ADMIXTURE MAPPING AND SUBSEQUENT FINEMAPPING SUGGESTS NOVEL LOCI FOR TYPE 2 DIABETES IN AFRICAN AMERICANS

By

Janina Maria Jeff

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Interdisciplinary Studies: Applied Statistics

December, 2012

Nashville, Tennessee

Approved by:

Dana C. Crawford, PhD

Jonathan Haines, PhD

TABLE OF CONTENTS

Page

LIST OF TABLES iii

LIST OF FIGURES iv

Chapter

I. INTRODUCTION 1

Diabetes mellitus 1
Epidemiology of Type 2 Diabetes 1
Risk Factors of Type 2 Diabetes 5
Genetics of Type 2 Diabetes 8
Candidate Gene and Genetic Linkage Studies 8
Genome-wide Association Studies 10
Beyond GWAS 11
Admixture Mapping as Tool to Identify Disease Loci 14

II. ADMIXTURE MAPPING REVEALS NOVEL LOCI FOR TYPE 2 DIABETES IN AFRICAN AMERICANS 19

Introduction 19
Methods 20
Study Population 20
Genotyping and SNP selection 24
Statistical Methods 24
Results 25
GWAS 25
Admixture Mapping 27
Discussion 34

III. FINEMAPPING THE *TCIRG1* REGION IN AFRICAN AMERICANS ... 39

Introduction 39
Methods 40
Study Population 40
SNP Selection and Genotyping 40
Statistical Methods 41
Results 41
Discussion 47

REFERENCES 51

LIST OF TABLES

Table Page

2.1 Population Characteristics 22

3.1 Study population demographics for 4,246 unrelated African Americans in BioVU 43

3.2 Fine mapping association results. 44

LIST OF FIGURES

Figure Page

1.1 Prevalence of T2D in 2000 and Projections for 2030, with Projected Percent Changes. 4

1.2 Global prevalence of diabetes in adults ages 25+ years, by WHO Region and World Bank income groups. 5

1.3 Racial and ethnic difference in diagnosed diabetes. 7

1.4 Effect sizes of common variants associated with T2D by candidate gene studies and GWAS 12

1.5 Chromosomal ancestry resulting from 8-20 generations of two-way admixture events. 16

1.6 Schematic of the pattern of chromosomal admixture around a disease locus. 18

2.1 Flow chart for Type 2 Diabetes case definition. 23

2.2 Flow chart for Type 2 Diabetes control definition....23

2.3 Genome-wide Association Results for T2D in African Americans from VGER/NUgene 26

2.4 Comparison of global European Ancestry in African American cases of T2D and controls....28

2.5 Local estimates of locus genome stastic across the genome for T2D cases and controls....29

2.6 Local admixture results on Chromosome 11... 32

2.7 Genes in the region under the admixture peak on chromomsome 11....33

2.8 Linkage disequilibrium around the 100kb around rs308328 in African American T2D cases and controls..... 38

3.1 Candidate region targeted for fine mapping..... 43

3.2 Linkage disequilibrium around the 100kb around rs308328 in African Americans T2D cases and controls..... 46

3.3 *TCIRG1* mRNA is highly expressed in human pancreas tissue....47

CHAPTER I

INTRODUCTION

Diabetes mellitus

Diabetes mellitus is a collection of chronic diseases that occurs when the pancreas cannot efficiently produce an adequate amount of insulin or when the body does not effectively use the insulin produced from the pancreas. Insulin is produced by islet cells in the pancreas and has profound effects on metabolism. Insulin is required for the uptake of glucose from the blood in liver, muscle, and fat cells. When insulin regulation is disrupted, cells fail to acquire glucose and the body begins to use fat as an energy source, resulting in potentially severe anabolic effects on the body. Diabetes mellitus takes on three main forms, all of which are characterized by high blood glucose. Type 1 diabetes is an autoimmunie disease where the body eventually no longer makes insulin. Type 2 diabetes (T2D), unlike type 1, is a result of insulin resistance. Gestational diabetes occurs during pregnancy in mothers that do not initially present with type 2 diabetes. The work presented here focuses on the most common form of diabetes melltus: T2D.

Epidemiology of Type 2 Diabetes

As of 2011, 346 million people worldwide have diabetes with over 90% of all diabetes cases being T2D (WHO, 2011). This is an alarming estimate particularly because in 2000 it was predicted that 366 million people will have diabetes in the year 2030, just 20 million more than the 2011 estimate (Figure 1.1) (Wild et al., 2004). There

has been a rapid increase of T2D in developing countries, as more individuals are adopting high fat diets and decreased physical inactivity (Figure 1.1) (Hossain et al., 2007). By 2030, the prevalence of T2D in the Middle East, India, Southeast Asia, and Sub-Saharan African is expected to increase by >150% (Figure 1.1) (Hossain et al., 2007). The interaction with T2D prevalence and poverty is complex. Low income countries tend to have lower rates of T2D compared to middle income countries (Figure 1.2). Yet in developed countries, impoverished communities suffer from T2D and have increased T2D risk factors compared to more affluent communities. It is suspected that this is due to the low cost and overconsumption high fat foods leading to increased obesity rates, a T2D risk factor (Hossain et al., 2007).

Type 2 diabetes also affects the industrialized and developed world. In the United States 25.8 million people have diabetes, which accounts for 8.3% of the total U.S. population (ADA, 2011). Consistent with global estimates, T2D accounts for 93% of all diabetes cases in the U.S. In 2007, diabetes was the underlying cause or contributing factor to about 230,000 deaths in the U.S. (ADA, 2011). It is predicted that the prevalence and consequently the mortality rate of T2D or related complications will double if not triple by the year 2030 (Figure 1.1) (ADA, 2011; Hossain et al., 2007).

The distribution of individuals with T2D is disproportionate across race/ethnicity. In the United States, the burden of diabetes is higher in minority populations compared to European descent populations (Mokdad et al., 2000). National survey data report 8.4% of Asian Americans, 11.8% of Hispanics, 12.6% of non-Hispanic blacks, and 16.1% of American Indians are diagnosed with diabetes at rates higher than non-Hispanic whites

(7.2%, a European descent population) (Centers for Disease Control and Prevention, 2011).

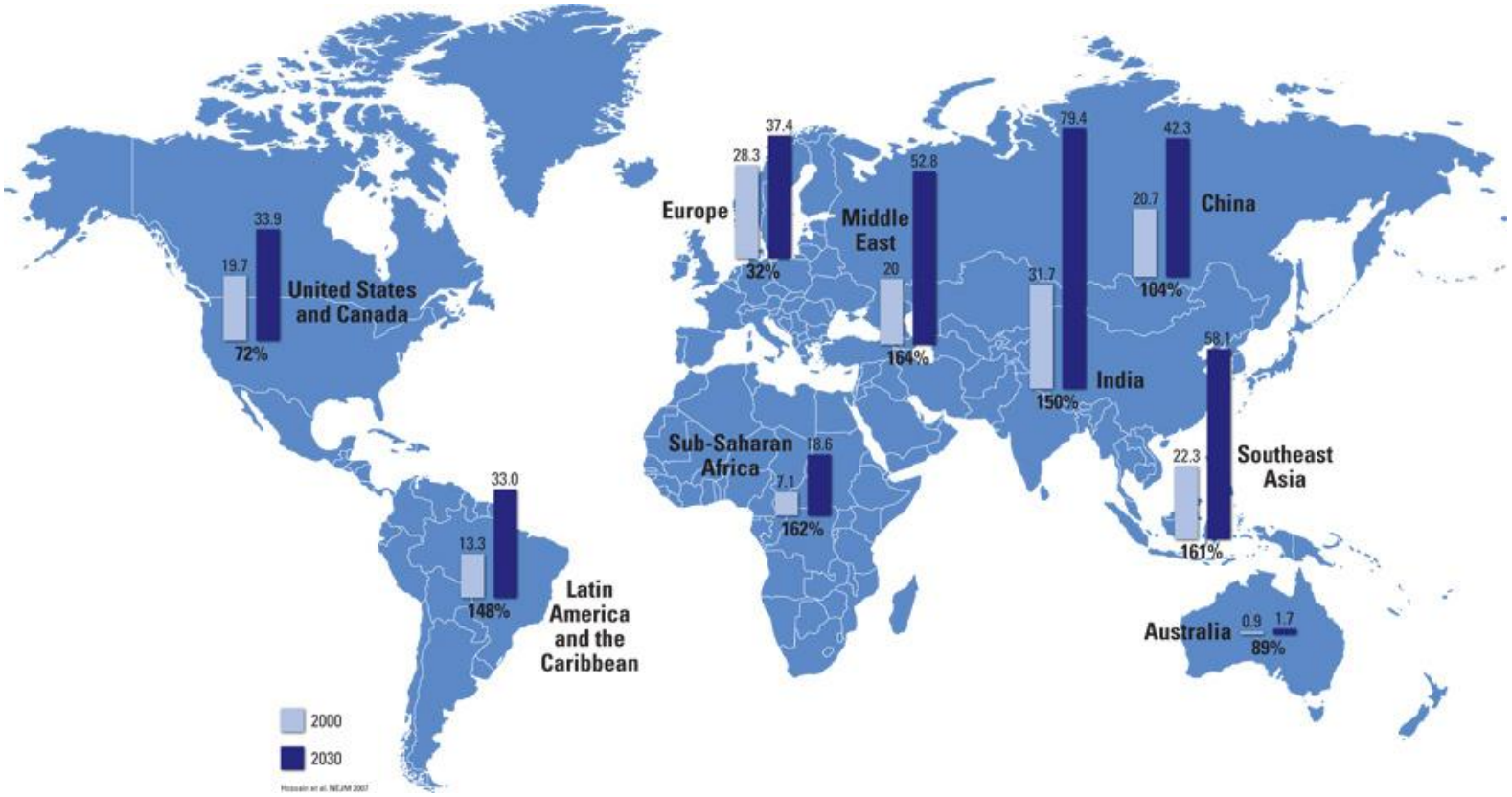


Figure 1.1 Prevalence of T2D in 2000 and Projections for 2030, with Projected Percent Changes. Numbers indicate the proportion (%) of the population with T2D. Modified from (Hossain et al., 2007).

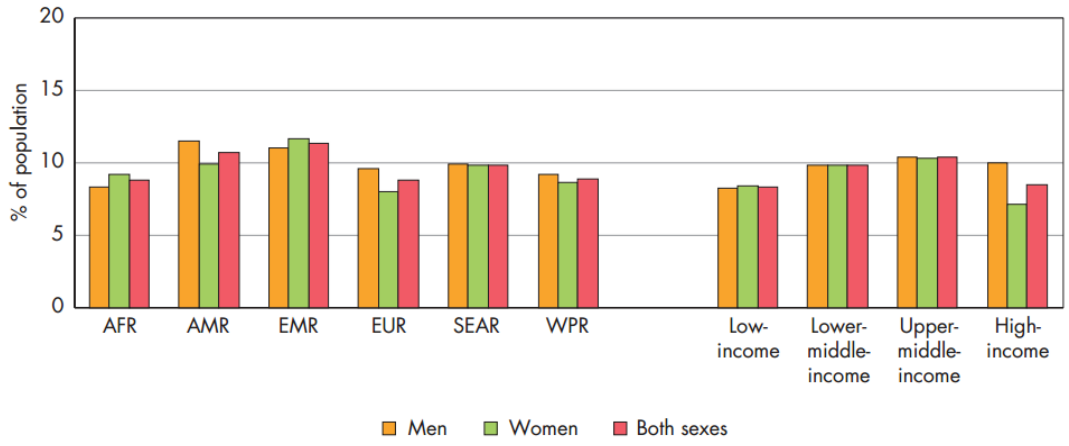


Figure 1.2 Global prevalence of diabetes in adults ages 25+ years, by WHO Region and World Bank income groups, 2008 (WHO, 2011). Abbreviations: AFR= African Region, AMR= Region of the Americas, EMR= Eastern Mediterranean Region, EUR= European Region, SEAR= South-East Asia Region, WPR= Western Pacific Region).

Risk Factors of Type 2 Diabetes

There are several risk factors that might explain the recent increase of T2D incidence. Metabolic diseases, such as T2D have a complex disease etiology and are likely caused by a combination of several environmental and genetic risk factors. Increased body mass index (BMI), as a result of poor diet and a sedentary lifestyle, is the major risk factor for T2D (Mozaffarian et al., 2009). Under the clinical definition of obesity (BMI >30), 85% of T2D cases are obese (National Institute of Diabetes and Kidney Disease, 2007). Beta cells in obese individuals become insulin resistant leading to increased blood glucose (Schinner et al., 2005). Obese individuals have higher lipid levels in the blood, resulting in an accumulation of triglycerides in muscle cells, therby reducing glucose uptake regardless of insulin levels (Schinner et al., 2005; Morino et al., 2006). Insulin insensitivity as a result of obesity is a clinical characteristic of T2D. To a lesser extent, risk factors such as smoking, stress, and lack of sleep are also associated

with T2D. All of these risk factors, including obesity, are modifiable (Mozaffarian et al., 2009).

Increased age, gender, and family history are additional risk factors of T2D that are not modifiable. In the U.S., 26.9% of all adults over the age of 65 have diagnosed diabetes compared to 11.3% under 65 (Centers for Disease Control and Prevention, 2011). Being female has been previously associated with obesity and increased T2D risk; however the role gender has on the clinical manifestation of T2D is still poorly understood and varies worldwide(Centers for Disease Control and Prevention, 2011). Family history also plays a major role in T2D risk. Having at least one parent with T2D increases the risk of developing T2D in 40% of individuals and in 70% of individuals with two affected parents, indicative a strong genetic component (Ahlqvist et al., 2011).

Together, known modifiable and unmodifiable risk factors only account for a small portion of T2D cases. Several investigators have identified unusual cases of obsese individuals that never develop T2D, suggestive of a large genetic component explaining T2D risk. Concordance rates in monozygotic twins with T2D have been reported to be as much as 70% compared to 20-30% in dizygotic twins (Kaprio et al., 1992; Newman et al., 1987). And, as previously mentioned, non-European populations have an increased prevalence of T2D. After adjusting for traditional T2D risk factors: age, sex, BMI, education, and physical activity; Native Americans, African Americans, and Lantinos still have increased risk of T2D compared to European Americans (Maskarinec et al., 2009). These findings give insight to the possibility that differences in genetic ancestry could explain increased disease prevalence as even ethnic and gender differences for T2D prevalence can be explained by genetic factors (Ahlqvist et al., 2011).

While all these studies insinuate a large genetic component involved in the etiology of T2D, the role genetic ancestry affects disease risk is poorly understood. Common approaches to identify genetic risk factors such as candidate gene studies, GWAS, and linkage studies have failed to explain the racial/ethnic disparity of T2D. Alternative methods such as admixture mapping, which comprehensively identifies differences in genetic ancestry associated with disease risk in admixed populations, are needed to identify genetic loci responsible for ancestry-specific risk for T2D.

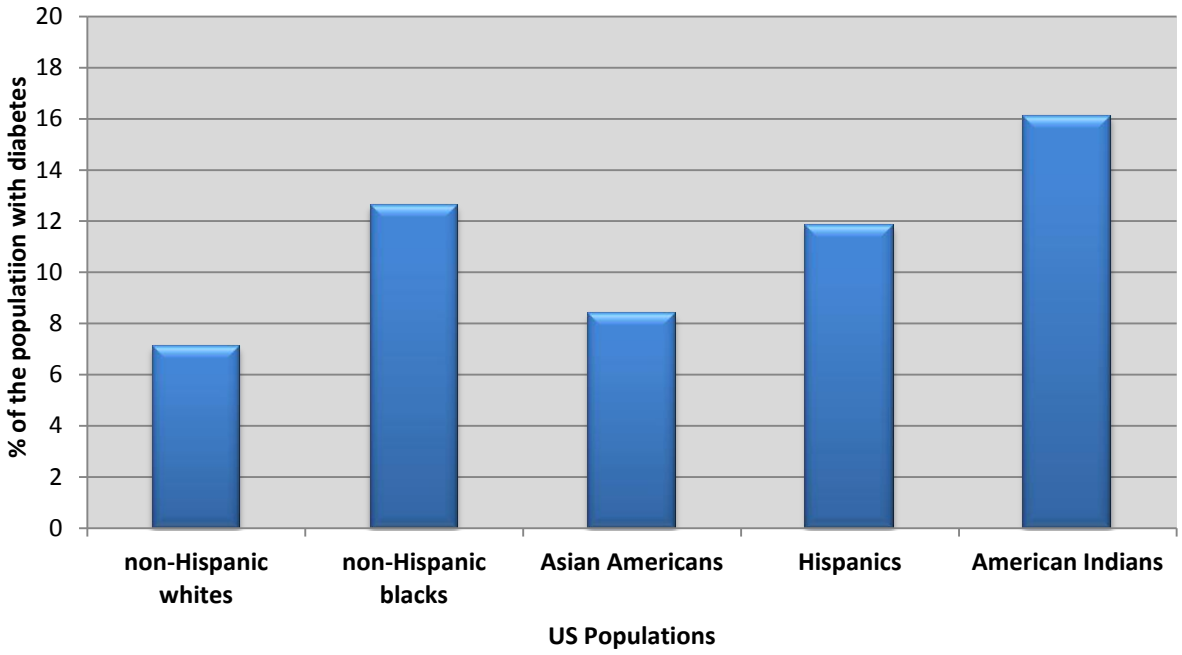


Figure 1.3 Racial and ethnic difference in diagnosed diabetes. Modified from (Centers for Disease Control and Prevention, 2011).

Genetics of Type 2 Diabetes

Candidate Gene and Genetic Linkage Studies

As of 2011, over 40 genes have been identified with an association with T2D; collectively these genes only account for 10% of the heritability (Ahlqvist et al., 2011; Herder and Roden, 2011; Köbberling, 1982; Voight et al., 2010). A genetic linkage study is a traditional approach used to identify disease causing genes using family data. Linkage methods have been very successful at identifying disease loci that are rare, monogenic, have high penetrance, and a known mode of inheritance. However, for common diseases such as T2D that are likely caused by a combination of several genes with smaller effects, interactions between genes and the environment, and have a complex mode of inheritance, these traditional methods are less useful.

Two genes have been identified by linkage studies, *TCF7L2* and *CAPN10* (Duggirala et al., 1999; Horikawa et al., 2000; Hanis et al., 1996; Tong et al., 2009; Grant et al., 2006; Reynisdottir et al., 2003). The transcription factor-7-like 2 is part of bipartite transcription factor and is encoded by the *TCF7L2* gene, located on chromosome 10. Several studies have identified single nucleotide polymorphisms (SNPs) in *TCF7L2* associated with increased risk of T2D, impaired insulin secretion, incretin effects and enhanced rate of hepatic glucose production (Grant et al., 2006; Lyssenko et al., 2007; lyssVoight et al., 2010; Saxena et al., 2007). The other gene identified by linkage studies, *CAPN10*, encodes a cysteine protease and is involved in glucose metabolism (Horikawa et al., 2000). Both genes have been replicated by association studies; however, only *TCF7L2* has been identified by genome-wide association studies (GWAS).

