigator (version 2.0), an integrated, searchable knowledge base of genetic associations and human genome epidemiology, funded by CDC's Office of Public Health Genomics (Yu, Gwinn, Clyne, Yesupriya, & Khoury, 2008). Phenopedia was selected from the home page and obesity was used as the disease criteria, which resulted in 3,048 publications related to 1,490 different genes. These genes were cross-referenced with candidate genes from the SLVDS and investigated for entry into this analysis using literature provided by the search. Genes and individual SNPs were further investigated using GeneCards, The Human Gene Compendium, created and maintained by Weizmann Institute of Science (Rebhan, Chalifa-Caspi, Prilusky, & Lancet, 1997) and Gene, a database provided by the National Center for Biotechnology Information.

Statistical Analysis

Hardy-Weinberg Equilibrium (HWE) was examined using the goodness-of-fit test. SNPs were coded as either 0 or 2 for homozygosity and 1 if they were heterozygous where

direction of risk was not determined. Candidate SNPs were entered into a PCA to establish patterns in the data where SNP variance was redistributed into factors. Factors were selected that retain most of the relevant information while decreasing the dimension of the data set. The selected factors containing SNP data were then analyzed using linear regression models to estimate their contribution to the variation of BMI, BMI change from baseline, and WC, while logistic regression was used for prevalence of T2D. A base model of age, gender, and smoking status was used to control for confounding effects. Regression models were run separately for Hispanics and NHWs. All analyses were carried out using SAS 9.3 (Cary, NC).

CHAPTER 4

Results

Participant Characteristics

Clinical data were collected from 837 participants of the SLVDS 1997-1998 examination. Demographic data were retrieved from the initial 1984-1985 examination, including the self-reported ethnicity indicator. Table 1 summarizes this information by gender and ethnicity. There were 353 Hispanics (153 males and 200 females) and 484 NHWs (228 males and 256 females) with an average age of 63.0 years old. During the 1997-1998 examination, Hispanic males were more likely to smoke than Hispanic females or whites with 22% being current smokers compared to only 11%-13%.

The average participant was overweight ($25 \leq BMI < 30$ kg/m^2) with Hispanic females having the highest average BMI of 28.0 (5.2) kg/m^2. Following them were white males with an average BMI of 27.5 (3.9) kg/m^2, while Hispanic males and white females had similar BMIs of 26.9 (5.4) kg/m^2. Hispanic females also had the greatest increase in BMI from the baseline examination with a 1.66 unit change, which was comparable to the 1.39 unit change for white females. On average, both Hispanic and white males had less than a one unit increase in BMI from baseline. The average WC of Hispanic and white males was 96.1 (9.7) cm and 98.8 (9.5) cm, respectively, while it was 89.1 (11.2) cm and 86.3 (10.7) cm for Hispanic and white females, respectively. Of the participants who did not have a diagnosis of diabetes at baseline, 20% of Hispanic females and 17% of Hispanic males were diagnosed with diabetes in the following 13 years, compared with only 11% of white males and females.

Table 1. Participant Characteristics (N=837)
1997-1998 examination

	Hispanic		NHW	
	Male N=153	Female N=200	Male N=228	Female N=256
Age (yrs), mean (sd)	62.9 (12.4)	61.6 (12.4)	63.4 (11.0)	63.9 (11.6)
Smoker, N (%)	34 (22%)	23 (12%)	30 (13%)	27 (11%)
BMI, mean (sd)	26.9 (4.5)	28.0 (5.2)	27.5 (3.9)	26.9 (5.9)
BMI change*	0.84 (2.08)	1.66 (3.07)	0.96 (2.11)	1.39 (3.09)
WC (cm), mean (sd)	96.1 (9.7)	89.1 (11.2)	98.8 (9.5)	86.3 (10.7)
Diabetes (yes), N (%)	26 (17%)	39 (20%)	26 (11%)	27 (11%)

*Change from SLVDS baseline examination (1984-1985)
BMI=body mass index, BMI change=change in BMI from baseline visit to third exam, WC=waist circumference

Genetic data

Genetic data were collected for 741 of the 837 participants (89%), which included 307 Hispanics and 434 NHW. Allele designation was recorded on 384 SNPs from 170 genes. After careful research and review of the literature, 22 of those genes were previously studied in obesity research, thus all SNPs for those genes were selected for this analysis, resulting in 107 total SNPs (Table 2). Each participant was classified for each SNP as being homozygous for its wildtype allele, homozygous for its compliment allele, or heterozygous with each type of allele. Population allele frequencies were not determined for this association study. However, minor allele frequencies (MAF) were defined as the lesser appearing allele within this sample and were calculated separately for Hispanic and NHW. No data were found for five of the selected SNPs (rs822396, rs4601580, rs2010994, rs3811787, rs10783219) and no variation occurred in two SNPs (rs2290200 and rs9823137) resulting in 100 SNPs for analysis.

The minor alleles for SNP rs8059937 (*A2BP1*) was 28% and SNP rs6822807 (*UCP1*) was 17% more prevalent in Hispanics than whites, accounting for the largest differences in the sample. Eleven other SNPs had minor alleles with at least 10% higher prevalence and 14 more had 5% higher prevalence in Hispanics than whites. Conversely, two SNPs rs11724758 (*FABP2*) and rs2239179 (*VDR*) had minor alleles that appeared 16% more often in whites than Hispanics. An additional 6 SNPs had minor alleles that appeared 10% more often and another 17 appeared 5% more often in whites than Hispanics. All other SNP alleles were of similar frequencies between Hispanics and whites in this sample. Allele frequencies were consistent with Hardy-Weinberg expectations (data not shown).

Table 2. Selected SNPs and Sample MAFs (N=741)

SNP	Gene Symbol	Gene Name	Chr	Sample MAF Hispanic N=307	Sample MAF NHW N=434
rs8059937	A2BP1	ataxin 2-binding protein 1	16	42.67%	15.09%
rs1501299	ADIPOQ	adiponectin	3	33.39%	26.15%
rs2241766	ADIPOQ	adiponectin	3	15.80%	13.02%
rs266729	ADIPOQ	adiponectin	3	26.22%	24.31%
rs3774261	ADIPOQ	adiponectin	3	49.35%	39.63%
rs822395	ADIPOQ	adiponectin	3	33.71%	39.29%
rs822396	ADIPOQ	adiponectin	3		
rs7539542	ADIPOR1	adiponectin receptor 1	1	38.27%	28.18%
rs12826079	ADIPOR2	adiponectin receptor 2	12	7.98%	8.87%
rs916055	ALOX15	lipoxygenase-15	17	37.13%	31.87%
rs10006877	FABP2	fatty acid binding protein 2	4	23.62%	32.56%
rs10034661	FABP2	fatty acid binding protein 2	4	29.97%	25.58%
rs11724758	FABP2	fatty acid binding protein 2	4	34.20%	50.00%
rs1397613	FABP2	fatty acid binding protein 2	4	40.07%	42.63%
rs1546503	FABP2	fatty acid binding protein 2	4	26.71%	35.19%
rs1799883	FABP2	fatty acid binding protein 2	4	29.64%	25.58%
rs2290200	FABP4	fatty acid binding protein 4	8	0.00%	0.00%
rs6992708	FABP4	fatty acid binding protein 4	8	29.58%	29.63%
rs8050136	FTO	fat mass and obesity associated	16	25.08%	38.23%
rs17467825	GC	vitamin D binding protein/Gc-globulin	4	20.36%	27.07%
rs222020	GC	vitamin D binding protein/Gc-globulin	4	27.69%	15.44%
rs222029	GC	vitamin D binding protein/Gc-globulin	4	25.33%	15.67%
rs2298849	GC	vitamin D binding protein/Gc-globulin	4	28.43%	19.59%
rs2069824	IL6	interleukin 6	7	10.59%	7.26%
rs1386821	IL6R	interleukin-6 receptor	1	10.75%	17.28%

