Genome-wide Association Study to Identify Single Nucleotide Polymorphisms Associated with Diabetic Nephropathy and estimated Glomerular Filtration Rate in FIND Study

By

Mohammad Hasanzadehkiabi

A Thesis Submitted to the

School of Graduate Studies

In partial fulfillment of the requirements for the degree of

Master of Science in Medicine

Department of Epidemiology/ Faculty of Medicine

Memorial University of Newfoundland

October 2016

Abstract

Diabetes mellitus is characterized by a chronic increase in blood glucose levels due to a dysfunction of carbohydrate, fat and protein metabolism, which in turn are the result of insulin resistance and/or insulin action disturbances. Its cardinal symptoms include polyuria, thirst and weight loss; its long term complications include neuropathy, nephropathy, and retinopathy. It is predicted that this rate will increase to 9.9% of the adult population by 2030. Although the rate of new cases of End Stage Renal Disease (ESRD) in 2011 decreased 4.2% compared to the year 2010, nephropathy due to diabetes continued to be the most common cause of ESRD in the USA. In this study a genome-wide association approach has been performed to identify loci associated with diabetic nephropathy (DN) and variation in estimated Glomerular Filtration Rate (eGFR) value in three distinct ethnic groups, i.e. African American, Mexican American and European American; additionally, the effect of sex has also been studied to reveal potential sex dependency on the loci associated with DN and eGFR value. Dataset provided in "The Family Investigation of Nephropathy and Diabetes" (FIND) Study, contains genotype of 1454 individuals with DN and 1168 individuals without DN .After conducting genome-wide association analyses, in this study, the strongest association with DN and/or eGFR values have been detected on the regions located on the 2p, 2q, 3q, 4q, 5q, 7q, 8q, 9q, 12q, 14q, 15q and Xq chromosomes.

Acknowledgments

I would like to thank my supervisor, Dr. Saeed Samet, for the opportunity he has provided for me to be able to accomplish this study and for all of his guidance and support through this study, both for the parts that worked well and those issues that did not in the way predicted, as well as for all of the knowledge and advice that he provided to me. I would also like to thank each of my supervisor committee members, Dr. Proton Rahman, and Dr. Gerard Farrell for their encouragement, guidance and suggestions. Lecture notes of a Genetic Epidemiology (MED6395) course offered at the winter semester of 2014, chaired by Dr. Guangiu Zhai, at MUN was my general guidance for preparing this study. Additionally, I would like to thank Dr. Guangiu Zhai for his continued guidance and support when some problems were not clear to me regarding the genome-wide association study in general. I would also like to thank Dr. J. Concepcion Loredo-Osti for all his comments and suggestions, and especially for his patience in answering my questions. I would also like to thank all the members of the FIND study and NIH for all the hard work they have done so far to provide a valuable opportunity for conducting further investigations based on the data collected by them, particularly Dr. William Knowler (Chief, Diabetes Epidemiology and Clinical Research Section, National Institute of Diabetes and Digestive and Kidney Diseases), Dr. Sudha Iyengar (director of the FIND coordinating center), and Dr. Barry Freedman (chair of the FIND steering committee) for their comments on my questions. In the literature review I read many papers and articles related to my study; however, the dissertation of Dr. Caitrin Wheeler

McDonough (McDonough 2010) for a PhD degree and specifically its reference list were very useful in preparing this script.

"The Family Investigation of Nephropathy and Diabetes (FIND) Study was conducted by the FIND Investigators and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). This document was not prepared in collaboration with Investigators of the Family Investigation of Nephropathy and Diabetes (FIND) study and does not necessarily reflect the opinions or views of the FIND study or the NIDDK. The datasets used for the analyses described in this certification were obtained from dbGaP at [http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=request_details_pi&filter=arid &arid=57176] through dbGaP access number [request # 24469-4] and [project #5841]".

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AA	African American
ACE	Angiotensin Converting Enzyme
ADA	American Diabetes Association
ADPRH	ADP-ribosylarginine hydrolase
ADTRP	androgen-dependent TFPI-regulating protein
AI	American Indian
Ala	Alanine
ALDR1	Aldose Reductase
ARIC	Atherosclerosis Risk In Communities
ASPs	Affected Sib-Pairs
BETA	Regression coefficient
BMI	Body Mass Index
BONF	Bonferroni single-step adjusted p-values
BP	Base-pair position
CAPN10	Calpain-10
CCNG1	cyclin G1
CD80	CD80 molecule
CDCV	Common-Disease Common-Variant
CHR	Chromosome
CLRN1	clarin 1
CMV	Cytomegalovirus
CPQ	carboxypeptidase Q
CR	Calorie Restriction
CRI	Chronic Renal Insufficiency
CTLA4	Cytotoxic T Lymphocyte Antigen 4
CVB	Coxsackievirus B
CVD	Cardiovascular Disease
DIAPH2	diaphanous related formin 2
DM	Diabetes Mellitus
DN	Diabetic Nephropathy
Dock180	Dedicator of Cytokinesis 180 kDa
DR	Diabetic Retinopathy
DZ	Dizygotic
EA	European American
ECM	Extracellular Matrices
eGFR	estimated Glomerular Filtration Rate
ELMO1	Engulfment and Cell Motility 1

eNOS	Endothelial Nitric Oxide Synthase
EPHB1	EPH receptor B1
ERBB4	v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 4
ESRD	End Stage Renal Disease
FDR	False Discovery Rate
FDR_BH	Benjamini & Hochberg (1995) step-up FDR control
FDR_BY	Benjamini & Yekutieli (2001) step-up FDR control
FERM	4.1 protein Ezrin, Radixin, Moesin
FERM3	FERM domain containing 3
FIND	the Family Investigation of Nephrology and Diabetes
FWER	Family-Wise Error Rate
GC	Genomic-control corrected p-values
GFR	Glomerular Filtration Rate
HDL	High Density Lipoprotein
HLA	Human Leukocyte Antigen
HMMR	hyaluronan-mediated motility receptor
HOLM	Holm (1979) step-down adjusted p-values
HWE	Hardy-Weinberg Equilibrium
IBD	identity-by-descent
IBS	identity-by-state
IDDM	Insulin Dependent Diabetes Mellitus
IDDM2	Insulin Dependent Diabetes Mellitus 2
IFG	Impaired Fasting Glucose
IGT	Impaired Glucose Tolerance
IL-1Ra	Interleukin 1 Receptor antagonist
IL-1β	Interleukin 1 β
IL-2RA	Interleukin 2 Receptor Alpha
IL-6	Interleukin 6
ILK	Integrin-Linked Kinase
INPP5D	inositol polyphosphate-5-phosphatase D
IRS	Insulin Receptor Substrate
IRS1	insulin receptor substrate 1
JHS	Jackson Heart Study
KCNJ6	potassium voltage-gated channel subfamily J member 6
LBW	Lower Birth Weight
LD	Linkage Disequilibrium
LINC01481	long intergenic non-protein coding RNA 1481
LOD	Logarithm of Odds

MA	Mexican American
MALD	Mapping by Admixture Linkage Disequilibrium
MAN1C1	mannosidase, alpha, class 1C, a member 1
MEC	Multiethnic Cohort
MHC	Major Histocompatibility Complex
Mn-SOD	Manganese Superoxide Dismutase
MODY	Maturity Onset Diabetes of Youth
MTCYBP27	mitochondrially encoded cytochrome b pseudogene 27
MTRNR2L7	MT-RNR2-like 7
MZ	Monozygotic
NAD	Nicotinamide Adenine Dinucleotide
NAM	Nicotinamide
Nampt	Nicotinamide Phosphoribosyl Transferase
NEFA	Non-Esterified Fatty Acids
NIDDM	Non-Insulin Dependent Diabetes Mellitus
NIH	National Institute of Health
NMISS	Number of non-missing individuals for this analysis
NMN	Nicotinamide Mononucleotide
NOS3	Nitric Oxide Synthase 3
NPAS3	neuronal PAS domain protein 3
NPL	Non-Parametric Linkage
NUP62CL	nucleoporin 62kDa C-terminal like
NXT2	nuclear transport factor 2-like export factor 2
OPCs	Organophosphorus Compounds
OSA	Ordered Subset Analysis
P	Asymptotic significance value for coefficient
PC1	Plasma Cell-1
PCA	Principal Component Analysis
POPs	Persistent Organic Pollutants
PPARγ	Peroxisome Proliferator-Activated Receptor gamma
Pro	Proline
PTPN22	Protein Tyrosine Phosphatase Nonreceptor type 22
Q-Q plot	Quantile-Quantile plot
R2	The regression r-squared (multiple correlation coefficient)
RBM41	RNA binding motif protein 41
RELN	reelin
RNU6-976P	RNA, U6 small nuclear 976, pseudogene
RORA	RAR related orphan receptor A

RPL23AP54	ribosomal protein L23a pseudogene 54
RR	Relative Risk
SASH1	SAM and SH3 domain containing 1
SDC2	syndecan 2
SE	Standard error of the coefficient
SEPT6	septin 6
SIDAK_SD	Sidak step-down adjusted p-values
SIDAK_SS	Sidak single-step adjusted p-values
SIR2	Silent Information Regulator-2
SIRT1, 2, …, 7	Sirtuin 1, 2, …, 7
SLC35G2	solute carrier family 35 member G2
SLC8A1-AS1	SLC8A1 antisense RNA 1
SNHG18	small nucleolar RNA host gene 18
SNP	Single Nucleotide Polymorphism
SOWAHD	sosondowah ankyrin repeat domain family member D
STAG1	stromal antigen 1
SULT1C2	sulfotransferase family 1C member 2
SULT1C3	sulfotransferase family 1C member 3
T	t-statistic for regression of phenotype on allele count
T1D	Type 1 Diabetes
T2D	Type 2 Diabetes
TAS2R1	taste 2 receptor member 1
TCF7L2	Transcription Factor 7-Like 2
TCRBC	T-cell Receptor β-chain
TDT	Transmission Disequilibrium Test
TMEM212	transmembrane protein 212
TMPRSS15	transmembrane protease, serine 15
TNF	Tumor Necrosis Factor
TNFα	Tumor Necrosis Factor-α
UAE	Urinary Albumin Excretion
UNADJ	Unadjusted p-value
WHO	World Health Organization
ZNF	zinc finger
ZNF236	Zinc-Finger gene 236
ZNF25	zinc finger protein 25
ZNF33A	zinc finger protein 33A
ZNF33CP	zinc finger protein 33C, pseudogene

Chapter I

Introduction

I.1. Diabetes Mellitus

Diabetes mellitus is characterized by a chronic increase in blood glucose level due to a dysfunction of carbohydrate, fat and protein metabolism, which in turn are the result of insulin resistance and/or insulin action disturbances. Its cardinal symptoms include polyuria, thirst and weight loss. Amputation and foot ulcers due to peripheral neuropathy, renal failure as the result of nephropathy, and blindness due to retinopathy are among its serious long-term complications (Alberti, Zimmet 1998). Considering the underlying cause, diabetes mellitus is classified in two main classes: Insulin Dependent Diabetes Mellitus (IDDM) or Type 1 Diabetes (T1D) and Non-Insulin Dependent Diabetes Mellitus (NIDDM) or Type 2 Diabetes (T2D). T1D is caused primarily by autoimmune processes, which leads to pancreatic β-cell islets destruction. Affected individuals are susceptible to ketoacidosis and require insulin to survive. T2D is caused either due to dysfunction of insulin secretion or to increased insulin resistance (Alberti, Zimmet 1998).

I.2. Clinical Presentation of T1D

Insulin secretion in T1D is very little or there is complete insulin deficiency and patients are dependent on exogenous insulin (Sherwin 2004). Majority of the patients are symptomatic and present with polyuria, polydipsia, polyphagia, and weight loss; they may also develop ketoacidosis. Evidence that supports an autoimmune nature of this type of diabetes include association with specific genes (e.g. HLA), presence of the

autoantibodies to insulin, glutamic acid decarboxylase, tyrosine phosphatases IA-2 and IA-2β and islet cells (Sherwin 2004, American Diabetes Association 2004). T1D is further sub-classified in autoimmune and idiopathic accounting for approximately 90% and 10% of all cases of T1D respectively (Maraschin 2012). Other autoimmune conditions such as Hashimoto's thyroiditis, Grave's disease and pernicious anemia may also be found in the patients with T1D (American Diabetes Association 2004).

I.3. Clinical Presentation T2D

Although patients with T2D have some insulin secretion, most of the patients with T2D are asymptomatic at evaluation; however fatigue, dizziness and blurred vision are common in these patients. Classical symptoms of hyperglycemia including polyuria, polydipsia and weight loss are seen in some of T2D (Sherwin 2004, American Diabetes Association 2004). Since T2D is a long lasting disease, patients may develop cardiovascular complications, retinopathy, neuropathy or nephropathy. Other common symptoms include erectile dysfunction in males and those related to vaginitis or generalized pruritus in females. Hypertension, dental decay and gingivitis are commonly seen in these patients. Most of the patients are overweight or obese. When significant insulin resistance exists, acanthosis nigricans is seen in some patients (Imam 2012). Although almost all of the children diagnosed as diabetic had type 1 diabetes in the past, recently it has been reported that the frequency of T2D is considerably increasing in this group. Most of these children are overweight or obese at the time of diagnosis and show glycosuria without classical symptoms (American Diabetes Association 2000).

I.4. Diagnosis of DM

American Diabetes Association described three ways for diagnosis of diabetes mellitus: 1) Random Plasma Glucose ≥200 mg/dl together with classical symptoms i.e. polyuria, polydipsia, and weight loss, 2) Fasting Plasma Glucose ≥126 mg/dl and 3) Oral Glucose Tolerance Test ≥200 mg/dl 2 hours after glucose is taken orally (American Diabetes Association 2004). Recently, measurement of the HbA1c has been recommended as an alternative method for the diagnosis of T2D by the World Health Organization and the American Diabetes Association (WHO 2011, American Diabetes Association 2010).

I.5. Complications of DM

The death rates due to heart disease and stroke have been reported to be 2 to 4 times higher in adults with diabetes compared to those without diabetes. It has also been estimated that 67% of adults with diabetes had high blood pressure, i.e. ≥ 140/90 mmHg (NDIC 2011). Diabetes mellitus is the leading cause of renal failure and blindness among adults; 44% and 28.5% of adults with diabetes developed diabetic nephropathy (DN) and diabetic retinopathy (DR) respectively. More than 200,000 people were on dialysis or had a kidney transplant in 2008 in the USA (NDIC 2011). Diabetes mellitus accounts for about 35% of cases with chronic renal failure. Approximately 30% of individuals with T2D have an increase in urinary albumin excretion (30-300 mg/24h) or overt albuminuria (>300 mg/24h) when they are diagnosed to have diabetes. DN is manifested initially by a rise in Glomerular Filtration Rate (GFR), then an increase in urinary albumin excretion and eventually decreased GFR (Estacio, Schrier 2001). Progressive renal function failure follows persistent proteinuria. Hypertension, nephrotic syndrome and end stage renal

failure are further clinical manifestations. Fluid retention, hyperkalemia, neurogenic bladder, and eventually coronary artery disease are among ESRD complications; the latter is the main cause of mortality in these patients (de Francisco 1999). Nervous system complications including pain or impaired sensation in the feet or hands, delayed digestion, erectile dysfunction and carpal tunnel syndrome have been found in 60% to 70% of diabetics. Periodontal disease and complications of pregnancy are also more common in individuals with diabetes. Diabetic ketoacidosis and nonketotic coma are acute life-threatening complications when diabetes is not controlled (NDIC 2011).

I.6. Epidemiology of DM

Following economic growth and urbanization which has led to change in lifestyle, particularly because lack of physical activity and an increase in the number of obese people, diabetes mellitus continues to be one of the most common chronic diseases in the world. The global prevalence rate of diabetes among adults has been estimated to be 8.3%, or 366 million adults, in 2011 (American Diabetes Association 2013). North America and the Caribbean countries were the second most prevalent regions after the Middle East and the North African countries. It is predicted that this rate will increase to 9.9% of the adult population, or 552 million adults, by 2030 (American Diabetes Association 2013). In Canada, there were 2.7 million adults, or 10.8% of the population, living with diabetes in 2011 and estimated to increase to 3.7 million people, i.e. 12.8% of the population, by 2030 (Whiting, Guariguata et al. 2011). According to a study of the American Diabetes Association, the total cost of diabetic care was estimated to be about $245 billion in 2012, with direct costs of $176 billion and indirect costs of $69 billion,

this shows a 41% increase when compared to the estimates of 2007 (American Diabetes Association 2013). In Canada, the total costs for diabetics are estimated to be more than CDN$8.14 billion in 2016 (Ohinmaa, Jacobs et al. 2004).

Currently T2D is more prevalent in the developing countries rather than in the developed countries compared to the past (Shaw, Sicree et al. 2010). In addition, individuals with T2D in the developing countries are younger compared to the type 2 diabetics in the developed countries (Shaw, Sicree et al. 2010). Not surprisingly, due to fast economic growth, nutrition transition and urbanization in a short time, Asia, particularly China and India, are encountered as the 'diabetic epicenter' worldwide (Chan, Malik et al. 2009). Individuals who emigrated from Middle East to Sweden have been reported to have greater prevalence of DM than native Swedes (Wandell, Johansson et al. 2008). Although usually T2D had been considered as an adulthood disorder, it is becoming more common in adolescents and children (Pinhas-Hamiel, Zeitler 2005). Younger individuals who have T2D are at more risk for developing early and chronic complications (Chen, Magliano et al. 2011). Risk factors for T2D have been divided to modifiable and non-modifiable. Modifiable risk factors include, but are not limited to, overweight or obesity, physical inactivity, dietary factors, hypertension, intrauterine environment and abnormal lipids. Family history of T2D, ethnicity, sex and age are among non-modifiable risk factors (Chen, Magliano et al. 2011).

I.7. Pathogenesis of T1D

T1D, also known as IDDM or juvenile-onset diabetes, accounts for approximately 5% of all diagnosed diabetes in adults (NDIC 2011); genetic, environmental factors and infections are known to be involved in its development (Acharjee, Ghosh et al. 2013).