Another traditional approach, the candidate gene study, has identified a number of genes associated with T2D. While several genes have been identified by candidate gene studies, only six genes have been consistently associated with T2D: *PPARG, WFS1, KCNJ11, IRS1, HFN1A,* and *HNF1B.* One of the first genes, *PPARG* encodes a PPAR-γ receptor that is a target for insulin-sensitizing drugs (Deeb et al., 1998). Having an extra exon in this gene increases insulin sensitivity in 15% of the European population and protects against T2D, a finding that has been consistently replicated (Deeb et al., 1998). The *WFS1* gene encodes a protein, Wolframin, and has been associated with T2D in a small family-based association study (Minton et al., 2002). Since, this association has been replicated in larger population-based studies as well as meta-analysis (Sandhu et al., 2007; Franks et al., 2008). An ATP-sensitive potassium channel is encoded by the *KCNJ11* gene and regulates membrane potential (Gloyn et al., 2004). This potassium channel regulates glucose-dependent insulin secretion in pancreatic beta cells, and mutations in this gene lead to T2D (Hani et al., 1998; Gloyn et al., 2003; van Dam et al., 2005; Saxena et al., 2007; Scott et al., 2007; Zeggini et al., 2007). The gene *IRS1* encodes a protein insulin receptor that plays an important role in insulin function (Almind et al., 1993). This association was originally identified in 1993 but failed to replicate for several years until recent GWAS. The genes *HNF1A* and *HNF1B* encode transcription factors and have been identified by candidate gene studies with an association with several forms of maturity onset diabetes of the young (MODY) (Winckler et al., 2007; Holmkvist et al., 2006; Holmkvist et al., 2008; Voight et al., 2010; Gudmundsson et al., 2007). MODY is often referred to as monogenic diabetes and having mutations in one of these genes

results in diabetes compared to T2D, which is polygenic (Goldstein and Müller-Wieland, 2008).

Genome-wide Association Studies

Unlike candidate gene studies, genome-wide association studies are considered hypothesis-free with respect to suspected or known biology. And, unlike linkage, GWA studies are more powerful in identifying common variants with small to moderate genetic effects typical of complex diseases such as T2D. As such, GWAS studies have been more successful for T2D than both linkage and candidate gene studies combined (Figure 1.4). Over 30 novel loci have been identified by GWAS for T2D or related phenotypes (Herder and Roden, 2011). These loci are common and have an allele frequency >25% with small single SNP effects (average odds ratio (OR) 1.0-1.3). The largest single SNP effect reported by GWAS is rs7903146, located in *TCF7L2*, which as originally identified by linkage analysis (Dupuis et al., 2010). Other loci such as *KCNJ11* and *IRS1* were also replicated by GWA studies (Saxena et al., 2007; Scott et al., 2007; Zeggini et al., 2007; Rung et al., 2009). Some of the novel findings identified by GWAS for T2D include *KCNQ1*, *CDKN2A/CDKN2B*, *IGF2BP2,* and *SLC3OA8* (Saxena et al., 2007; Scott et al., 2007; Sladek et al., 2007; WTCC, 2007).

At the time GWAS on T2D were limited to a case-control study design and even with the wave of meta-analyses, were subject to limited power. Often for common diseases the use of quantitative measurements can improve power to detect an association. Quantitative measurements of T2D include fasting glucose, 2-hour glucose tolerance test, glycolated hemoglobin A1c (HbA1c), fasting insulin, markers of insulin resistance and beta cell function, body mass index, and fat mass. One meta-analysis

identified 16 loci associated with fasting glucose that explain about 10% of the trait variability for these traits (Dupuis et al., 2010). Only seven loci were actually associated with T2D outcome in replication studies (Dupuis et al., 2010). These findings suggest that disease loci may contribute to quantitative measurements such as fasting glucose but not to T2D, a common limitation to quantitative study designs (Dupuis et al., 2010). The *FTO* gene that encodes the enzyme alpha-ketoglutarate-dependent dioxygenase is strongly associated with obesity (Dina et al., 2007; Frayling et al., 2007). This association, unlike most associations with quantitative measurements thought to be involved in the pathophysiology of T2D, is also associated with T2D in studies not matched on BMI (Vimaleswaran and Loos, 2010).

Beyond GWAS

GWAS have identified several loci associated with T2D but these variants are common, have small effects, and explain a small portion of the heritability. Thus, other approaches besides GWAS are needed to potentially explain the missing heritability (Manolio et al., 2009). One such approach would be to examine rare variants with modest to large effects missed by current methods. Rare variants can be missed by linkage studies for T2D where a significant amount of genetic heterogeneity exists and by GWAS if they are rare in the target population (Ahlqvist et al., 2011). With the emergence of high-throughput next generation sequencing and custom genotyping chips, the identification of rare disease variants is expected to increase.

Type 2 diabetes has a large environmental component and is likely caused by several genes. Therefore it is highly plausible that gene-gene and gene-environment interactions will explain some of the missing heritability for T2D. To date, little is known

about the effects these interactions have on T2D. In 2011, Bell et al identified espistatic interactions with *TCF7L2* and *FTO*, *TSPAN8* and *CDKAL1* but these studies were underpowered and have yet to be replicated (Bell et al., 2011). There have been few studies that identify gene-environment interactions associated with T2D or related phenotypes. This is mainly due to the imprecise and unreliable ways to measure environmental factors in large populations. Despite these limitations one study has identified an association between dietary whole-grain intake and fasting glucose (Nettleton et al., 2010).

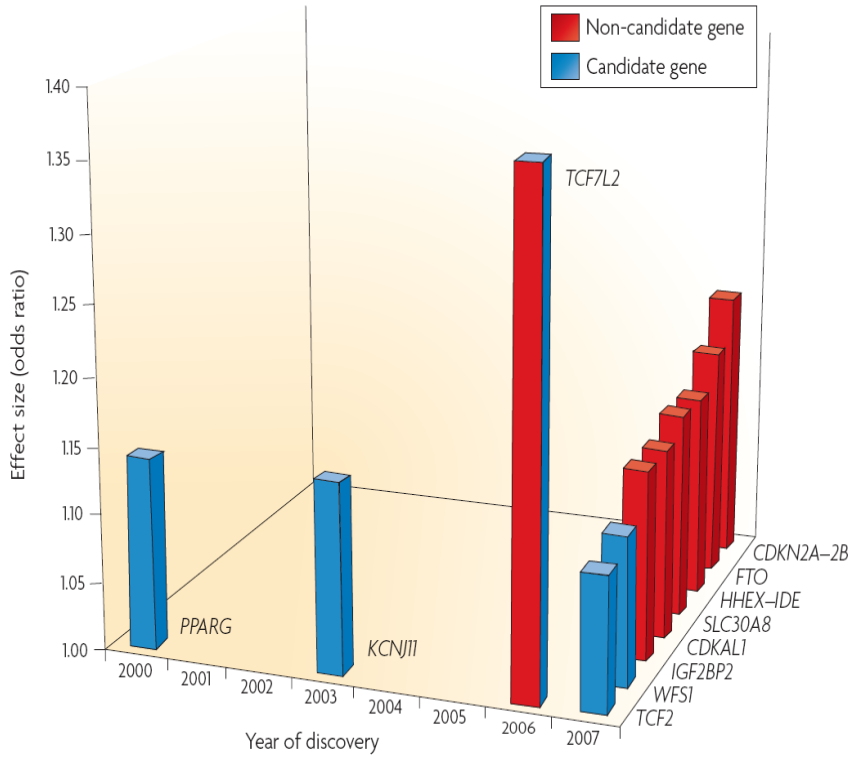


Figure 1.4 Effect sizes of common variants associated with T2D by candidate gene studies and GWAS. Obtained from (Frayling, 2007)

A portion of the missing heritability could also be explained by expanding association studies to more diverse populations. There are several genetic differences observed across populations such as differences in allele frequency and linkage disequilibrium (LD) patterns (Rosenberg et al., 2010). Linkage disequilibrium is the non-random association of alleles in the genome. Numerous factors can influence LD patterns such as mutations and recombination. Because the human species originated in Africa, African-descent populations are more likely to have gone through more recombination events and have accumulated more mutations over time compared with European-descent populations (Rosenberg et al., 2010; Teo et al., 2010; Teo et al., 2009). Thus, African descent populations have more genetic variation and lower spans of LD in the genome compared to Europeans.

Although genetic association studies have been performed in African descent populations, most studies are performed in European populations. These studies rely on the sometimes extensive LD patterns in European populations to identify single nucleotide polymorphisms (SNPs) associated with common diseases or traits. SNPs associated with disease are likely tagSNPs rather than the true causal variant. This property of LD in disease mapping is often referred to as a double-edge sword: the extensive LD allows the identification of assoications but makes it difficult to identify the true causal variant behind the association.

One approach to identifying the true causal variant(s) is to use African descent populations to fine-map the region where an association has been identified in a European population. This is often achieved by densely genotyping a region of interest in an African descent population, where more genetic variation exists. Another benefit to

using diverse populations such as African Americans is to assess the extent to which results originally identified in European populations can be generalized to other populations.

Admixture Mapping as Tool to Identify Disease Loci

Both fine-mapping and generalization in diverse populations rely on previous studies to identify regions of interest associated with a disease or trait of interest. Genetic discovery in diverse populations often involves the traditional study designs mentioned above such as linkage or GWAS. More recently, with the advent of cost-effective genome-wide genotyping, admixture mapping methodology has become a popular tool in the growing arsenal of genetic association discovery methods.

Admixture is the result of gene flow over time between genetically distinct human populations which leads to a mosaic of chromosomal segments representing each population (Figure 1.5). The gene flow can be a single event or a continual event over many generations (Winkler et al., 2010). The proportion of admixture is determined by the rate, direction, and duration of gene flow in the population (Winkler et al., 2010). In an admixture event large haplotype blocks from each population are observed in the first few generations. Haplotype blocks are in high linkage disequilibrium (LD) and decrease in size as generations of admixture increase (Figure 1.5).

There are several admixed populations that exist today. Particularly in the United States there are two large admixed populations, African Americans and Latinos. African Americans were derived from an admixture event dating back to the 16th century during the trans-Atlantic slave trade. Today African Americans on average have gametes that are

~80% African derived and ~20% European derived (Patterson et al., 2004; Reiner et al., 2005). American Latino populations are admixed and are the result of a three way admixture event between European, African, and Native American ancestry (Bertoni et al., 2003; Collins-Schramm et al., 2004; Bonilla et al., 2004; Madrigal et al., 2001). The Latino population is larger and much more diverse compared to African Americans and the proportions of admixture vary significantly by geographical region (Bertoni et al., 2003; Collins-Schramm et al., 2004; Bonilla et al., 2004; Madrigal et al., 2001).

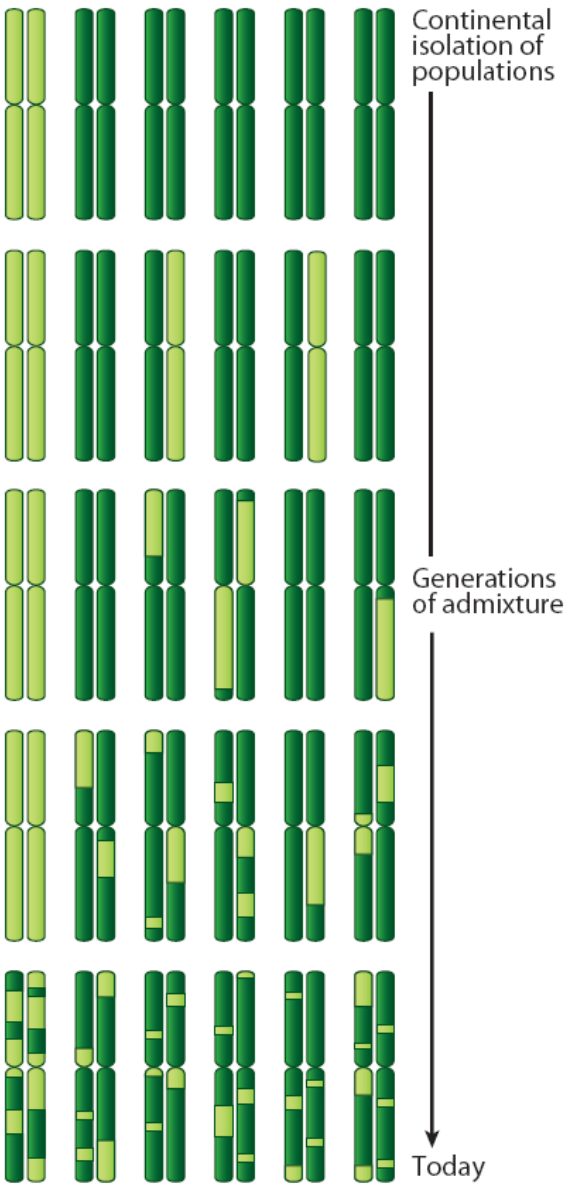


Figure 1.5. Chromosomal ancestry resulting from 8-20 generations of two-way admixture events. Each shade of green represents each ancestral population. Over time (each row) recombination produces chromosomal blocks of different ancestries. The present day admixed population has blocks of ancestry that vary and size. Obtained from (Winkler, 2010 28261 /id).

In 1959, Rife proposed the idea that the large blocks linkage disequilibrium created by admixture events can be used to assign traits to linkage groups (RIFE, 1954). Since the late 1980's there has been a strong interest in developing statistical methods to study this possibility. One of the first was mapping admixture linkage disequilibrium (MALD) (Stephens et al., 1994). This method was based on the association between a marker allele and a trait to assign genes to a linkage group (Stephens et al., 1994). Recently, more attention has been focused to determine if admixture mapping could also be used to identify potential disease loci (Figure 1.6). To do so, alternative methods have been developed that test for an association of the disease with parental ancestry. These methods are based on local chromosomal ancestry rather than LD like previous methods and can be applied to a case/control study design (Price et al., 2009). The application of admixture mapping to identify disease loci is based on the hypothesis that disease loci in the admixed population will be more frequent in chromosomal segments derived from the ancestral population with higher disease prevalence (Winkler et al., 2010). For example, a disease such as hypertension, which is more prevalent in African descent populations, will have cases with an increased amount of African derived alleles at the disease causing locus (Figure 1.5). Current methods are designed for populations derived from two-way admixture events and dichotomous disease outcomes.

Admixture mapping has successfully identified disease loci in African Americans for several common diseases such as end stage renal disease, white blood cell count, and prostate cancer(Freedman et al., 2006; Haiman et al., 2007; Kao et al., 2008; Gower et al., 2002; Reich et al., 2009). While GWAS have identified over 30 loci for T2D, the majority of the studies have been performed in European descent populations and these

variants collectively account for a small fraction of the genetic component of the disease. Since T2D disproportionately affects admixed populations such as African Americans; using admixture mapping to identify potential disease loci is a promising tool to identify additional and possibly population-specific variants relevant to T2D risk.

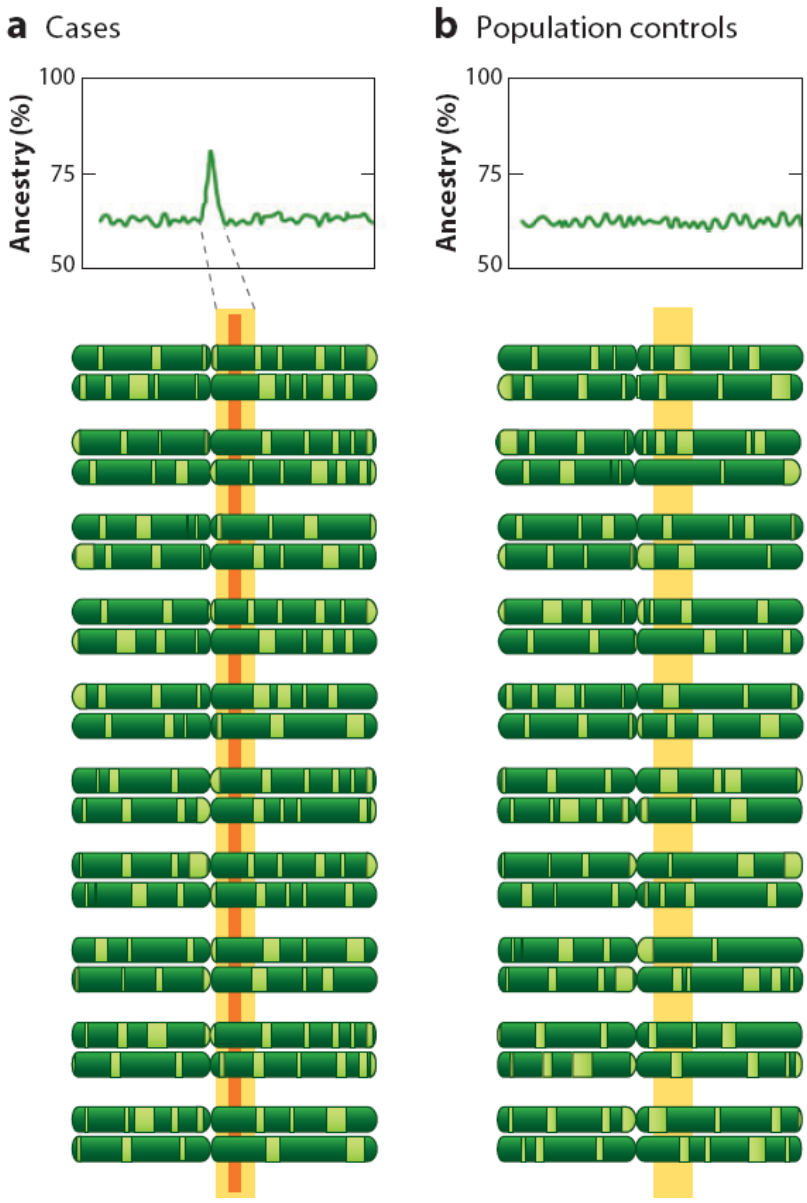


Figure 1.6. Schematic of the pattern of chromosomal admixture around a disease locus. Obtained from (Winkler et al., 2010)

CHAPTER II

ADMIXTURE MAPPING REVEALS POTENTIAL NOVEL LOCI FOR TYPE 2 DIABETES IN AFRICAN AMERICANS

Introduction

As previously mentioned, T2D disproportionately affects African Americans compared to individuals of European descent. To date there have been several GWAS for T2D in Europeans. To confirm associations identified from these studies we will need to assess their role in other populations, especially populations with increased prevalence of diabetes, such as South Asian, African American and Mexican American populations (Frayling, 2007). In African Americans several genetic association studies of T2D related phenotypes such as hypertension, BMI, and fat mass have been performed (Adeyemo et al., 2009; Cheng et al., 2009; Helgadottir et al., 2006; Hooper et al., 1999; Norris et al., 2005; Plant et al., 2006; Van Norstrand et al., 2008). Genetic association studies have not been extensively performed in African Americans for T2D as an outcome variable (Haiman et al., 2012).

Until recently it has been difficult to use identify disease loci in recently admixed populations since the methods to study admixture are new and underdeveloped. Admixture mapping is an alternative method that detects differences in ancestry in admixed individuals and the association these differences may have with the disease. Admixed populations, such as African Americans, have mosaic genomes with varying

portions derived from either European or African ancestry (Figure 1.5). Admixture mapping compares differences in allele frequencies between ancestral populations at each locus between cases and controls (Figure 1.6). Performing an admixture scan for diseases such as T2D, for which the disease prevalence is twofold higher in African Americans, is ideal and will lead to the identification of novel disease loci. This method may also be more powerful when compared to traditional approaches that require large sample sizes to detect an association. We hypothesize T2D associated loci will have increased African ancestry in T2D cases compared to controls.

A complementary approach to the use of alternative statistical approaches to increase power to detect association with disease is the use of alterntive methods for ascertainment. Assessing the use of electronic medical records (EMRs) systems coupled to DNA repositories as a resource for genome science is one of the primary objectives for the National Human Genome Research Institute's electronic MEdical Records and GEnomics (eMERGE) Network (McCarty et al., 2011). Studies from eMERGE have demonstrated that EMR-based genetic association studies replicate existing findings and discover novel associations (Denny et al., 2011; Denny et al., 2010; Crosslin et al., 2011; Turner et al., 2011). In the present study, we sought to first generalize previous GWAS associations and identify novel disease loci for T2D in African Americans. To access whether genetic ancestry is associated with T2D risk, we then performed a genome-wide admixture scan.