SNP	Gene Symbol	Gene Name	Chr	Sample MAF Hispanic N=307	Sample MAF NHW N=434
rs4075015	IL6R	interleukin-6 receptor	1	42.35%	44.34%
rs4329505	IL6R	interleukin-6 receptor	1	13.19%	18.89%
rs4601580	IL6R	interleukin-6 receptor	1		
rs4845625	IL6R	interleukin-6 receptor	1	34.20%	38.54%
rs8192284	IL6R	interleukin-6 receptor	1	49.02%	41.80%
rs17782313	LOC342784	near melanocortin 4 receptor	18	16.94%	23.39%
rs11090819	PPARA	peroxisome proliferator-activated receptor α	22	7.33%	6.57%
rs11703495	PPARA	peroxisome proliferator-activated receptor α	22	9.31%	10.16%
rs12330015	PPARA	peroxisome proliferator-activated receptor α	22	20.20%	10.94%
rs135538	PPARA	peroxisome proliferator-activated receptor α	22	44.63%	42.84%
rs135547	PPARA	peroxisome proliferator-activated receptor α	22	26.22%	31.57%
rs135550	PPARA	peroxisome proliferator-activated receptor α	22	21.01%	28.46%
rs4253623	PPARA	peroxisome proliferator-activated receptor α	22	17.43%	15.44%
rs4253655	PPARA	peroxisome proliferator-activated receptor α	22	14.01%	18.59%
rs4253701	PPARA	peroxisome proliferator-activated receptor α	22	9.45%	10.60%
rs4253754	PPARA	peroxisome proliferator-activated receptor α	22	18.63%	19.35%
rs4253755	PPARA	peroxisome proliferator-activated receptor α	22	7.33%	11.64%
rs4253772	PPARA	peroxisome proliferator-activated receptor α	22	13.36%	10.83%
rs4823613	PPARA	peroxisome proliferator-activated receptor α	22	33.88%	25.69%
rs6007662	PPARA	peroxisome proliferator-activated receptor α	22	24.84%	25.93%
rs8138102	PPARA	peroxisome proliferator-activated receptor α	22	18.57%	22.80%
rs9615264	PPARA	peroxisome proliferator-activated receptor α	22	4.23%	8.87%
rs2076167	PPARD	peroxisome proliferator-activated receptor δ	6	24.67%	24.31%
rs2076169	PPARD	peroxisome proliferator-activated receptor δ	6	7.68%	12.24%
rs2267665	PPARD	peroxisome proliferator-activated receptor δ	6	14.50%	19.40%
rs6457816	PPARD	peroxisome proliferator-activated receptor δ	6	10.26%	6.81%
rs7744392	PPARD	peroxisome proliferator-activated receptor δ	6	4.56%	4.50%
rs9470001	PPARD	peroxisome proliferator-activated receptor δ	6	9.45%	5.99%
rs1151996	PPARG	peroxisome proliferator-activated receptor γ	3	27.04%	39.70%
rs1175540	PPARG	peroxisome proliferator-activated receptor γ	3	27.32%	37.18%
rs1801282	PPARG	peroxisome proliferator-activated receptor γ	3	12.75%	12.90%
rs3856806	PPARG	peroxisome proliferator-activated receptor γ	3	12.54%	15.13%
rs4684846	PPARG	peroxisome proliferator-activated receptor γ	3	36.48%	25.52%
rs709149	PPARG	peroxisome proliferator-activated receptor γ	3	26.87%	39.98%
rs9823137	PPARG	peroxisome proliferator-activated receptor γ	3	0.00%	0.00%
rs9829551	PPARG	peroxisome proliferator-activated receptor γ	3	2.61%	0.12%
rs12650562	PPARGC1A	PPARG coactivator 1 α	4	34.20%	45.51%
rs1873532	PPARGC1A	PPARG coactivator 1 α	4	33.50%	40.21%
rs2932965	PPARGC1A	PPARG coactivator 1 α	4	15.15%	20.55%
rs2932976	PPARGC1A	PPARG coactivator 1 α	4	41.69%	27.07%
rs2946385	PPARGC1A	PPARG coactivator 1 α	4	36.48%	45.84%
rs3736265	PPARGC1A	PPARG coactivator 1 α	4	8.47%	5.66%
rs3755863	PPARGC1A	PPARG coactivator 1 α	4	35.34%	40.21%
rs3774902	PPARGC1A	PPARG coactivator 1 α	4	13.68%	5.66%
rs4361373	PPARGC1A	PPARG coactivator 1 α	4	29.15%	15.78%
rs4619879	PPARGC1A	PPARG coactivator 1 α	4	24.10%	34.76%
rs6838600	PPARGC1A	PPARG coactivator 1 α	4	42.83%	29.68%
rs7657071	PPARGC1A	PPARG coactivator 1 α	4	22.64%	33.87%

SNP	Gene Symbol	Gene Name	Chr	Sample MAF Hispanic N=307	Sample MAF NHW N=434
rs7665116	PPARGC1A	PPARG coactivator 1 α	4	8.99%	11.75%
rs7672915	PPARGC1A	PPARG coactivator 1 α	4	48.69%	45.08%
rs7677000	PPARGC1A	PPARG coactivator 1 α	4	20.92%	14.52%
rs7682765	PPARGC1A	PPARG coactivator 1 α	4	1.31%	7.83%
rs8192678	PPARGC1A	PPARG coactivator 1 α	4	28.10%	34.91%
rs2010994	PPARGC1B	PPARG coactivator 1 β	5		
rs2052490	PPARGC1B	PPARG coactivator 1 β	5	12.38%	15.82%
rs741581	PPARGC1B	PPARG coactivator 1 β	5	14.01%	9.91%
rs11187545	RBP4	retinol binding protein 4	10	4.40%	7.37%
rs13376835	RBP4	retinol binding protein 4	10	17.59%	18.55%
rs13376898	RBP4	retinol binding protein 4	10	0.16%	0.00%
rs17484721	RBP4	retinol binding protein 4	10	16.12%	18.36%
rs3758538	RBP4	retinol binding protein 4	10	13.36%	14.50%
rs3758539	RBP4	retinol binding protein 4	10	18.40%	16.94%
rs12502572	UCP1	uncoupling protein 1	4	45.41%	30.65%
rs3811787	UCP1	uncoupling protein 1	4		
rs3811790	UCP1	uncoupling protein 1	4	22.77%	13.90%
rs6536991	UCP1	uncoupling protein 1	4	33.06%	22.35%
rs6818140	UCP1	uncoupling protein 1	4	30.78%	17.32%
rs6822807	UCP1	uncoupling protein 1	4	42.67%	25.64%
rs6829571	UCP1	uncoupling protein 1	4	33.53%	21.61%
rs7687015	UCP1	uncoupling protein 1	4	22.64%	12.90%
rs7688743	UCP1	uncoupling protein 1	4	33.82%	19.24%
rs11602906	UCP2	uncoupling protein 2	11	2.93%	6.68%
rs643064	UCP2	uncoupling protein 2	11	13.84%	12.33%
rs655717	UCP2	uncoupling protein 2	11	48.53%	43.20%
rs659366	UCP2	uncoupling protein 2	11	43.81%	36.57%
rs1800849	UCP3	uncoupling protein 2	11	19.54%	23.21%
rs2075577	UCP3	uncoupling protein 2	11	43.32%	47.81%
rs10783219	VDR	vitamin D receptor	12		
rs2239179	VDR	vitamin D receptor	12	30.13%	46.06%
rs3782905	VDR	vitamin D receptor	12	24.27%	31.68%
rs4334089	VDR	vitamin D receptor	12	19.06%	27.83%
rs4760648	VDR	vitamin D receptor	12	42.51%	44.47%

Principal Components Analysis (PCA)

Due to the ordinal nature of allele assignment with genetic data and the fact that PCA is based on Pearson correlations, the SNP data were put through a transformation process. The PRINQUAL procedure was run in SAS, which uses the method of alternating least squares to optimize properties of the transformed variables' correlation matrix. It

transforms ordinal variables monotonically by scoring the ordered categories so that order is weakly preserved. This procedure also estimates missing values to optimize the covariance matrix, using variable means as initial estimates.