I.7.1. Genetic Factors

The role of genetics in T1D has been revealed by many family and twin studies. For example, high concordance rate has been reported among twin pairs in a study conducted in Finland (Hyttinen, Kaprio et al. 2003). HLA genes located on the chromosome 6 have been identified to be associated with the development of T1D; HLA DR3 and DR4 alleles increase the risk of T1D (Erlich, Valdes et al. 2008). HLA molecules are involved in the binding procedure of antigens to T lymphocytes (Bluestone, Herold et al. 2010). IDDM2 gene and CTLA-4 gene, by affecting insulin gene expression in the thymus and impact on T cell function and regulation respectively, are possibly involved in the development of T1D (Sherwin 2004). Parts of innate immune system including B cells, natural killer cells and macrophages have been reported to be involved in the development of T1D through autoimmunity processes and their interaction with adaptive immune system (Ting, Bansal et al. 2012). Genes in the Human Leukocyte Antigen (HLA) region within the major histocompatibility complex (MHC) on chromosome 6 have been reported to be correlated with T1D (Concannon, Erlich et al. 2005). Disparities in alleles frequencies could explain the genetic susceptibility variation in different races, e.g. the statistically significant high frequency of protective heterozygous and homozygous alleles have been reported in Chinese 72% and 22% respectively. In White Americans the frequency of the

homozygous allele, which is associated with greater risk, has been found to be about 96%. These findings are consistent with the observed incidence rates of T1D in these populations, i.e. 0.7 and 15.8 persons per 100,000 per year in Chinese and White Americans, respectively (Bao, Wang et al. 1989). It has also been reported that a polymorphic locus near the insulin gene (Bell, Horita et al. 1984), and polymorphism in the protein tyrosine phosphatase nonreceptor type 22 (PTPN22) gene (Bottini, Musumeci et al. 2004) are associated with T1D. There is evidence that changes in the cytotoxic T lymphocyte antigen 4 gene (CTLA4), probably through their effects on T-cell activity, could be involved in the causation of some autoimmune disorders such as IDDM (Ueda, Howson et al. 2003). The interleukin 2 receptor alpha (IL2RA) gene region has also been determined to be associated with T1D (Lowe, Cooper et al. 2007). Since the incidence rate of T1D differs considerably not only between countries, e.g. in Finland the risk is about 35 times higher than in Japan (LaPorte, Tajima et al. 1985), but also within very similar ethnic groups that are located in different geographical locations, e.g. three times more in Finland than in Estonia, environmental and geographic factors could have particular impacts on its development (Zimmet, Alberti et al. 2001).

I.7.2. Environmental Factors

Since the pairwise concordance rates among identical twins vary between 13% and 33%, environmental factors must play essential role in the progression of the T1D (Kaprio, Tuomilehto et al. 1992). These factors include diet, viral infections and toxins. The potential role of cow's milk, which contains bovine insulin, in the development of T1D has been investigated in different studies. It is supposed that it acts through β-cell

autoimmunity process by exposing the β-cell and insulin autoantigens (Sabbah, Savola et al. 2000, Vaarala, Atkinson et al. 2008). Coxsackievirus B (CVB), mumps, rubella, cytomegalovirus (CMV) and enterovirus infections, possibly because of their effects on the β-cell autoimmunity, have been reported to be associated with an increased incidence of T1D (Ting, Bansal et al. 2012). It has been documented that while T1D incidence rate was low in the first half of the 20th century, it increased in the second half (Gale 2002). This increase may not only be explained by diagnosis improvements; either the increased survival of children with T1D resulted in pooling putative genes, or changes in environmental factors are considered as possible mechanisms (Krolewski, Warram et al. 1987). In animal models, diet, specifically gluten-free has been determined to have an impact on the incidence of T1D (Buschard 2011). The protective effect of a gluten-free diet at the risk of autoimmune disorders in humans has also been investigated (Cosnes, Cellier et al. 2008). Interaction between environmental factors and MHC might explain the difference in the incidence rate of T1D before and after 1980 (Badenhoop, Kahles et al. 2009). It has been shown that there is an association between enteroviral infection and the incidence of T1D (Dahlquist, Ivarsson et al. 1995). It has also been reported that the incidence rate of T1D significantly increased after an outbreak of coxsackievirus B5 infection (Wagenknecht, Roseman et al. 1991).

I.8. Pathogenesis of T2D

T2D, also known as NIDDM or adult-onset diabetes, is a common complex disease that is caused by a disturbance in glucose metabolism, which initially starts due to the resistance to insulin and eventually may lead to a dysfunction of the pancreas; T2D

accounts for about 90% to 95% of diabetic cases in adults (NDIC 2011). Like other complex diseases several factors are involved in the development of T2D; the impacts of lifestyle, environmental and genetic factors have been well investigated.

I.8.1. Genetic Factors

Family and twin studies have been used to identify genetic predisposition in T2D. In one study, investigators concluded that there is 100% concordance rate for type 2 diabetes among monozygotic twins (Barnett, Eff et al. 1981). On the other hand, only less than 50% of obese individuals who are insulin resistant develop diabetes; this suggests that only some of the patients who are at risk are susceptible to develop the disease (Stumvoll, Goldstein et al. 2007). In monogenic forms of diabetes such as Maturity Onset Diabetes of Youth (MODY), the genetic basis has been revealed although the majority of patients are polygenic. Peroxisome proliferator activated receptor gamma (PPAR-γ), insulin gene, calpain 10, and intestinal fatty acid binding protein 2 (FABP 2) are some of the candidate genes correlated with the development of T2D (Sherwin 2004). The prevalence rate of T2D has been demonstrated to be different within and between migrant groups suggesting the effect of the nature/nurture phenomenon (Abouzeid, Philpot et al. 2013). The same genotypes might result in different phenotypes among individuals with the same origin who live in different countries, indicating genetic and/or environmental components are involved for the same trait. For example, it has been detected that diabetic "Japanese Americans" with a "Pro/Pro" genotype of the peroxisome proliferator-activated receptor (PPARγ) had greater BMI and fasting serum insulin levels compared to diabetic "native Japanese". On the other hand, the expression of the "Ala-allele" genotype

was the same between two groups (Nemoto, Sasaki et al. 2002, Leonetti, Fujimoto 1989). In contrast, the prevalence rate of diabetes was similar between Indians who immigrated to Britain and those who lived in India (Patel, Vyas et al. 2006). African ancestry in three US population cohorts, i.e. Atherosclerosis Risk in Communities (ARIC) Study, Multiethnic Cohort (MEC) and the Jackson Heart Study (JHS), has been reported to be significantly associated with a higher odds ratio of diabetes, even after adjusted for age, sex, study centre, BMI and socioeconomic status (Cheng, Reich et al. 2012). Native American ancestry in Hispanic Americans (Parra, Hoggart et al. 2004) and in Mexicans (Martinez-Marignac, Valladares et al. 2007) has been identified to be correlated with higher risk of T2D. European ancestry among central Mexicans and Colombians has been shown to be associated with an increased risk of T2D (Florez, Price et al. 2009).

I.8.2. Environmental Factors

The fact that diabetes genotype, i.e. individual who has susceptibility genes for T2D, does not result in diabetes phenotype in all cases implies that clinical presentation of T2D depends on environmental factors (Leahy 2005). Association of obesity and lack of physical activity with T2D have been well studied (Hu, Manson et al. 2001). Modern lifestyle, characterized by high calorie intake and less physical activity, increases the incidence rate of T2D; it has been discussed that these factors affect on glucose homeostasis by increasing insulin resistance or β-cell dysfunction. In the presence of genetic predisposition, these factors increase the risk of hyperglycemia (Leahy 2005). Migrant studies support the concept that environmental factors are involved in the development of T2D. Moving from countryside and rural areas to urban areas has been

reported to be associated with severe obesity followed by an increase in the incidence of T2D; interestingly, it has been shown that the frequency of T2D decreased by moving back to prior lifestyle. Examples include Australian Aboriginals, and Pima Indians (O'Dea 1984). Increase in the incidence of T2D in China and India, due to the movement to the metropolitans, is another example (Weiss, Dziura et al. 2004). Insulin resistance occurs when the effects of insulin are impaired (Dinneen, Gerich et al. 1992). Many patients with impaired fasting glucose (IFG) or with T2D have hyperinsulinemia; therefore it has been proposed that the insulin resistance results in hyperglycemia (Stumvoll, Goldstein et al. 2007). Insulin resistance is associated with overweight and obesity; in obese individuals nonesterified fatty acids (NEFA), which are originated in adipocytes, are increased and result in more resistance to the insulin effects (Boden 1997). Insulin bonds to receptor that shows tyrosine kinase activity, and in turn activates the insulin receptor substrate (IRS) proteins. Serine and/or threonine phosphorylation results in the reduction of the tyrosine kinase activity of the insulin receptor which in turns can inhibit IRS protein activity; this mechanism is considered to be involved in the development of insulin resistance (Stumvoll, Goldstein et al. 2007). Adipose tissue is the main source of the NEFA, tumor necrosis factor-α (TNFα) and interleukin 6 (IL-6); increased levels of these products impair insulin signal process and can cause insulin resistance (Rajala, Scherer 2003, Ravussin, Smith 2002). β-cell dysfunction results in insulin secretion below the normal levels in patients with T2D. In the presence of insulin resistance (resulted from obesity or lack of physical activity) the pancreas secretory function is decreased that could cause impaired glucose tolerance (IGT) or diabetes

(Stumvoll, Goldstein et al. 2007). There is evidence indicating a linkage between fetal life and diseases developed at adolescence. It has been reported that a lower birth weight (LBW) was related to the development of T2D later in life (Barker, Hales et al. 1993, Whincup, Kaye et al. 2008); in premature infants, regardless of their weight for gestational age, insulin sensitivity had been reduced (Hofman, Regan et al. 2004). Statistically significant association has been found between famine exposure during fetal life and hyperglycemia in adulthood (Li, He et al. 2010). A change of lifestyle from traditional which included more activities to modern ways (i.e. more static and sedentary) could be associated with an increase of insulin resistance that in turn predisposes one to the development of obesity and T2D (O'Dea 1991); urbanization and "western food" have also been reported to be correlated with a higher prevalence of T2D (Ostbye, Welby et al. 1989). On the other hand, the effectiveness of lifestyle interventions in reducing the incidence rate of T2D has also been well documented (Knowler, Barrett-Connor et al. 2002, Bo, Ciccone et al. 2007, Diabetes Prevention Program Research Group, Knowler et al. 2009). Although familial aggregation supports genetic contribution to T2D, it should be noted that many studies have shown the protective effects of increased physical activity, appropriate diet and lifestyle modification in proceeding from IGT to diabetes (Pan, Li et al. 1997, Tuomilehto, Lindstrom et al. 2001, Kosaka, Noda et al. 2005, Kontogianni, Liatis et al. 2012, Ramachandran, Snehalatha et al. 2013). Reviews of studies conducted to evaluate the impact of environmental chemicals and/or pollutants on the risk of T2D show that there is evidence suggesting a high exposure of persistent organic pollutants (POPs) (Taylor, Novak et al. 2013), arsenic (Maull, Ahsan et al. 2012)

mercury (He, Xun et al. 2013) and Organophosphorus Compounds (OPCs) (Pournourmohammadi, Farzami et al. 2005) could increase the risk of diabetes.

I.9. Genetics of DM

Different epidemiologic studies including international variation, migrant/admixture studies, ethnic/racial differences, the effects of age, sex and time on incidence rate, familial aggregation and twin/adoption studies are used to investigate the genetic component of complex disorders such as diabetes (Thomas 2004). It has been shown that polymorphisms in the calpain-10 (CAPN10) regions are associated with the risk of T2D (Horikawa, Oda et al. 2000); however, there is considerable difference in the observed allele frequencies in CAPN10 between African and non-African populations, suggesting that this difference could be partly the result of adaptation and be explained by natural selection (Fullerton, Bartoszewicz et al. 2002). Regardless of ethnicity, geographic and environmental diversities, the transcription factor 7-like 2 (TCF7L2) gene has been demonstrated to be associated with the development of T2D, which supports the existence of genetic predisposition as a potential risk factor (Cauchi, El Achhab et al. 2007). A considerably lower frequency of a protective haplotype of human immunoglobulin G system, i.e. $Gm^{3;5,13,14}$, in Pima and Papago Indians ancestors strongly supports the effect of ethnicity and genetic mixture on the prevalence of T2D (Knowler, Williams et al. 1988). It has been estimated that 15.7 million or 10.2% of all non-Hispanic Whites, ages 20 years or older, and 4.9 million (18.7%) of all Non-Hispanic Blacks, ages 20 years or older, suffered from diabetes in the United States in 2010 (NDIC 2011), indicating considerable prevalence variation between different racial and ethnic

groups. In the multiethnic cohort study of diet and cancer, T2D prevalence was more than two times higher in native Hawaiians, Latinos, AA and Japanese compared to EA (Maskarinec, Grandinetti et al. 2009). However, allele frequencies of risk markers did not show any greater genetic risk in these populations (Waters, Stram et al. 2010); therefore, the effect size of these variants is controversial. Asian-Australians in the Fremantle Diabetes Study had a similar T2D prevalence compared to the general population although they had significantly lower BMI and more paternal history of diabetes (Tan, Davis et al. 2013). In contrast, the prevalence of T2D has been shown to be significantly higher among south Asian migrants in Norway (Jenum, Holme et al. 2005).

I.9.1. Population-Based Studies

Since families share environment and genes, familial clustering of T2D and its traits could be used to determine genetic contribution in T2D. It has been shown that in addition to lower high density lipoprotein (HDL) cholesterol concentrations, greater fasting plasma insulin concentration and BMI, Caucasian first degree relatives of diagnosed diabetic individuals living in the UK had T2D prevalence rate four times higher, which could be the result of genetic and/or environmental predisposition among the relatives of probands (Shaw, Purdie et al. 1999). In the Framingham Offspring Study, it has been reported that when both parents were diabetics the risk of T2D among offspring was six fold higher than those without any parental diabetes history, which supports the genetic transmission pattern of T2D (Meigs, Cupples et al. 2000). In a cohort of young Chinese, the Hong Kong Family Diabetes Study, the risk of Impaired Glucose Tolerance (IGT) and diabetes were 1.5 and 4.3 higher respectively among siblings of

probands; obesity, dyslipidaemia and hypertension rates were also more frequent in these relatives compared to the general population (Li, Ng et al. 2006). In a study in Tunisia, it has been detected that the prevalence of T2D in the siblings of probands was 45% compared to 8.9% in the general population; 34% of diagnosed diabetics had one affected parent, and maternal transmission was two times greater than paternal (Arfa, Abid et al. 2007, Whiting, Guariguata et al. 2011). In a large-scale study, conducted in Sweden, the genetic basis of familial aggregation has been investigated. The relative risk (RR) was 2.03 in offspring of one affected parent similar to the risk for singleton siblings (2.77) but increased to 5.35 when two parents were affected. When biological parents of adoptees were probands, the risk for adoptees was 2.16 while the risk did not change in adoptees when their adopted parents were affected (Hemminki, Li et al. 2010).

I.9.2. Twin Studies

To distinguish between genetic (liability or proportion due to common genes) and environmental (environmentality or the proportion due to a shared environment) components involved in the development of a complex disease, twins are identified among affected individuals to compare the concordance/discordance rates of monozygotic (MZ) and dizygotic (DZ) to estimate the heritability (Thomas 2004).
In the first study, conducted in England, in which twins had been differentiated according to the etiology of diabetes, i.e. T1D vs. T2D, and tested for glucose tolerance, it was found that 90.6% of MZ twins had been concordance for T2D (48 out of 53, proband concordance rate: 64%); the authors concluded that, by considering the age of the affected twins, genetic predisposition was more involved in the development of T2D

(Barnett, Eff et al. 1981). In a study of 250 MZ American male twins, it has been reported that the concordance rate for T2D in MZ twins was 58% and the concordance rate of impaired glucose tolerance has been detected in 65% of non-diabetic MZ twins; therefore, MZ twin brothers had six fold higher risk for developing T2D suggesting a strong genetic correlation among affected persons (Newman, Selby et al. 1987). First large scale study outside Western countries revealed that 38 out of 46(83%) and 4 out of 10 (40%) MZ and DZ diabetic pairs were concordance for T2D respectively in Japan (Committee on Diabetic Twins 1988) and the probandwise concordances for T2D have been reported to be 31.6% and 10.0% among MZ and DZ male twins respectively in a Finnish study (Kaprio, Tuomilehto et al. 1992). In the Danish Twin Register, the probandwise and pairwise concordance rates were significantly higher among MZ pairs than DZ just when IGT had been considered in addition to T2D. The study showed significantly high heritability for BMI (r = 0.68 for MZ versus r = 0.28 for DZ) and weight (r = 0.78 for MZ versus r = 0.39 for DZ), i.e. h^2 equals to 0.80 and 0.78 respectively, indicating a major role of genetic predisposition in the variations of these traits; r specifies "Interclass Correlation" that measures resemblance within twin pairs and h^2 identifies "Heritability" that states the fraction of the whole variation of a trait attributable to genetic variation, which is calculated as $h^2 = 2 \times (r_{MZ} - r_{DZ})$ (Poulsen, Kyvik et al. 1999). In a more recent study the probandwise concordances reported to be 0.58 and 0.11 among MZs and DZs respectively and the genetic risk derived from concordance ratio has been estimated to be 5.3 for T2D (Condon, Shaw et al. 2008).

I.9.3. Mode of Inheritance

Different inheritance patterns in T1D have been discussed in different studies. In one study, it has been reported that 55% to 60% of affected type 1 diabetic pairs of siblings shared two identical HLA haplotypes, 40% shared one haplotype, and only very few shared zero haplotypes. Supposing that the related gene is frequent in the population, this pattern follows autosomal recessive inheritance (Rubinstein, Ginsberg-Fellner et al. 1981). In contrast, autosomal dominant pattern fitted better in studying the inheritance of Juvenile diabetes in Caucasian and American Blacks (MacDonald 1980). In another study, it has been proposed that the phenotype of T1D could be best explained by the gene dosage effect; it has been estimated that the penetrance rate was 0.71 in individuals having two haplotypes, compared to 0.065 in cases with one haplotype (Spielman, Baker et al. 1980). However, many studies support heterogeneity concept regarding the mode of inheritance in T1D (Rotter 1981). It has been shown that there is strong association between one allele of ACE gene and T2D in Indian population; the best mode of inheritance has been reported to be recessive and multiplicative (Singh, Naz et al. 2006). Although Plasma Cell-1 (PC1) gene and its polymorphism have been reported to be associated with T2D and insulin resistance, neither a recessive nor an additive mode of inheritance could be found in the Moroccan population. However, in the presence of obesity, its polymorphism showed association with T2D in both models (El Achhab, Meyre et al. 2009). Additive mode of inheritance has been identified between TCF7L2 gene allele, and the treatment result in type 2 diabetics (Holstein, Hahn et al. 2011). Mendelian inheritance model with a single major gene effect was the best model, which

explained the aggregation of overt albuminuria in Pima Indian families (Imperatore, Hanson et al. 1997). A Mendelian major gene effect and multifactorial inheritance has been identified in the clustering of albuminuria and urinary albumin excretion (UAE) in Caucasian families with T2D (Fogarty, Hanna et al. 2000).