Methods

Study Population

African American subjects were collected from the Vanderbilt and Northwestern University biobanks. The Vanderbilt biobank, BioVU, is a collection of DNA samples from discarded blood samples collected for routine clinical care linked to de-identified electronic medical records (EMRs) (Roden et al., 2008). The Northwestern biobank, NUgene, combines DNA samples from consented participants with enrollment questionnaire and longitudinal data from the EMR (McCarty et al., 2011). Both biobanks were approved by Institutional Review Boards at their respective sites.

Patients with available DNA samples were selected from BioVU and/or NUgene. BioVU subjects were African American as indicated by observer reported ancestry, which is highly concordant with genetic ancestry (Dumitrescu et al., 2010). African American ancestry was self-reported for Northwestern subjects (McCarty et al., 2011), which is also known to be highly concordant with genetic ancestry (Yaeger et al., 2008). Our algorithm was carefully designed by expert clinicians experienced with T2D diagnosis. T2D cases were defined as having the following in their EMR: a T2D ICD-9 medical billing code, information about insulin medication, abnormal glucose or HbA1c levels, or more than two diagnoses of T2D by a clinican. All T2D cases with an ICD-9 code for T1D were removed from further analyses. All control subjects had to have at least 2 clinical visits, at least one blood glucose measurement, normal blood glucose or HbA1c levels, no ICD-9 codes for T2D or any related condition, no history of being on insulin or any diabetes related medication, and no family history of T1D or T2D. Using these criteria, an automated method selected cases and matched controls and is described in Figures 2.1 and 2.2. Study population demographics and differences between cases and controls are described in Table 2.1.

Table 2.1. Population Characteristics. We calculated descriptive statistics in African American T2D cases and controls separately. We also compared differences between cases and controls for these characteristics using either Student's T-test or Chi^2 in STATA.

Variable	Cases (n=736)	Controls (n= 827)	P-value
Mean Age (yrs)	53 (± 13.41)	43 (±14.55)	<0.00001
% female	62%	67%	<0.04
Mean BMI (kg/m^2)	35 (±8.88)	30 (±7.61)	<0.0001
Mean % European Ancestry	20.3% (±11.64)	20.6% (± 14.26)	0.36

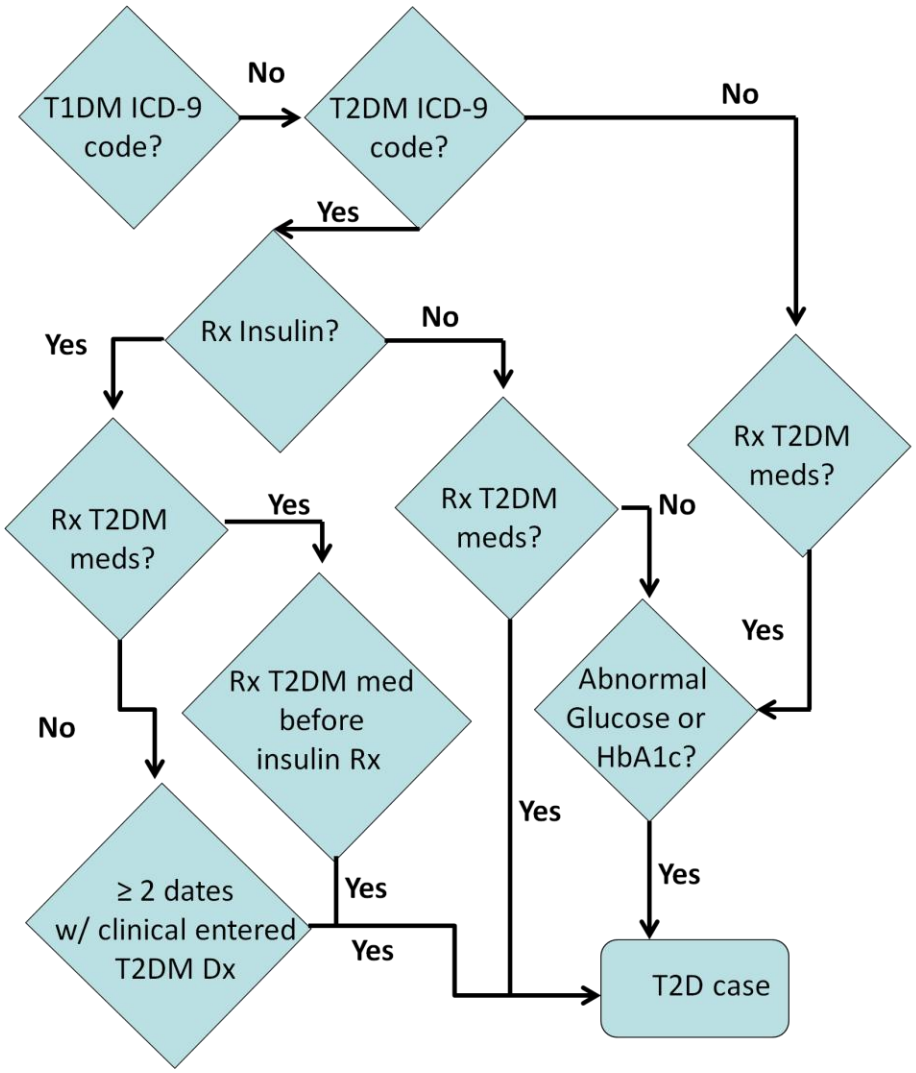


Figure 2.1 Flow chart for Type 2 Diabetes case definition.

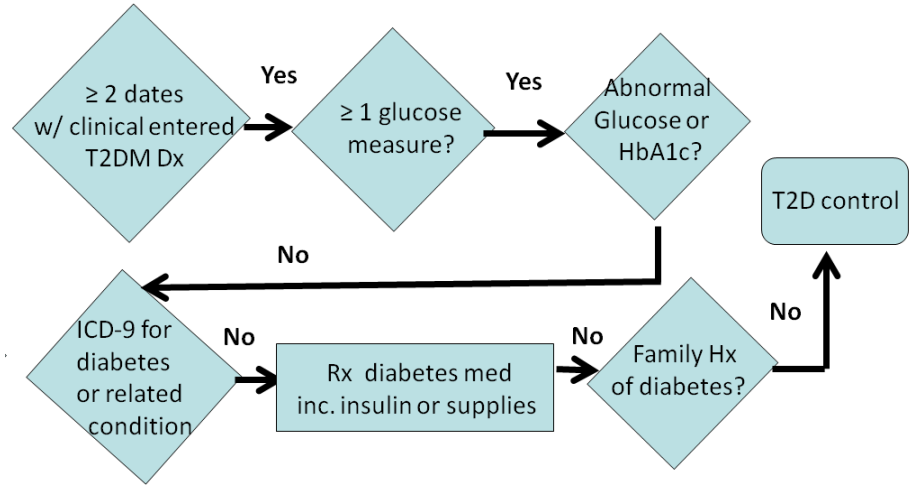


Figure 2,2 Flow chart for Type 2 Diabetes control definition

Genotyping and SNP selection

All individuals that met the inclusion criteria were genotyped for >1.1 million SNPs using the Illumina 1M BeadChip at the Broad Institute. Data were cleaned by the eMERGE QC pipeline (Zuvich et al., 2011). Individuals with cryptic relatedness, ancestry inconsistent with observer- or self-reported ancestry, anomalous X-chromosome heterozygosity or poor genotyping efficiency were removed from further analysis (n = 46). All markers that were intensity probes, had technical failures, were monomorphic, had genotyping efficiencies <99%, had discordant calls with duplicates, had a Hardy Weinberg p-value <1 x 10-4, and had >0 Mendelian errors were removed. A total of 930,000 SNPs remained for association analyses after quality control.

Statistical Methods

We first performed a genome wide association study (GWAS) with >900,000 SNPs genotyped on the Illumina 1M Bead Chip. Assuming an additive genetic model, we performed single SNPs tests of association in 736 T2D cases and 827 controls using logistic regression in PLINK (Purcell et al., 2007). All tests were adjusted for age, sex, and the first three principle components.

Using our GWAS genotype data we also performed an admixture scan on a subset of 4,333 autosomal ancestry informative markers (AIMs) on the Illumina 1M BeadChip (Tandon et al., 2011). Using ANCESTRYMAP we tested AIMs for disease risk variants that differ in frequency across ancestral populations in our African American study population (Patterson et al., 2004). This method uses a Bayesian-likelihood ratio test to identify disease associations throughout the genome (Patterson et al., 2004). Genome-wide ancestry estimates for each individual are averaged across all individuals. A region where enhanced ancestry from one ancestral population is observed compared to the

average across all individuals is indicative of disease association (Patterson et al., 2004). Two statistics are used to determine disease association: the locus genome statistic (LGS, >2.0 significant) and the case-control statistic (CCS, >4.0, significant). In this study we used ANCESTRYMAP to identify average proportion of European ancestry for each individual and to identify disease loci throughout the genome. Pairwise linkage disequilibrium (r^2) was calculated and plotted using Haploview (Barrett et al., 2005). Power calculations were performed using Quanto (Gauderman and Morrison, 2006).

Results

GWAS

There was only one SNP, rs7903146, that met the genome-wide significance at $p<10^{-8}$ (OR=1.7, $p=1.17 \times 10^{-8}$) in our African American study population. This intronic SNP is located on chromosome 10 in the transcription factor 7 –like 2 (*TCF7L2)* gene and has been previously identified with an association with T2D and related phenotypes. There were three novel SNPs that were significant at the $p<10^{-6}$ level (Figure 2.3). Intergenic SNP, rs9347819, located on chromosome 6 increases the risk of T2D (OR= 1.9, $p=5.63 \times 10^{-7}$). Intronic *PRUNE2* rs11177982 located on chromosome 9 is associated with increased risk for T2D (OR = 2.15, $p= 7.112 \times 10^{-6}$). One SNP, rs1048317, is located intron of *LARGE* on chromosome 22 and is less frequent in cases compared to controls. There were no other SNPs that reached genome-wide significance at $p<10^{-8}$ or suggest a possible association at $p<10^{-6}$.

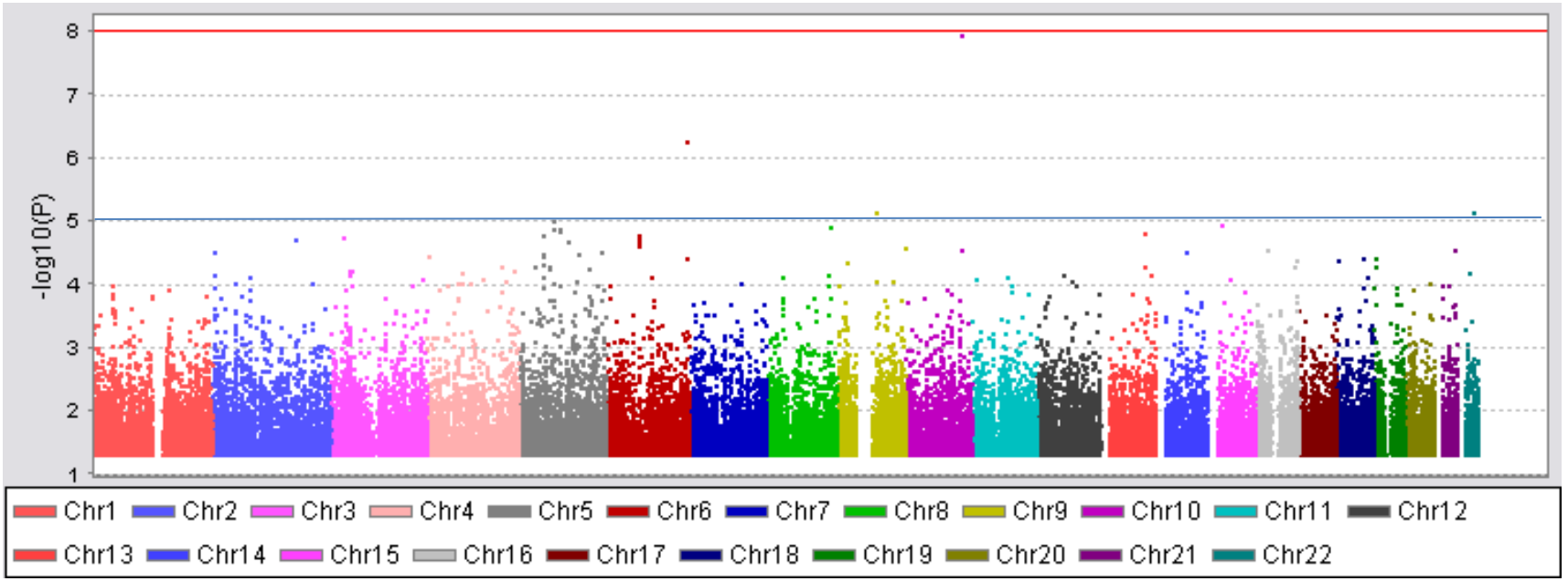


Figure 2.3 Genome-wide Association Results for T2D in African Americans from VGER/NUgene. We performed single SNP test of association with >930K SNPs across the genome using logistic regression in 736 T2D cases and 827 controls. The red line indicates the genome-wide significance level $p<10^{-8}$ and the blue line is indicate of suggestive associations at $p<10^{-6}$.

Admixture Mapping

We performed an admixture scan on a subset of 4,016 ancestry informative markers (AIMs) that passed QC and were genotyped on the Illuminia 1M Bead Chip using ANCESTRYMAP. Global estimates of European ancestry across the genome were calculated for T2D cases and controls. As expected there were no differences in European ancestry between cases and controls ($p = 0.71$). On average cases and controls have 20.3% and 20.5% their genome of European ancestry (Figure 2.4).

In addition to global admixture estimates, we also sought to localize T2D risk variants in our African American study population. We tested AIMs across entire genome for variants with a higher percentage ancestry from one ancestral population in cases compared to controls. The global LGS across the all autosomes was 0.05 indicating most loci in the genome were not associated T2D status in this study population (Figure 2.5). The lowest LGS (averaged across the entire chromosome) was -0.89, on chromosome 7 and the highest estimate was on chromosome 11 (Figure 2.5). Chromosome 11 was the only region in the genome with a significant LGS (LOD = 2.30). Variants in this region had a local LGS ranging from 2.05 to 2.34 and a CCS ranging from 3.2 to 3.7 (Figure 2.6). This region spans about 2,800 kb and encompasses over 90 genes (Figure 2.6). One SNP, rs308328 had the highest CCS (3.7) and is located in intron of the *UNC93B1*. This gene has no known function in humans; however, it is about 50kb upstream of *TCIRG1*, which is expressed in the pancreas.