The PRINCOMP procedure was run in SAS on the transformed variables to summarize the SNP data into a smaller set of derived components. This procedure redistributed the total variance in the SNP frequencies in such a way that the derived components explain a descending amount of the variance, each component being a linear combination of the original variables. Additionally, the covariance between each pair of components is zero, which reduces collinearity and increases stability of the pending model. The principal axis method was used to extract the components, and this was followed by a varimax (orthogonal) rotation. One hundred components were derived from the original 100 SNPs selected for analysis. Using the eigenvalue-one criterion, 32 components contained more information than any single variable alone displaying eigenvalues greater than 1 (figure 1).

The first component contained 17.2% of the total variance in genetic data, followed by 6.2% for the second component, and 4.8% for the third. Seventeen SNPs loaded onto the first component at a level of 0.24 before dropping off to a loading of 0.04. These SNPs were from loci on the following genes: *PPARA, PPARD, PPARG, PPARGC1A, PPARGC1B, FTO, FABP2, ADIPOR1, UCP1*, and *VDR*. By including the first 32 components, we have accounted for 77% of the total variation in SNP allele frequencies (figure 2). In comparison, the last 30 components only accounted for 0.57% of the total variation. Therefore, the 100 SNPs selected for analysis of their association with obesity outcomes was successfully reduced to 32 orthogonal variables without sacrificing accuracy.

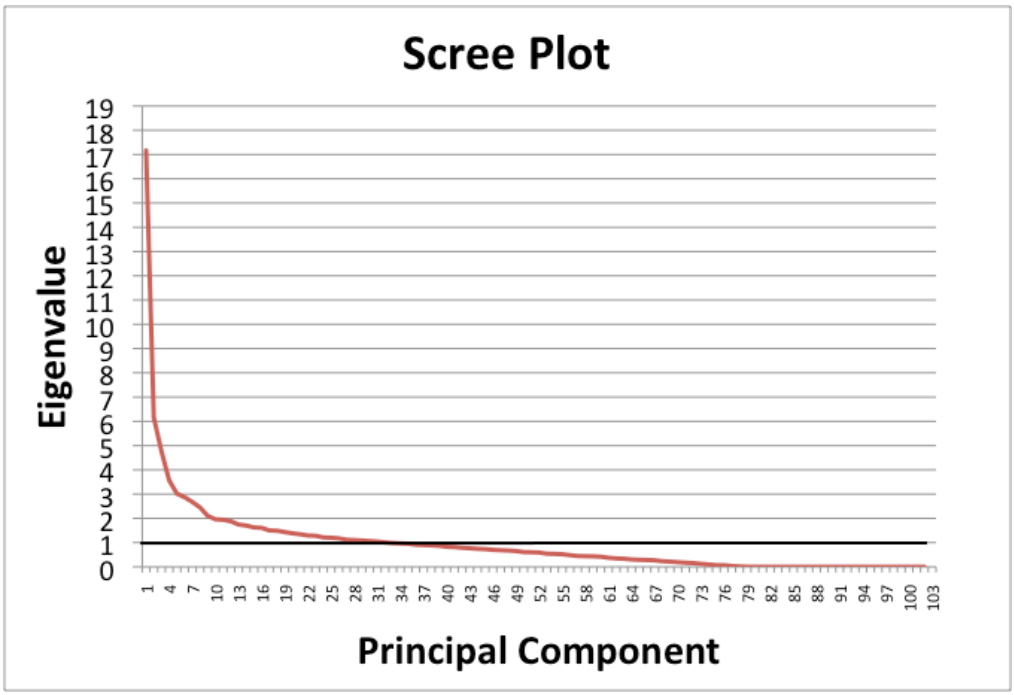

Figure 1. Scree Plot of Descending Eigenvalues for Each Component

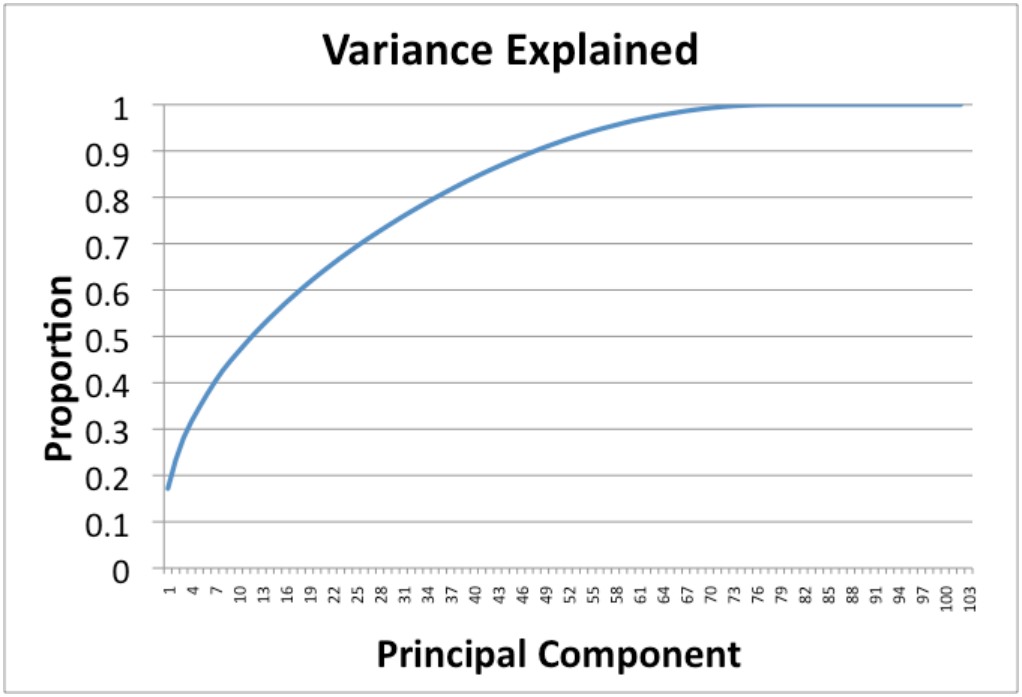

Figure 2. Plot of Cumulative Variance Explained by Each Component

Predicting obesity and diabetes

A base model was established to account for potential confounding factors when predicting obesity measures and prevalence of T2D. This model included age, gender, and smoking status. The 32 principal components were added to the base model to measure added explanatory power accounted for by the genetic information. Each model was assessed separately for Hispanics and NHWs for the outcomes of BMI, change in BMI from baseline, WC, and prevalence of diabetes.