I.10. Genetics of Diabetic Nephropathy

It has been reported that 41% of the siblings of probands with T1D and diabetic renal disease developed ESRD while none of the siblings of probands without DN had such extreme renal dysfunction (Seaquist, Goetz et al. 1989). In Pima Indians, protienuria was found in 45.9%, 22.9% and 14.3% of the offspring if both, one or neither parents had T2D and proteinuria, respectively (Pettitt, Saad et al. 1990). The siblings of insulin dependent diabetics with nephropathy were five times more likely to develop elevated microalbuminuria or clinical nephropathy (Borch-Johnsen, Norgaard et al. 1992). 19% of non-insulin dependent African American diabetics with ESRD had a first degree relative with ESRD; if second and third degree relatives were also included, the risk of developing ESRD increased eight times (Freedman, Tuttle et al. 1995). The risk of diabetes in offspring of Pima Indian parents with DN, independent of age at the onset of parental diabetes, has been found to be 2.5 times higher than offspring of diabetic parents without diabetic renal disease suggesting that DN could be a positive predictor for developing diabetes in the offspring (McCance, Hanson et al. 1995). Familial clustering of ESRD, with different causes, has also been identified in White Americans; if a first, second or third degree relative had ESRD, the individual risk of ESRD was increased 3.5 times (Spray, Atassi et al. 1995). Persistent proteinuria incidence rate was 2.5 times

higher in the siblings of T1D probands with advanced diabetic renal disease compared to the siblings of probands without renal complication or with microalbuminuria (Quinn, Angelico et al. 1996).

I.11. Strategies of Gene Identification

There are different methods to identify genes, which are correlated with diabetes and DN.

I.11.1. Candidate Gene Studies

Candidate gene studies rely mostly on the present understanding of causal relationship between the potential causative gene and the disease condition; Freedman et al. and Bowden reviewed these candidate genes (Bowden 2003, Freedman, Bostrom et al. 2007). Endothelial Nitric Oxide Synthase (eNOS) gene by hemodynamic mechanisms, possibly through intraglomerular hypertension, has been demonstrated to be involved in the development of ESRD in Japanese population (Suzuki, Nagase et al. 2000). Growth factor and cytokine genes have also been reported that play a role in the development and progression of diabetic nephropathy in type 2 diabetics. Kallikrein genes, with the effect on release of bradykinin, have been shown to be associated with DN in AA population (Bowden 2003). The role of inflammation in the pathogenesis of DN has also been studied, and genes encoding cytokines such as Tumor Necrosis Factor (TNF), Interleukin 1 receptor antagonist (IL-1Ra) and interleukin 1 β (IL-1β) are other candidate genes that might be associated with DN (Lee, Lee et al. 2005). Manganese superoxide dismutase (Mn-SOD) gene, by its impact on the oxidative stress mechanism, has been identified to be associated with albuminuria in Korean type 2 diabetics (Lee, Choi et al. 2006).

Angiotensin Converting Enzyme (ACE), through renal renin-angiotensin system, has been reported to be related with DN (Ng, Tai et al. 2005, Ng, Placha et al. 2006). Engulfment and Cell Motility 1 (ELMO1) gene, possibly through the accumulation of extracellular matrix proteins in the glomeruli, has been found to be associated with chronic renal dysfunction in Japanese (Shimazaki, Kawamura et al. 2005, Shimazaki, Tanaka et al. 2006); variants of ELMO1 gene have also been reported to be associated with ESRD in AA (Leak, Perlegas et al. 2009).

I.11.2. Genome-Wide Linkage Studies

In the linkage studies about 400 genetic markers, which are uniformly located across the genome, are genotyped in the relatives with DN. Statistical methods are used to identify chromosomal regions that are segregated with DN; the basic concept is that DN contributing genes could be found within these regions regardless of their previous known biological function. The logarithm of odds (LOD) score is used to determine the probability that these locations are being inherited with the disease of interest (Bowden 2003, Freedman, Bostrom et al. 2007). Linkage analyses, both two-point and multi-point, of Pima Indian sib-pairs with T2D revealed that two loci on chromosome 7, i.e. D7S500 and D7S1804, were strongly linked with nephropathy; aldose reductase (ALDR1), T-cell receptor β-chain (TCRBC) and constitutive endothelial nitric oxide synthase 3 (NOS3) genes are located in this region that might be involved in the development of DN (Imperatore, Hanson et al. 1998). In Turkish families, genome wide scan showed strong linkage with DN with maximum LOD score of 6.1 in a region located on the chromosome 18q when all participants included in the analysis. A region between

D18S43 and D18S50 markers on chromosome 18 has been reported to have a strong linkage with DN in type 2 diabetes; the Krupple-like zinc-finger gene 236 (ZNF236), the expression of which is glucose dependent, was located on this region (Vardarli, Baier et al. 2002). A genome scan of African American family members with NIDDM and ESRD revealed three susceptible loci on chromosomes 3q, 7p and 18q linked with diabetic nephropathy; with maximum LOD score of 1.43, the region on chromosome 7p showed the strongest linkage with DN in the entire sample. In addition, when Ordered Subset Analysis (OSA) with phenotypic traits was performed (age at ESRD onset), a region on the chromosome 3q has been found to have a combined LOD score of 4.55 in a subset of families (Bowden, Colicigno et al. 2004). Non-parametric linkage analysis (NPL) and OSA revealed strong linkage in African Americans with renal disease on the chromosomes 2q (individuals with earliest age at ESRD onset) and 10q (individuals with latest age at ESRD onset) with the LOD score of 3.05 and 2.47 respectively (Freedman, Bowden et al. 2005). Familial aggregation of renal function has been studied through a genome-wide linkage scan of patients with T2D and revealed significant heritability, $(h^2 = 0.29 - 0.47)$ of estimated (GFR) among diabetic relatives. Regions within chromosome 2 and chromosome 10 have been reported to be linked with GFR aggregation with LOD score of 4.1 and 3.1 respectively (Placha, Poznik et al. 2006). The results of these studies are summarized in the table 1.

Table 1: The results of the linkage studies

Study	DM	Sample Size	Marker	CHR	Method	p-value	LOD
Imperatore et al.	T2D	135	D7S500	7q	M-NPL	2.00E-04	2.73
Vardarli et al.	T2D	368	D18S469	18q	M-NPL	9.00E-03	6.14
Bowden et al.	T2D	355	D7S3051	7p	S-NPL	1.00E-02	1.43
Freedman et al.	T2D	1023	D13S796*	13q	M-NPL/OSA	2.40E-03	4.94
			D2S1391**	2q		1.68E-02	3.05
			D10S1248***	10q		4.86E-02	2.47
Placha et al.	T2D	406	D2S1384	2q	M-VCL	6.20E-06	4.01

M-NPL: Multiple Non Parametric Linkage, S-NPL: Single Non Parametric Linkage, OSA: Ordered Subsets Analysis, M-VCL: Multiple Variance Component Linkage, LOD: Logarithm of Odds, *subset by BMI (High), **subset by age at diagnosis of ESRD (early onset), *** subset by age at diagnosis of ESRD (late onset)

I.11.3. Genome-Wide Association Studies

In Genome-Wide Association Study (GWAS) the allele frequency is compared in individuals with the condition of interest, such as disease status or specific phenotype (i.e. cases), and those without this condition (i.e. controls) to identify associations between single nucleotide polymorphisms (SNPs) and traits of interest. Association studies could be conducted at population level to compare unrelated individuals, indeed the correlations of observed phenotypes and observed marker alleles would be investigated. The power of these studies is good only if common causal variants are under investigation i.e. the minor allele frequency (MAF) is not too small, Common-Disease Common-Variant (CDCV) hypothesis is hold and population is on Hardy-Weinberg Equilibrium (HWE) (Balding 2006, Balding, Bishop et al. 2007, Ziegler, Konig 2010). In the GoKinD collection, a genome-wide association scan of T1D revealed 11 SNPs variants of four loci located on chromosomes 7p, 9q, 11p and 13q to be associated with DN; rs10868025 on chromosome 9q, near the 4.1 protein ezrin, radixin, moesin (FERM) gene, showed the strongest association; in addition, some variations in ELMO1 gene have also been

reported to be associated with DN in Caucasians (Pezzolesi, Katavetin et al. 2009), (Pezzolesi, Skupien et al. 2010). In another study conducted in African Americans, 25 SNPs have been identified to be associated with ESRD in T2D; rs6930576 ($p = 7.04 \times 10^{-7}$) showed the strongest association with ESRD and is located on the chromosome 6q within the SAM and SH3 domain containing 1 (SASH1) gene (McDonough, Palmer et al. 2011). In an attempt to replicate the results of a previous study conducted in European T1D, which found 3 SNPs associated with DN, in Japanese population only one (rs7588550, p = 0.0126, odds ratio (OR) = 0.79, 95 % confidence interval: 0.65–0.95) of 3 SNPs has been reported to be significantly associated with DN in Japanese T2D, although the effect direction was reverse; this SNP is located on the chromosome 2q near the erb-b2 receptor tyrosine kinase 4 (ERBB4) gene that encodes one of the type 1 receptor tyrosine kinase subfamily (Maeda, Imamura et al. 2013). The results of these three studies are summarized in the table 2.

Table 2: Strongest SNPs associated with DN in GWA Studies

Study*	Type of DM	Sample Size	SNP	Gene	CHR	Method	p-value
Pezzolesi et al.	T1DM	1705	rs10868025	FRMD3	9q	GWAS	5.00E−07
McDonough et al.	T2DM	2993	rs6930576	SASH1	6q	GWAS	7.04E−07
Maeda et al. (replication study)	T2DM	2300	rs7588550	ERBB4	2q	Logistic Regression Analysis	1.26E−02

*these studies were case-control designed

I.11.4. Present Study

In this study The Family Investigation of Nephrology and Diabetes (FIND) study dataset has been used to perform a genome-wide association approach to identify loci associated with DN as a dichotomous trait and changes in estimated GFR (eGFR) which is a

quantitative variable in three distinct ethnic groups including 1454 cases with diabetes and kidney disease selected from FIND probands, and 1168 controls from unrelated FIND family members.

The goal of this study is to determine genetic contribution in the development of diabetic nephropathy and variation in eGFR across the whole genome among those having T2D. The main hypothesis is that since T2D is a polygenic disorder (Imam 2012, Thomas 2004), abundant SNPs located on the various genes could be found that are associated with DN and eGFR values in type 2 diabetics. In addition, because the dataset used in this study includes information of three distinct ethnicities, differences could also be determined to identify whether the potential SNP(s) and/or gene(s) associated with DN/eGFR are located on the same regions, or presented in different loci in each race, i.e. AA, MA and EA; finally, sex dependency in each group could also be detected. Therefore, the assumptions are that SNP(s)/gene(s) do exist in Type 2 Diabetics of FIND study cases, although they might be located in different loci and chromosomes in different ethnic group, and it could be possible to detect sex dependency as well. Several quality control metrics were applied to this data to avoid biases due to study design and errors in genotype calling, which could potentially lead to systematic biases in genetic case-control association studies and result in increasing the number of false-positive and false-negative associations. Since the dataset which is used in this study provided information for each race separately, quality control metrics would also be applied for each race separately.

The FIND investigators have used a case-control design with inclusion/exclusion criteria for their study, and I performed GWA analyses to detect potential association(s) between SNP(s) and DN or eGFR values. Although no SNPs were found significantly associated with diabetic nephropathy in any ethnic groups, the strongest ones associated with DN and eGFR values, reported in this study, are summarized in the table 3 and table 4 respectively.

Table 3: Top Associated SNPs with DN in different races and sexes

Ethnicity	Sample Size	SNP	Gene	CHR*	χ^2	p-value	GC**
African Americans							
All Together	864	rs1285582	NUP62CL	Xq	29.55	5.46E-08	9.23E-08
Females	573	rs6705592	SLC8A1-AS1	2p	21.80	3.02E-06	4.27E-06
Males	291	rs3857190	In. variant***	4q	23.80	1.07E-06	1.30E-06
European Americans							
All Together	763	rs4774390	RORA	15q	27.67	1.44E-07	1.69E-07
Females	414	rs2069347	CCNG1	5q	24.98	5.79E-07	5.79E-07
Males	349	rs1239908	In. variant***	12q	26.89	2.16E-07	3.94E-07
Mexican Americans							
All Together	1000	rs17067207	In. variant***	5q	23.42	1.30E-06	1.81E-06
Females	603	rs2923173	In. variant***	5q	19.00	1.31E-05	1.60E-05
Males	397	rs1106228	In. variant***	9q	24.74	6.55E-07	6.77E-07

*CHR: Chromosome, **Genomic Control (GC) is an easy way to detect, and adjust for population stratification, ***Intergenic Variant

Table 4: Top Associated SNPs with eGFR variation in different races and sexes

Ethnicity	Sample Size	SNP	Gene	CHR	B*	p-value	GC**
African Americans							
All Together	864	rs1285582	NUP62CL	Xq	14.07	9.97E-08	1.10E-07
Females	573	rs17820651	ds. variant***	2q	24.91	3.57E-06	3.76E-06
Males	291	rs243294	NPAS3	14q	23.04	9.81E-07	1.03E-06
European Americans							
All Together	763	rs4774390	RORA	15q	13.15	3.64E-07	3.73E-07
Females	414	rs4675095	IRS1	2q	-36.36	5.62E-06	5.91E-06
Males	349	rs2118063	In. variant****	8q	-17.60	7.46E-07	9.94E-07
Mexican Americans							
All Together	1000	rs4680068	CLRN1	3q	62.21	1.83E-08	1.83E-08
Females	603	rs10949778	ds. variant***	7q	63.23	9.11E-08	9.11E-08
Males	397	rs4904196	In. variant****	14q	22.28	3.75E-07	3.75E-07

*B: Beta i.e. regression coefficient, **Genomic Control (GC), **Downstream Gene Variant, ***Intergenic Variant

Chapter II

Genome-wide association study to identify single nucleotide polymorphisms associated with diabetic nephropathy in FIND study

II.1. Introduction

Since in the FIND study, data is available for three distinct ethnicities (AA, MA and EA), in this work a genome-wide association study has been performed to identify loci associated with diabetic nephropathy in three distinct ethnic groups. A genome-wide association study (GWAS) has been done in multi-ethnic groups including 1454 cases with diabetes and kidney disease selected from FIND probands and 1168 controls from unrelated FIND family members.

Although the rate of new cases of ESRD in 2011 decreased 4.2% compared to the year 2010, nephropathy due to diabetes continued to be the most common cause of ESRD in the USA accounting for 44.7% of new cases. The incident rate of ESRD in diabetics varied from 122.3 per million of the population to 405.6 per million between different races; African Americans had the highest rate followed by Hispanics; the rate was ≈ 3.32 times higher than in whites and ≈ 1.28 times higher than in Hispanics (USRDS 2013). Considerable variation regarding prevalence rate of diabetes and diabetic nephropathy between different racial and ethnic groups has also been reported (NDIC 2011).

II.2. Materials and Methods

The Family Investigation of Nephrology and Diabetes (FIND)

The Family Investigation of Nephrology and Diabetes (FIND) has been established by the National Institute of Health (NIH) to investigate the genetic contribution in the

development and progress of diabetic nephropathy. Eight U.S. clinical centers have been involved for recruitment of eligible participants among four distinct ethnic groups. Two main strategies for detecting susceptibility genes include: 1) a family-based linkage study, and 2) mapping by admixture linkage disequilibrium (MALD) in case-control group. Individuals entered to the study have been classified in two groups: "Family" and "MALD" (Family Investigation of Nephropathy and Diabetes Research Group 2003). The "Family" study has been used for linkage analysis among those with DN who either their both parents are alive or have a diabetic sib. On the other hand, in the "MALD" study a case-control design has been used for mapping by admixture linkage disequilibrium (Family Investigation of Nephropathy and Diabetes Research Group 2003, Knowler, Coresh et al. 2005).

Admixture mapping or Mapping by Admixture Linkage disequilibrium (MALD) method is based on the concept that long haplotype blocks are generated by gene flow during admixture, and is used to identify linkage disequilibrium (LD) of MALD markers (≈2000-3000) and disease genes. Some points are required to be considered in MALD method: the disease incidence rate should be very different between the parental populations, frequency difference of disease-causing alleles should be measureable between the parental populations, at least admixture should be two generations old, and markers should be selected appropriately. Since in AAs the relative risk of NIDD and ESRD differs between the parental populations (i.e. African ancestry and European ancestry) MALD approach could be useful for African-America population to identify

disease genes following linkage studies and HapMap-based association scans (Smith, O'Brien 2005).

Linkage Disequilibrium (LD) is a property that identifies the tendency of an allele of one SNP to be inherited with an allele of another SNP in the population, when these two alleles (at nearby loci) are seen together on the same chromosome more often than expected by chance the loci are said to be in disequilibrium. A specific set of alleles on the chromosome is known as a haplotype. D′ and $r^2$are the two most common statistics of LD. D′ measures the recombination events between markers and ranges from 0 to 1, where 0 indicates complete linkage equilibrium (frequent recombination happened) and 1 indicates complete LD (there is no historical recombination). $r^2$is defined as the squared correlation coefficient between 2 SNPs and is correlated with the allele frequencies, r^2ranges also from 0 to 1. If r^2equals to 1, it would indicate perfect LD. In this case, only one of the two markers can be genotyped to identify the allelic variation (Ardlie, Kruglyak et al. 2002). LD can be assessed by using HapMap genotype database. The International HapMap Project provided a genome-wide database of common SNPs variants, which are required to identify genetic association with clinical phenotypes; samples of 11 human populations have been genotyped for more than 1.5 million SNPs (International HapMap Consortium 2005).

The most common used methods to correct the multiple- testing results are described briefly: 1) In Control of Family-Wise Error Rate (FWER) or Bonferroni correction the significant level is adjusted considering the number of independent tests; to reach the traditional significant level of 0.05, when there are *m* independent tests, α is calculated

either as $\alpha = 1 - (1 - 0.05)^{1/m}$ known as Sidak correction, or as $\alpha \approx 0.05/m$ known as Bonferroni correction. 2) In permutation test, cases and controls are randomly changed by each other and all *m* tests will be calculated again; this procedure will be repeated typically 1000 times. 3) In Bayesian method, Bayes' rule is used to assess the probability that null hypothesis is true when a test is significant; it depends on the study power, false positive rate and the prior probability that null hypothesis is true. 4) Benjamini and Hochberg introduced False Discovery Rate (FDR) method; in this method α is set at an appropriate level (driven from data), which is used for controlling the expected number of false discoveries in *m* tests (So, Sham 2011).