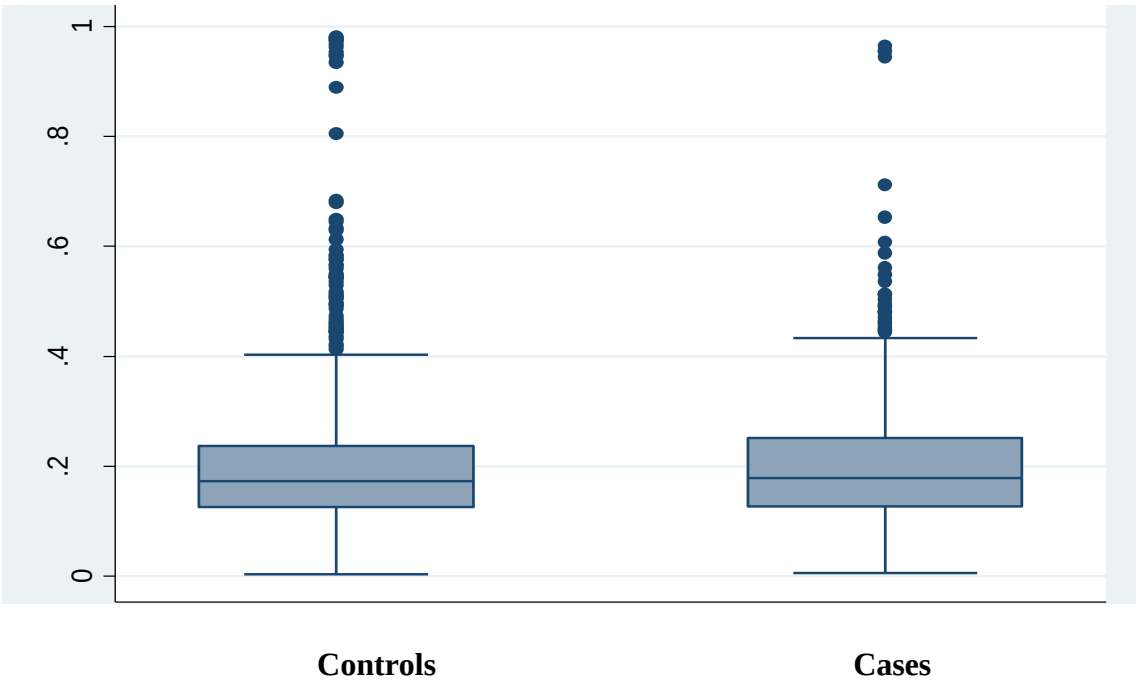
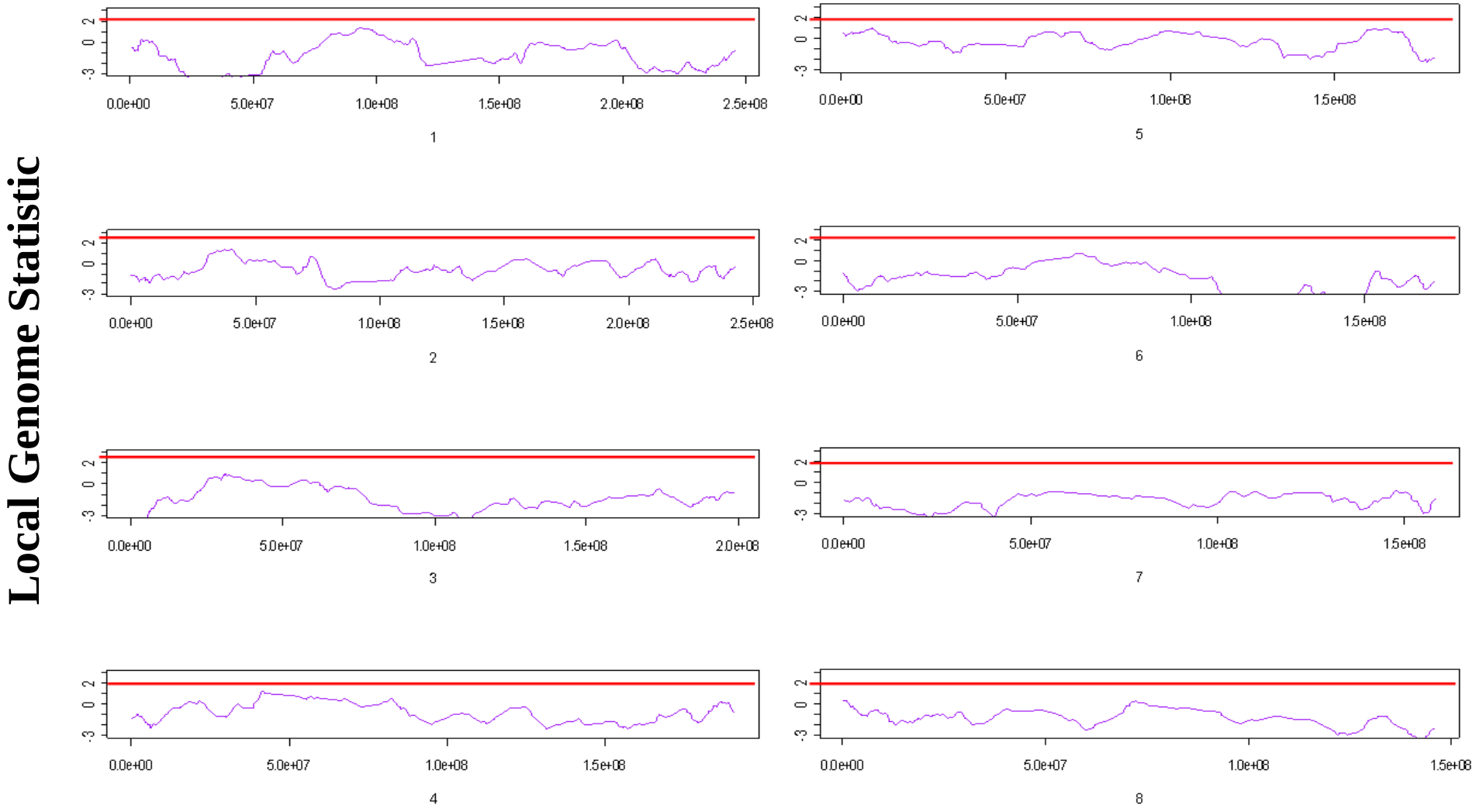
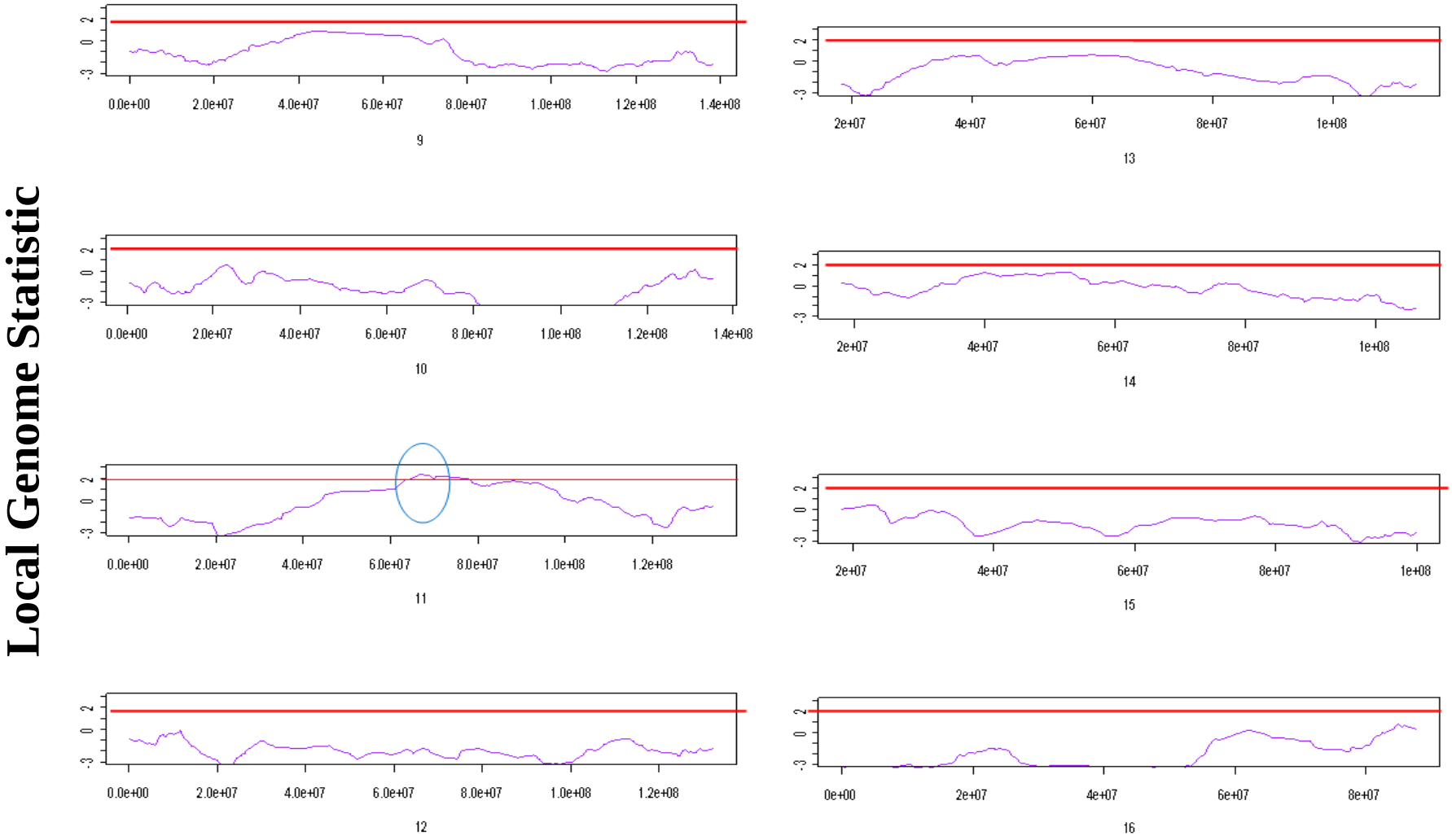


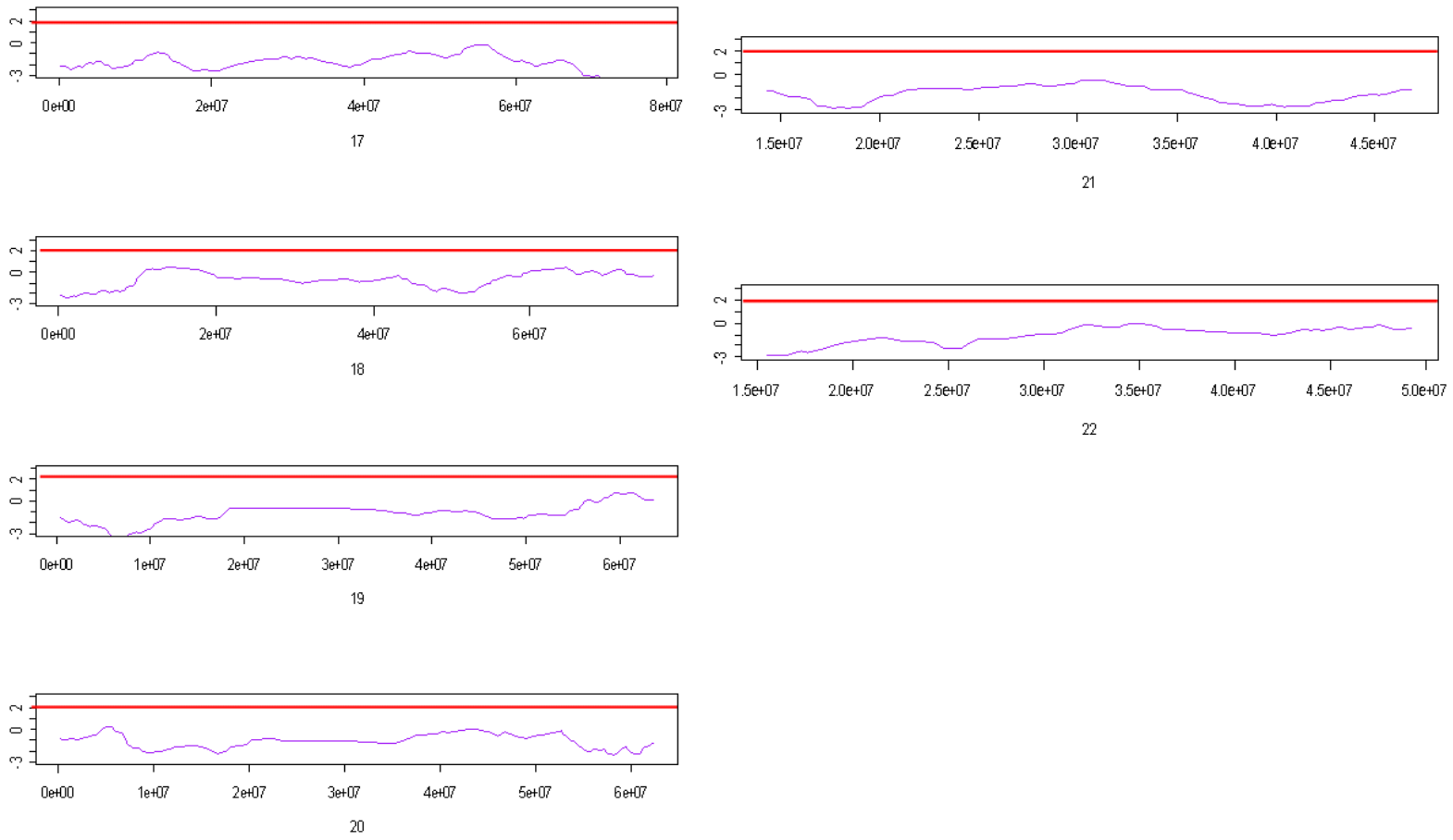
Figure 2.4 Comparison of global European Ancestry in African American cases of T2D and controls. Genome wide estimates of percent European ancestry were calculated for each individual using ANCESTRYMAP. Average European ancestry was compared between cases and controls (p= 0.72).



(A) Admixture mapping results chromosome 1-8.



(B) Admixture mapping results for chromosomes 9-16.



(C) Admixture mapping results for chromosomes 17-22.

Figure 2.5 Local estimates of locus genome stastic across the genome for T2D cases and controls. We performed a genome-wide admixture scan in T2D cases and controls. Plotted on the y-axis is the local LGS by the location of each variant on the x-axis. The red line indicates the level of signifigance (LGS >2.0) and the circled region is the significant region identified from the admixture scan.

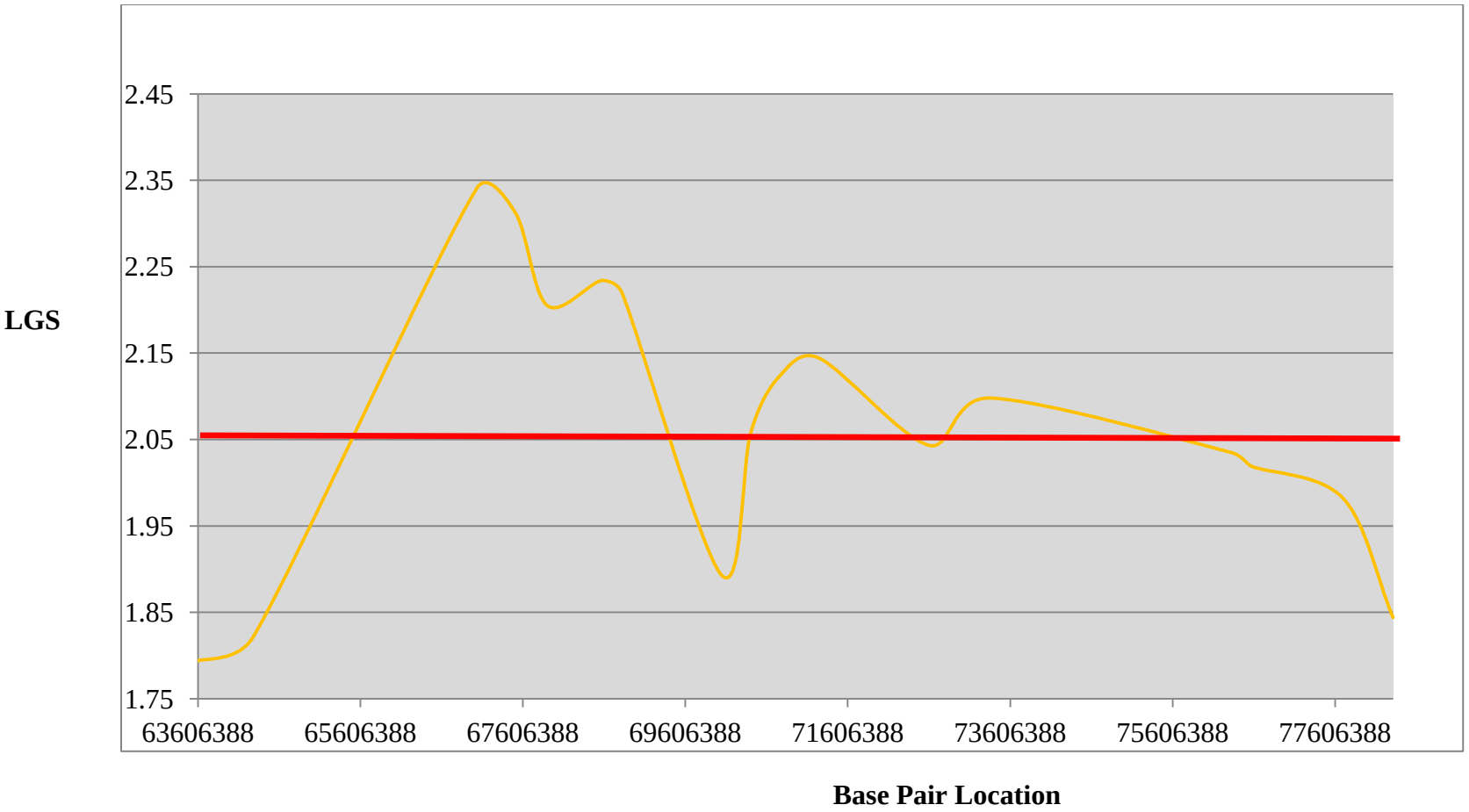


Figure 2.6. Local admixture results on Chromosome 11. Regional view of chromosome 11 with significant LGS score of 2.0 or greater. Plotted on the y-axis is the local LGS by variant location on the x-axis.

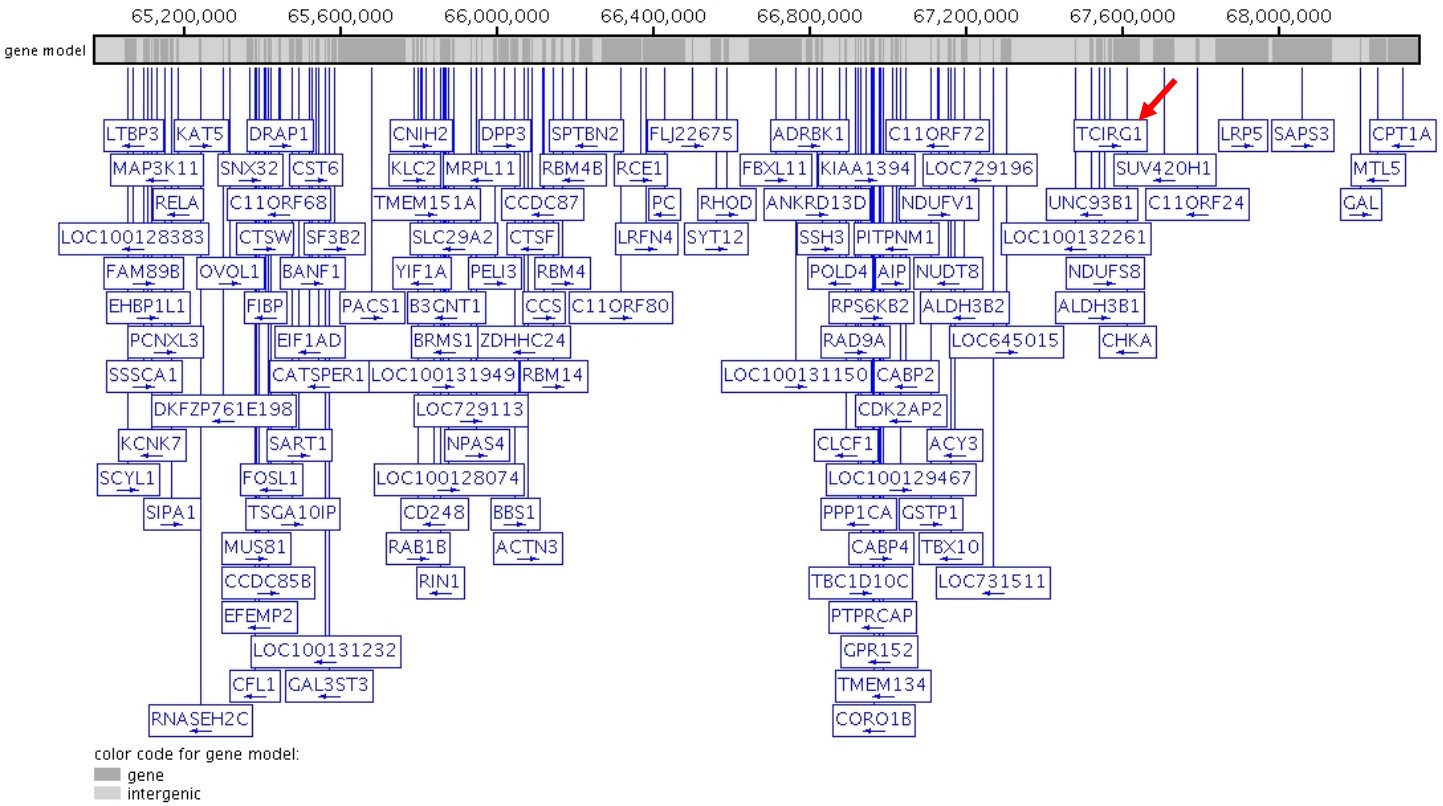


Figure 2.7 Genes in the region under the admixture peak on chromomsome 11. Using Seattle SNPs genome browser we identified all of the genes located in the significant region on chromosome 11 http://pga.gs.washington.edu/. The *TCIRG1* gene indicated by the arrow was selected for as a candidate gene to fine-mapping.

Discussion

We successfully identified several associations with T2D in African Americans using samples derived frm a biorepository linked to EMRs. Notably, we were able to replicate a known association, rs7903146, in our African American study population. Similar to European populations (MAF = 0.29, OR = 1.26) (Tong et al., 2009), this SNP is common in African Americans (MAF = 0.30) and is associated with T2D risk (OR = 1.69). We were also able to identify potential novel disease genes by admixture mapping that can be used for future fine mapping studies.

There were three associations identified in our initial GWAS that suggest an association with T2D in African Americans (Figure 2.3) Intergenic SNP rs9347819 increases the risk of T2D and is common African Americans compared to European descent populations (HapMap CEU MAF = 0.13 and 0.04, respectively). The nearest gene, *QKI* encodes a RNA binding protein and plays a role in schizophrenia, but to date has never been identified for T2D (Aberg et al., 2006b; Aberg et al., 2006a; Lindholm et al., 2001). The other SNPs rs11177982 (*PRUNE2*) and rs1048317 (*LARGE*) also have never been reported to have an association with T2D. As opposed to mutations, the effects of common variation are less clear. All of the SNPs identified in this GWAS are located in the intronic regions with no obvious biological relevance. This region also has low LD in African Americans and associated SNPs are not correlated with presumed functional variants. Despite this finding, reporting these results is still important as intronic SNPs such as rs7903146 (*TCF7L2*) have functional relevance to T2D (Grant et al., 2006; Lyssenko et al., 2007; Tong et al., 2009).