Results of the prediction analysis for Hispanics are presented in table 3. The base model alone accounted for 6.3% of variation seen in BMI measurements. When 32 components of genetic information were added to the base model, only 4.2% of the variation in BMI was accounted for based on adjusted-R^2 values. Therefore, predictive power of BMI in Hispanics was weakened when genetic data were added to the equation. Similarly, predictive power of WC in Hispanics was weakened from 10.8% to 9.2% when including genetic data in the linear model. Thus, genetic information did not contribute to variation in these outcomes, while the models were penalized for 32 additional variables added. However, 5.5% of predictive power was gained when estimating change in BMI from baseline to the final examination, increasing from 18.0% to 23.5%. Thus, more than one-twentieth of the variation in BMI change can be attributed to 100 genetic SNPs summarized in 32 principal components. For the case of diabetes prevalence, the max-rescaled R^2 went up 17.9 percentage points, increasing from 3.5% to 21.4%, indicating almost 18% of information is gained by adding genetic data to the model.

Significance of individual components for each obesity and diabetes outcome in Hispanics is presented in table 4. Although many statistical textbooks use the absolute

value of 0.40 as a strong loading on any component, this analysis considered 0.30 or greater to be strong due to the small number of loadings above 0.40. In the Hispanic population, PC15 was significant for all three obesity outcomes (p=0.0103 for BMI, p=0.0471 for BMIchg, p=0.0101 for WC) and was the only significant component for BMI and WC. Two SNPs contributed strongly to PC15, rs2267665 from the *PPARD* gene and rs13376898 from the *RBP4* gene, with loadings of 0.43 and 0.41, respectively. PC16 was significant for both BMI change (p=0.0428) and diabetes status (p=0.0298) with a loading of 0.33 from rs3758538 and 0.31 from rs11187545, both on the *RBP4* gene sequence. Additionally, change in BMI from baseline was significantly associated with PC4 (p=0.0205), which had five strong loadings of greater than 0.30 coming from SNPs on the *UCP1* gene, and PC31 (p=0.0035) with the strongest loadings coming from SNPs on genes *PPARA* and *PPARGC1B*. Lastly, diabetes status was also associated with PC7 (p=0.0410), which had strong loadings from SNPs on the *RBP4* and *FABP2* genes.

Table 3. Predicting Obesity and Diabetes in Hispanics

Outcome*	Base§ adj-R^2	Base + PCs§ adj-R^2	Difference	% Diff
BMI	0.0629	0.0415	-0.0214	-2.14%
BMIchg	0.1796	0.2353	0.0557	5.57%
WC	0.1079	0.0919	-0.0160	-1.60%
Diabetes status†	0.0354	0.2144	0.1790	17.90%

BMI=body mass index, BMIchg=change in BMI from baseline visit to third exam, WC=waist circumference, PC=principal components.
*General linear models were used for BMI, BMI chg, and WC. Logistic regression model was used for diabetes status.
§ Base model included age, gender, and smoking status. Base + PC model included base variables plus 32 components.
†Max-rescaled R^2 was used for comparison.

Table 4. Components Associated with Obesity Outcomes in Hispanics

			absolute value (loading)	BMI	BMIchg	WC	Diabetes status
PC4					p=0.0205		
	rs12502572	UCP1	0.36				
	rs7688743	UCP1	0.35				
	rs6536991	UCP1	0.33				
	rs6829571	UCP1	0.32				
	rs7687015	UCP1	0.30				
PC7							p=0.0410
	rs17484721	RBP4	0.35				
	rs13376835	RBP4	0.34				
	rs3758539	RBP4	0.32				
	rs10034661	FABP2	0.30				
	rs1799883	FABP2	0.30				
PC15				p=0.0103	p=0.0471	p=0.0101	
	rs2267665	PPARD	0.43				
	rs13376898	RBP4	0.41				
PC16					p=0.0428		p=0.0298
	rs3758538	RBP4	0.33				
	rs11187545	RBP4	0.31				
PC31					p=0.0035		
	rs4253655	PPARA	0.37				
	rs741581	PPARGC1B	0.33				

BMI=body mass index, BMIchg=change in BMI from baseline visit to third exam, WC=waist circumference, PC=principal component.

Results of regression models for NHWs are summarized in table 5. Looking at only the base model, 1.7% of variation in BMI is explained, while 12.2% of variation in BMI change is explained. This indicates that age, gender, and smoking status are descent predictors of weight change over the 13-year study period. When genetic information is added, 7.5% and 1.1% more predictive power is added to BMI and BMI change measurements, respectively. Thus, while the base model is a good predictor of BMI change, genetic information adds more value to current BMI predictability. Percent of variation explained in WC is 27.9% with the base model alone and 33.0% with the base plus genetic data model, demonstrating a 5.1% increase in explanatory power. Similar to Hispanics,

18.5% information is gained for the status of diabetes in NHW, increasing from 2.7% to 21.2% for the base and full models, respectively.

Ten components were significantly associated with BMI in NHWs, most notably PC9, PC23, and PC30 with p-values less than 0.01 (table 6). SNPs that loaded strongly on these components were related to the *ADIPOQ, PPARG, UCP2, UCP3, ALOX15*, and *VDR* genes, however *UCP3* did not have any loadings greater than 0.30. Components with high representation of the peroxisome proliferator-activated receptors were associated with BMI, including PC2, PC3, PC10, and PC18. The *UCP1* gene provided five SNPs that strongly loaded onto PC4, a component that was significant for BMI, WC, and diabetes status in NHW. This is in contrast to PC4's only significance with BMIchg for Hispanics. Change in BMI from baseline for NHW was associated with only one component of genetic information, PC28 (p=0.0071), which reported just one strong loading of 0.40 from a SNP on the vitamin D binding gene. In comparison, Hispanics had four significant components for the BMIchg outcome. In addition to PC4, WC was associated with PC9, PC10, PC12, PC23, and PC30, all of which were jointly associated with BMI. This overlap strengthens the connection between genes loading strongly on those components with obesity, which include PPARs and UCPs. Lastly, diabetes status was associated with strong loadings from SNPs on the *PPARGC1A, UCP1, RBP4*, and *FABP2* genes in NHWs in this study.

Table 5. Predicting Obesity and Diabetes in Non-Hispanic Whites

Outcome*	Base§ adj-R^2	Base + PCs§ adj-R^2	Difference	% Diff
BMI	0.0172	0.0923	0.0751	7.51%
BMIchg	0.1222	0.1330	0.0108	1.08%
WC	0.2789	0.3301	0.0512	5.12%
Diabetes status†	0.0269	0.2122	0.1853	18.53%

BMI=body mass index, BMIchg=change in BMI from baseline visit to third exam, WC=waist circumference, PC=principal components.

*General linear models were used for BMI, BMI chg, and WC. Logistic regression model was used for diabetes status.

§ Base model included age, gender, and smoking status. Base + PC model included base variables plus 32 components.

†Max-rescaled R^2 was used for comparison.