The main goal of FIND study has been to detect and find genes that are involved in the development of DN. To be able to consider genetic differences, African Americans (AAs), European Americans (EAs), Mexican Americans (MAs), and American Indians were recruited; mostly, FIND collected family members information, along with case and control subjects. Three approaches have been defined to detect disease-causing alleles: genome-wide linkage study by investigating diabetic family members concordant or discordant for DN, MALD and analysis the associations of certain alleles with DN in a population (Knowler, Coresh et al. 2005).

The results of the first linkage analysis, using a microsatellite marker-based genome scan in FIND participants, have shown linkage to DN in regions of chromosomes 7q, 10p, 14q and 18q (Iyengar, Abboud et al. 2007). It has also been reported that regions on chromosomes 1q, 2p, 7q, 8q and 18q were linked to changes in eGFR in Mexican Americans population; however, in American Indians the linkage was seen on

chromosome 11p, and in African Americans it was on chromosome 15q (Schelling, Abboud et al. 2008). Interval on chromosome 6p has also been identified to be linked with DN among European Americans (Igo, Iyengar et al. 2011). Another study revealed evidence of linkage between eGFR variation and regions on chromosomes 20q (SNP: rs736264) and 15q (SNP: rs2928714) in Mexican Americans and European Americans, respectively (Thameem, Igo et al. 2013).

The sample provided by NIH is a population study. Initially, it was supposed to detect potential impact of polymorphisms in the engulfment and cell motility 1 (ELMO1) and silent information regulator 1 (SIRT1) genes on diabetic nephropathy in different ethnic groups through authorized access to the dataset granted by NIH (please refer to the appendix 1 for this approval); however, due to resources limitations, specially funding the project, and since the title of the FIND database i.e. "The Family Investigation of Nephropathy and Diabetes" is broad, in the present study, I conducted a GWAS to determine potential SNPs associated with DN and eGFR values.

The characteristics of participants are presented in table 5. This sample includes information about 2622 individuals (1454 cases with DN and 1168 controls without DN) who participated in the FIND study. Family members who received insulin and/or oral hypoglycemic medicines, and those without diabetes history with hemoglobin A1C (HbA1C) ≥6.0% were categorized as diabetic. Overt proteinuria with proteinuria ≥500 mg/24h or albuminuria ≥300 mg/24h, together with a protein/creatinine ratio ≥0.5 or urine albumin/creatinine ration ≥0.3 was defined as nephropathy (Knowler, Coresh et al. 2005).

The dataset provides information such as medical history (5 variables: diabetes, high blood pressure, kidney failure, leg amputation, and retinopathy), anthropometry measurements (4 variables: current and maximum height and weight), laboratory tests (18 variables: history and laboratory urine protein excretion, serum creatinine, urine protein, urine albumin sign, urine albumin, urine creatinine, PCR, ACR, and GFR), sociodemographic data including sex, race, and age; medication history (2 variables: taking Angiotensin-converting enzyme (ACE) inhibitors or Angiotensin II Receptor Blockers (ARBs) and taking insulin); high blood pressure; duration of diabetes, kidney failure and high blood pressure; and the age of onset of diabetes and kidney failure (for details of variables please see appendix 2).

Table 5: Participants Characteristics

	DN (+) RF (-)	DN (+) RF (+)	DN (-) RF (-)
Male / Female (%)	110(46.4)/127(53.6)	535(44)/682(56)	387(33.1)/781(66.9)
Age (SD)	58.39 (11.01)	61.32 (11.02)	59.53 (10.22)
Ethnicity			
African American (%)	36 (15.2)	545 (44.8)	278 (23.8)
European American (%)	57 (24.1)	282 (23.2)	424 (36.3)
Mexican American (%)	144 (60.8)	390 (32)	466 (39.9)
Total	237	1217	1168
Diabetes Age of Onset (SD)	40.60 (12.02)	38.57 (12.78)	41.83 (11.33)
RF Age of Onset (SD)	----------	58.23 (11.31)	----------
Diabetes Duration (SD)	17.79 (7.72)	22.82 (9.19)	17.44 (7.70)
RF Duration (SD)	----------	2.94 (2.59)	----------
Lab_HbA1C (SD)	8.01 (2.14)	7.01 (1.65)	7.96 (1.88)
Lab_GFR (SD)	46.20 (30.97)	14.82 (20.85)	86.64 (39.35)
Lab_Serum Creatinine (SD)	2.08 (1.28)	7.18 (3.47)	0.90 (0.24)
Lab_Urine Protein (SD)	414.06 (488.54)	189.09 (253.06)	8.88 (8.89)
Lab_Urine Albumin (SD)	228.02 (331.12)	81.25 (117.11)	3.45 (8.42)
Lab_Urine Creatinine (SD)	84.92 (59.03)	98.64 (71.49)	94.31 (67.06)
Lab_PCR (SD)	7.12 (24.48)	2.72 (3.60)	0.49 (7.45)
Lab_ACR (SD)	6.44 (55.83)	1.54 (1.91)	0.05 (0.86)

DN: Diabetic Nephropathy, RF: Renal Failure, Lab: Laboratory, SD: Standard Deviation

II.3. Subjects

In this case-control study, two criteria were drawn up to select case subjects, one with firm inclusion criteria and one with moderate inclusion. The FIND principal investigators, coinvestigators, program coordinator in eight clinical centers and a Genetic Analysis and Data Coordinating Center well-defined the inclusion and exclusion criteria as follow (Family Investigation of Nephropathy and Diabetes Research Group 2003, Knowler, Coresh et al. 2005).

The firm cases were selected from FIND probands or siblings if they had DN or met one or more of the following criteria:

a) Albumin-to-Creatinine ratio (ACR) ≥ 0.3

b) Protein-Creatinine ratio (PCR) ≥ 0.5

c) Glomerular Filtration Rate (MDRD formula) ≤ 60

d) Serum Creatinine ≥ 1.6 mg/dL for Males, or ≥ 1.4 mg/dL for Females

The moderated cases were selected from FIND probands or siblings if they had azotemic and normal proteinuria, based upon the following criteria:

a) ACR ≤ 0.03

b) GFR ≤ 60, Serum Creatinine ≥ 1.6 mg/dL for Males, or ≥ 1.4 mg/dL for Females.

Controls for this case-control study were selected from the unrelated FIND family members who were long term ($\geq$10 years) diabetics with healthy kidney function. The method included selection based on one or more of the following criteria:

a) ACR ≤ 0.03, PCR ≤ 0.05, GFR ≥ 60

b) Serum Creatinine ≤ 1.6 mg/dL for Males, or ≤ 1.4mg/dL for Females.

Based on these subject selection criteria, participants were divided into 1454 cases (with DN) and 1168 controls (without DN).

II.4. Genotyping and Quality Control Methods

DNA samples for eligible cases and controls were genotyped with Affymetrix 6.0 chip. Several quality control (QC) metrics were applied to this data to avoid biases due to study design (i.e. case-control) and errors in genotype calling, which potentially could lead to systematic biases into genetic case-control association studies and result in an increased number of false-positive and false-negative associations; to decrease the false positive associations, individuals and SNPs with high error rates are removed before applying statistical analyses. These QC steps would be implemented both at the individual level and marker level (Anderson, Pettersson et al. 2010).

In this study the thresholds used for filtering are as below: at individual level, those with sex inconsistency (discordant sex information), a missing rate of more than 3% (missing genotype or heterozygosity rate outliers) are not included in the analyses. Linkage disequilibrium (LD) is calculated to measure the dependency between SNPs, and is mostly reported as r^2which is the square of the Pearson correlation coefficient. If X_A is the value of one allelic locus and X_B is the value of the second allelic locus, the standard transform of the covariance in allelic value between the loci is called the Pearson correlation coefficient and calculated as $r_{AB} = \frac{Cov(X_A, X_B)}{\sqrt{Var(X_A)Var(X_B)}} = \frac{D_{AB}}{\sqrt{f_A f_a f_B f_b}}$; therefore $r^2 = \frac{D^2}{f_A f_a f_B f_b}$, where D is the covariance in allelic value between the alleles of a pair of

loci, f_A is the marginal allele frequency of allele *"A"* and f_a is the marginal allele frequency of allele *"a"* at locus one, f_B is the marginal allele frequency of allele *"B"* and f_b is the marginal allele frequency of allele *"b"* at locus two (Balding, Bishop et al. 2007). Sharing a common ancestry is related to the observed level of LD. High level of LD is found when loci are almost linked, and low level of LD is seen in unlinked loci with independent ancestry when r^2 equals to 1, it means there is perfect LD (Balding, Bishop et al. 2007). Two-locus Identity by Descent (IBD) could be used to measure the shared ancestry between two loci and refers to relatedness, it measures the probability of having a single common ancestor at both loci when two chromosomes are selected randomly and is expressed as $\Pr(Q_i^k \equiv Q_{i'}^{k'} \mid G)$, where Q_i^k is allele *k* in individual *I*, G is a set of observed genotypes on the pedigree and ≡ denotes IBD, $i \neq i'$ and $k \neq k'$ (Balding, Bishop et al. 2007). Theoretically, IBD = 1 shows monozygotic twins or duplicates, in first degree relatives IBD equals to 0.5, IBD = 0.25 is expected in second degree relatives and IBD = 0.125 indicates third degree relatives. In this study IBD greater than 0.1875 is considered to identify duplicated or related individuals; it means halfway between third and second degree relatives (Anderson, Pettersson et al. 2010). Therefore, duplicated or related individuals (with r^2 greater than 0.1875) are also excluded in analyses. To detect population stratification or divergent ancestry, between group IBS differences by using permutation test, IBS similarity matrix and multidimensional scaling plots are performed (Anderson, Pettersson et al. 2010, Purcell 2009, Purcell 2010, Purcell, Neale et al. 2007). Population stratification could be detected by using Devlin and Roeder method (Devlin, Roeder 1999). In this method, named Genomic Control

(GC), inflation factor λ value would identify the presence or absence of population substructure; i.e. the expected λ will be 1 if there is no population stratification and λ greater than 1 indicates population structure. A robust estimate of λ is obtained by $\hat{\lambda} = \frac{median(\chi_1^2,\chi_2^2,\ldots,\chi_L^2)}{0.456}$, where $\chi_1^2,\ \chi_2^2,\ldots,\chi_L^2$ are Pearson χ^2tests of L unlinked markers (Devlin, Roeder 1999); therefore genomic inflation factor (λ) based on median χ^2 is also reported. Finally, from the "nearest neighbor" scores distribution, individuals with less than 4 standard deviation units are considered "outliers" and removed before analyses (Purcell, Neale et al. 2007, Purcell 2010).

At marker level, markers with excessive missing genotype rate (SNPs with a call rate of less than 95%), markers that show a significant deviation from Hardy-Weinberg equilibrium (HWE) in controls (SNPs with a P value threshold of less than 0.0001 for HWE) and those with a low minor allele frequency (SNPs with MAF of less than 0.05) are not included in analyses. There was not any SNP with significantly ($p < 0.00001$) different missing genotype rates between cases and controls (Anderson, Pettersson et al. 2010, Purcell 2010).

A pruned subset of SNPs was generated that were in near linkage equilibrium with each other based on the variance inflation factor (VIF). The VIF equals $\frac{1}{1-R^2}$,where R^2 is the multiple correlation coefficient for a SNP regressed on all other SNPs. Three parameters were used to create this subset: 1) window size in SNPs = 50, 2) the number of SNPs to shift the window at each step = 5 and 3) the VIF threshold =2. (Purcell 2010).

After removing individuals with sex discordant, the sex codes were imputed based on the SNP data. To identify duplicated or related individuals, the pairwise IBS metrics, without imposing any thresholds, were calculated.

Since the data provided in the FIND study does not differentiate T1D from T2D (Knowler, Coresh et al. 2005), and the peak incidence of T1D onset is in adolescence and most of these patients are insulin dependent and symptomatic (Herold, Vignali et al. 2013), individuals with age of onset less than 20 who received insulin were excluded before statistical analyses to make the study population more homogenous considering T2D although it could decrease the power of the study; additionally, one female European American with Lab-GFR equivalent to 1112.418 was detected as an extreme outlier and excluded from analyses. To determine SNPs that could be sex dependent, all statistical analyses were conducted once for both sexes together and once for each sex separately. Finally, because the dataset, which is used in this study, provides information for each race separately, quality control metrics would also be applied for each race separately.

II.5. Statistical Analyses

For statistical analyses I used P-link (Purcell 2009), R (R Core Team 2013) and SPSS. In P-link, the functions used for analyses include: check-sex, missing, genome (IBD), remove (sub setting), impute-sex, indep (pruning), extract, exclude (sub setting), hardy (HWE), nonfounders, freq (frequency), pheno (quantitative phenotype i.e. eGFR), ibs (IBS group-difference empirical p-values), cluster, mds-plot, assoc (association) and adjust. In R, the following functions were used to generate the plots and sub sets: read

table, sub setting, scaling plots and creating text files. To create some tables, mean, max/min, standard deviation, frequency functions were used in SPSS.

After filtering applied to the dataset, loci associated with diabetic nephropathy among T2D are identified using several statistical analyses in plink software (Purcell 2009, Purcell, Neale et al. 2007) for both categorical and quantitative traits although the original dataset is based on the dichotomous outcome variable. In the FIND study participants are divided into cases and controls regarding the categorical trait of diabetic nephropathy as outcome variable; therefore, to be able to use a quantitative outcome variable as well, additionally, the data provided in the phenotypic files of FIND study are used; the eGFR values was computed as a continuous (quantitative) outcome variable. Several equations have been developed to estimate GFR; in Cockcroft-Gault method, eGFR in mL/min/1.73 m^2 is computed based on the creatinine clearance in mL/min adjusted to 1.73 m^2. Creatinine clearance (C_{cr}) is derived by using the equation: $C_{cr} = \frac{(140-age)\times(weight)\times(0.85\ if\ female)}{72 \times S_{cr}\ (mg/100\ mL)}$, where C_{cr} is in mL/min, age in years, weight in kilograms, and S_{cr} (serum creatinine) in mg/dL (Cockcroft, Gault 1976).

In the Modification of Diet and Renal Disease (MDRD) study, a four variable formula was developed to estimate GFR as: $GFR = 186 \times (S_{cr})^{-1.154} \times (age)^{-0.203} \times (0.742\ if\ female) \times (1.210\ if\ AA)$, or: $GFR = 175 \times (Standardized\ S_{cr})^{-1.154} \times (age)^{-0.203} \times (0.742\ if\ female) \times (1.210\ if\ AA)$, where GFR is in mL/min/1.73 m^2, S_{cr} in mg/dL, age in years, and AA indicates African American (Levey, Bosch et al. 1999, Levey, Coresh et al. 2007). In the FIND study, four variable MDRD formula was

used to compute eGFR; the eGFR value for those with kidney failure has been set to 5 mL/min/1.73 m^2 (Thameem, Igo et al. 2013).

In this study, association between SNPs and the dichotomous or categorical trait has been analyzed through the Chi-square test; correction for multiple testing has also been considered.

Regarding quantitative outcome, a linear regression model was considered; then, association tests have been conducted for quantitative outcome i.e. eGFR to detect the top associated SNPs with the continuous trait and by using a regression coefficient (β) and a multiple correlation coefficient (R^2) contingency table method, i.e. a Chi-square test, the associations have been adjusted for multiple testing.

To prevent elongation, only the log files of p-link analyses in African Americans are stated in appendix 3. The other log files are available upon request.

II.6. Results

II.6.1. Genome-wide association scans for single SNPs associated with diabetic nephropathy in type 2 diabetics - FIND study:

II.6.1.1. African Americans

The number of SNPs and individuals before and after filtering that are indicated in the section "II.4." are summarized in table 6.

Table 6: The Results of Quality Control Metrics in AAs

	Before Filtering	After Filtering
Number of SNPs	932534	782810
Number of Cases	581	426
Number of Controls	278	209
Number of Males	291	199
Number of Females	573	436

If cases and controls are selected from a population with different ancestries and allele and disease frequency rates differ in subpopulations, the results could be false positive (Marchini, Cardon et al. 2004, Wu, DeWan et al. 2011); therefore, population stratification is a main concern in GWAS case-control studies. As mentioned in section "II.4.", permutation test to detect between group IBS differences and multidimensional scaling plots have been used in this study to assess population stratification (Purcell, Neale et al. 2007, Anderson, Pettersson et al. 2010, Purcell 2010). To test the population structure, permutation test for between group IBS differences was conducted; after calculating the pairwise IBS distances among all individuals, assuming a binary phenotype, the differences between and within cases and controls could be verified (Purcell 2009, Purcell 2010). The result is reported in the table 7, and figure 1 shows the multidimensional scaling plots.

Table 7: Permutation Test, IBS Differences among all AAs

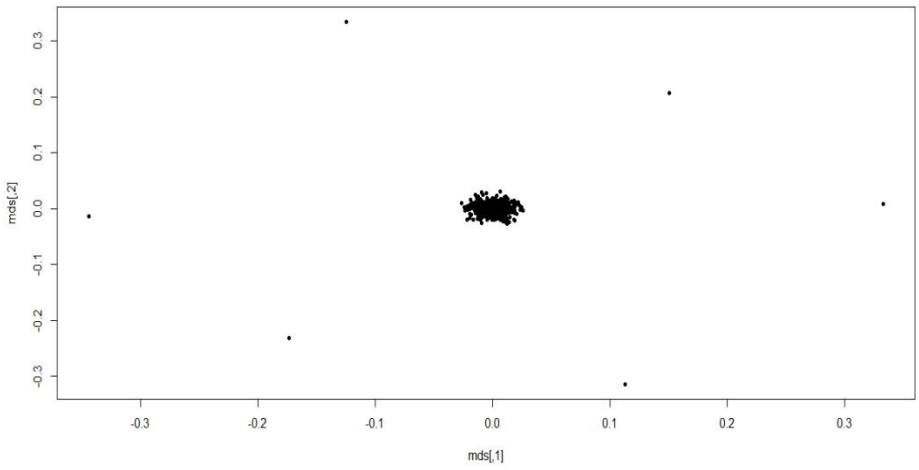
Between-group IBS (mean, SD) = 0.817981, 0.00336465 In-group (2) IBS (mean, SD) = 0.81839, 0.00293358 In-group (1) IBS (mean, SD) = 0.817598, 0.00367504 Approximate proportion of variance between group = 0.00155461		
IBS group-difference empirical p-values		
No	Description	Empirical p-value
T1	Case/control less similar	0.0137299
T2	Case/control more similar	0.98628
T3	Case/case less similar than control/control	0.98632
T4	Case/case more similar than control/control	0.0136899
T5	Case/case less similar	0.98635
T6	Case/case more similar	0.0136599
T7	Control/control less similar	0.0136299
T8	Control/control more similar	0.98638
T9	Case/case less similar than case/control	0.9863
T10	Case/case more similar than case/control	0.0137099
T11	Control/control less similar than case/control	0.9816
T12	Control/control more similar than case/control	0.0184098
Between-group: case/control, group 1: control/control, group 2: case/case		

Using a fixed 10,000 permutations, P-link permutes case/control label, and recalculates between-group metrics considering average IBS within that group. The "Permutation Test, IBS Differences" tables show mean and standard deviation of each group first; and then, the proportion of variance between groups is computed following by 12 separate tests with self-description labels. This table could be used to identify whether cases and controls are from two distinct populations or not (Purcell 2009, Purcell 2010).