Using admixture mapping we were able to detect a significant admixture peak on chromosome 11. The region under the peak was large and included over 90 genes (Figure 2.7). Our most significant finding rs308328 is located in the *UNC93B1* gene. This gene encodes the 12 membrane spanning protein UNC-93B and is involved in the exogenous antigen presentation and the signaling of toll-like receptors (Tabeta et al., 2006). Mutations in *UNC93B1* in humans have led to impaired production of interferon which is necessary to fight herpetic virus infection (Casrouge et al., 2006). To our knowledge this gene has no direct correlation with T2D, however *UNC93B1* is located upstream of *TCIRG1* which is highly expressed in the pancreas. While this region is gene rich, there are only 146 common SNPs reported in Yoruban (YRI) HapMap samples within 100kb of the *UNC93B1* gene (The International HapMap Project, 2003). To determine if rs308328 is correlated with SNPs in the *TCIRG1* gene and thus representing this effect we calculated pairwise LD (r^2) in this region (Figure 2.8). There were only 59 SNPs in this region genotyped in our study population, and there were high levels of linkage disequilibrium in our African Americans study population (Figure 2.8). However there is only one SNP rs308351, located in an intergenic region, which is in moderate LD with rs308328 in our data. Further fine-mapping and/or deep re-sequencing will be required to uncover the functional variant in this region in African-descent populations.

There were several limitations to the present study. Particularly, for our GWAS we were underpowered to detect effects smaller than 1.6 (OR) for SNPs with common SNPs (MAF >0.05). Our study population compared to recently published GWAS for T2D is relatively small and underpowered. Also, most GWAS fixed-content products are biased to common variation and based on LD patterns observed in European populations

(Spencer et al., 2009). Therefore, the SNPs on these platforms capture the majority of common variation in Europeans but perform worse in African-descent populations. Also rare variation and/or population-specific variation are underrepresented in GWAS fixed-content products. These limitations impact our power to detect associations for T2D in diverse populations.

Further limitations include the fact that this admixture scan was limited to ancestry informative markers (AIMs) genotyped on the Illumina 1M. A total of 4,016 AIMs were used, which only represents 0.37% of the markers genotyped in the entire dataset. The AIMs were previously identified for several GWAS platforms by Tandon et al (Tandon et al., 2011). It is possible that are several SNPs genotyped on these platforms that do not meet the criteria outlined by Tandon et al but are informative and possibly associated with T2D status (Tandon et al., 2011). Additionally, we only included SNPs that were unlinked ($r^2 < 0.80$), a requirement of the ANCESTRYMAP software package (Patterson et al., 2004). Other methods have recently been developed that can accomodate correlated SNPs, which increase power and give better estimates of local ancestry estimates in the genome (Price et al., 2009).

We used a case-control study design for both the GWAS and the admixture scan. However, using quantitative traits is an excellent alternative to using clinical outcomes for complex diseases. Quantitative measurements can be more robust to phenotypic heterogeneity and are often more uniform compared to disease outcomes. Given the complexity of T2D, it is likely that different genetic etiologies explain several traits or intermediate phenotypes associated with T2D. Quantitative measurements such as BMI and fat mass have been used to identify genetic risk factors of T2D. One example is with

SNPs in the *FTO* gene and obesity; an association that has been well-characterized (Dina et al., 2007; Frayling et al., 2007; Scuteri et al., 2007). We were unable to reproduce this finding using a case-control study design in our analysis, which is likely a result of power (p-value > 0.40) (Figure 2.3).

Alternatively, one would argue that using a disease outcome is more informative compared to quantitative measurements, given the direct and obvious correlation to the disease. One major limitation to using intermediate phenotypes (such as quantitative measurements) is the assumption that the associations identified are also associated with the disease. We successfully identified several novel associations across the genome using T2D disease outcome status in our GWAS (at the $p<10^{-6}$ level) and genome-wide admixture scan. Our admixture scan revealed a region on chromosome 11 that might be associated with T2D. These data suggest that the functional variants have yet to be identified for T2D and emphasize the need for future fine mapping studies in a larger African-descent population. Subsequent fine-mapping studies will provide a better understanding these associations with T2D.

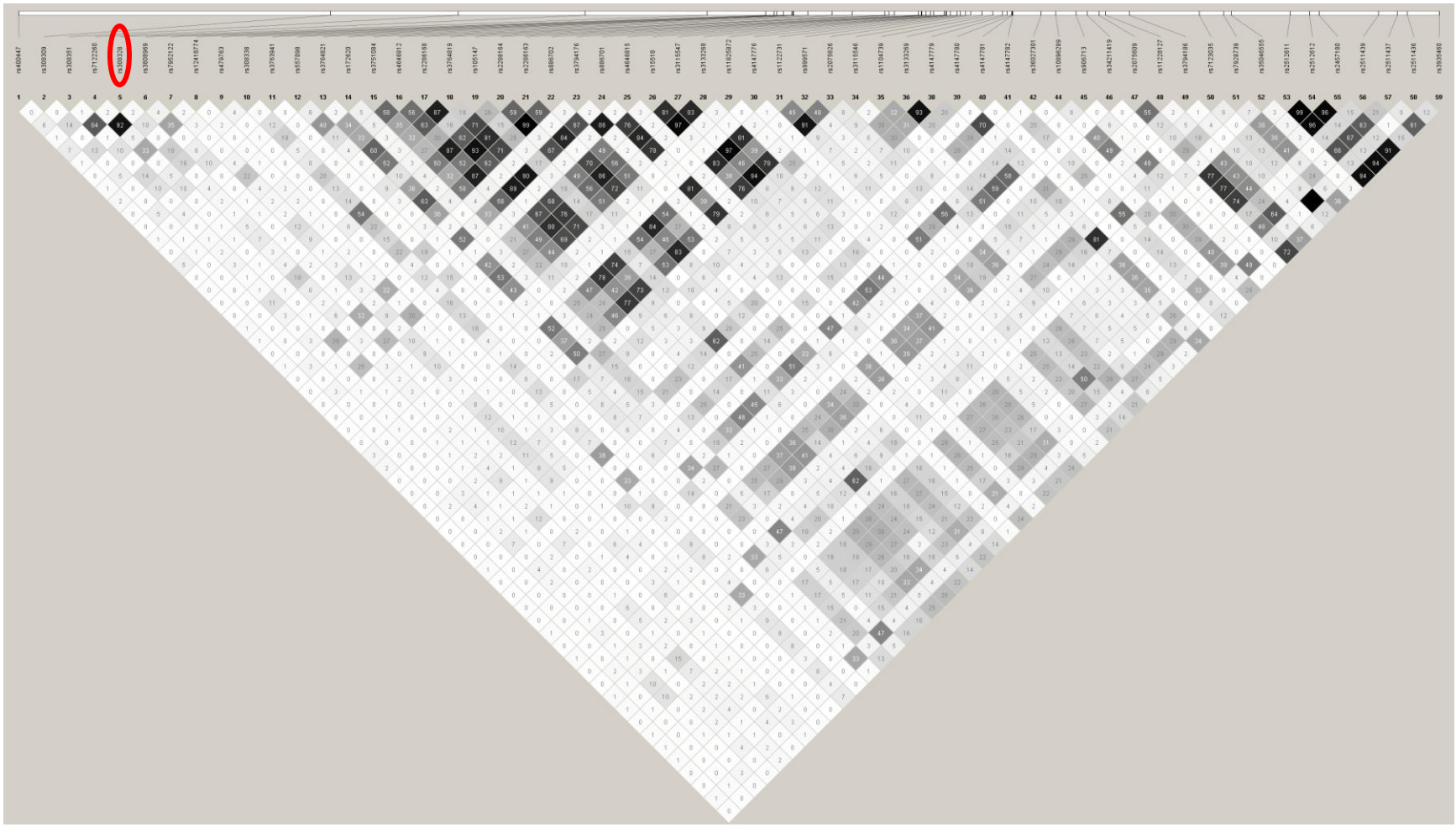


Figure 2.8. Linkage disequilibrium around the 100kb around rs308328 in African American T2D cases and controls. We calculated pairwise LD around our most significant finding (circled in red) from the admixture scan in our AfricanAmerican study population using Haploview (Barrett et al., 2005).

CHAPTER III

FINEMAPPING THE *TCIRG1* REGION IN AFRICAN AMERICANS

Introduction

In Chapter II we identified a potential disease region on chromosome 11 that includes over 90 genes (Figure 2.7). Our most significant association, rs308328, is located in *UNC93B1* which has no direct biological relevance to T2D. Further investigation of this region revealed a potential candidate gene, *TCIRG1*, which is expressed in the pancreas; a vital organ involved in the pathogenesis of T2D. Our initial analysis was limited to AIMs and warrants fine mapping studies to elucidate the role this region has with T2D.

Fine-mapping is a complementary approach to gene discovery methods such as GWAS and admixture mapping to identify potential 'causal' variants among highly correlated SNPs. This method requires densely genotyping or sequencing the region of interest. Fine-mapping studies are usually hypothesis-driven where a gene is selected within the associated region that is suspected to be biologically relevant to the phenotype being studied. For example, this approach would be useful to follow-up the *UNC93B1* region, where the SNP rs308328 is associated with T2D, but has no obvious biological relevance to T2D. Performing fine-mapping studies in this region using a large sample of African Americans may help identify the putative casual variants in *UNC93B1* and other genes identified by GWAS with unknown biological function. Therefore, using additional

African Americans in BioVU we sought to fine-map the *UNC93B1* gene region in T2D cases and controls.

Methods

Study Population

All subjects were observer-reported African American, at least 18 years of age, and were collected from the Vanderbilt University biobank, BioVU (see Chapter II, page 20**).** The algorithms used to define cases and controls have been previously described in Chapter II, figures 2.2 and 2.2. Since BioVU accrues new samples each week, we re-ran the algorithm for this follow-up study in June of 2011 (~a year later). We identified a total of 1,659 cases and 3,372 controls for this fine-mapping study, of which 1,272 samples overlap with the discovery study described in Chapter II. Study population characteristics are described in Table 3.1.

SNP Selection and Genotyping

We genotyped 51 tagSNPs approximately 100kb upstream and downstream of the *TCIRG1* gene, previously identified by admixture mapping (see Chapter II, figure 3.1). HapMap Yoruban population linkage disequilibrium patterns were used to select tagSNPs in this region (Figure 3.1 http://pga.gs.washington.edu/). All tagSNPs with a minor allele frequency >0.08 were selected for this analysis. For quality control purposes, we purposely included 20 (out of 1276) SNP pairs that are in high linkage disequilibrium (LD) (r^2>0.80) with each other. Genotyping was performed at the Vanderbilt University Center for Human Genetics Research DNA Resources Core. The quality control metrics described in Chapter II were also used for this analysis.

Statistical Methods

Using logistic regression, we performed single SNP tests of association assuming an additive genetic model for SNPs that passed QC in PLINK (Purcell et al., 2007). Analyses were performed unadjusted and adjusted for age and sex. All results were plotted using Synthesis-View (Pendergrass et al., 2010). To account for multiple testing, we employed Bonferonni correction and declared a significance threshold of $p<0.001$. Linkage disequilibrium was calculated by measuring the correlation coefficient r^2 in Haploview (Barrett et al., 2005).

Results

We detected several novel associations near the candidate region identified by admixture mapping in Chapter II. There were seven SNPs that met our conservative Bonferroni significance threshold of $p<0.001$, all of which are located in the *CHKA* gene. Having one or more copies of the minor allele increases the risk of having T2D in African Americans for six of these SNPs (OR= 1.08- 1.37). One SNP rs10791957 has a protective effect decreasing the risk of T2D (OR= 0.83). There were an additional twelve SNPs within this region that were significant at the $p<0.05$ level, suggestive of an association. There were seven SNPs in the *CHKA* gene, three SNPs in the *ALDH3B1*, and two SNPs in the *TCIRG1* that trended towards significance (Table 3.2). Of the twelve associations having one or more copies of the minor allele decreseases the risk of T2D for half (six SNPs) while the other half increases the risk of T2D in African Americans (Table 3.2).

As previously mentioned, we selected several SNPs in high LD for quality control purposes. To distinguish between independent and redundant results, we calculated pairwise LD between all SNPs in our African American study population (Figure 3.2). There were a total of 22 highly correlated SNP-SNP pairs (r^2>0.80), 20 of which were expected based on our assay design (Figure 3.2). Taking into account these correlations, there were only three independent associations (r^2< 0.80) that met our significance threshold. However even these associations are in moderate LD (r^2 <0.80 and >0.40).

Table 3.1. Study population demographics for 4,246 unrelated African Americans in BioVU. We calculated descriptive statistics and tested whether there were differences between cases and controls for these characteristics in using either Student's T-test or Chi^2 in STATA.

Variable	Cases (n= 1,602)	Controls (n=2,644)	P-value
Mean Age (yrs)	52 (±15.83)	46 (±18.84)	<0.00001
Sex (% female)	67%	67%	0.23

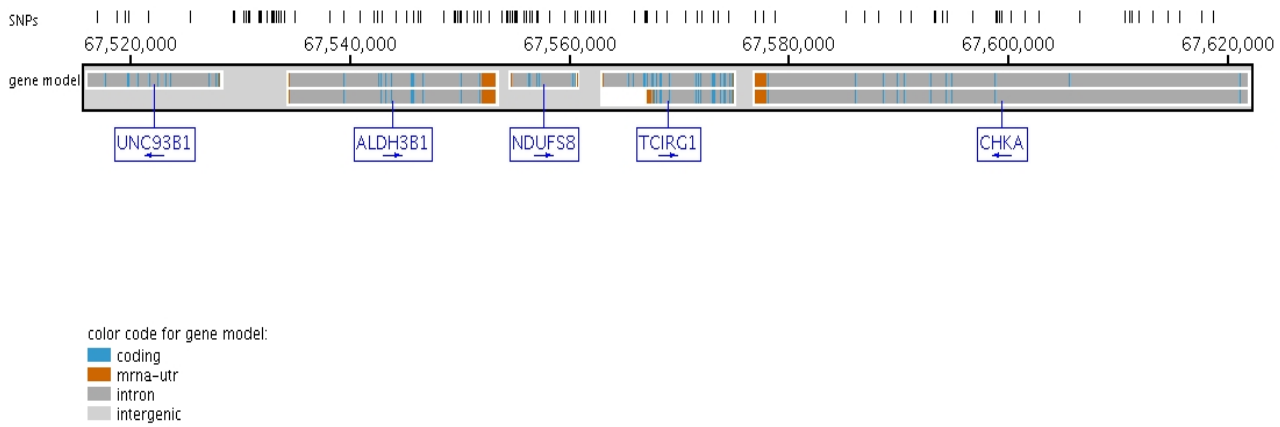


Figure 3.1 Candidate region targeted for fine mapping. Using Seattle SNPs genome browser the candidate genes located within 100kb of the *TCIRG1* gene, their orientation, and gene structure are displayed. SNPs annotated for these genes are located at the top of the figure denoted by hash marks. (Image generated from http://pga.gs.washington.edu/).

Table 3.2. Fine mapping association results. We performed logistic regression with 48 SNPs within 100kb of *TCIRG1* gene with T2D case/control status in 4,246 African Americans. All tests were adjusted for age and sex, assuming an additive model. Bolded are significant results p-value < 1.0 E^{-3}.

SNP	Gene/ location	Coded Allele	OR	95% CI	SE	P
rs105147	*ALDH3B1* intron	A	0.89	(0.80, 1.00)	0.06	0.050
rs2286164	*ALDH3B1* intron	C	1.09	(0.98, 1.22)	0.05	0.100
rs2286169	*ALDH3B1* intron	G	1.15	(1.03, 1.29)	0.06	0.014
rs308335	*ALDH3B1* intron	G	0.95	(0.82, 1.09)	0.07	0.455
rs308337	*ALDH3B1* intron	A	0.94	(0.81, 1.09)	0.07	0.404
rs308338	*ALDH3B1* intron	G	0.93	(0.84, 1.04)	0.05	0.204
rs557098	*ALDH3B1* intron	G	0.96	(0.83, 1.10)	0.07	0.562
rs2286163	*ALDH3B1* Ser→ Ser	A	1.16	(1.03, 1.31)	0.06	0.013
rs11825872	*ALDH3B1*downstream	C	0.90	(0.73, 1.10)	0.10	0.302
rs3751082	*ALDH3B1*Leu→Leu	T	1.07	(0.93, 1.23)	0.07	0.315
rs12418774	*ALDH3B1*upstream	G	1.16	(0.89, 1.51)	0.14	0.280
rs3763942	*ALDH3B1*upstream	C	1.00	(0.83, 1.19)	0.09	0.976
rs7928739	*CHKA* intron	A	0.87	(0.78, 0.97)	0.06	0.011
rs3794186	*CHKA* 3' UTR	T	1.05	(0.87, 1.26)	0.09	0.636
rs11228145	*CHKA* downstream	T	0.87	(0.76, 1.00)	0.07	0.052
rs1547888	*CHKA* downstream	C	1.17	(1.03, 1.32)	0.06	0.016
rs1547889	*CHKA* downstream	T	1.14	(0.99, 1.30)	0.07	0.062
rs2511469	***CHKA* downstream**	**A**	**1.21**	**(1.09, 1.34)**	**0.05**	**4.46E-04**
rs2511470	*CHKA* downstream	C	1.14	(1.01, 1.27)	0.06	0.029
rs2511472	*CHKA* downstream	C	0.87	(0.78, 0.97)	0.06	0.010
rs2512623	*CHKA* downstream	T	0.86	(0.77, 0.95)	0.06	0.005
rs7944372	***CHKA* downstream**	**C**	**1.23**	**(1.11, 1.37)**	**0.05**	**1.04E-04**
rs10791957	***CHKA* intron**	**A**	**0.83**	**(0.74, 0.92)**	**0.05**	**4.71E-04**

rs2511437	***CHKA* intron**	**G**	**1.20**	**(1.08, 1.33)**	**0.05**	**5.52E-04**
rs2511439	*CHKA* intron	T	1.12	(0.96, 1.31)	0.08	0.159
rs2512612	***CHKA* intron**	**C**	**1.22**	**(1.10, 1.36)**	**0.05**	**1.61E-04**
rs4930557	***CHKA* intron**	**C**	**1.22**	**(1.10, 1.35)**	**0.05**	**2.18E-04**
rs6591331	*CHKA* intron	A	1.15	(1.03, 1.28)	0.06	0.016
rs6591333	***CHKA* intron**	**G**	**1.20**	**(1.08, 1.34)**	**0.05**	**8.88E-04**
rs7123035	*CHKA* intron	A	0.97	(0.82, 1.15)	0.08	0.739
rs7952122	intergenic (*UNC93B1 ALDH3B1*)	A	1.04	(0.93, 1.16)	0.06	0.497
rs308351	LOC100132261 downstream	C	1.05	(0.92, 1.20)	0.07	0.436
rs308309	LOC100132261 upstream	G	0.93	(0.79, 1.09)	0.08	0.362
rs4147776	*NDUFS8* 5' UTR	C	1.18	(0.88, 1.58)	0.15	0.277
rs11824781	*NDUFS8* intron	C	0.97	(0.84, 1.12)	0.07	0.676
rs3115545	*NDUFS8* intron	T	0.94	(0.77, 1.14)	0.10	0.532
rs10896288	*NDUFS8* upstream	G	0.95	(0.82, 1.09)	0.07	0.448
rs11823975	*NDUFS8* upstream	C	0.88	(0.73, 1.05)	0.09	0.165
rs3115546	*NDUFS8*intron	G	0.94	(0.83, 1.06)	0.06	0.332
rs7949541	*NDUFS8*intron	T	1.01	(0.80, 1.28)	0.12	0.931
rs3133269	*NDUFS8*upstream	C	1.01	(0.90, 1.13)	0.06	0.821
rs7951010	*TCIRG1* upstream	T	0.93	(0.76, 1.14)	0.10	0.479
rs884826	*TCIRG1* upstream	A	0.87	(0.78, 0.97)	0.06	0.015
rs2075609	*TCIRG1* intron	G	1.16	(1.05, 1.29)	0.05	0.004
rs2240387	*UNC93B1*	G	1.05	(0.86, 1.13)	0.06	0.380
rs3808969	*UNC93B1*downstream	C	1.03	(0.87, 1.22)	0.09	0.757
rs308328	*UNC93B1*intron	A	0.98	(0.86, 1.13)	0.07	0.796