Table 6. Components Associated with Obesity Outcomes in Non-Hispanic Whites

			absolute value (loading)	BMI	BMIchg	WC	Diabetes status
PC2				p=0.0353			
	rs7682765	PPARGC1A	0.392565				
	rs7665116	PPARGC1A	0.392562				
	rs2076167	PPARD	0.392528				
	rs222029	GC	0.392482				
	rs2076169	PPARD	0.392466				
	rs7677000	PPARGC1A	0.392449				
PC3				p=0.0474			p=0.0456
	rs1873532	PPARGC1A	0.387502				
	rs3755863	PPARGC1A	0.382077				
	rs12650562	PPARGC1A	0.379946				
	rs8192678	PPARGC1A	0.364995				
PC4				p=0.0297		p=0.0467	p=0.0332
	rs12502572	UCP1	0.362761				
	rs7688743	UCP1	0.346076				
	rs6536991	UCP1	0.326283				
	rs6829571	UCP1	0.323687				
	rs7687015	UCP1	0.304011				
PC7							p=0.0180
	rs17484721	RBP4	0.346197				
	rs13376835	RBP4	0.34047				
	rs3758539	RBP4	0.320776				
	rs10034661	FABP2	0.303308				
	rs1799883	FABP2	0.30282				
PC9				p=0.0080		p=0.0131	
	rs3774261	ADIPOQ	0.355404				
	rs1175540	PPARG	0.325168				
	rs1501299	ADIPOQ	0.316256				
	rs709149	PPARG	0.309365				
PC10				p=0.0233		p=0.0052	
	rs709149	PPARG	0.367799				
	rs1175540	PPARG	0.356077				
PC12*				p=0.0120		p=0.0050	
	rs643064	UCP2	0.29657				
PC18				p=0.0449			
	rs7672915	PPARGC1A	0.320308				
PC23*				p=0.0084		p=0.0019	
	rs2075577	UCP3	0.266811				
PC26				p=0.0353			
	rs916055	ALOX15	0.405822				
	rs822395	ADIPOQ	0.374624				
PC28					p=0.0071		
	rs17467825	GC	0.404464				
PC30				p=0.0089		p=0.0130	
	rs916055	ALOX15	0.407087				
	rs3782905	VDR	0.33833				
	rs11602906	UCP2	0.314923				

*These models did not have any loadings greater than 0.30.
BMI=body mass index, BMIchg=change in BMI from baseline visit to third exam, WC=waist circumference, PC=principal component.

Regression models were examined further, increasing the principal components in the full model from 32 to 50, with no increase in predictive power. In the same way, principal components in the full model were reduced from 32 to 15 and still did not prevail over the 32 component model when accounting for variation is obesity and diabetes outcome measures (data not shown).

CHAPTER 5

Discussion

In the current study of SLVDS participants, we attempted to examine genetic links to variation seen in typical obesity measurements and prevalence of T2D. One hundred seven loci were selected from 22 gene sequences for their suspected association. We were able to summarize 77% of the genetic data into 32 individual components. When added to demographic data, these components added valuable information for determining why we see such a wide range of obesity and diabetes outcomes.

The percent of variation in obesity measurements explained by our PCA models were low, ranging from 1% to 33%. However, the added explanatory power from only 100 SNP variants was considerable in relation to the amount of genetic information contained in the human genome. It can be argued that models with low r-squared values can still be practical; giving statistically significant results and is good representation of an outcome despite a large amount of noise present. Given what we know about the complexity in obesity outcomes, a small r-squared is expected with any predictive model focusing on just one aspect of the disease, in this case genetics. A study performed by *Peterson et al* found that only 4.13% of BMI variance was explained by their genetic risk sum score (GRSS) model, which was just 0.66% more than the base model. Their analysis included non-Hispanic European- and African-Americans, 56 SNP variants, and a similar base model to the one in the present analysis. Thus, our summary form of 100 SNP variants, which added 7.5% explanatory power for BMI in NHWs, surpassed their predictions by a multiple of 11 (Peterson et al., 2011). A notable difference in their study was the calculation of a genetic

score reflecting an increased appearance of risk alleles, while our study looked at associations in obesity measures with condensed information from multiple alleles. Common SNPs between the two studies are rs8050136 (*FTO*) and rs17782313 (near the *MC4R* gene). Similarly, the population-based European Prospective Investigation into Cancer and Nutrition (EPIC) study, conducted by *Li et al*, found significant effect sizes of risk alleles from 12 genetic loci on BMI, though their combined effect explained only 0.9% of BMI variation (Li et al., 2010). The EPIC study also included rs17782313, however no other SNPs overlaped with those in the present study, including one from the FTO gene.

Genetic data in this study were significant in explaining differences between Hispanic and NHW susceptibility to T2D. With approximately 18% predictive ability gained in each group, the genetic information retained in our 32 components was helpful in estimating an outcome of diabetes in subjects who did not have diabetes at the 1984-1985-baseline examination. This knowledge could be useful in a clinical setting to implement a more aggressive prevention strategy through lifestyle modifications. Ideally, knowing one's risk of developing T2D could motivate behavior change toward regulating food intake and increasing moderate exercise to curtail manifestation of the disease. A study conducted by *Iwata et al* showed that a genetic risk score composed of 14 SNPs had significantly stronger association with T2D than any of the single SNPs alone in a Japanese population (Iwata et al., 2012). They counted the number of risk alleles for each individual and found the sum to be a useful indicator of early onset diabetes with an odds ratio of 1.26 towards development of the disease. Their analysis incorporated two of the same SNPs as the present analysis, including rs8050136 (*FTO*) and rs1801282 (*PPARG*), indicating a strong relationship between T2D with fat mass and lipid homeostasis. In the Iwata study,

these SNPs were only significant when included in a genetic risk score, while in our study they were recognized when combined with other SNPs within a factor. Another similarity was their use of age and gender as the primary covariates combined with genetic information. An important aspect of the *Iwata et al* study as well as the current study is the significance of these SNPs in multiple ethnic and racial populations, which have already been confirmed to be susceptibility loci in those of European origin. This characteristic makes it possible for clinicians to develop a screening process for T2D, which uses a group of SNPs that is relevant for all individuals, though their expected results may be different.

One likely reason the genetic data in the present study was better able to predict an outcome of diabetes than obesity was the fact that the SLVDS was initiated as a diabetes study and thus had that as their primary outcome of interest. Therefore, when collecting genetic information, the investigators focused on associations with T2D and its related complications such as cardiovascular disease. This aspect makes it difficult for anyone using the same genetic information to assess obesity outcomes. Also, using BMI and related measurements as a diagnosis of obesity can be inexact and may not reflect body fat in different types of people. In the TIGER study, researchers point out the scale was created based on Caucasian men and women and does not account for differences in body composition between various racial or ethnic groups (Jackson, Ellis, McFarlin, Sailors, & Bray, 2009). The study found that overweight and obesity was underestimated in Hispanic women based on the usual BMI cut-offs, resulting in bias conclusions. Conversely, diabetes status is viewed as having a more concrete diagnosis algorithm that is universal among different groups and individuals.

Another possible reason for the perceived greater predictability of diabetes status over obesity measures is the use of the max-rescaled R^2 values. The R^2 value after modeling a dichotomous outcome such as disease status is not directly comparable to the R^2 value after modeling a continuous variable outcome. Simply speaking, an outcome with only two possible values does not have the typical variability seen with continuous data, thus its R^2 is not equivalent to the percent of variance explained. The max-rescaled R^2 value is a transformation of the ratio of two likelihood functions comparing the full-proposed model with a model containing only an intercept (Menard, 2000). Nonetheless, it is a measure of good fit and can be interpreted as the amount of information gained when including predictors into the model in comparison with a null model. While not directly comparable to the obesity outcome measures, using R^2 to assess predictability of diabetes status when genetic data are added to the model is a good assessment of the data's usefulness.