To display any potential clustering within a population, "$--mds-plot\ k$" function together with "$--cluster$" function, embedded in P-link, creates a file that contains one row per individual including Family ID, Individual ID, Position on the first dimension, Position on the second dimension and so on, depending on the defined number of

dimensions "k" to be extracted. Plotting the values on one dimension against the values on the other one will generate a scatter plot that could be useful to detect any clustering; it is supposed that when population stratification exists, separate clusters would be formed (Purcell 2009, Purcell 2010). In “Multidimensional scaling plots” figures, “X” axis represents the position on the first dimension denoted by “mds[, 1]” and “Y” axis is the position on the second dimension denoted by “mds[, 2].

Figure 1: Multidimensional scaling plots in all AAs



Considering the empirical p-values, approximate proportion of variance between group, multidimensional plots, which reveals one cluster, and the value of λ, equals to 1.03567, it could be stated that population stratification does not exist in this sample.

In GWAS, multiple tests to assess associations between phenotype and genotype, multiple genetic loci and multiple phenotypes result in testing a large number of hypotheses that in turn could cause increase of the type I error rate (So, Sham 2011). Therefore, several methods developed for multiple-testing correction to estimate the

significant level in GWAS; The Wellcome Trust Case Control Consortium: 5×10^{-7}(Wellcome Trust Case Control Consortium 2007), Dudbridge and Gusnanto: 7.2×10^{-8} (Dudbridge, Gusnanto 2008) and Bonferroni correction: 5×10^{-8}(So, Sham 2011).

The aim of this study is to detect SNPs that are significantly associated with DN and or eGFR, considering Bonferroni correction, although no SNP showed significant association after correction for multiple testing; the top associated SNP was (rs1285582, χ^2: 29.55, unadjusted p-value: 5.46×10^{-8}, OR: 0.42) located on the long arm of chromosome X, the genomic inflation factor (λ: 1.03567), MAF: 0.14 and 0.27 in cases and controls respectively; its function consequence is the nc transcript variant, the utr variant 3 prime; the related gene is (NUP62CL) nucleoporin 62kDa C-terminal like (NCBI 2016). In single-point analysis, the OR compares individual SNP frequencies between cases and controls, OR with value greater than one supports that the SNP is causative for the trait under investigation and OR with value less than one indicates that the SNP is protective considering the disease status; since OR is less than 1, it would be interpreted that exposure to this SNP could be protective. The association analysis identified 14 SNPs with an unadjusted $p < 1 \times 10^{-5}$ (table 8) associated with DN; of those, 6 SNPs are located on chromosome X, 2 SNPS are located on chromosome 6 and 2 SNPs are located on chromosome 3 (table 8). Other significance levels are reported when using "$--adjust$" function in P-link (Purcell 2009, Purcell 2010) including false discovery rate (FDR); Sidak correction, i.e. $\alpha = 1 - (1 - 0.05)^{1/m}$ where m is the number of independent tests and α is the significance level (So, Sham 2011); and non-

parametric stepwise test recommended by Holm (Holm 1979, Dmitrienko, Wiens et al. 2006).

Table 8: Associated SNPs with DN and Unadjusted $p < 1 \times 10^{-5}$ in AAs

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
23	rs1285582	5.46E-08	9.23E-08	0.0427	0.0427	0.0418	0.0418	0.0427	0.6042
23	rs1285715	2.54E-07	4.08E-07	0.1988	0.1988	0.1803	0.1803	0.09939	1
17	rs9904264	1.29E-06	1.96E-06	1	1	0.635	0.635	0.2797	1
4	rs4833161	1.59E-06	2.40E-06	1	1	0.7114	0.7114	0.2797	1
13	rs7330933	1.79E-06	2.69E-06	1	1	0.7531	0.7531	0.2797	1
23	rs2782219	2.98E-06	4.42E-06	1	1	0.9033	0.9033	0.292	1
23	rs2528724	2.98E-06	4.42E-06	1	1	0.9033	0.9033	0.292	1
23	rs2528720	2.98E-06	4.42E-06	1	1	0.9033	0.9033	0.292	1
14	rs1951981	3.60E-06	5.29E-06	1	1	0.9402	0.9402	0.313	1
6	rs9367077	4.70E-06	6.85E-06	1	1	0.9747	0.9747	0.3676	1
6	rs7741325	6.49E-06	9.36E-06	1	1	0.9938	0.9938	0.4575	1
3	rs1828671	8.84E-06	1.26E-05	1	1	0.999	0.999	0.4575	1
3	rs17421075	8.84E-06	1.26E-05	1	1	0.999	0.999	0.4575	1
23	rs5921139	9.50E-06	1.35E-05	1	1	0.9994	0.9994	0.4575	1

CHR: Chromosome number, SNP: SNP identifier, UNADJ: Unadjusted asymptotic significance p-value, GC: Genomic-control corrected p-values, BONF: Bonferroni single-step adjusted p-values, HOLM: Holm (1970) step-down adjusted p-values, SIDAK_SS: Sidak single-step adjusted p-values, SIDAK_SD: Sidak step-down adjusted p-values, FDR_BH: Benjamini & Hochberg (1995) step-up FDR control, FDR_BY: Benjamini & Yekutieli (2001) step-up FDR control.

II.6.1.2. Female/Male African Americans

The population stratification has been investigated among females and males African Americans separately; the results are demonstrated in table 9, table 10, figure 2 and figure 3.

Table 9: Permutation Test, IBS Differences among female AAs

Between-group IBS (mean, SD) = 0.818024, 0.00336845 In-group (2) IBS (mean, SD) = 0.818529, 0.00287222 In-group (1) IBS (mean, SD) = 0.817551, 0.00370617 Approximate proportion of variance between group = 0.00135679		
IBS group-difference empirical p-values		
No	Description	Empirical p-value
T1	Case/control less similar	0.0115199
T2	Case/control more similar	0.98849
T3	Case/case less similar than control/control	0.98916
T4	Case/case more similar than control/control	0.0108499
T5	Case/case less similar	0.989
T6	Case/case more similar	0.0110099
T7	Control/control less similar	0.0104599
T8	Control/control more similar	0.98955
T9	Case/case less similar than case/control	0.98885
T10	Case/case more similar than case/control	0.0111599
T11	Control/control less similar than case/control	0.00944991
T12	Control/control more similar than case/control	0.99056
Between-group: case/control, group 1: control/control, group 2: case/case		

Figure 2: Multidimensional scaling plots in female AAs

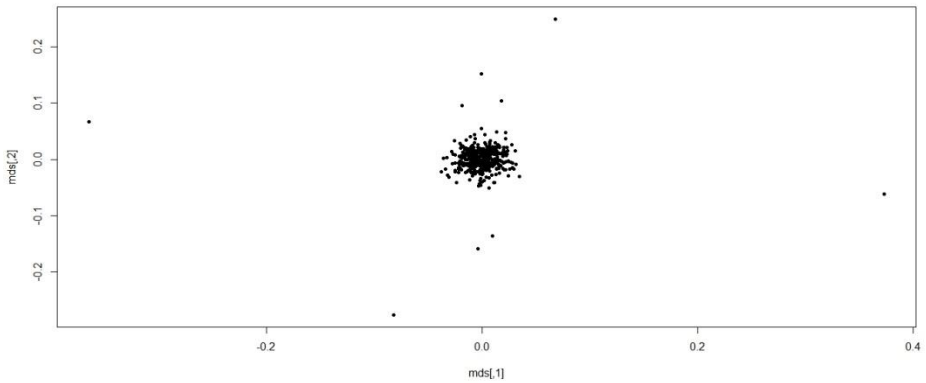
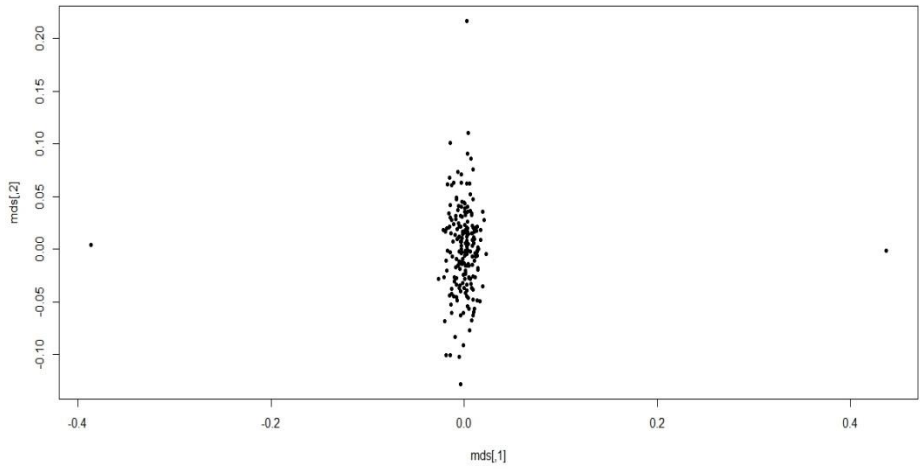


Table 10: Permutation Test, IBS Differences among male AAs

Between-group IBS (mean, SD) = 0.81796, 0.003327 In-group (2) IBS (mean, SD) = 0.818153, 0.00302344 In-group (1) IBS (mean, SD) = 0.817795, 0.00356257 Approximate proportion of variance between group = 0.000629879		
IBS group-difference empirical p-values		
No	Description	Empirical p-value
T1	Case/control less similar	0.275467
T2	Case/control more similar	0.724543
T3	Case/case less similar than control/control	0.710103
T4	Case/case more similar than control/control	0.289907
T5	Case/case less similar	0.715663
T6	Case/case more similar	0.284347
T7	Control/control less similar	0.309337
T8	Control/control more similar	0.690673
T9	Case/case less similar than case/control	0.719433
T10	Case/case more similar than case/control	0.280577
T11	Control/control less similar than case/control	0.743833
T12	Control/control more similar than case/control	0.256177
Between-group: case/control, group 1: control/control, group 2: case/case		

Figure 3: Multidimensional scaling plots in male AAs



Genomic inflation factor (based on median chi-squared) or λ is 1.03136 and 1.01609 in female AAs and male AAs respectively; considering the tests, multidimensional plots and

genomic inflation factor it seems that there is not population stratification in these samples although the empirical p-values in males AAs are not significant.

To determine SNP(s) that might be sex dependent, association tests were conducted on each sex separately; after correction for multiple testing, no SNP was significantly associated. The top associated SNP was (rs6705592, χ^2: 21.80, unadjusted p-value: 3.02×10^{-6}, OR: 2.16) an intron variant located on the SLC8A1 antisense RNA 1 (SLC8A1-AS1) gene on the short arm of chromosome 2 (NCBI 2016), the genomic inflation factor (λ: 1.03136), MAF: 0.34 and 0.19 in cases and controls respectively in females; and (rs3857190, χ^2: 23.80, unadjusted p-value: 1.07×10^{-6}, OR: 0.26) an intergenic variant located on the long arm of chromosome 4 (NCBI 2016), the genomic inflation factor (λ: 1.01609), MAF: 0.12 and 0.34 in cases and controls respectively in males. SNPs with unadjusted $p < 1 \times 10^{-5}$ and are listed in table 11 and table 12.

Table 11: Associated SNPs with DN and Unadjusted $p < 1 \times 10^{-5}$in Female AAs

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
2	rs6705592	3.02E-06	4.27E-06	1	1	0.9058	0.9058	0.6185	1
22	rs2058119	3.50E-06	4.92E-06	1	1	0.9351	0.9351	0.6185	1
21	rs858044	3.69E-06	5.18E-06	1	1	0.9443	0.9443	0.6185	1
2	rs1541568	4.73E-06	6.60E-06	1	1	0.9753	0.9753	0.6185	1
21	rs2406176	9.01E-06	1.23E-05	1	1	0.9991	0.9991	0.6185	1
8	rs10090978	9.48E-06	1.30E-05	1	1	0.9994	0.9994	0.6185	1

Table 12: Associated SNPs with DN and Unadjusted $p < 1 \times 10^{-5}$ in Male AAs

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
4	rs3857190	1.07E-06	1.30E-06	0.8353	0.8353	0.5663	0.5663	0.6324	1
1	rs2786864	2.55E-06	3.06E-06	1	1	0.864	0.864	0.6324	1
1	rs2744778	2.55E-06	3.06E-06	1	1	0.864	0.864	0.6324	1
2	rs6437092	6.24E-06	7.39E-06	1	1	0.9924	0.9924	0.6324	1
2	rs6437089	6.24E-06	7.39E-06	1	1	0.9924	0.9924	0.6324	1
2	rs6437097	6.24E-06	7.39E-06	1	1	0.9924	0.9924	0.6324	1
2	rs4284824	6.24E-06	7.39E-06	1	1	0.9924	0.9924	0.6324	1
5	rs12656464	6.46E-06	7.65E-06	1	1	0.9936	0.9936	0.6324	1
9	rs967294	8.51E-06	1.00E-05	1	1	0.9987	0.9987	0.6658	1
3	rs1020042	8.51E-06	1.00E-05	1	1	0.9987	0.9987	0.6658	1

II.6.1.3. European Americans

The number of SNPs and individuals before and after filtering are summarized in table 13.

Table 13: The Results of Quality Control Metrics in EAs

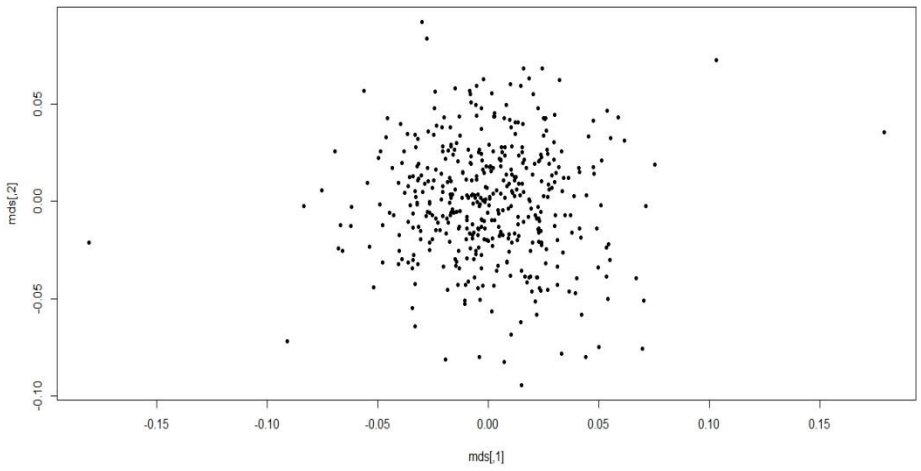
	Before Filtering	After Filtering
Number of SNPs	932534	676504
Number of Cases	339	187
Number of Controls	424	253
Number of Males	349	193
Number of Females	414	247

Permutation test for between group IBS differences was conducted, and the result is described in the table 14. The multidimensional scaling plots are illustrated in figure 4.

Table 14: Permutation Test, IBS Differences among EAs

Between-group IBS (mean, SD) = 0.873521, 0.00169252 In-group (2) IBS (mean, SD) = 0.87357, 0.00146153 In-group (1) IBS (mean, SD) = 0.873474, 0.0018746 Approximate proportion of variance between group = 1.43504e-005		
IBS group-difference empirical p-values		
No	Description	Empirical p-value
T1	Case/control less similar	0.639104
T2	Case/control more similar	0.360906
T3	Case/case less similar than control/control	0.656983
T4	Case/case more similar than control/control	0.343027
T5	Case/case less similar	0.659943
T6	Case/case more similar	0.340067
T7	Control/control less similar	0.345317
T8	Control/control more similar	0.654693
T9	Case/case less similar than case/control	0.671813
T10	Case/case more similar than case/control	0.328197
T11	Control/control less similar than case/control	0.348577
T12	Control/control more similar than case/control	0.651433
Between-group: case/control, group 1: control/control, group 2: case/case		

Figure 4: Multidimensional scaling plots in EAs



None of the empirical p-values of IBS group-difference is significant although the approximate proportion of variance between groups equals to 1.43504e-005. The sample

shows, almost, one cluster in multidimensional plots, and the genomic inflation factor (λ) approximates to 1.01138. Therefore, it could be concluded that population stratification does not exist.

The top associated SNP was (rs4774390,χ^2: 27.67, unadjusted p-value:1.44×10^{-7}, OR: 0.47), an intron variant located in RAR related orphan receptor A (RORA) gene on the long arm of chromosome 15 (NCBI 2016), the genomic inflation factor (λ: 1.01138), MAF: 0.28 and 0.46 in cases and controls respectively; the association was not significant after correction for multiple testing; SNPs with unadjusted $p < 1 \times 10^{-5}$ are listed in the table 15.

Table 15: Associated SNPs with DN and Unadjusted $p < 1 \times 10^{-5}$in EAs

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
15	rs4774390	1.44E-07	1.69E-07	0.09735	0.09735	0.09276	0.09276	0.09735	1
5	rs6884502	6.22E-07	7.19E-07	0.4207	0.4207	0.3434	0.3434	0.2103	1
1	rs4532841	3.09E-06	3.51E-06	1	1	0.8762	0.8762	0.6963	1
14	rs11620655	4.16E-06	4.71E-06	1	1	0.9399	0.9399	0.7031	1
23	rs5905843	5.92E-06	6.68E-06	1	1	0.9818	0.9818	0.8009	1
8	rs919659	7.15E-06	8.05E-06	1	1	0.9921	0.9921	0.8059	1

II.6.1.4. Female/Male European Americans

Population stratification was verified, the results are summarized in table 16, table 17, figure 5, and figure 6.