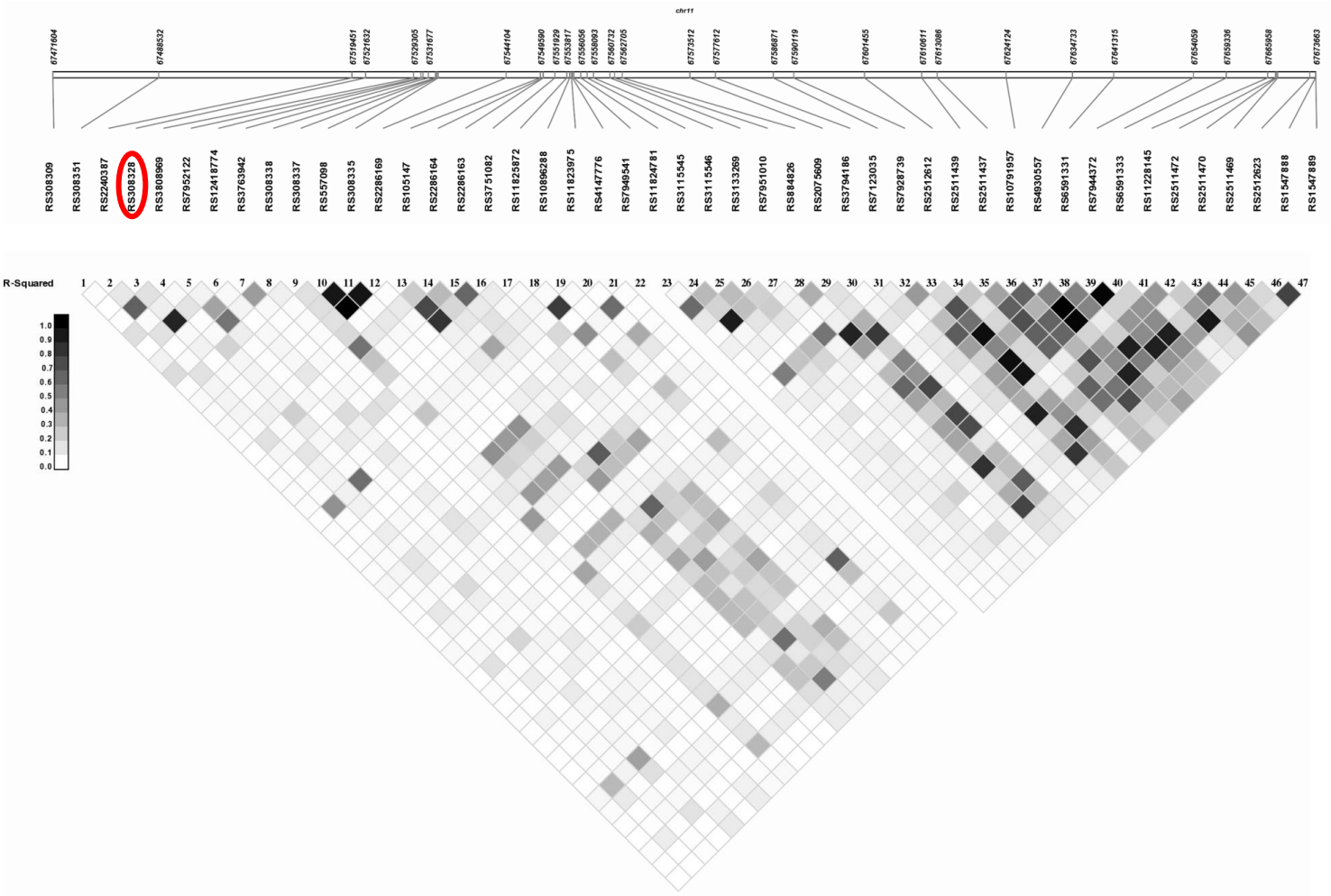


Figure 3.2. Linkage disequilibrium around the 100kb around rs308328 in African Americans T2D cases and controls. We calculated pairwise LD around our most significant finding (circled in red) from the admixture scan in our AfricanAmerican study population using Haploview (Barrett et al., 2005).

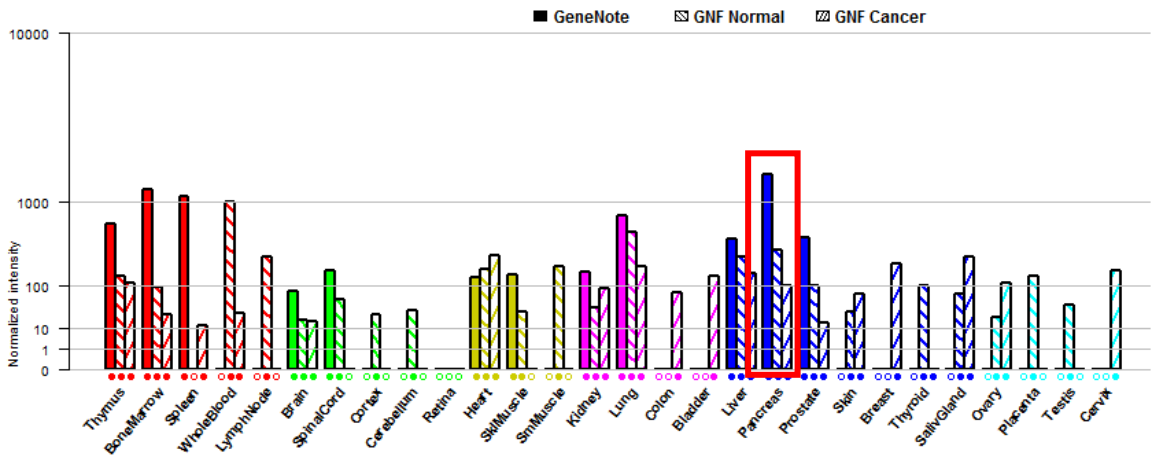


Figure 3.3. *TCIRG1* mRNA is highly expressed in human pancreas tissue. Using GeneNote (http://bioinfo2.weizmann.ac.il/cgi-bin/genenote/home_page.pl), we identified the mRNA expression levels in various human tissues for *TCIRG1*.

Discussion

A review article published in 2007 reported that the next steps (after GWAS) for finding diabetes susceptibility loci was to first fine-map the novel T2D loci (Frayling et al., 2007). Based on the results from the genome-wide admixture scan performed in Chapter II, we sought to fine-map associations with T2D on chromosome 11 in a large study population of African Americans. This region included five genes; one of which *TCIRG1* is highly expressed in the pancreas and yet to be assoicated with T2D. We observed a total of seven significant associations in this region at $p<0.001$, two were in *TCIRG1* at the $p<0.05$ level.

In Chapter II rs308328 was our most significant finding, and the region under the >2,500 kb admixture peak included over 90 genes. A 100 kb window surrounding *TCIRG1* was chosen for fine-mapping based on the possible biological relevance of this

gene with T2D. The T-cell immune regulator, ATPase, H+ transporting, lysosomal V0 subnit A3 gene, *TCIRG1*, is highly expressed in white blood cells (lymphocytes and monocytes) but is also expressed in the adrenal cortex and the pancreas in normal human cells (Figure 3.3.). There have been reports of an association with the innate immune system and T2D but the exact pathway is still unclear (Pickup, 2004). The pancreas is an organ in the endocrine system that produces several hormones including insulin and is directly involved in insulin regulation and thus T2D risk.

We were able to identify several associations in this region; however, there were no significant findings in the *TCIRG1* gene met our significance threshold (Table 3.2). In fact all of the significant associations are located in the choline kinase alpha (*CHKA*) gene, which encodes a protein essential in the phospholipid biosynthesis and may contribute to tumor cell growth (Gallego-Ortega et al., 2009; Ramirez de et al., 2008). To our knowledge, we are the first to report this association with T2D diabetes.

This work emphasizes the importance of genetic association studies in African American populations. Compared to our discovery analysis in Chapter II, which was powered to detect moderate to large associations (≥1.7 OR), the fine-mapping study was sufficiently powered (>80%) to detect effects as small as 1.3 (OR) given our sample size. Furthermore, fine mapping the *TCIRG1* improved our power to directly detect an association in this candidate region (Table 3.3.). Of the 51 SNPs selected for fine mapping only 29 were directly assayed and genotyped in the original GWAS reported in Chapter II. We identified eight independent associations at $p<0.05$, five were also included in the GWAS. None of these SNPs were significant in our GWAS ($p<0.05$) which might be due to an increase in power (larger sample size) but also indicating that

admixture mapping followed by fine mapping is a powerful method in the identification of novel disease loci.

The major limitation of this study is process by which SNPs were selected. First, we limited the fine mapping study to one gene region from approximately 90 genes located within the admixture peak. We based our gene selection on known biological function and presumed relevance to T2D, but it is possible that other genes within the admixture peak are responsible for the association with T2D. Second, we implemented a tagSNP approach to select specific SNPs for genotyping. The use of tagSNPs is a cost effective alternative to genotyping all known common variants regardless of linkage disequilibrium patterns throughout the genome. The caveat is that there is a possibility that the associated SNPs are not the functional or true "risk" SNP but are in linkage disequilibrium with the un-genotyped functional variant. Also, a fraction of SNPs must be assayed directly as no SNPs will serve as sufficient tags or proxies. We selected common tagSNPs with a minor allele frequency of at least 0.08 within 100kb of the *TCIRG1* gene (Figure 3.1). Selected SNPs had to have an $r^2>0.80$ with other SNPs in the region to be selected for this analysis. Limiting our analysis to only common tagSNPs we fail to capture rare variants which may have large effects on T2D risk. Additionally our high r^2 threshold resulted in several redundant associations due to moderate to high LD in this gene region (Table 3.3, Figure 2.8). Finally, it is possible that we failed to assay a sufficient proxy or tagSNP for the true "causal" variant associated with T2D.

Although we were unable to detect significant associations in the *TCIRG1* gene at $p<0.001$, we detected associations with the *CHKA* gene. We were also able increase our power to detect SNPs with small effects in this region by increasing our sample size and

fine mapping this gene region. The variants we identified are likely tagSNPs as they have no obvious biological function; therefore deep sequencing of this region is warranted to identify rare variants and the affect they may have on T2D disease risk. Overall, these data contribute to the understanding of the genetic etiology of T2D in African Americans and serve as a starting point for future sequencing and molecular studies to better understand the biological processes that lead to T2D and other complex human diseases.

REFERENCES

Aberg,K., Saetre,P., Jareborg,N., and Jazin,E. (2006a). Human QKI, a potential regulator of mRNA expression of human oligodendrocyte-related genes involved in schizophrenia. Proc. Natl. Acad. Sci. U. S. A *103*, 7482-7487.

Aberg,K., Saetre,P., Lindholm,E., Ekholm,B., Pettersson,U., Adolfsson,R., and Jazin,E. (2006b). Human QKI, a new candidate gene for schizophrenia involved in myelination. Am. J Med. Genet. B Neuropsychiatr. Genet. *141B*, 84-90.

ADA. Diabetes Statistics. 1-26-2011.
Ref Type: Report

Adeyemo,A., Gerry,N., Chen,G., Herbert,A., Doumatey,A., Huang,H., Zhou,J., Lashley,K., Chen,Y., Christman,M., and Rotimi,C. (2009). A genome-wide association study of hypertension and blood pressure in African Americans. PLoS. Genet. *5*, e1000564.

Ahlqvist,E., Ahluwalia,T.S., and Groop,L. (2011). Genetics of type 2 diabetes. Clin. Chem. *57*, 241-254.

Almind,K., Bjorbaek,C., Vestergaard,H., Hansen,T., Echwald,S., and Pedersen,O. (1993). Aminoacid polymorphisms of insulin receptor substrate-1 in non-insulin-dependent diabetes mellitus. Lancet *342*, 828-832.

Barrett,J.C., Fry,B., Maller,J., and Daly,M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. *21*, 263-265.

Bell,J.T., Timpson,N.J., Rayner,N.W., Zeggini,E., Frayling,T.M., Hattersley,A.T., Morris,A.P., and McCarthy,M.I. (2011). Genome-wide association scan allowing for epistasis in type 2 diabetes. Ann. Hum. Genet *75*, 10-19.

Bertoni,B., Budowle,B., Sans,M., Barton,S.A., and Chakraborty,R. (2003). Admixture in Hispanics: distribution of ancestral population contributions in the Continental United States. Hum. Biol *75*, 1-11.

Bonilla,C., Shriver,M.D., Parra,E.J., Jones,A., and Fernandez,J.R. (2004). Ancestral proportions and their association with skin pigmentation and bone mineral density in Puerto Rican women from New York city. Hum. Genet *115*, 57-68.

Casrouge,A., Zhang,S.Y., Eidenschenk,C., Jouanguy,E., Puel,A., Yang,K., Alcais,A., Picard,C., Mahfoufi,N., Nicolas,N., Lorenzo,L., Plancoulaine,S., Senechal,B., Geissmann,F., Tabeta,K., Hoebe,K., Du,X., Miller,R.L., Heron,B., Mignot,C., de Villemeur,T.B., Lebon,P., Dulac,O., Rozenberg,F., Beutler,B., Tardieu,M., Abel,L., and Casanova,J.L. (2006). Herpes simplex virus encephalitis in human UNC-93B deficiency. Science *314*, 308-312.

Centers for Disease Control and Prevention. National Diabetes Fact Sheet: National Estimates and General Information on Diabetes and Prediabetes in the United States 2011. 2011. Atlanta, GA:U.S., U.S. Department of Health and Human Services.
Ref Type: Report

Cheng,C.Y., Kao,W.H., Patterson,N., Tandon,A., Haiman,C.A., Harris,T.B., Xing,C., John,E.M., Ambrosone,C.B., Brancati,F.L., Coresh,J., Press,M.F., Parekh,R.S., Klag,M.J., Meoni,L.A., Hsueh,W.C., Fejerman,L., Pawlikowska,L., Freedman,M.L., Jandorf,L.H., Bandera,E.V., Ciupak,G.L., Nalls,M.A., Akylbekova,E.L., Orwoll,E.S., Leak,T.S., Miljkovic,I., Li,R., Ursin,G., Bernstein,L., Ardlie,K., Taylor,H.A., Boerwinckle,E., Zmuda,J.M., Henderson,B.E., Wilson,J.G., and Reich,D. (2009). Admixture mapping of 15,280 African Americans identifies obesity susceptibility loci on chromosomes 5 and X. PLoS. Genet. *5*, e1000490.

Collins-Schramm,H.E., Chima,B., Morii,T., Wah,K., Figueroa,Y., Criswell,L.A., Hanson,R.L., Knowler,W.C., Silva,G., Belmont,J.W., and Seldin,M.F. (2004). Mexican American ancestry-informative markers: examination of population structure and marker characteristics in European Americans, Mexican Americans, Amerindians and Asians. Hum. Genet *114*, 263-271.

Crosslin,D.R., McDavid,A., Weston,N., Nelson,S.C., Zheng,X., Hart,E., de,A.M., Kullo,I.J., McCarty,C.A., Doheny,K.F., Pugh,E., Kho,A., Hayes,M.G., Pretel,S., Saip,A., Ritchie,M.D., Crawford,D.C., Crane,P.K., Newton,K., Li,R., Mirel,D.B., Crenshaw,A., Larson,E.B., Carlson,C.S., and Jarvik,G.P. (2011). Genetic variants associated with the white blood cell count in 13,923 subjects in the eMERGE Network. Hum. Genet.

Deeb,S.S., Fajas,L., Nemoto,M., Pihlajamaki,J., Mykkanen,L., Kuusisto,J., Laakso,M., Fujimoto,W., and Auwerx,J. (1998). A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. Nat Genet *20*, 284-287.

Denny,J.C., Crawford,D.C., Ritchie,M.D., Bielinski,S.J., Basford,M.A., Bradford,Y., Chai,H.S., Bastarache,L., Zuvich,R., Peissig,P., Carrell,D., Ramirez,A.H., Pathak,J., Wilke,R.A., Rasmussen,L., Wang,X., Pacheco,J.A., Kho,A.N., Hayes,M.G., Weston,N., Matsumoto,M., Kopp,P.A., Newton,K.M., Jarvik,G.P., Li,R., Manolio,T.A., Kullo,I.J., Chute,C.G., Chisholm,R.L., Larson,E.B., McCarty,C.A., Masys,D.R., Roden,D.M., and de,A.M. (2011). Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. Am. J. Hum. Genet. *89*, 529-542.

Denny,J.C., Ritchie,M.D., Crawford,D.C., Schildcrout,J.S., Ramirez,A.H., Pulley,J.M., Basford,M.A., Masys,D.R., Haines,J.L., and Roden,D.M. (2010). Identification of genomic predictors of atrioventricular conduction: using electronic medical records as a tool for genome science. Circulation *122*, 2016-2021.

Dina,C., Meyre,D., Gallina,S., Durand,E., Korner,A., Jacobson,P., Carlsson,L.M., Kiess,W., Vatin,V., Lecoeur,C., Delplanque,J., Vaillant,E., Pattou,F., Ruiz,J., Weill,J., Levy-Marchal,C., Horber,F., Potoczna,N., Hercberg,S., Le,S.C., Bougneres,P., Kovacs,P., Marre,M., Balkau,B., Cauchi,S., Chevre,J.C., and Froguel,P. (2007). Variation in FTO contributes to childhood obesity and severe adult obesity. Nat Genet *39*, 724-726.

Duggirala,R., Blangero,J., Almasy,L., Dyer,T.D., Williams,K.L., Leach,R.J., O'Connell,P., and Stern,M.P. (1999). Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. Am. J Hum. Genet. *64*, 1127-1140.

Dumitrescu,L., Ritchie,M.D., Brown-Gentry,K., Pulley,J.M., Basford,M., Denny,J.C., Oksenberg,J.R., Roden,D.M., Haines,J.L., and Crawford,D.C. (2010). Assessing the accuracy of observer-reported ancestry in a biorepository linked to electronic medical records. Genet. Med.