Interestingly, SNPs from the *FTO* and near the *MC4R* gene did not show association with obesity outcomes in this analysis, which have both proven to be associated with obesity in previous studies (Frayling et al., 2007; Mejia-Benitez et al., 2013). Their effect sizes may be weakened when other genetic data are present that contain all or some of the same information. It is worth noting that SNP rs8050136 on the *FTO* gene was included in the first component, which accounted for 17% of the total variation in SNP data, though PC1 was not significantly associated with obesity outcomes in either group. However, this analysis did uncover several potential SNPs to be included in future research of genetics and obesity, including rs12502572, rs7688743, rs6536991, rs6829571, and rs7687015

from the *UCP1* gene, rs17484721, rs13376835, and rs3758539 from the *RBP4* gene, and rs10034661 and rs1799883 from the *FABP2* gene.

This study is an application of the approach suggested by *Wang and Abbott* on the use of principal components in genetic association studies (Wang & Abbott, 2008). They propose a general method for reducing the dimension and collinearity of genetic data, which is then entered into a regression model to determine association with an outcome. Although this method is common in other fields, such as econometrics, it has not been widely employed in disease-related studies. One drawback of condensing genetic data is that some capability to directly associate a single SNP with the outcome variable is lost. However, the goal of this study was to assess the effective association of a group of carefully selected SNPs on each outcome rather than restricting the focus to individual SNPs, which has proven to be fruitless (Li et al., 2010).

With genetics, particularly SNP data, there is an abundance of information, which creates a need for condensing or summarizing the data into an appropriate format for analysis. We have shown that PCA is a practical way of doing this when individual variable contributions are not desired. Associations can be made between phenotype and a large group of genetic data. The challenge now lies in the method of choosing an ideal group of potential SNPs that together will have maximum power of explaining variations in typical markers for obesity, such as BMI and WC, and T2D. This method allows usage of many common SNPs within obesity related gene sequences like *FTO* and *PPARG*.

Another advantage of this approach over other genetic models is the ability to include covariates to the regression model. Incorporating demographic and lifestyle data along with genetics are important for creating a clear picture of complex diseases.

Additionally, this method does not rely on directionality of the risk allele, allowing flexibility in choosing groups of SNPs without detailed population statistics. However, risk direction would be useful in follow-up analyses.

Aside from PCA, unique to this study was its large population of Hispanics. When looking at the regression models, very different results were seen between Hispanics and NHWs for both BMI and WC. This proves that genetic information is valuable when explaining phenotypic variations seen within different populations. The Viva La Familia Study was developed to genetically map childhood obesity and its metabolic consequences in Hispanics (Butte et al., 2006). They found a strong genetic contribution to the high prevalence of childhood obesity in Mexican-American families with significant heritability for obesity-related traits. This large cohort of overweight and obese Hispanic families would serve as a great resource for identifying genes or a group of SNPs to include in a genetic model such as the one in the present study.

A limitation of this study is the omission of lifestyle variables, including energy intake and physical activity, as well as the lack of availability of more obesity SNP variants. Future studies should devote more time to identification of genes and individual SNPs related to weight status through the use of GWAS and current literature. Also, there could be confounding due to population stratification since allele frequencies can vary across ethnic groups due to differences in ancestry. This event may explain differences observed in allele frequencies between Hispanic and NHW for individual SNPs and may be a factor in differences seen with their effects on obesity and diabetes outcomes. However, substantial variation in allele frequencies or disease rates across groups was not thought to hinder this analysis.

In conclusion, this study presents an effective method for incorporating a large amount of genetic information into a simplified predictability model. It provides further information regarding genetic variation as a crucial part of explaining phenotypic and metabolic differences in Hispanic and NHW populations. With an ideal set of starting variables, PCA could summarize many SNPs into a few powerful components. When this group of SNPs is found and if genetic data were readily available, individuals would have a better understanding of their risks. Further investigation of obesity related SNPs is needed for future analysis.

REFERENCES

American Diabetes Association. (2011). Standards of medical care in diabetes--2011. *Diabetes Care*, *34*(Suppl 1), S11-S61.

Belsky, D. W., Moffitt, T. E., & Houts, R. (2012). Polygenic risk, rapid childhood growth, and the development of obesity: Evidence from a 4-decade longitudinal study polygenic risk for adult obesity. *Arch Pediatr Adolesc Med*, *166*(6), 515-521.

Berry, W. D., & Feldman, S. (1985). *Multiple regression in practice*. Newbury Park, CA: Sage Publications.

Bouchard, C. (2007). The biological predisposition to obesity: Beyond the thrifty genotype scenario. *Int J Obes (Lond)*, *31*(9), 1337-1339.

Bouchard, C., Tremblay, A., Després, J., Nadeau, A., Lupien, P. J., Thériault, G., ... Fournier, G. (1990). The response to long-term overfeeding in identical twins. *N Engl J Med*, *322*(21), 1477-1482.

Butte, N. F., Cai, G., Cole, S. A., & Comuzzie, A. G. (2006). Viva la Familia Study: Genetic and environmental contributions to childhood obesity and its comorbidities in the Hispanic population. *Am J Clin Nutr*, *84*(3), 646-654.

Cauchi, S., Stutzmann, F., Cavalcanti-Proença, C., Durand, E., Pouta, A., Hartikainen, A. L., ... Froguel, P. (2009). Combined effects of MC4R and FTO common genetic variants on obesity in European general populations. *J Mol Med (Berl)*, *87*(5), 537-546.

Choquet, H., & Meyre, D. (2011). Genetics of obesity: What have we learned? *Curr Genomics*, *12*(3), 169-179.

Cole, S. A., Mitchell, B. D., Hsueh, W. C., Pineda, P., Beamer, B. A., Shuldiner, A. R., ... Hixson, J. E. (2000). The Pro12Ala variant of peroxisome proliferator-activated receptor-gamma2 (PPAR-gamma2) is associated with measures of obesity in Mexican Americans. *Int J Obes Relat Metab Disord*, *24*(4), 522-524.

Comuzzie, A. G., Cole, S. A., Laston, S. L., Voruganti, V. S., Haack, K., Gibbs, R. A., & Butte, N. F. (2012). Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS One*, *7*(12), E51954.

Damcott, C. M., Moffett, S. P., Feingold, E., Barmada, M. M., Marshall, J. A., Hamman, R. F., & Ferrell, R. E. (2004). Genetic variation in fatty acid-binding protein-4 and peroxisome proliferator-activated receptor γ interactively influence insulin sensitivity and body composition in males. *Metabolism*, *53*(3), 303-309.

Damcott, C. M., Moffett, S. P., Feingold, E., Barmada, M. M., Marshall, J. A., Hamman, R. F., & Ferrell, R. E. (2004). Genetic variation in fatty acid-binding protein-4 and peroxisome proliferator-activated receptor gamma interactively influence insulin sensitivity and body composition in males. *Metabolism*, *53*(3), 303-309.

Diabetes data & trends. (2011). *Diabetes Surveillance System*. Retrieved from http://apps.nccd.cdc.gov/DDTSTRS/default.aspx

Dohoo, I. R., Ducrot, C., Fourichon, C., Donald, A., & Hurnik, D. (1997). An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies. *Prev Vet Med*, *29*(3), 221-239.

Dorjgochoo, T., Shi, J., Gao, Y. T., Long, J., Delahanty, R., Xiang, Y. B., ... Shu, X. O. (2012). Genetic variants in vitamin D metabolism-related genes and body mass index: Analysis of genome-wide scan data of approximately 7000 Chinese women. *Int J Obes (Lond)*, *36*(9), 1252-1255.