Table 16: Permutation Test, IBS Differences in female EAs

Between-group IBS (mean, SD) = 0.873696, 0.001642 In-group (2) IBS (mean, SD) = 0.873947, 0.00103401 In-group (1) IBS (mean, SD) = 0.87344, 0.00205764 Approximate proportion of variance between group = 0.00118486		
IBS group-difference empirical p-values		
No	Description	Empirical p-value
T1	Case/control less similar	0.95383
T2	Case/control more similar	0.0461795
T3	Case/case less similar than control/control	0.949571
T4	Case/case more similar than control/control	0.0504395
T5	Case/case less similar	0.947681
T6	Case/case more similar	0.0523295
T7	Control/control less similar	0.0494195
T8	Control/control more similar	0.95059
T9	Case/case less similar than case/control	0.930731
T10	Case/case more similar than case/control	0.0692793
T11	Control/control less similar than case/control	0.0485195
T12	Control/control more similar than case/control	0.95149
Between-group: case/control, group 1: control/control, group 2: case/case		

Figure 5: Multidimensional scaling plots in female EAs

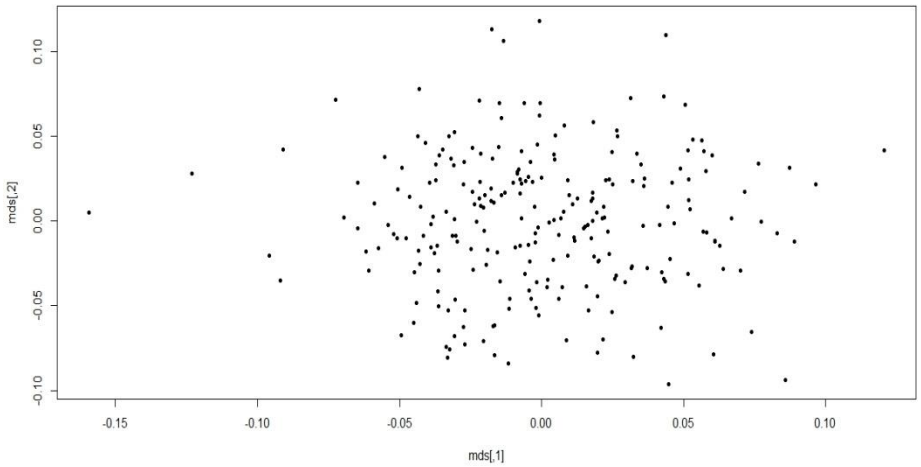
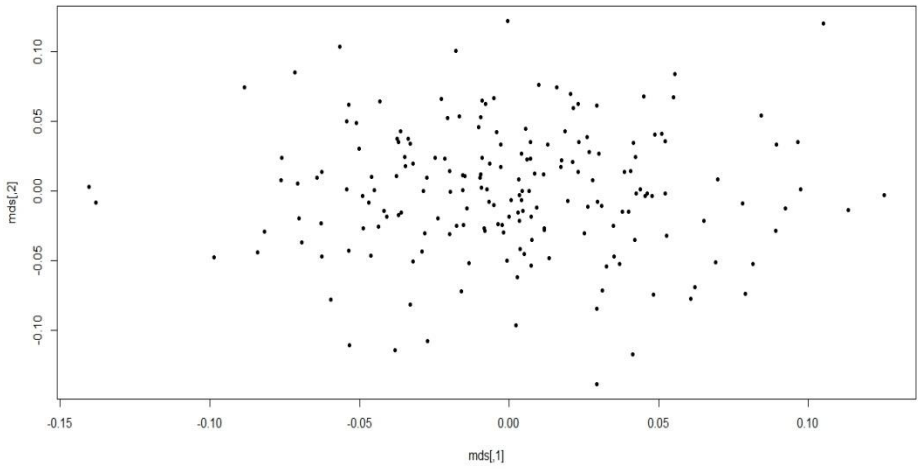


Table 17: Permutation Test, IBS Differences in male EAs

Between-group IBS (mean, SD) = 0.873362, 0.00166999 In-group (2) IBS (mean, SD) = 0.873231, 0.00168415 In-group (1) IBS (mean, SD) = 0.873538, 0.00153057 Approximate proportion of variance between group = 0.000104152		
IBS group-difference empirical p-values		
No	Description	Empirical p-value
T1	Case/control less similar	0.0164398
T2	Case/control more similar	0.98357
T3	Case/case less similar than control/control	0.170078
T4	Case/case more similar than control/control	0.829932
T5	Case/case less similar	0.188898
T6	Case/case more similar	0.811112
T7	Control/control less similar	0.846402
T8	Control/control more similar	0.153608
T9	Case/case less similar than case/control	0.237398
T10	Case/case more similar than case/control	0.762612
T11	Control/control less similar than case/control	0.870571
T12	Control/control more similar than case/control	0.129439
Between-group: case/control, group 1: control/control, group 2: case/case		

Figure 6: Multidimensional scaling plots in male EAs



Genomic inflation factor (λ) was assessed, which is reported to be 1.00 and 1.04522 in female EAs and male EAs respectively. Although male EAs show more scattering

compared to female EAs in multidimensional, considering the λ values, it does not seem that population stratification exists in these samples.

To investigate sex dependent SNP(s), association tests were conducted on each sex separately; after correction for multiple testing, no SNP was significantly associated. The top associated SNPs were (rs2069347, χ^2: 24.98, unadjusted p-value:5.79×10^{-7}, OR: 2.60) an intron variant located on cyclin G1 (CCNG1) gene on the long arm of chromosome 5 (NCBI 2016), the genomic inflation factor (λ: 1.00), MAF: 0.64 and 0.41 in cases and controls respectively in female EAs; and (rs1239911 and rs1239908, χ^2: 26.89, unadjusted p-value:2.16×10^{-7}, OR: 3.04) two intron variants located on the long intergenic non-protein coding RNA 1481 (LINC01481) gene on the long arm of chromosome 12 (NCBI 2016), the genomic inflation factor (λ: 1.04522), MAF: 0.53 and 0.27 in cases and controls in male EAs. SNPs with unadjusted $p < 1 \times 10^{-5}$ are listed in table 18 and table 19.

Table 18: Associated SNPs with DN and Unadjusted $p < 1 \times 10^{-5}$ in female EAs

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
5	rs2069347	5.79E-07	5.79E-07	0.3918	0.3918	0.3241	0.3241	0.2084	1
2	rs4675095	6.57E-07	6.57E-07	0.4447	0.4447	0.359	0.359	0.2084	1
5	rs3756648	1.05E-06	1.05E-06	0.7108	0.7108	0.5088	0.5088	0.2084	1
8	rs4535723	1.23E-06	1.23E-06	0.8335	0.8335	0.5655	0.5655	0.2084	1
8	rs4531033	4.85E-06	4.85E-06	1	1	0.9625	0.9625	0.6565	1

Table 19: Associated SNPs with DN and Unadjusted $p < 1 \times 10^{-5}$ in male EAs

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
12	rs1239911	2.16E-07	3.94E-07	0.146	0.146	0.1358	0.1358	0.073	1
12	rs1239908	2.16E-07	3.94E-07	0.146	0.146	0.1358	0.1358	0.073	1
2	rs10932732	3.00E-06	4.91E-06	1	1	0.8686	0.8686	0.5323	1
2	rs6743834	5.03E-06	8.05E-06	1	1	0.9666	0.9666	0.5323	1
8	rs2118063	5.30E-06	8.47E-06	1	1	0.9723	0.9723	0.5323	1
8	rs1526347	7.01E-06	1.11E-05	1	1	0.9913	0.9913	0.5323	1
12	rs11178059	7.22E-06	1.14E-05	1	1	0.9925	0.9925	0.5323	1
8	rs2693804	8.15E-06	1.28E-05	1	1	0.996	0.996	0.5323	1
8	rs750342	8.63E-06	1.35E-05	1	1	0.9971	0.9971	0.5323	1
8	rs750343	8.66E-06	1.36E-05	1	1	0.9971	0.9971	0.5323	1
8	rs2704255	8.66E-06	1.36E-05	1	1	0.9971	0.9971	0.5323	1

II.6.1.5. Mexican Americans

The number of SNPs and individuals before and after filtering are summarized in table 20.

Table 20: The Results of Quality Control Metrics in MAs

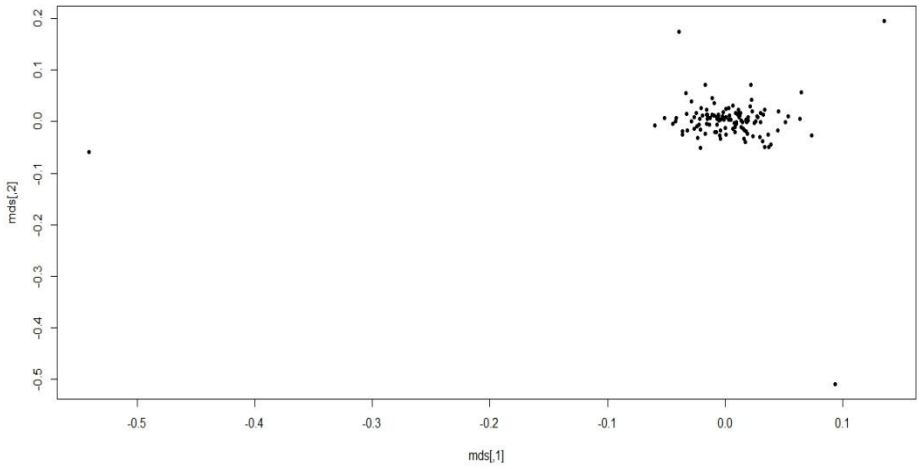
	Before Filtering	After Filtering
Number of SNPs	932534	683140
Number of Cases	534	81
Number of Controls	466	38
Number of Males	397	56
Number of Females	603	63

Population stratification was investigated and the results are presented in table 21 and figure 7.

Table 21: Permutation Test, IBS Differences in MAs

Between-group IBS (mean, SD) = 0.87321, 0.00294932 In-group (2) IBS (mean, SD) = 0.873021, 0.00298606 In-group (1) IBS (mean, SD) = 0.873323, 0.0028272 Approximate proportion of variance between group = 0.000512768		
IBS group-difference empirical p-values		
No	Description	Empirical p-value
T1	Case/control less similar	0.698663
T2	Case/control more similar	0.301347
T3	Case/case less similar than control/control	0.331847
T4	Case/case more similar than control/control	0.668163
T5	Case/case less similar	0.324377
T6	Case/case more similar	0.675633
T7	Control/control less similar	0.651613
T8	Control/control more similar	0.348397
T9	Case/case less similar than case/control	0.316237
T10	Case/case more similar than case/control	0.683773
T11	Control/control less similar than case/control	0.0607294
T12	Control/control more similar than case/control	0.939281
Between-group: case/control, group 1: control/control, group 2: case/case		

Figure 7: Multidimensional scaling plots in MAs



The genomic inflation factor (λ) is equivalent to 1.02805; considering the λ value and one cluster illustrated in multidimensional scaling plot, it could be concluded that the sample does not include population stratification.

The top associated SNP was (rs17067207, χ^2: 23.42, unadjusted p-value: 1.30×10^{-6}, OR: 0.20), the genomic inflation factor (λ: 1.02805), MAF: 0.10 and 0.37 in cases and controls respectively. This SNP is an intergenic variant between RP11-67M9.1 and CTB-7E3.1 genes located on the long arm of the chromosome 5 (NCBI 2016); the association was not significant after correction for multiple testing. SNPs with unadjusted $p < 1 \times 10^{-5}$ in this group are listed in the table 22.

Table 22: SNPs Associated with DN and Unadjusted $p < 1 \times 10^{-5}$in MAs

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
5	rs17067207	1.30E-06	1.81E-06	0.8882	0.8882	0.5886	0.5886	0.2994	1
5	rs10516017	1.97E-06	2.71E-06	1	1	0.7393	0.7393	0.2994	1
5	rs2964303	2.16E-06	2.98E-06	1	1	0.772	0.772	0.2994	1
10	rs190411	4.01E-06	5.43E-06	1	1	0.9353	0.9353	0.2994	1
7	rs11977436	4.64E-06	6.26E-06	1	1	0.9581	0.9581	0.2994	1
3	rs2222630	5.00E-06	6.73E-06	1	1	0.9672	0.9672	0.2994	1
3	rs10511393	5.00E-06	6.73E-06	1	1	0.9672	0.9672	0.2994	1
3	rs16830002a	5.00E-06	6.73E-06	1	1	0.9672	0.9672	0.2994	1
3	rs16830002b	5.00E-06	6.73E-06	1	1	0.9672	0.9672	0.2994	1
3	rs16830004	5.00E-06	6.73E-06	1	1	0.9672	0.9672	0.2994	1
3	rs4407429	5.00E-06	6.73E-06	1	1	0.9672	0.9672	0.2994	1
5	rs2161413	5.70E-06	7.64E-06	1	1	0.9796	0.9796	0.2994	1
5	rs349608	5.70E-06	7.64E-06	1	1	0.9796	0.9796	0.2994	1
3	rs6806039	8.62E-06	1.14E-05	1	1	0.9972	0.9972	0.3694	1
3	rs4680068	8.62E-06	1.14E-05	1	1	0.9972	0.9972	0.3694	1
5	rs2923173	8.65E-06	1.15E-05	1	1	0.9973	0.9973	0.3694	1

II.6.1.6. Female/Male Mexican Americans

Population stratification was assessed and the results are reported in the table 23, table 24, figure 8 and figure 9.

Table 23: Permutation Test, IBS Differences in female MAs

Between-group IBS (mean, SD) = 0.87255, 0.00274349 In-group (2) IBS (mean, SD) = 0.872487, 0.00318981 In-group (1) IBS (mean, SD) = 0.872496, 0.00216365 Approximate proportion of variance between group = 0.000112904		
IBS group-difference empirical p-values		
No	Description	Empirical p-value
T1	Case/control less similar	0.572214
T2	Case/control more similar	0.427796
T3	Case/case less similar than control/control	0.482315
T4	Case/case more similar than control/control	0.517695
T5	Case/case less similar	0.471855
T6	Case/case more similar	0.528155
T7	Control/control less similar	0.501555
T8	Control/control more similar	0.498485
T9	Case/case less similar than case/control	0.460255
T10	Case/case more similar than case/control	0.539755
T11	Control/control less similar than case/control	0.328897
T12	Control/control more similar than case/control	0.671113
Between-group: case/control, group 1: control/control, group 2: case/case		

Figure 8: Multidimensional scaling plots in female MAs

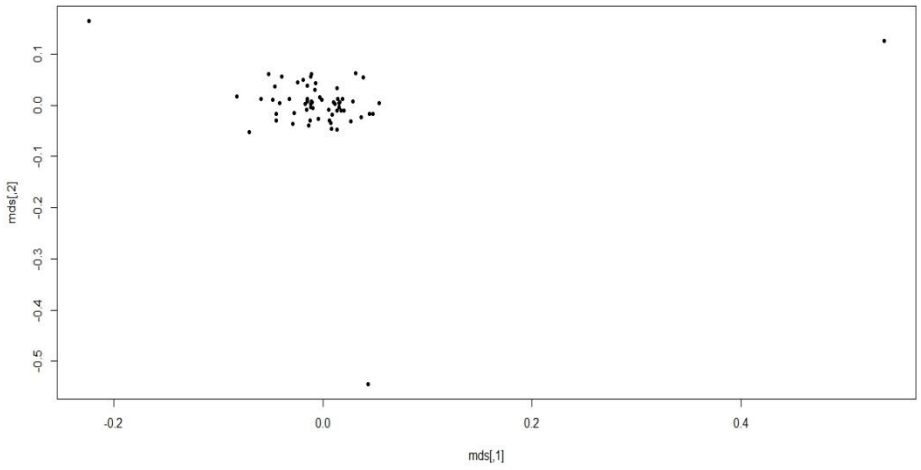
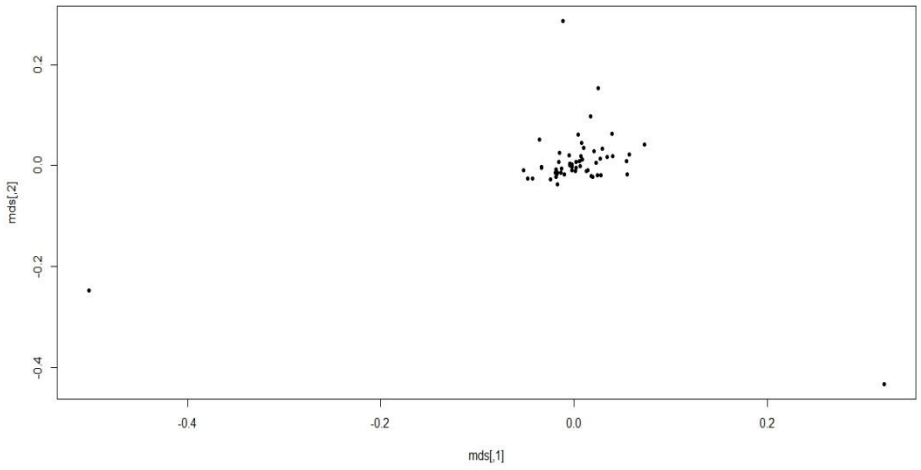


Table 24: Permutation Test, IBS Differences in male MAs

Between-group IBS (mean, SD) = 0.874189, 0.00316636 In-group (2) IBS (mean, SD) = 0.873506, 0.00267958 In-group (1) IBS (mean, SD) = 0.874757, 0.00357518 Approximate proportion of variance between group = 0.00873541		
IBS group-difference empirical p-values		
No	Description	Empirical p-value
T1	Case/control less similar	0.870051
T2	Case/control more similar	0.129959
T3	Case/case less similar than control/control	0.146329
T4	Case/case more similar than control/control	0.853681
T5	Case/case less similar	0.140839
T6	Case/case more similar	0.859171
T7	Control/control less similar	0.837142
T8	Control/control more similar	0.162868
T9	Case/case less similar than case/control	0.136329
T10	Case/case more similar than case/control	0.863681
T11	Control/control less similar than case/control	0.099259
T12	Control/control more similar than case/control	0.900751
Between-group: case/control, group 1: control/control, group 2: case/case		

Figure 9: Multidimensional scaling plots in male MAs



The genomic inflation factor (λ) equals to 1.02048 and 1.00251 in female and male MAs respectively. The λ value and clustering are in the favor of no population stratification in the samples although the number of individuals included in the analyses reduced considerably after constraining quality control metrics.