Dupuis,J., Langenberg,C., Prokopenko,I., Saxena,R., Soranzo,N., Jackson,A.U., Wheeler,E., Glazer,N.L., Bouatia-Naji,N., Gloyn,A.L., Lindgren,C.M., Magi,R., Morris,A.P., Randall,J., Johnson,T., Elliott,P., Rybin,D., Thorleifsson,G., Steinthorsdottir,V., Henneman,P., Grallert,H., Dehghan,A., Hottenga,J.J., Franklin,C.S., Navarro,P., Song,K., Goel,A., Perry,J.R., Egan,J.M., Lajunen,T., Grarup,N., Sparso,T., Doney,A., Voight,B.F., Stringham,H.M., Li,M., Kanoni,S., Shrader,P., Cavalcanti-Proenca,C., Kumari,M., Qi,L., Timpson,N.J., Gieger,C., Zabena,C., Rocheleau,G., Ingelsson,E., An,P., O'Connell,J., Luan,J., Elliott,A., McCarroll,S.A., Payne,F., Roccasecca,R.M., Pattou,F., Sethupathy,P., Ardlie,K., Ariyurek,Y., Balkau,B., Barter,P., Beilby,J.P., Ben-Shlomo,Y., Benediktsson,R., Bennett,A.J., Bergmann,S., Bochud,M., Boerwinkle,E., Bonnefond,A., Bonnycastle,L.L., Borch-Johnsen,K., Bottcher,Y., Brunner,E., Bumpstead,S.J., Charpentier,G., Chen,Y.D., Chines,P., Clarke,R., Coin,L.J., Cooper,M.N., Cornelis,M., Crawford,G., Crisponi,L., Day,I.N., de Geus,E.J., Delplanque,J., Dina,C., Erdos,M.R., Fedson,A.C., Fischer-Rosinsky,A., Forouhi,N.G., Fox,C.S., Frants,R., Franzosi,M.G., Galan,P., Goodarzi,M.O., Graessler,J., Groves,C.J., Grundy,S., Gwilliam,R., Gyllensten,U., Hadjadj,S., Hallmans,G., Hammond,N., Han,X., Hartikainen,A.L., Hassanali,N., Hayward,C., Heath,S.C., Hercberg,S., Herder,C., Hicks,A.A., Hillman,D.R., Hingorani,A.D., Hofman,A., Hui,J., Hung,J., Isomaa,B., Johnson,P.R., Jorgensen,T., Jula,A., Kaakinen,M., Kaprio,J., Kesaniemi,Y.A., Kivimaki,M., Knight,B., Koskinen,S., Kovacs,P., Kyvik,K.O., Lathrop,G.M., Lawlor,D.A., Le,B.O., Lecoeur,C., Li,Y., Lyssenko,V., Mahley,R., Mangino,M., Manning,A.K., Martinez-Larrad,M.T., McAteer,J.B., McCulloch,L.J., McPherson,R., Meisinger,C., Melzer,D., Meyre,D., Mitchell,B.D., Morken,M.A., Mukherjee,S., Naitza,S., Narisu,N., Neville,M.J., Oostra,B.A., Orru,M., Pakyz,R., Palmer,C.N., Paolisso,G., Pattaro,C., Pearson,D., Peden,J.F., Pedersen,N.L., Perola,M., Pfeiffer,A.F., Pichler,I., Polasek,O., Posthuma,D., Potter,S.C., Pouta,A., Province,M.A., Psaty,B.M., Rathmann,W., Rayner,N.W., Rice,K., Ripatti,S., Rivadeneira,F., Roden,M., Rolandsson,O., Sandbaek,A., Sandhu,M., Sanna,S., Sayer,A.A., Scheet,P., Scott,L.J., Seedorf,U., Sharp,S.J., Shields,B., Sigurethsson,G., Sijbrands,E.J., Silveira,A., Simpson,L., Singleton,A., Smith,N.L., Sovio,U., Swift,A., Syddall,H., Syvanen,A.C., Tanaka,T., Thorand,B., Tichet,J., Tonjes,A., Tuomi,T., Uitterlinden,A.G., van Dijk,K.W., van,H.M., Varma,D., Visvikis-Siest,S., Vitart,V., Vogelzangs,N., Waeber,G., Wagner,P.J., Walley,A., Walters,G.B., Ward,K.L., Watkins,H., Weedon,M.N., Wild,S.H., Willemsen,G., Witteman,J.C., Yarnell,J.W., Zeggini,E., Zelenika,D., Zethelius,B., Zhai,G., Zhao,J.H., Zillikens,M.C., Borecki,I.B., Loos,R.J., Meneton,P., Magnusson,P.K., Nathan,D.M., Williams,G.H., Hattersley,A.T., Silander,K., Salomaa,V., Smith,G.D., Bornstein,S.R., Schwarz,P., Spranger,J., Karpe,F., Shuldiner,A.R., Cooper,C., Dedoussis,G.V., Serrano-Rios,M., Morris,A.D., Lind,L., Palmer,L.J., Hu,F.B., Franks,P.W., Ebrahim,S., Marmot,M., Kao,W.H., Pankow,J.S., Sampson,M.J., Kuusisto,J., Laakso,M., Hansen,T., Pedersen,O., and Pramstaller,P.P. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet *42*, 105-116.

Franks,P.W., Rolandsson,O., Debenham,S.L., Fawcett,K.A., Payne,F., Dina,C., Froguel,P., Mohlke,K.L., Willer,C., Olsson,T., Wareham,N.J., Hallmans,G., Barroso,I., and Sandhu,M.S. (2008). Replication of the association between variants in WFS1 and risk of type 2 diabetes in European populations. Diabetologia *51*, 458-463.

Frayling,T.M., Timpson,N.J., Weedon,M.N., Zeggini,E., Freathy,R.M., Lindgren,C.M., Perry,J.R., Elliott,K.S., Lango,H., Rayner,N.W., Shields,B., Harries,L.W., Barrett,J.C., Ellard,S., Groves,C.J., Knight,B., Patch,A.M., Ness,A.R., Ebrahim,S., Lawlor,D.A., Ring,S.M., Ben-Shlomo,Y., Jarvelin,M.R., Sovio,U., Bennett,A.J., Melzer,D., Ferrucci,L., Loos,R.J., Barroso,I., Wareham,N.J., Karpe,F., Owen,K.R., Cardon,L.R., Walker,M., Hitman,G.A., Palmer,C.N., Doney,A.S., Morris,A.D., Smith,G.D., Hattersley,A.T., and 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*, 889-894.

Frayling,T.M. (2007). Genome-wide association studies provide new insights into type 2 diabetes aetiology. Nat Rev Genet *8*, 657-662.

Freedman,M.L., Haiman,C.A., Patterson,N., McDonald,G.J., Tandon,A., Waliszewska,A., Penney,K., Steen,R.G., Ardlie,K., John,E.M., Oakley-Girvan,I., Whittemore,A.S., Cooney,K.A., Ingles,S.A., Altshuler,D., Henderson,B.E., and Reich,D. (2006). Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc. Natl. Acad. Sci. U. S. A *103*, 14068-14073.

Gallego-Ortega,D., Ramirez de,M.A., Ramos,M.A., Valdes-Mora,F., Barderas,M.G., Sarmentero-Estrada,J., and Lacal,J.C. (2009). Differential role of human choline kinase alpha and beta enzymes in lipid metabolism: implications in cancer onset and treatment. PLoS. One. *4*, e7819.

Gauderman, W. J. and Morrison, J. M. QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies. 2006.
Ref Type: Computer Program

Gloyn,A.L., Pearson,E.R., Antcliff,J.F., Proks,P., Bruining,G.J., Slingerland,A.S., Howard,N., Srinivasan,S., Silva,J.M., Molnes,J., Edghill,E.L., Frayling,T.M., Temple,I.K., Mackay,D., Shield,J.P., Sumnik,Z., van,R.A., Wales,J.K., Clark,P., Gorman,S., Aisenberg,J., Ellard,S., Njolstad,P.R., Ashcroft,F.M., and Hattersley,A.T. (2004). Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. N. Engl. J Med. *350*, 1838-1849.

Gloyn,A.L., Weedon,M.N., Owen,K.R., Turner,M.J., Knight,B.A., Hitman,G., Walker,M., Levy,J.C., Sampson,M., Halford,S., McCarthy,M.I., Hattersley,A.T., and Frayling,T.M. (2003). Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. Diabetes *52*, 568-572.

Goldstein,B.J. and Müller-Wieland,D. (2008). Type 2 diabetes: Principles and Practice. CRC Press), p. 529.

Gower,B.A., Granger,W.M., Franklin,F., Shewchuk,R.M., and Goran,M.I. (2002). Contribution of insulin secretion and clearance to glucose-induced insulin concentration in african-american and caucasian children. J Clin. Endocrinol. Metab *87*, 2218-2224.

Grant,S.F., Thorleifsson,G., Reynisdottir,I., Benediktsson,R., Manolescu,A., Sainz,J., Helgason,A., Stefansson,H., Emilsson,V., Helgadottir,A., Styrkarsdottir,U., Magnusson,K.P., Walters,G.B., Palsdottir,E., Jonsdottir,T., Gudmundsdottir,T., Gylfason,A., Saemundsdottir,J., Wilensky,R.L., Reilly,M.P., Rader,D.J., Bagger,Y., Christiansen,C., Gudnason,V., Sigurdsson,G.,

Thorsteinsdottir,U., Gulcher,J.R., Kong,A., and Stefansson,K. (2006). Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat. Genet. *38*, 320-323.

Gudmundsson,J., Sulem,P., Steinthorsdottir,V., Bergthorsson,J.T., Thorleifsson,G., Manolescu,A., Rafnar,T., Gudbjartsson,D., Agnarsson,B.A., Baker,A., Sigurdsson,A., Benediktsdottir,K.R., Jakobsdottir,M., Blondal,T., Stacey,S.N., Helgason,A., Gunnarsdottir,S., Olafsdottir,A., Kristinsson,K.T., Birgisdottir,B., Ghosh,S., Thorlacius,S., Magnusdottir,D., Stefansdottir,G., Kristjansson,K., Bagger,Y., Wilensky,R.L., Reilly,M.P., Morris,A.D., Kimber,C.H., Adeyemo,A., Chen,Y., Zhou,J., So,W.Y., Tong,P.C., Ng,M.C., Hansen,T., Andersen,G., Borch-Johnsen,K., Jorgensen,T., Tres,A., Fuertes,F., Ruiz-Echarri,M., Asin,L., Saez,B., van,B.E., Klaver,S., Swinkels,D.W., Aben,K.K., Graif,T., Cashy,J., Suarez,B.K., van Vierssen,T.O., Frigge,M.L., Ober,C., Hofker,M.H., Wijmenga,C., Christiansen,C., Rader,D.J., Palmer,C.N., Rotimi,C., Chan,J.C., Pedersen,O., Sigurdsson,G., Benediktsson,R., Jonsson,E., Einarsson,G.V., Mayordomo,J.I., Catalona,W.J., Kiemeney,L.A., Barkardottir,R.B., Gulcher,J.R., Thorsteinsdottir,U., Kong,A., and Stefansson,K. (2007). Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat Genet *39*, 977-983.

Haiman,C.A., Fesinmeyer,M.D., Spencer,K.L., Buzkova,P., Voruganti,V.S., Wan,P., Haessler,J., Franceschini,N., Monroe,K.R., Howard,B.V., Jackson,R.D., Florez,J.C., Kolonel,L.N., Buyske,S., Goodloe,R.J., Liu,S., Manson,J.E., Meigs,J.B., Waters,K., Mukamal,K.J., Pendergrass,S.A., Shrader,P., Wilkens,L.R., Hindorff,L.A., Ambite,J.L., North,K.E., Peters,U., Crawford,D.C., Le,M.L., and Pankow,J.S. (2012). Consistent directions of effect for established type 2 diabetes risk variants across populations: the population architecture using Genomics and Epidemiology (PAGE) Consortium. Diabetes *61*, 1642-1647.

Haiman,C.A., Patterson,N., Freedman,M.L., Myers,S.R., Pike,M.C., Waliszewska,A., Neubauer,J., Tandon,A., Schirmer,C., McDonald,G.J., Greenway,S.C., Stram,D.O., Le,M.L., Kolonel,L.N., Frasco,M., Wong,D., Pooler,L.C., Ardlie,K., Oakley-Girvan,I., Whittemore,A.S., Cooney,K.A., John,E.M., Ingles,S.A., Altshuler,D., Henderson,B.E., and Reich,D. (2007). Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet *39*, 638-644.

Hani,E.H., Boutin,P., Durand,E., Inoue,H., Permutt,M.A., Velho,G., and Froguel,P. (1998). Missense mutations in the pancreatic islet beta cell inwardly rectifying K+ channel gene (KIR6.2/BIR): a meta-analysis suggests a role in the polygenic basis of Type II diabetes mellitus in Caucasians. Diabetologia *41*, 1511-1515.

Hanis,C.L., Boerwinkle,E., Chakraborty,R., Ellsworth,D.L., Concannon,P., Stirling,B., Morrison,V.A., Wapelhorst,B., Spielman,R.S., Gogolin-Ewens,K.J., Shepard,J.M., Williams,S.R., Risch,N., Hinds,D., Iwasaki,N., Ogata,M., Omori,Y., Petzold,C., Rietzch,H., Schroder,H.E., Schulze,J., Cox,N.J., Menzel,S., Boriraj,V.V., Chen,X., Lim,L.R., Lindner,T., Mereu,L.E., Wang,Y.Q., Xiang,K., Yamagata,K., Yang,Y., and Bell,G.I. (1996). A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. Nat. Genet. *13*, 161-166.

Helgadottir,A., Manolescu,A., Helgason,A., Thorleifsson,G., Thorsteinsdottir,U., Gudbjartsson,D.F., Gretarsdottir,S., Magnusson,K.P., Gudmundsson,G., Hicks,A., Jonsson,T., Grant,S.F., Sainz,J., O'Brien,S.J., Sveinbjornsdottir,S., Valdimarsson,E.M., Matthiasson,S.E., Levey,A.I., Abramson,J.L., Reilly,M.P., Vaccarino,V., Wolfe,M.L., Gudnason,V., Quyyumi,A.A., Topol,E.J., Rader,D.J., Thorgeirsson,G., Gulcher,J.R., Hakonarson,H., Kong,A.,

and Stefansson,K. (2006). A variant of the gene encoding leukotriene A4 hydrolase confers ethnicity-specific risk of myocardial infarction. Nat. Genet. *38*, 68-74.

Herder,C. and Roden,M. (2011). Genetics of type 2 diabetes: pathophysiologic and clinical relevance. Eur. J Clin. Invest *41*, 679-692.

Holmkvist,J., Almgren,P., Lyssenko,V., Lindgren,C.M., Eriksson,K.F., Isomaa,B., Tuomi,T., Nilsson,P., and Groop,L. (2008). Common variants in maturity-onset diabetes of the young genes and future risk of type 2 diabetes. Diabetes *57*, 1738-1744.

Holmkvist,J., Cervin,C., Lyssenko,V., Winckler,W., Anevski,D., Cilio,C., Almgren,P., Berglund,G., Nilsson,P., Tuomi,T., Lindgren,C.M., Altshuler,D., and Groop,L. (2006). Common variants in HNF-1 alpha and risk of type 2 diabetes. Diabetologia *49*, 2882-2891.

Hooper,W.C., Lally,C., Austin,H., Benson,J., Dilley,A., Wenger,N.K., Whitsett,C., Rawlins,P., and Evatt,B.L. (1999). The relationship between polymorphisms in the endothelial cell nitric oxide synthase gene and the platelet GPIIIa gene with myocardial infarction and venous thromboembolism in African Americans. Chest *116*, 880-886.

Horikawa,Y., Oda,N., Cox,N.J., Li,X., Orho-Melander,M., Hara,M., Hinokio,Y., Lindner,T.H., Mashima,H., Schwarz,P.E., del Bosque-Plata,L., Horikawa,Y., Oda,Y., Yoshiuchi,I., Colilla,S., Polonsky,K.S., Wei,S., Concannon,P., Iwasaki,N., Schulze,J., Baier,L.J., Bogardus,C., Groop,L., Boerwinkle,E., Hanis,C.L., and Bell,G.I. (2000). Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat. Genet. *26*, 163-175.

Hossain,P., Kawar,B., and El,N.M. (2007). Obesity and diabetes in the developing world--a growing challenge. N. Engl. J Med. *356*, 213-215.

Kao,W.H., Klag,M.J., Meoni,L.A., Reich,D., Berthier-Schaad,Y., Li,M., Coresh,J., Patterson,N., Tandon,A., Powe,N.R., Fink,N.E., Sadler,J.H., Weir,M.R., Abboud,H.E., Adler,S.G., Divers,J., Iyengar,S.K., Freedman,B.I., Kimmel,P.L., Knowler,W.C., Kohn,O.F., Kramp,K., Leehey,D.J., Nicholas,S.B., Pahl,M.V., Schelling,J.R., Sedor,J.R., Thornley-Brown,D., Winkler,C.A., Smith,M.W., and Parekh,R.S. (2008). MYH9 is associated with nondiabetic end-stage renal disease in African Americans. Nat Genet *40*, 1185-1192.

Kaprio,J., Tuomilehto,J., Koskenvuo,M., Romanov,K., Reunanen,A., Eriksson,J., Stengard,J., and Kesaniemi,Y.A. (1992). Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. Diabetologia *35*, 1060-1067.

Köbberling,J.T.H. (1982). Empirical risk figures for first-degree relatives of non-insulin dependent diabetics. The genetics of diabetes mellitus.

Lindholm,E., Ekholm,B., Shaw,S., Jalonen,P., Johansson,G., Pettersson,U., Sherrington,R., Adolfsson,R., and Jazin,E. (2001). A schizophrenia-susceptibility locus at 6q25, in one of the world's largest reported pedigrees. Am. J Hum. Genet. *69*, 96-105.

Lyssenko,V., Lupi,R., Marchetti,P., Del,G.S., Orho-Melander,M., Almgren,P., Sjogren,M., Ling,C., Eriksson,K.F., Lethagen,A.L., Mancarella,R., Berglund,G., Tuomi,T., Nilsson,P., Del,P.S., and Groop,L. (2007). Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. J Clin. Invest *117*, 2155-2163.

Madrigal,L., Ware,B., Miller,R., Saenz,G., Chavez,M., and Dykes,D. (2001). Ethnicity, gene flow, and population subdivision in Limon, Costa Rica. Am. J Phys. Anthropol. *114*, 99-108.

Manolio,T.A., Collins,F.S., Cox,N.J., Goldstein,D.B., Hindorff,L.A., Hunter,D.J., McCarthy,M.I., Ramos,E.M., Cardon,L.R., Chakravarti,A., Cho,J.H., Guttmacher,A.E., Kong,A., Kruglyak,L., Mardis,E., Rotimi,C.N., Slatkin,M., Valle,D., Whittemore,A.S., Boehnke,M., Clark,A.G., Eichler,E.E., Gibson,G., Haines,J.L., Mackay,T.F., McCarroll,S.A., and Visscher,P.M. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747-753.

Maskarinec,G., Grandinetti,A., Matsuura,G., Sharma,S., Mau,M., Henderson,B.E., and Kolonel,L.N. (2009). Diabetes prevalence and body mass index differ by ethnicity: the Multiethnic Cohort. Ethn. Dis. *19*, 49-55.

McCarty,C.A., Chisholm,R.L., Chute,C.G., Kullo,I.J., Jarvik,G.P., Larson,E.B., Li,R., Masys,D.R., Ritchie,M.D., Roden,D.M., Struewing,J.P., and Wolf,W.A. (2011). The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC. Med. Genomics *4*, 13.

Minton,J.A., Hattersley,A.T., Owen,K., McCarthy,M.I., Walker,M., Latif,F., Barrett,T., and Frayling,T.M. (2002). Association studies of genetic variation in the WFS1 gene and type 2 diabetes in U.K. populations. Diabetes *51*, 1287-1290.

Mokdad,A.H., Ford,E.S., Bowman,B.A., Nelson,D.E., Engelgau,M.M., Vinicor,F., and Marks,J.S. (2000). Diabetes trends in the U.S.: 1990-1998. Diabetes Care *23*, 1278-1283.

Morino,K., Petersen,K.F., and Shulman,G.I. (2006). Molecular mechanisms of insulin resistance in humans and their potential links with mitochondrial dysfunction. Diabetes *55 Suppl 2*, S9-S15.

Mozaffarian,D., Kamineni,A., Carnethon,M., Djousse,L., Mukamal,K.J., and Siscovick,D. (2009). Lifestyle risk factors and new-onset diabetes mellitus in older adults: the cardiovascular health study. Arch. Intern. Med *169*, 798-807.