Drong, A. W., Lindgren, C. M., & McCarthy, M. I. (2012). The genetic and epigenetic basis of type 2 diabetes and obesity. *Clin Pharmacol Ther*, *92*(6), 707-715.

Duran-Gonzalez, J., Ortiz, I., Gonzales, E., Ruiz, N., Ortiz, M., Gonzalez, A., ... Nair, S. (2011). Association study of candidate gene polymorphisms and obesity in a young Mexican-American population from south Texas. *Arch Med Res*, *42*(6), 523-531.

Emanuela, F., Grazia, M., Marco de, R., Maria Paola, L., Giorgio, F., & Marco, B. (2012). Inflammation as a link between obesity and metabolic syndrome. *J Nutr Metab*, *2012*, 476380.

Fall, T., & Ingelsson, E. (2012). Genome-wide association studies of obesity and metabolic syndrome. *Mol Cell Endocrinol*, *Sep 3*.

Finkelstein, E. A., Trogdon, J. G., Cohen, J. W., & Dietz, W. (2009). Annual medical spending attributable to obesity: Payer-and service-specific estimates. *Health Affairs*, *28*(5), W822-W831.

Flegal, K. M., Carroll, M. D., Ogden, C. L., & Curtin, L. R. (2010). Prevalence and trends in obesity among US adults, 1999-2008. *JAMA*, *303*(3), 235-241.

Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., ... McCarthy, M. I. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, *316*(5826), 889-894.

Gamboa-Melendez, M. A., Huerta-Chagoya, A., Moreno-Macias, H., Vazquez-Cardenas, P., Ordonez-Sanchez, M. L., Rodriquez-Guillen, R., … Tusie-Luna, M. T. (2012).

Contribution of common genetic variation to the risk of type 2 diabetes in the Mexican Mestizo population. *Diabetes*, *61*(12), 3314-3321.

Grau, K., Hansen, T., Holst, C., Astrup, A., Saris, W. M., Arner, P., ... Sørensen, T. A. (2009). Macronutrient-specific effect of FTO rs9939609 in response to a 10-week randomized hypo-energetic diet among obese Europeans. *Int J Obes (Lond)*, *33*(11), 1227-1234.

Guo, S. S., & Chumlea, W. C. (1999). Tracking of body mass index in children in relation to overweight in adulthood. *Am J Clin Nutr*, *70*, S145-S148.

Hamman, R. F., Marshall, J. A., Baxter, J., Kahn, L. B., Mayer, E. J., Orleans, M., … Lezotte, D. C. (1989). Methods and prevalence of non-insulin-dependent diabetes mellitus in a biethnic Colorado population. The San Luis Valley Diabetes Study. *Am J Epidemiol*, *129*(2), 295-311.

Health United States, 2011. (2012). *Centers for Disease Control and Prevention*. Retrieved from http://www.cdc.gov/nchs/data/hus/hus11.pdf

Hill, J. O. (2006). Understanding and addressing the epidemic of obesity: An energy balance perspective. *Endocr Rev*, *27*(7), 750-761.

Hu, E., Liang, P., & Spiegelman, B. M. (1996). AdipoQ is a novel adipose-specific gene dysregulated in obesity. *J Biol Chem*, *271*(18), 10697-10703.

Hultin, H., Edfeldt, K., Sundbom, M., & Hellman, P. (2010). Left-shifted relation between calcium and parathyroid hormone in obesity. *J Clin Endocrinol Metab*, *95*(8), 3973-3981.

Isganaitis, E., & Lustig, R. H. (2005). Fast food, central nervous system insulin resistance, and obesity. *Arterioscler Thromb Vasc Biol*, *25*(12), 2451-2462.

Iwata, M., Maeda, S., Kamura, Y., Takano, A., Kato, H., Murakami, S., … Tobe, K. (2012). Genetic risk score constructed using 14 susceptibility alleles for type 2 diabetes is associated with the early onset of diabetes and may predict the future requirement of insulin injections among Japanese individuals. *Diabetes Care*, *35*(8), 1763-1770.

Jackson, A. S., Ellis, K. J., McFarlin, B. K., Sailors, M. H., & Bray, M. S. (2009). Body mass index bias in defining obesity of diverse young adults: The Training Intervention and Genetics of Exercise Response (TIGER) study. *Br J Nutr*, *102*(7), 1084-1090.

Jackson, J. E. (2004). *A user's guide to principal components*. Hoboken, NJ: John Wiley & Sons.

Jiang, H., Xiong, D. H., Guo, Y. F., Shen, H., Xiao, P., Yang, F., ... Deng, H. W. (2007). Association analysis of vitamin D-binding protein gene polymorphisms with variations of

obesity-related traits in Caucasian nuclear families. *Int J Obes (Lond)*, *31*(8), 1319-1324.

Ke, Y. H., Xiao, W. J., He, J. W., Zhang, H., Yu, J. B., Hu, W. W., … Zhang, Z. L. (2012). Association of ALOX15 gene polymorphisms with obesity-related phenotypes in Chinese nuclear families with male offspring. *Acta Pharmacol Sin*, *33*(2), 201-207.

Khalyfa, A., Bhushan, B., Hegazi, M., Kim, J., Kheirandish-Gozal, L., Bhattacharjee, R., ... Gozal, D. (2010). Fatty-acid binding protein 4 gene variants and childhood obesity: Potential implications for insulin sensitivity and CRP levels. *Lipids Health Dis*, *9*, 18.

Kosoy, R., Nassir, R., Tian, C., White, P. A., Butler, L. M., Silva, G., ... Seldin, M. F. (2009). Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Hum Mutat*, *30*(1), 69-78.

Li, S., Zhao, J. H., Luan, J., Luben, R. N., Rodwell, S. A., Khaw, K., ... Loos, R. J. (2010). Cumulative effects and predictive value of common obesity-susceptibility variants identified by genome-wide association studies. *Am J Clin Nutr*, *91*(1), 184-190.

Lin, H., Bermudez, O. I., & Tucker, K. L. (2003). Dietary patterns of Hispanic elders are associated with acculturation and obesity. *J Nutr*, *133*(11), 3651-3657.

Loos, R. J., & Bouchard, C. (2003). Obesity--is it a genetic disorder? *J Intern Med*, *254*(5), 401-425.

Lyon, H. N., & Hirschhorn, J. N. (2005). Genetics of common forms of obesity: A brief overview. *Am J Clin Nutr*, *82*(1 Suppl), 215S-217S.

Ma, L., Hanson, R. L., Traurig, M. T., Muller, Y. L., Kaur, B. P., Perez, J. M., ... Baier, L. J. (2010). Evaluation of A2BP1 as an obesity gene. *Diabetes*, *59*(11), 2837-2845.

Martinez-Hervas, S., Mansego, M. L., De Marco, G., Martinez, F., Alonso, M. P., Morcillo, S., ... Chaves, F. J. (2012). Polymorphisms of the UCP2 gene are associated with body fat distribution and risk of abdominal obesity in Spanish population. *Eur J Clin Invest*, *42*(2), 171-178.

Martinez-Lopez, E., Garcia-Garcia, M. R., Gonzalez-Avalos, J. M., Maldonado-Gonzalez, M., Ruiz-Madrigal, B., Vizmanos, B., … Panduro, A. (2013). Effect of Ala54Thr polymorphism of FABP2 on anthropometric and biochemical variables in response to a moderate-fat diet. *Nutrition*, *29*(1), 46-51.

McCarthy, M. I., & Zeggini, E. (2009). Genome-wide association studies in type 2 diabetes. *Curr Diab Rep*, *9*(2), 164-171.