To discover SNP(s) that might be sex dependent, association tests were conducted on each sex separately; after correction for multiple testing, no SNP was significantly associated with DN. The top associated SNPs were (rs2923173,χ^2: 19.00, unadjusted p-value: 1.31×10^{-5}, OR: 0.13), which is an intergenic variant between RP11-67M9.1 and CTB-7E3.1 genes located on the long arm of the chromosome 5 (NCBI 2016), the genomic inflation factor (λ: 1.02048), MAF: 0.08 and 0.40 in cases and controls respectively in females; and (rs1106228,χ^2: 24.74, unadjusted p-value: 6.55×10^{-7}, OR: 0.03) an intergenic variant located on the long arm of the chromosome 9 between RP11-343J18.2 and RP11-343J18.1 genes (Ensembl 2016), the genomic inflation factor (λ:

1.00251), MAF: 0.01 and 0.32 in cases and controls respectively in males. Only, one SNP in male MAs was detected with unadjusted $p < 1 \times 10^{-5}$ which is listed in table 25.

Table 25: SNP with Unadjusted $p < 1 \times 10^{-5}$ Associated with DN in male MAs

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
3	Rs1106228	6.55E-07	6.77E-07	0.4476	0.4476	0.3608	0.3608	0.4476	1

The top associated SNPs with DN (unadjusted $p < 1 \times 10^{-5}$) in each ethnic group, in this study, are listed in the table 26 although no SNP reached Bonferroni significance level.

Table 26: Top Associated SNPs with DN in AAs, EAs and MAs

Population	SNP	Variant	CHR	Position	Gene/Nearest gene(s)
African Americans					
All together	rs1285582	nc transcript variant,utr variant 3	Xq	107.1	NUP62CL
Females	rs6705592	intron variant	2p	40.0	SLC8A1-AS1
Males	rs3857190	intergenic variant	4q	7.1	FLJ36777, SORCS2
European					
All together	rs4774390	intron variant	15q	61.2	RORA
Females	rs2069347	intron variant	5q	163.4	CCNG1
Males	rs1239908	intron variant	12q	70.1	LINC01481
Mexican Americans					
All together	rs17067207	Intergenic variant	5q	166.6	RP11-67M9.1, CTB-7E3.1
Females	rs2923173	intergenic variant	5q	116.5	RP11-67M9.1, CTB-7E3.1
Males	rs1106228	intergenic variant	9q	126.2	RP11-343J18.2, RP11-343J18.1

II.6.2. Genome-wide association scans for single SNPs associated with glomerular filtration rate (GFR) in T2D - FIND study:

To be able to detect association with a quantitative outcome variable (i.e. GFR value), the values measured as laboratory eGFR provided in the phenotypic files have been used in a

linear regression model for further analyses; Bonferroni correction is considered as significance level.

II.6.2.1. African Americans

The top associated SNP with eGFR value was rs1285582 (unadjusted p-value: 9.97×10^{-8}, β: 14.07; β indicates regression coefficient, and p-value is asymptotic significance value for coefficient); nc transcript variant, utr variant 3 prime located on the nucleoporin 62kDa C-terminal like (NUP62CL) gene located on the long arm of chromosome X (NCBI 2016). The results of the analysis of this SNP and those with p-value < 1×10^{-5} are summarized in table 27.

Table 27: SNPs with Unadjusted $p < 1 \times 10^{-5}$ Associated with eGFR in AAs

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
23	rs1285582	9.97E-08	1.10E-07	0.07801	0.07801	0.07505	0.07505	0.07801	1
23	rs1285715	9.25E-07	1.01E-06	0.7242	0.7242	0.5153	0.5153	0.3621	1
23	rs12559704	2.22E-06	2.40E-06	1	1	0.8239	0.8239	0.4415	1
9	rs7031414	2.79E-06	3.01E-06	1	1	0.8871	0.8871	0.4415	1
4	rs4833161	3.51E-06	3.79E-06	1	1	0.9359	0.9359	0.4415	1
3	rs9843448	4.23E-06	4.56E-06	1	1	0.9634	0.9634	0.4415	1
10	rs11011417	5.73E-06	6.16E-06	1	1	0.9887	0.9887	0.4415	1
10	rs11817782	6.20E-06	6.67E-06	1	1	0.9922	0.9922	0.4415	1
10	rs11011437	6.20E-06	6.67E-06	1	1	0.9922	0.9922	0.4415	1
10	rs11011415	6.20E-06	6.67E-06	1	1	0.9922	0.9922	0.4415	1
10	rs12267847	6.20E-06	6.67E-06	1	1	0.9922	0.9922	0.4415	1
10	rs11011299	8.41E-06	9.03E-06	1	1	0.9986	0.9986	0.4887	1
6	rs4413625	8.69E-06	9.32E-06	1	1	0.9989	0.9989	0.4887	1
12	rs7970296	8.74E-06	9.38E-06	1	1	0.9989	0.9989	0.4887	1

II.6.2.2. Sex-specific association with GFR in African Americans

To identify potential SNPs that their expression might be sex dependent, each sex was evaluated separately and the results for those SNPs with unadjusted p-value $< 1 \times 10^{-5}$ are presented in table 28 and table 29. Although no SNP was significantly associated with eGFR value after correction for multiple testing (Bonferroni method), the top associated one was rs17820651 with unadjusted p-value: 3.57×10^{-6}, a downstream gene variant located on the sulfotransferase family 1C member 3 (SULT1C3) gene on the long arm of the chromosome 2, and rs243294 with unadjusted p-value:9.81×10^{-7}, an intron variant on the neuronal PAS domain protein 3 (NPAS3) gene located on the long arm of chromosome 14 in females and males respectively.

Table 28: SNPs Associated with eGFR/ Female AAs with Unadjusted $p < 1 \times 10^{-5}$

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
2	rs17820651	3.57E-06	3.76E-06	1	1	0.9389	0.9389	0.4927	1
9	rs1759414	5.55E-06	5.83E-06	1	1	0.987	0.987	0.4927	1
10	rs11011417	6.56E-06	6.89E-06	1	1	0.9941	0.9941	0.4927	1
22	rs16998591	6.93E-06	7.28E-06	1	1	0.9956	0.9956	0.4927	1
10	rs12267847	7.86E-06	8.25E-06	1	1	0.9979	0.9979	0.4927	1
10	rs11011415	7.86E-06	8.25E-06	1	1	0.9979	0.9979	0.4927	1
10	rs11817782	7.86E-06	8.25E-06	1	1	0.9979	0.9979	0.4927	1
10	rs11011299	7.86E-06	8.25E-06	1	1	0.9979	0.9979	0.4927	1
10	rs11011437	7.86E-06	8.25E-06	1	1	0.9979	0.9979	0.4927	1
8	rs318873	8.16E-06	8.56E-06	1	1	0.9983	0.9983	0.4927	1
1	rs12117655	8.31E-06	8.72E-06	1	1	0.9985	0.9985	0.4927	1
23	rs9887370	8.99E-06	9.43E-06	1	1	0.9991	0.9991	0.4927	1
10	rs12266900	9.26E-06	9.71E-06	1	1	0.9993	0.9993	0.4927	1

Table 29: SNPs Associated with eGFR/ Male AAs with Unadjusted $p < 1 \times 10^{-5}$

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
14	rs243294	9.81E-07	1.03E-06	0.7681	0.7681	0.5361	0.5361	0.4479	1
12	rs28739424	1.17E-06	1.22E-06	0.9123	0.9123	0.5984	0.5984	0.4479	1
4	rs3857190	1.72E-06	1.79E-06	1	1	0.7391	0.7391	0.4479	1
20	rs6035372	5.02E-06	5.22E-06	1	1	0.9803	0.9803	0.5385	1
2	rs4284824	5.50E-06	5.72E-06	1	1	0.9865	0.9865	0.5385	1
2	rs6437089	5.50E-06	5.72E-06	1	1	0.9865	0.9865	0.5385	1
2	rs6437092	5.50E-06	5.72E-06	1	1	0.9865	0.9865	0.5385	1
2	rs6437097	5.50E-06	5.72E-06	1	1	0.9865	0.9865	0.5385	1
2	rs4973069	7.44E-06	7.72E-06	1	1	0.997	0.997	0.5702	1
2	rs17027920	7.91E-06	8.21E-06	1	1	0.9979	0.9979	0.5702	1
3	rs13322971	8.01E-06	8.32E-06	1	1	0.9981	0.9981	0.5702	1

II.6.2.3. European Americans

When both sexes in European Americans group were analyzed together, the top associated one was rs4774390, unadjusted p-value: 3.64×10^{-7}, an intron variant located on the RAR related orphan receptor A (RORA) gene on the long arm of chromosome 15; table 30 listed those SNPs with $p < 1 \times 10^{-5}$ although no SNP was significantly associated after Bonferroni correction for multiple testing (NCBI 2016, Ensembl 2016).

Table 30: SNPs Associated with eGFR in EAs with Unadjusted $p < 1 \times 10^{-5}$

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
15	rs4774390	3.64E-07	3.73E-07	0.2461	0.2461	0.2182	0.2182	0.1981	1
3	rs1373814	9.29E-07	9.50E-07	0.6287	0.6287	0.4667	0.4667	0.1981	1
3	rs1373813	9.29E-07	9.50E-07	0.6287	0.6287	0.4667	0.4667	0.1981	1
3	rs2120925	1.23E-06	1.26E-06	0.8316	0.8316	0.5647	0.5647	0.1981	1
3	rs9874828	1.48E-06	1.51E-06	1	1	0.6327	0.6327	0.1981	1
3	rs11715833	1.76E-06	1.79E-06	1	1	0.6954	0.6954	0.1981	1
15	rs930847	2.08E-06	2.12E-06	1	1	0.7545	0.7545	0.2006	1
3	rs10935157	3.21E-06	3.27E-06	1	1	0.8858	0.8858	0.2411	1
3	rs9877871	3.21E-06	3.27E-06	1	1	0.8858	0.8858	0.2411	1

II.6.2.4. Sex Specific European Americans

The SNPs associated with eGFR value with unadjusted p-value less than 1×10^{-5} in female European Americans are listed in the table 31; top associated one is rs4675095, an intron variant located on the insulin receptor substrate 1 (IRS1) gene on the long arm of chromosome 2 (NCBI 2016, Ensembl 2016).

Table 31: SNPs Associated with eGFR and Unadjusted $p < 1 \times 10^{-5}$ in EA Females

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
2	rs4675095	5.62E-06	5.91E-06	1	1	0.9777	0.9777	0.9011	1
15	rs4774390	5.66E-06	5.95E-06	1	1	0.9782	0.9782	0.9011	1
5	rs2069347	8.65E-06	9.08E-06	1	1	0.9971	0.9971	0.9011	1

In male European Americans, SNPs associated with GFR having unadjusted p-value less than 1×10^{-5}, are listed in Table 32; no SNP was significantly associated after Bonferroni correction for multiple testing. The top associated one was rs2118063, p-

value: 7.46×10^{-7}; an intergenic variant between syndecan 2 (SDC2) and carboxypeptidase Q (CPQ) genes located on the long arm of chromosome 8.

Table 32: SNPs Associated with eGFR in Male EAs, Unadjusted $p < 1 \times 10^{-5}$

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
8	rs2118063	7.46E-07	9.94E-07	0.5049	0.5049	0.3964	0.3964	0.183	1
8	rs10086882	1.02E-06	1.34E-06	0.6864	0.6864	0.4966	0.4966	0.183	1
9	rs10781011	1.04E-06	1.37E-06	0.6999	0.6999	0.5034	0.5034	0.183	1
17	rs7226040	1.83E-06	2.39E-06	1	1	0.7095	0.7095	0.183	1
8	rs2704255	1.90E-06	2.48E-06	1	1	0.723	0.723	0.183	1
8	rs750343	1.90E-06	2.48E-06	1	1	0.723	0.723	0.183	1
8	rs750342	2.03E-06	2.65E-06	1	1	0.7471	0.7471	0.183	1
8	rs2018041	2.16E-06	2.82E-06	1	1	0.7687	0.7687	0.183	1
21	rs2835611	7.77E-06	9.86E-06	1	1	0.9948	0.9948	0.5256	1
21	rs2835583	7.77E-06	9.86E-06	1	1	0.9948	0.9948	0.5256	1

II.6.2.5. Mexican Americans

The association test in Mexican Americans doesn't detect any SNPs significantly associated with GFR. The SNPs with unadjusted p value less than 1×10^{-5} are summarized in table 33, 34 and 35. The top associated one, when both sexes were included, was rs6806039, an intron variant on the clarin 1 (CLRN1) gene with p-value: 1.83×10^{-8} located on the long arm of chromosome 3(Ensembl 2016, NCBI 2016).

II.6.2.6. Sex Specific Mexican Americans

The primary sex specific analysis showed that seven SNPs achieved unadjusted p-value less than 1×10^{-5} in females; the top associated SNP in this group was rs10949778 with

p-value: 9.11×10^{-8}, a downstream gene variant located on the long arm chromosome 7; in male Mexican Americans, twelve SNPs were associated with GFR with p-value$<$ 1×10^{-5}; the top one was rs4904196, an intergenic variant located on the long arm of chromosome 14 with unadjusted p-value:3.75×10^{-7}.

Table 33: SNPs with Unadjusted $p < 1 \times 10^{-5}$Associated with eGFR in MAs

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
3	rs6806039	1.83E-08	1.83E-08	0.01252	0.01252	0.01244	0.01244	0.006261	0.08772
3	rs4680068	1.83E-08	1.83E-08	0.01252	0.01252	0.01244	0.01244	0.006261	0.08772
6	rs2235406	1.75E-06	1.75E-06	1	1	0.6982	0.6982	0.2886	1
5	rs9313277	4.29E-06	4.29E-06	1	1	0.9465	0.9465	0.2886	1
5	rs12233967	4.29E-06	4.29E-06	1	1	0.9465	0.9465	0.2886	1
7	rs11769954	4.72E-06	4.72E-06	1	1	0.9602	0.9602	0.2886	1
10	rs11015025	5.72E-06	5.72E-06	1	1	0.9799	0.9799	0.2886	1
3	rs2222630	5.81E-06	5.81E-06	1	1	0.981	0.981	0.2886	1
3	rs4407429	5.81E-06	5.81E-06	1	1	0.981	0.981	0.2886	1
3	rs10511393	5.81E-06	5.81E-06	1	1	0.981	0.981	0.2886	1
3	rs16830002a	5.81E-06	5.81E-06	1	1	0.981	0.981	0.2886	1
3	rs16830002b	5.81E-06	5.81E-06	1	1	0.981	0.981	0.2886	1
3	rs16830004	5.81E-06	5.81E-06	1	1	0.981	0.981	0.2886	1
5	rs12187731	6.24E-06	6.24E-06	1	1	0.9859	0.9859	0.2886	1
6	rs2235412	6.34E-06	6.34E-06	1	1	0.9868	0.9868	0.2886	1

Table 34: SNPs Associated with eGFR, Unadjusted $p < 1 \times 10^{-5}$ in Female MAs

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
7	rs10949778	9.11E-08	9.11E-08	0.06219	0.06219	0.0603	0.0603	0.06219	0.8714
7	rs10272697	7.06E-07	7.06E-07	0.4822	0.4822	0.3826	0.3826	0.2411	1
20	rs17789761	3.06E-06	3.06E-06	1	1	0.8759	0.8759	0.5331	1
14	rs2152977	3.61E-06	3.61E-06	1	1	0.915	0.915	0.5331	1
7	rs7785498	3.90E-06	3.90E-06	1	1	0.9304	0.9304	0.5331	1
7	rs41434053	4.76E-06	4.76E-06	1	1	0.9612	0.9612	0.5417	1
3	rs7643629	7.51E-06	7.51E-06	1	1	0.9941	0.9941	0.6575	1

Table 35: SNPs Associated with eGFR in Male MAs, Unadjusted $p < 1 \times 10^{-5}$

CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH	FDR_BY
14	rs4904196	3.75E-07	3.75E-07	0.2565	0.2565	0.2262	0.2262	0.2196	1
1	rs12409665	7.50E-07	7.50E-07	0.5123	0.5123	0.4009	0.4009	0.2196	1
8	rs7008234	9.65E-07	9.65E-07	0.6588	0.6588	0.4825	0.4825	0.2196	1
3	rs10934434	2.21E-06	2.21E-06	1	1	0.7797	0.7797	0.3392	1
3	rs13433683	3.67E-06	3.67E-06	1	1	0.9182	0.9182	0.3392	1
3	rs1872239	3.67E-06	3.67E-06	1	1	0.9182	0.9182	0.3392	1
3	rs13314357	3.67E-06	3.67E-06	1	1	0.9182	0.9182	0.3392	1
8	rs17667096	3.97E-06	3.97E-06	1	1	0.9337	0.9337	0.3392	1
6	rs932600	6.43E-06	6.43E-06	1	1	0.9876	0.9876	0.4447	1
3	rs6790274	6.70E-06	6.70E-06	1	1	0.9897	0.9897	0.4447	1
15	rs2243453	9.29E-06	9.29E-06	1	1	0.9982	0.9982	0.4447	1
13	rs17635528	9.37E-06	9.37E-06	1	1	0.9983	0.9983	0.4447	1

The top associated SNPs with eGFR (unadjusted $p < 1 \times 10^{-5}$) in each ethnic group, in this study, are listed in the table 36 although no SNP reached Bonferroni significance level.

Table 36: Top Associated SNPs with eGFR in AAs, EAs and MAs

Population	SNP	Variant	CHR	Position (Mb)	Gene/Nearest gene(s)
African Americans					
All together	rs1285582	nc transcript variant,utr variant 3	Xq	107.1	NUP62CL
Females	rs17820651	Downstream gene variant	2q	108.3	SULT1C3, SULT1C2
Males	rs243294	Intron variant	14q	33.3	NPAS3
European Americans					
All together	rs4774390	Intron variant	15q	61.2	RORA
Females	rs4675095	intron variant	2q	226.8	IRS1
Males	rs2118063	intergenic variant	8q	96.6	SDC2, CPQ
Mexican Americans					
All together	rs4680068	intron variant	3q	150.9	CLRN1
Females	rs10949778	Downstream gene variant	7q	155.4	AC008060.8
Males	rs4904196	Intergenic variant	14q	84.7	MTCYBP27, RNU6-976P

II.6.3. Region Analyses

Further analyses have been performed to identify potential regions that might be plausible for more investigation to detect an association with DN or variation in eGFR value that could not be found primarily by association tests conducted as genome-wide association. The SNPs that were found to be associated with DN or eGFR changes with unadjusted $p - value < 1 \times 10^{-5}$ have been reassessed considering their position and p-values and will be discussed for each group.