National Institute of Diabetes and Kidney Disease. National Diabetes Information: Prediabetes, What You Need to Know. 2007. National Institues of Health of the U.S. Department of Health and Human Services.
Ref Type: Report

Nettleton,J.A., McKeown,N.M., Kanoni,S., Lemaitre,R.N., Hivert,M.F., Ngwa,J., van Rooij,F.J., Sonestedt,E., Wojczynski,M.K., Ye,Z., Tanaka,T., Garcia,M., Anderson,J.S., Follis,J.L., Djousse,L., Mukamal,K., Papoutsakis,C., Mozaffarian,D., Zillikens,M.C., Bandinelli,S., Bennett,A.J., Borecki,I.B., Feitosa,M.F., Ferrucci,L., Forouhi,N.G., Groves,C.J., Hallmans,G., Harris,T., Hofman,A., Houston,D.K., Hu,F.B., Johansson,I., Kritchevsky,S.B., Langenberg,C., Launer,L., Liu,Y., Loos,R.J., Nalls,M., Orho-Melander,M., Renstrom,F., Rice,K., Riserus,U., Rolandsson,O., Rotter,J.I., Saylor,G., Sijbrands,E.J., Sjogren,P., Smith,A., Steingrimsdottir,L., Uitterlinden,A.G., Wareham,N.J., Prokopenko,I., Pankow,J.S., van Duijn,C.M., Florez,J.C., Witteman,J.C., Dupuis,J., Dedoussis,G.V., Ordovas,J.M., Ingelsson,E., Cupples,L.A., Siscovick,D.S., Franks,P.W., and Meigs,J.B. (2010). Interactions of dietary whole-grain intake with fasting glucose- and insulin-related genetic loci in individuals of European descent: a meta-analysis of 14 cohort studies. Diabetes Care *33*, 2684-2691.

Newman,B., Selby,J.V., King,M.C., Slemenda,C., Fabsitz,R., and Friedman,G.D. (1987). Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. Diabetologia *30*, 763-768.

Norris,J.M., Langefeld,C.D., Scherzinger,A.L., Rich,S.S., Bookman,E., Beck,S.R., Saad,M.F., Haffner,S.M., Bergman,R.N., Bowden,D.W., and Wagenknecht,L.E. (2005). Quantitative trait loci for abdominal fat and BMI in Hispanic-Americans and African-Americans: the IRAS Family study. Int. J Obes. (Lond) *29*, 67-77.

Patterson,N., Hattangadi,N., Lane,B., Lohmueller,K.E., Hafler,D.A., Oksenberg,J.R., Hauser,S.L., Smith,M.W., O'Brien,S.J., Altshuler,D., Daly,M.J., and Reich,D. (2004). Methods for high-density admixture mapping of disease genes. Am. J Hum. Genet. *74*, 979-1000.

Pendergrass, S. A., Dudek, S., Crawford, D. C. C., and Ritchie, M. Synthesis-View: visualization and interpretation of SNP association results for multi-cohort, multi-phenotype data and meta-analysis. BioData Mining . 2010.
Ref Type: In Press

Pickup,J.C. (2004). Inflammation and activated innate immunity in the pathogenesis of type 2 diabetes. Diabetes Care *27*, 813-823.

Plant,L.D., Bowers,P.N., Liu,Q., Morgan,T., Zhang,T., State MW, Chen,W., Kittles,R.A., and Goldstein,S.A. (2006). A common cardiac sodium channel variant associated with sudden infant death in African Americans, SCN5A S1103Y. J Clin. Invest *116*, 430-435.

Price,A.L., Tandon,A., Patterson,N., Barnes,K.C., Rafaels,N., Ruczinski,I., Beaty,T.H., Mathias,R., Reich,D., and Myers,S. (2009). Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. PLoS. Genet *5*, e1000519.

Purcell,S., Neale,B., Todd-Brown,K., Thomas,L., Ferreira,M.A., Bender,D., Maller,J., Sklar,P., de Bakker,P.I., Daly,M.J., and Sham,P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J Hum. Genet. *81*, 559-575.

Ramirez de,M.A., Gallego-Ortega,D., Sarmentero-Estrada,J., Lagares,D., Gomez Del,P.T., Bandres,E., Garcia-Foncillas,J., and Lacal,J.C. (2008). Choline kinase as a link connecting phospholipid metabolism and cell cycle regulation: implications in cancer therapy. Int. J Biochem. Cell Biol *40*, 1753-1763.

Reich,D., Nalls,M.A., Kao,W.H., Akylbekova,E.L., Tandon,A., Patterson,N., Mullikin,J., Hsueh,W.C., Cheng,C.Y., Coresh,J., Boerwinkle,E., Li,M., Waliszewska,A., Neubauer,J., Li,R., Leak,T.S., Ekunwe,L., Files,J.C., Hardy,C.L., Zmuda,J.M., Taylor,H.A., Ziv,E., Harris,T.B., and Wilson,J.G. (2009). Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. PLoS. Genet *5*, e1000360.

Reiner,A.P., Ziv,E., Lind,D.L., Nievergelt,C.M., Schork,N.J., Cummings,S.R., Phong,A., Burchard,E.G., Harris,T.B., Psaty,B.M., and Kwok,P.Y. (2005). Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. Am. J Hum. Genet *76*, 463-477.

Reynisdottir,I., Thorleifsson,G., Benediktsson,R., Sigurdsson,G., Emilsson,V., Einarsdottir,A.S., Hjorleifsdottir,E.E., Orlygsdottir,G.T., Bjornsdottir,G.T., Saemundsdottir,J., Halldorsson,S.,

Hrafnkelsdottir,S., Sigurjonsdottir,S.B., Steinsdottir,S., Martin,M., Kochan,J.P., Rhees,B.K., Grant,S.F., Frigge,M.L., Kong,A., Gudnason,V., Stefansson,K., and Gulcher,J.R. (2003). Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34-q35.2. Am. J Hum. Genet. *73*, 323-335.

RIFE,D.C. (1954). Populations of hybrid origin as source material for the detection of linkage. Am. J Hum. Genet *6*, 26-33.

Roden,D.M., Pulley,J.M., Basford,M.A., Bernard,G.R., Clayton,E.W., Balser,J.R., and Masys,D.R. (2008). Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin. Pharmacol. Ther. *84*, 362-369.

Rosenberg,N.A., Huang,L., Jewett,E.M., Szpiech,Z.A., Jankovic,I., and Boehnke,M. (2010). Genome-wide association studies in diverse populations. Nat. Rev. Genet. *11*, 356-366.

Rung,J., Cauchi,S., Albrechtsen,A., Shen,L., Rocheleau,G., Cavalcanti-Proenca,C., Bacot,F., Balkau,B., Belisle,A., Borch-Johnsen,K., Charpentier,G., Dina,C., Durand,E., Elliott,P., Hadjadj,S., Jarvelin,M.R., Laitinen,J., Lauritzen,T., Marre,M., Mazur,A., Meyre,D., Montpetit,A., Pisinger,C., Posner,B., Poulsen,P., Pouta,A., Prentki,M., Ribel-Madsen,R., Ruokonen,A., Sandbaek,A., Serre,D., Tichet,J., Vaxillaire,M., Wojtaszewski,J.F., Vaag,A., Hansen,T., Polychronakos,C., Pedersen,O., Froguel,P., and Sladek,R. (2009). Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. Nat Genet *41*, 1110-1115.

Sandhu,M.S., Weedon,M.N., Fawcett,K.A., Wasson,J., Debenham,S.L., Daly,A., Lango,H., Frayling,T.M., Neumann,R.J., Sherva,R., Blech,I., Pharoah,P.D., Palmer,C.N., Kimber,C., Tavendale,R., Morris,A.D., McCarthy,M.I., Walker,M., Hitman,G., Glaser,B., Permutt,M.A., Hattersley,A.T., Wareham,N.J., and Barroso,I. (2007). Common variants in WFS1 confer risk of type 2 diabetes. Nat Genet *39*, 951-953.

Saxena,R., Voight,B.F., Lyssenko,V., Burtt,N.P., de Bakker,P.I., Chen,H., Roix,J.J., Kathiresan,S., Hirschhorn,J.N., Daly,M.J., Hughes,T.E., Groop,L., Altshuler,D., Almgren,P., Florez,J.C., Meyer,J., Ardlie,K., Bengtsson,B.K., Isomaa,B., Lettre,G., Lindblad,U., Lyon,H.N., Melander,O., Newton-Cheh,C., Nilsson,P., Orho-Melander,M., Rastam,L., Speliotes,E.K., Taskinen,M.R., Tuomi,T., Guiducci,C., Berglund,A., Carlson,J., Gianniny,L., Hackett,R., Hall,L., Holmkvist,J., Laurila,E., Sjogren,M., Sterner,M., Surti,A., Svensson,M., Svensson,M., Tewhey,R., Blumenstiel,B., Parkin,M., Defelice,M., Barry,R., Brodeur,W., Camarata,J., Chia,N., Fava,M., Gibbons,J., Handsaker,B., Healy,C., Nguyen,K., Gates,C., Sougnez,C., Gage,D., Nizzari,M., Gabriel,S.B., Chirn,G.W., Ma,Q., Parikh,H., Richardson,D., Ricke,D., and Purcell,S. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science *316*, 1331-1336.

Schinner,S., Scherbaum,W.A., Bornstein,S.R., and Barthel,A. (2005). Molecular mechanisms of insulin resistance. Diabet. Med. *22*, 674-682.

Scott,L.J., Mohlke,K.L., Bonnycastle,L.L., Willer,C.J., Li,Y., Duren,W.L., Erdos,M.R., Stringham,H.M., Chines,P.S., Jackson,A.U., Prokunina-Olsson,L., Ding,C.J., Swift,A.J., Narisu,N., Hu,T., Pruim,R., Xiao,R., Li,X.Y., Conneely,K.N., Riebow,N.L., Sprau,A.G., Tong,M., White,P.P., Hetrick,K.N., Barnhart,M.W., Bark,C.W., Goldstein,J.L., Watkins,L., Xiang,F., Saramies,J., Buchanan,T.A., Watanabe,R.M., Valle,T.T., Kinnunen,L., Abecasis,G.R., Pugh,E.W., Doheny,K.F., Bergman,R.N., Tuomilehto,J., Collins,F.S., and Boehnke,M. (2007). A

genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science *316*, 1341-1345.

Scuteri,A., Sanna,S., Chen,W.M., Uda,M., Albai,G., Strait,J., Najjar,S., Nagaraja,R., Orru,M., Usala,G., Dei,M., Lai,S., Maschio,A., Busonero,F., Mulas,A., Ehret,G.B., Fink,A.A., Weder,A.B., Cooper,R.S., Galan,P., Chakravarti,A., Schlessinger,D., Cao,A., Lakatta,E., and Abecasis,G.R. (2007). Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. PLoS. Genet *3*, e115.

Sladek,R., Rocheleau,G., Rung,J., Dina,C., Shen,L., Serre,D., Boutin,P., Vincent,D., Belisle,A., Hadjadj,S., Balkau,B., Heude,B., Charpentier,G., Hudson,T.J., Montpetit,A., Pshezhetsky,A.V., Prentki,M., Posner,B.I., Balding,D.J., Meyre,D., Polychronakos,C., and Froguel,P. (2007). A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature *445*, 881-885.

Spencer,C.C., Su,Z., Donnelly,P., and Marchini,J. (2009). Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. PLoS. Genet. *5*, e1000477.

Stephens,J.C., Briscoe,D., and O'Brien,S.J. (1994). Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. Am. J Hum. Genet *55*, 809-824.

Tabeta,K., Hoebe,K., Janssen,E.M., Du,X., Georgel,P., Crozat,K., Mudd,S., Mann,N., Sovath,S., Goode,J., Shamel,L., Herskovits,A.A., Portnoy,D.A., Cooke,M., Tarantino,L.M., Wiltshire,T., Steinberg,B.E., Grinstein,S., and Beutler,B. (2006). The Unc93b1 mutation 3d disrupts exogenous antigen presentation and signaling via Toll-like receptors 3, 7 and 9. Nat Immunol *7*, 156-164.

Tandon,A., Patterson,N., and Reich,D. (2011). Ancestry informative marker panels for African Americans based on subsets of commercially available SNP arrays. Genet. Epidemiol. *35*, 80-83.

Teo,Y.Y., Small,K.S., Fry,A.E., Wu,Y., Kwiatkowski,D.P., and Clark,T.G. (2009). Power consequences of linkage disequilibrium variation between populations. Genet. Epidemiol. *33*, 128-135.

Teo,Y.Y., Small,K.S., and Kwiatkowski,D.P. (2010). Methodological challenges of genome-wide association analysis in Africa. Nat. Rev. Genet. *11*, 149-160.

The International HapMap Project (2003). The International HapMap Project. Nature *426*, 789-796.

Tong,Y., Lin,Y., Zhang,Y., Yang,J., Zhang,Y., Liu,H., and Zhang,B. (2009). Association between TCF7L2 gene polymorphisms and susceptibility to type 2 diabetes mellitus: a large Human Genome Epidemiology (HuGE) review and meta-analysis. BMC. Med Genet. *10*, 15.

Turner,S.D., Berg,R.L., Linneman,J.G., Peissig,P.L., Crawford,D.C., Denny,J.C., Roden,D.M., McCarty,C.A., Ritchie,M.D., and Wilke,R.A. (2011). Knowledge-driven multi-locus analysis reveals gene-gene interactions influencing HDL cholesterol level in two independent EMR-linked biobanks. PLoS. One. *6*, e19586.

van Dam,R.M., Hoebee,B., Seidell,J.C., Schaap,M.M., de Bruin,T.W., and Feskens,E.J. (2005). Common variants in the ATP-sensitive K+ channel genes KCNJ11 (Kir6.2) and ABCC8 (SUR1)

in relation to glucose intolerance: population-based studies and meta-analyses. Diabet. Med. *22*, 590-598.

Van Norstrand,D.W., Tester,D.J., and Ackerman,M.J. (2008). Overrepresentation of the proarrhythmic, sudden death predisposing sodium channel polymorphism S1103Y in a population-based cohort of African-American sudden infant death syndrome. Heart Rhythm. *5*, 712-715.

Vimaleswaran,K.S. and Loos,R.J. (2010). Progress in the genetics of common obesity and type 2 diabetes. Expert. Rev Mol. Med. *12*, e7.

Voight,B.F., Scott,L.J., Steinthorsdottir,V., Morris,A.P., Dina,C., Welch,R.P., Zeggini,E., Huth,C., Aulchenko,Y.S., Thorleifsson,G., McCulloch,L.J., Ferreira,T., Grallert,H., Amin,N., Wu,G., Willer,C.J., Raychaudhuri,S., McCarroll,S.A., Langenberg,C., Hofmann,O.M., Dupuis,J., Qi,L., Segre,A.V., van,H.M., Navarro,P., Ardlie,K., Balkau,B., Benediktsson,R., Bennett,A.J., Blagieva,R., Boerwinkle,E., Bonnycastle,L.L., Bengtsson,B.K., Bravenboer,B., Bumpstead,S., Burtt,N.P., Charpentier,G., Chines,P.S., Cornelis,M., Couper,D.J., Crawford,G., Doney,A.S., Elliott,K.S., Elliott,A.L., Erdos,M.R., Fox,C.S., Franklin,C.S., Ganser,M., Gieger,C., Grarup,N., Green,T., Griffin,S., Groves,C.J., Guiducci,C., Hadjadj,S., Hassanali,N., Herder,C., Isomaa,B., Jackson,A.U., Johnson,P.R., Jorgensen,T., Kao,W.H., Klopp,N., Kong,A., Kraft,P., Kuusisto,J., Lauritzen,T., Li,M., Lieverse,A., Lindgren,C.M., Lyssenko,V., Marre,M., Meitinger,T., Midthjell,K., Morken,M.A., Narisu,N., Nilsson,P., Owen,K.R., Payne,F., Perry,J.R., Petersen,A.K., Platou,C., Proenca,C., Prokopenko,I., Rathmann,W., Rayner,N.W., Robertson,N.R., Rocheleau,G., Roden,M., Sampson,M.J., Saxena,R., Shields,B.M., Shrader,P., Sigurdsson,G., Sparso,T., Strassburger,K., Stringham,H.M., Sun,Q., Swift,A.J., Thorand,B., Tichet,J., Tuomi,T., van Dam,R.M., van Haeften,T.W., van,H.T., van Vliet-Ostaptchouk,J.V., Walters,G.B., Weedon,M.N., Wijmenga,C., Witteman,J., Bergman,R.N., Cauchi,S., Collins,F.S., Gloyn,A.L., Gyllensten,U., Hansen,T., Hide,W.A., Hitman,G.A., Hofman,A., Hunter,D.J., Hveem,K., Laakso,M., Mohlke,K.L., Morris,A.D., Palmer,C.N., Pramstaller,P.P., Rudan,I., Sijbrands,E., Stein,L.D., Tuomilehto,J., Uitterlinden,A., Walker,M., Wareham,N.J., Watanabe,R.M., Abecasis,G.R., Boehm,B.O., Campbell,H., Daly,M.J., Hattersley,A.T., Hu,F.B., Meigs,J.B., Pankow,J.S., Pedersen,O., Wichmann,H.E., Barroso,I., Florez,J.C., Frayling,T.M., Groop,L., Sladek,R., Thorsteinsdottir,U., Wilson,J.F., Illig,T., Froguel,P., van Duijn,C.M., Stefansson,K., Altshuler,D., Boehnke,M., and McCarthy,M.I. (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat. Genet. *42*, 579-589.

WHO. World Health Organization Fact Sheet. 2011.
Ref Type: Report

Wild,S., Roglic,G., Green,A., Sicree,R., and King,H. (2004). Global Prevalence of Diabetes:Estimates for the Year 2000 and Projections for 2030. Diabetes Care *29*, 2569.

Winckler,W., Weedon,M.N., Graham,R.R., McCarroll,S.A., Purcell,S., Almgren,P., Tuomi,T., Gaudet,D., Bostrom,K.B., Walker,M., Hitman,G., Hattersley,A.T., McCarthy,M.I., Ardlie,K.G., Hirschhorn,J.N., Daly,M.J., Frayling,T.M., Groop,L., and Altshuler,D. (2007). Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. Diabetes *56*, 685-693.

Winkler,C.A., Nelson,G.W., and Smith,M.W. (2010). Admixture mapping comes of age. Annu. Rev Genomics Hum. Genet *11*, 65-89.

WTCC (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661-678.

Yaeger,R., vila-Bront,A., Abdul,K., Nolan,P.C., Grann,V.R., Birchette,M.G., Choudhry,S., Burchard,E.G., Beckman,K.B., Gorroochurn,P., Ziv,E., Consedine,N.S., and Joe,A.K. (2008). Comparing genetic ancestry and self-described race in african americans born in the United States and in Africa. Cancer Epidemiol. Biomarkers Prev. *17*, 1329-1338.

Zeggini,E., Weedon,M.N., Lindgren,C.M., Frayling,T.M., Elliott,K.S., Lango,H., Timpson,N.J., Perry,J.R., Rayner,N.W., Freathy,R.M., Barrett,J.C., Shields,B., Morris,A.P., Ellard,S., Groves,C.J., Harries,L.W., Marchini,J.L., Owen,K.R., Knight,B., Cardon,L.R., Walker,M., Hitman,G.A., Morris,A.D., Doney,A.S., McCarthy,M.I., and Hattersley,A.T. (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science *316*, 1336-1341.

Zuvich,R.L., Armstrong,L.L., Bielinski,S.J., Bradford,Y., Carlson,C.S., Crawford,D.C., Crenshaw,A.T., de,A.M., Doheny,K.F., Haines,J.L., Hayes,M.G., Jarvik,G.P., Jiang,L., Kullo,I.J., Li,R., Ling,H., Manolio,T.A., Matsumoto,M.E., McCarty,C.A., McDavid,A.N., Mirel,D.B., Olson,L.M., Paschall,J.E., Pugh,E.W., Rasmussen,L.V., Rasmussen-Torvik,L.J., Turner,S.D., Wilke,R.A., and Ritchie,M.D. (2011). Pitfalls of merging GWAS data: lessons learned in the eMERGE network and quality control procedures to maintain high data quality. Genet. Epidemiol. *35*, 887-898.