McCarthy, M. I. (2010). Genomics, type 2 diabetes, and obesity. *N Engl J Med*, *363*(24), 2339-2350.

Mei, H., Chen, W., Jiang, F., He, J., Srinivasan, S., Smith, E. N., ... Murray, S. (2012). Longitudinal replication studies of GWAS risk SNPs influencing body mass index over the course of childhood and adulthood. *PLoS One*, *7*(2), E31470.

Mejia-Benitez, A., Klunder-Klunder, M., Yengo, L., Meyre, D., Aradillas, C., Cruz, E., … Bonnefond, A. (2013). Analysis of the contribution of FTO, NPC1, ENPP1, NEGR1, GNPDA2 and MC4R genes to obesity in Mexican children. *BMC Med Genet*, *14*(1), 21.

Menard, S. (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statistician*, *54*(4), 17-24.

Mitchell, J. A., Church, T. S., Rankinen, T., Earnest, C. P., Sui, X., & Blair, S. N. (2010). FTO genotype and the weight loss benefits of moderate intensity exercise. *Obesity (Silver Spring)*, *18*(3), 641-643.

Mitchell, N. S., Catenacci, V. A., Wyatt, H. R., & Hill, J. O. (2011). Obesity: Overview of an epidemic. *Psychiatr Clin North Am*, *34*(4), 717-732.

Munkhtulga, L., Nagashima, S., Nakayama, K., Utsumi, N., Yanagisawa, Y., Gotoh, T., ... Iwamoto, S. (2010). Regulatory SNP in the RBP4 gene modified the expression in adipocytes and associated with BMI. *Obesity (Silver Spring)*, *18*(5), 1006-1014.

National Center for Biotechnology Information. (n.d.). *Genes and mapped phenotypes*. Retrieved from http://www.ncbi.nlm.nih.gov/gene

National Hispanic Caucus of State Legislators. Hispanic obesity: An american crisis. (2010). *Policy Brief*.

Nelson, T. L., Bessesen, D. H., & Marshall, J. A. (2008). Relationship of abdominal obesity measured by DXA and waist circumference with insulin sensitivity in Hispanic and non-Hispanic white individuals: The San Luis Valley diabetes study. *Diabetes Metab Res Rev*, *24*(1), 33-40.

Ochoa, M. C., Santos, J. L., Azcona, C., Moreno-Aliaga, M. J., Martinez-Gonzalez, M. A., Martinez, J. A., & Marti, A. (2007). Association between obesity and insulin resistance with UCP2–UCP3 gene variants in Spanish children and adolescents. *Mol Genet Metab*, *92*(4), 351-358.

Ochs-Balcom, H. M., Chennamaneni, R., Millen, A. E., Shields, P. F., Marian, C., & Trevisan, M. (2011). Vitamin D receptor gene polymorphisms are associated with adiposity phenotypes. *Am J Clin Nutr*, *93*(1), 5-10.

Ogden, C. L., Carroll, M. D., Kit, B. K., & Flegal, K. M. (2012). Prevalence of obesity in the United States, 2009–2010. NCHS data brief, no 82. *Hyattsville, MD: National Center for Health Statistics*.

Perry, J. R., Voight, B. F., Yengo, L., Amin, N., Dupuis, J., Ganser, M., … Cauchi, S. (2012). Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases. *PLoS Genet*, *8*(5).

Peterson, R. E., Maes, H. H., Holmans, P., Sanders, A. R., Levinson, D. F., Shi, J., ... Webb, B. T. (2011). Genetic risk sum score comprised of common polygenic variation is associated with body mass index. *Hum Genet*, *129*(2), 221-230.

Rebhan, M., Chalifa-Caspi, V., Prilusky, J., & Lancet, D. (1997). GeneCards: Integrating information about genes, proteins and diseases. *Trends in Genetics*, *13*(163). www.genecards.org

Rewers, M., Shetterly, S. M., Baxter, J., Marshall, J. A., & Hamman, R. F. (1992). Prevalence of coronary heart disease in subjects with normal and impaired glucose tolerance and non-insulin-dependent diabetes mellitus in a biethnic Colorado population. The San Luis Valley Diabetes Study. *Am J Epidemiol*, *135*(12), 1321-1330.

Siitonen, N., Pulkkinen, L., Lindstrom, J., Kolehmainen, M., Eriksson, J. G., Venojarvi, M., … Uusitupa, M. (2011). Association of ADIPOQ gene variants with body weight, type 2 diabetes and serum adiponectin concentrations: The Finnish Diabetes Prevention Study. *BMC Med Genet*, *12*, 5.

Slomko, H., Heo, H. J., & Einstein, F. H. (2012). Minireview: Epigenetics of obesity and diabetes in humans. *Endocrinology*, *153*(3), 1025-1030.

Stunkard, A. J., Harris, J. R., Pedersen, N. L., & McClearn, G. E. (1990). The body-mass index of twins who have been reared apart. *N Engl J Med*, *322*(21), 1483-1487.

Summary health statistics for U.S. adults: 2010. (2012). *Centers for Disease Control and Prevention*. Retrieved from http://www.cdc.gov/nchs/data/series/sr_10/sr10_252.pdf

Temelkova-Kurktschiev, T., & Stefanov, T. (2012). Lifestyle and genetics in obesity and type 2 diabetes. *Exp Clin Endocrinol Diabetes*, *120*(1), 1-6.

Van Abeelen, A. F., De Krom, M., Hendriks, J., Grobbee, D. E., Adan, R. A., & Van der Schouw, Y. T. (2008). Variations in the uncoupling protein-3 gene are associated with specific obesity phenotypes. *Eur J Endocrinol*, *158*(5), 669-676.

Wake, M., Baur, L. A., Gerner, B., Gibbons, K., Gold, L., Gunn, J., ... Ukoumunne, O. C. (2009). Outcomes and costs of primary care surveillance and intervention for overweight or obese children: The LEAP 2 randomised controlled trial. *BMJ*, *Sept 3*(339), B3308.

Wallenius, V., Wallenius, K., Ahren, B., Rudling, M., Carlsten, H., Dickson, S. L., ... Jansson, J. O. (2002). Interleukin-6-deficient mice develop mature-onset obesity. *Nat Med*, *8*(1), 75-79.

Wang, K., & Abbott, D. (2008). A principal components regression approach to multilocus genetic association studies. *Genetic Epidemiology*, *32*(2), 108-118.

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

Whitaker, K. L., Jarvis, M. J., Beeken, R. J., Boniface, D., & Wardle, J. (2010). Comparing maternal and paternal intergenerational transmission of obesity risk in a large population-based sample. *Am J Clin Nutr*, *91*(6), 1560-1567.

Wolford, J. K., Colligan, P. B., Gruber, J. D., & Bogardus, C. (2003). Variants in the interleukin 6 receptor gene are associated with obesity in Pima Indians. *Mol Genet Metab*, *80*(3), 338-343.

Yu, W., Gwinn, M., Clyne, M., Yesupriya, A., & Khoury, M. J. (2008). A navigator for human genome epidemiology. *Nature Genetics*, *40*(2), 124-125.
www.hugenavigator.net

Zhang, C., Qi, L., Hunter, D. J., Meigs, J. B., Manson, J. E., Van Dam, R. M., & Hu, F. B. (2006). Variant of transcription factor 7-like 2 (TCF7L2) gene and the risk of type 2 diabetes in large cohorts of U.S. women and men. *Diabetes*, *55*(9), 2645-2648.