II.6.3.1. Assessing regions that might be associated with DN

II.6.3.1.1. African Americans all together:

rs1828671 and rs17421075 (table 8) are intron variants in the transmembrane protein 212 (TMEM212) gene located on the long arm of chromosome 3 (NCBI 2016); previously, Ng & Hester reported that one SNP (i.e. rs6794092) which is located near TMEM212 was associated with BMI values in African Americans (Ng, Hester et al. 2012). There are

evidences that TMEM212 is likely expressed in the brain, one SNP in this gene was determined to be significantly associated with the fusiform face area activation (Brown, Jensen et al. 2012). It was detected as one of the completely differentially regulated genes in response to hypoxia (Baghbani, Raoofian et al. 2013). It has been identified that TMEM212 is associated with blood pressure response to angiotensin II receptor blockers (Turner, Bailey et al. 2012); this gene was found to be, sex dependently, associated with reactivity variables of blood pressure (Nagel 2014).

rs1285715 (table 8) is an utr variant 3 prime, associated with RNA binding motif protein 41 (RBM41) gene located on the long arm of chromosome X (NCBI 2016); RBM41 protein has been found to be associated with epidermal growth factor receptor (Tong, Taylor et al. 2008).

rs1285582 (table 8) is a nc transcript variant, utr variant 3 prime, associated with nucleoporin 62kDa C-terminal like (NUP62CL) gene located on the long arm of chromosome X; the encoded protein is found in nuclear pore complexes (NCBI 2016). It was shown that in breast cancer the inactive X chromosome was epigenetically variable and adjacent genes could perform differently in a cancer environment; for example, while the NUP62CL remained silent in basal-like carcinoma cells, its near neighbor the nuclear transport factor 2-like export factor 2 (NXT2) gene displayed atypical transcription (Chaligné, Popova et al. 2015). It was reported as one of the ciliary candidate genes due to its expression in cells with motile cilia and its function associated likely with such cells (Ivliev, 't Hoen et al. 2012). Both silencing and overexpression of syndecan-1 could alter NUP62CL expression (Szatmári, Mundt et al. 2012).

rs2528720, rs2528724 and rs2782219 (table 8) are Upstream gene variants associated with the sosondowah ankyrin repeat domain family member D (SOWAHD) gene located on the long arm of chromosome X (Ensembl 2016). In the human genome, SOWAHD is head-to-head linked to septin 6 (SEPT6) gene as the result of a single retroposition event (Maeso, Irimia et al. 2012). SEPT6 is required for cytokinesis (NCBI 2016).

II.6.3.1.2. Female African Americans

rs6705592 and rs1541568 (table 11) are intron variants associated with SLC8A1 antisense RNA 1 (SLC8A1-AS1) gene located on the short arm of chromosome 2 (NCBI 2016); this gene has been found to be related with allergic diseases (Ortiz, Barnes 2015). rs858044 and rs2406176 (table 11) are intron variants located on the potassium voltage-gated channel subfamily J member 6 (KCNJ6) gene and transmembrane protease, serine 15 (TMPRSS15) gene respectively; both are positioned on the long arm of chromosome 21 (Ensembl 2016, NCBI 2016). It has been reported that performance of KCNJ6 gene, a putative ATP-sensitive K-channel subunit, is related to insulin release from βcells (Sakura, Bond et al. 1995); this gene is involved in pancreatic functions and has been found to be associated with T2D in Australian Aboriginal population (Anderson, Cordell et al. 2015). TMPRSS15 encodes an enzyme that is involved in the trypsin production (NCBI 2016); it is one of the taxol resistance gene candidates that its expression decreases by androgen receptor silencing (Sun, Huang et al. 2015). TMPRSS15 is related to hirschsprung disease, and its mutations may result in enterokinase deficiency (Liu, Lee et al. 2011).; it is one of the proteases that is expressed essentially in duodenum and jejunum in gastro intestinal tract by encoding enteropeptidase (Faller, Gautschi et al.

2014).; it has been shown that its expression is associated with human cytomegalovirus (HCMV) infection (Michaelis, Barth et al. 2012).

II.6.3.1.3. Male African Americans

rs2744778 and rs2786864 (table 12) are intron variants associated with mannosidase, alpha, class 1C, a member 1 (MAN1C1) gene located on the short arm of chromosome 1 (NCBI 2016). It has been reported that MAN1C1, by encoding a Golgi mannosidase, is associated with risk and size of the uterine leiomayomas in AAs (Aissani, Zhang et al. 2015); it is also involved in the N-glycosylation pathway (Mkhikian, Grigorian et al. 2011), and is related to preeclampsia (Moslehi, Mills et al. 2013). MAN1C1 expression is increased by T cell receptor signaling (Chen, Li et al. 2009).

rs6437092, rs6437089, rs6437097 and rs4284824 (table 12) are intron variants on the inositol polyphosphate-5-phosphatase D (INPP5D) gene located on the long arm of chromosome 2 (Ensembl 2016); immune system malignancies and defects could be occurred due to mutations in this gene (NCBI 2016). This, biologically relevant, gene has been identified to be common in all multiple sclerosis phenotypes (Mahurkar, Moldovan et al. 2013). It has been found that INPP5D expression was enhanced in T2D (Marselli, Thorne et al. 2010). It has been shown that 5-ptases such as INPP5D are involved in macrophage phagocytosis regulation and haemopoietic cell proliferation (Ooms, Horan et al. 2009).

II.6.3.1.4. Female European Americans

rs2069347 and rs3756648 (table 18) are intron variants located on the CCNG1 gene and hyaluronan-mediated motility receptor (HMMR) gene respectively on the long arm of

chromosome 5 (Ensembl 2016). It has been reported that one SNP close to CCNG1 was significantly associated with systolic blood pressure (Org, Eyheramendy et al. 2009). HMMR is expressed in breast tissue and is involved in cell motility (NCBI 2016). Locus on HMMR was found to be associated with increased risk of breast cancer (Sinha, Gruber et al. 2008, Guo, Gao et al. 2011). Atypical expression of this gene was detected in primary and/or advanced prostate cancer (Willard, Koochekpour 2012). In animal models, it has been identified that HMMR is involved in adipogenesis regulation and diabetes pathogenesis (Liu, Bera et al. 2012, Yang, Gurung et al. 2010).

rs4535723 and rs4531033 (table 18) are intergenic variants located between RP11-381I15.1 gene and CTD-2501M5.1 gene on the long arm of chromosome 8 (Ensembl 2016).

II.6.3.1.5. Male European Americans

rs1239911, rs1239908 and rs11178059 (table 19) are intron variants located in the long intergenic non-protein coding RNA 1481 (LINC01481) gene on the long arm of chromosome 12 (Ensembl 2016). rs1526347 and rs2693804 (table 19) are downstream gene variants close to ribosomal protein L23a pseudogene 54 (RPL23AP54) gene on the short arm of chromosome 8. rs2118063, rs750342, rs750343 and rs2704255 (table 19) are intergenic variants between syndecan 2 (SDC2) gene and carboxypeptidase Q (CPQ) gene, located on the long arm chromosome 8 (Ensembl 2016). It has been shown that SDC2 gene is involved in extracellular-receptor matrix interactions and cell adhesion molecules (Olsson, Volkov et al. 2014); its function is correlated with different growth factors pathways, inflammation, and lipid metabolism (Kelder, Verschuren et al. 2014,

Gotte 2003). SDC2 has been identified as one of the DNA methylation-based candidate genes for early detection of colorectal cancer (Mikeska, Craig 2014); it may also play role in whole-body energy metabolism regulation (De Luca, Klimentidis et al. 2010). In one study, interactions between offspring SNPs and maternal surgical (gastrointestinal bypass surgery for obesity) status was reviewed; CPQ was reported as one of the genes that showed significant interactions for multiple transcripts (Guénard, Lamontagne et al. 2015).

II.6.3.1.6. Mexican Americans

rs2222630 (table 22) is an intron variant located on the CD80 molecule (CD80) gene, rs6806039 and rs4680068 (table 22) are intron variants in the clarin 1 (CLRN1) gene; five SNPs (rs4407429, rs10511393, rs16830002a, rs16830002b, and rs16830004; see table 22) which are intergenic variants between the CD80 and ADP-ribosylarginine hydrolase (ADPRH) genes, are located on the long arm of chromosome 3 (NCBI 2016, Ensembl 2016). rs17067207, rs10516017, rs2964303, rs2161413, rs349608 and rs2923173 (table 22) are intergenic variants located between CTB-7E3.1 gene and CTB-63M22.1 gene on the long arm of chromosome 5 (Ensembl 2016).

It has been reported that CD80 expression increases in obese humans, specifically on cells stimulating T cell proliferation (Deng, Lyon et al. 2013). Numbers of monocytes have been found to be increased in diabetics; greater expression of CD80 in the infiltrating monocytes implies a proinflammatory role for these cells (Espinoza-Jiménez, Peón et al. 2012, Bradshaw, Raddassi et al. 2009). It has been identified that CD80 gene expression enhanced in human keratinocytes in the presence of allergens and irritants

(Wakem, Burns et al. 2000). It has been described that CD80 plays role in the activation of T lymphocytes in hepatocarcinoma cells (Li, Zhu et al. 2006). A polymorphism of CD80 has been reported to be associated with lower risk of multiple sclerosis (Wagner, Sobczynski et al. 2015).

CLRN1 function is associated with the development of the retina and its expression is increased in non-pigmented epithelium compared with pigmented epithelium (Janssen, Gorgels et al. 2012). Mutation in CLRN1 gene has been demonstrated to be associated with Usher's syndrome and related to autosomal recessive retinitis pigmentosa (Khan, Kersten et al. 2011, Zhang, Sanders et al. 2014). It has been reported that this gene plays role in $CD4^+$ T cell activation and Th2 differentiation (Techasintana, Davis et al. 2015). Mutations in this gene have been detected to be associated with higher prevalence of Usher syndrome in the Ashkenazi population (Ness, Ben-Yosef et al. 2003, Ben-Yosef, Ness et al. 2003, Guha, Rosenfeld et al. 2012). Considering its expression in inner ear, its function in the development of inner ear and actin organization, CLRN1 has been identified as one of the candidate genes that probably are involved in non-syndromic hearing loss phenotype (Accetturo, Creanza et al. 2010).

ADPRH is expressed in the cytoplasm, and encodes an enzyme that is involved in removal of ADP-ribose from arginine residues (Laing, Unger et al. 2011). Its expression has been found to be increased in patients, with sleep disturbances, who self-reported improved sleep (Livingston, Rusch et al. 2015). This gene was one of the genes that were only differentially expressed in ankylosing spondylitis compared with systemic lupus erythematosus and systemic sclerosis (Assassi, Reveille et al. 2011)

II.6.3.2. Assessing regions that might be associated with eGFR values

II.6.3.2.1. African Americans

rs1285582 (table 27) is a 3 prime UTR variant in the NUP62CL gene, rs1285715 (table 27) is a 3 prime UTR variant in the RBM41 gene and rs12559704 (table 27) is an intron variant of diaphanous related formin 2 (DIAPH2) gene, which all are located on the long arm of chromosome X; the latter one is related to ovary function (Ensembl 2016, NCBI 2016).

It has been identified that mutations in DIAPH2 are involved in infertility that are associated with hypergonadotrophic hypogonadism and premature ovarian failure (Bione, Sala et al. 1998, The Evian Annual Reproduction (EVAR) Workshop Group 2010, The Evian Annual Reproduction (EVAR) Workshop Group 2010 et al. 2011, Layman 2013); it is also linked to age related macular degeneration (Zheng, Joo et al. 2007, Tolppanen, Nevalainen et al. 2009).

Six SNPs (table 27), located very close to centromere, on the short arm of chromosome 10, have been identified to be associated with eGFR values: rs11011417 an intron variant in RP11-393J16.4 gene, rs11817782 and rs11011437 intron variants in zinc finger protein 33A (ZNF33A) gene, rs11011415 intron variant in zinc finger protein 25 (ZNF25) gene, rs12267847 intergenic variant between zinc finger protein 33C, pseudogene (ZNF33CP) gene and ZNF25 gene, and rs11011299 an intergenic variant between MT-RNR2-like 7 (MTRNR2L7) gene and RP11-258F22.1 gene (Ensembl 2016).

Loci in ZNF33A and ZNF25 have been identified to be associated with bipolar disorder and type 2 diabetes (Ross 2011). Gene ontology analyses revealed that ZNF33A

expression was downregulated in the presence of diabetic nephropathy (Wanic, Krolewski et al. 2013); another study identified that its expression was upregulated in the patients with gastroesophageal reflux disease who stopped proton pump inhibitor treatment (De Vries, Ter Linde et al. 2009). It has been reported that its expression could be sex dependent (Maher, Fu et al. 2009). ZNF25 belongs to one of the two related ZNF gene clusters in pericentromeric region of chromosome 10; these two clusters are located opposite to each other around the centromere that could be occurred as the result of gene duplication and chromosome rearrangement (Tunnacliffe, Liu et al. 1993).

II.6.3.2.2. Female African Americans

Seven SNPs (table 28) that are located on the short arm of chromosome 10 were associated with eGFR values in female AAs. Six out of these are the same as those detected when both sexes were included in analysis (section: II.6.3.2.1.); and the seventh one, i.e. rs12266900, is an intergenic variant between MTRNR2L7 gene and RP11-258F22.1 gene (Ensembl 2016).

II.6.3.2.3.Male African Americans

Five intron variants (rs4284824, rs6437089, rs6437092, rs6437097, and rs4973069; see table 29) in the INPP5D gene were detected to be associated with eGFR values in male AAs; this gene is located on the long arm of chromosome 2 (Ensembl 2016). Four out of these five variants have also been reported to be associated with DN in male AAs in this study (section: II.6.3.1.3.).

II.6.3.2.4. European Americans

Association analysis in this group revealed three intergenic variants (rs1373814, rs1373813, and rs2120925; see table 30) between EPH receptor B1 (EPHB1) gene and RP11-657O9.1 gene, two downstream gene variants (rs10935157 and rs9877871; see table 30) of EPHB1 gene, and two intron variants (rs9874828 and rs11715833; see table 30) on the EPHB1 gene; these SNPs are located on the long arm of chromosome 3 (Ensembl 2016).

It has been stated that EPHB1 expression is associated with schizophrenia in Japanese population (Kushima, Nakamura et al. 2012). It was detected that its expression enhanced in diabetic patients (Broquères-You, Leré-Déan et al. 2012); however, loss of its expression was strongly correlated with colorectal cancer progression (Batlle, Bacani et al. 2005).

II.6.3.2.5.Male European Americans

rs2118063, rs2704255, rs750343, rs750342, and rs2018041 (table 32) are intergenic variants located on the long arm of chromosome 8, between SDC2 gene and CPQ gene; rs10086882 (table 32) is an upstream gene variant to CPQ gene (Ensembl 2016). Four out of the five intergenic variants have also been described to be associated with DN in male EAs in this study (section: II.6.3.1.5.).

II.6.3.2.6. Mexican Americans

rs6806039 and rs4680068 (table 33) are intron variants in the CLRN1 gene, rs2222630 (table 33) is an intron variant in the CD80 gene; rs4407429, rs10511393, rs16830002a, rs16830002b and rs16830004 (table 33) are intergenic variants between CD80 gene and ADPRH gene. These SNPs are located on the long arm of chromosome 3. rs16830002a

and rs16830002b are not listed in any databases; however, rs16830002 is stated to be a regulatory region variant (Ensembl 2016). These SNPs have also been found to be associated with DN in MAs in this study (section: II.6.3.1.6.).

rs9313277 and rs12233967 (table 33) are intergenic variants located on the short arm of chromosome 5, between small nucleolar RNA host gene 18 (SNHG18) gene and taste 2 receptor member 1 (TAS2R1) gene (Ensembl 2016). It was shown that there was association between TAS2R1 gene expression and type 2 diabetes in Amish families (Dotson, Zhang et al. 2008).

rs2235406 and rs2235412 (table 33) are intron variants in the androgen-dependent TFPI-regulating protein (ADTRP) gene located on the short arm of chromosome 6 (Ensembl 2016). The association between ADTRP expression and coronary artery disease has been investigated in several studies (Wang, Xu et al. 2011, Guo, Gu et al. 2012, Huang, Peng et al. 2015); it might be engaged in metabolic disorders and insulin-dependent metabolism control (English, Mandour et al. 2000, Muller, van den Beld et al. 2004).

II.6.3.2.7.Female Mexican Americans

rs10272697 and rs41434053 (table 34) are intron variants located on the long arm of chromosome 7, in the reelin (RELN) gene (Ensembl 2016).

It was identified that RELN gene was associated with Crohn disease (He, Fuller et al. 2013), the age of onset of multiple sclerosis (Baranzini, Wang et al. 2009), subtype of T cell leukemia (Zhang, Ding et al. 2012); and sex-dependently was involved in schizophrenia (Shifman, Johannesson et al. 2008). It was shown that promoter

hypermethylation and decreased expression of this gene was occurred at puberty (Lintas, Persico 2010).

II.6.3.2.8.Male Mexican Americans

rs13433683, rs1872239 and rs13314357 (table 35) are intergenic variants located on the long arm of chromosome 3, between stromal antigen 1 (STAG1) gene and solute carrier family 35 member G2 (SLC35G2) gene (Ensembl 2016).

It was reported that one of the SNPs, significantly associated with three correlated inflammatory biomarker traits, was located in STAG1 gene (Benjamin, Dupuis et al. 2007); it was indicated as one of the candidate genes associated with high density lipoprotein (HDL) concentration (Chasman, Paré et al. 2009).

Although none of the SNPs found in this study has been reported previously to be associated with DN or eGFR values, considering the genes functions and SNPs variants, the regions and the genes that might be more likely to be associated with DN or changes in eGFR values in this study are summarized in table 37 and table 38. These genes could be potential candidate genes for DN or variation in eGFR value, and further investigation, e.g. linkage analysis, is required to assess their association.

Table 37: Genes That Might Be Associated With DN

Region Associated with Diabetic Nephropathy					
Ethnicity	Start (BP)	End (BP)	Length (Kb)	CHR	Gene(s)
AAs	173,064,267	173,064,907	1	3q	TMEM212
	106,195,072	106,253,758	59	Xq	RMB41, NUP62CL
	118,772,536	118,773,182	0.65	Xq	SOWAHD
Female AAs	40,078,519	40,079,478	1	2q	SLC8A1-AS1
	18,641,297	38,073,521	19,432	21q	TMPRSS15, KCNJ6
Male AAs	25,885,839	25,886,611	1	1p	MAN1C1
	233,667,683	233,671,494	4	2q	INPP5D
Female EAs	162,799,773	162,824,472	25	5q	CCNG1, HMMR
	132,279,108	132,299,463	20	8q	RP11-381I15.1, CTD-2501M5.1
Male EAs	68,727,655	68,729,751	2	12q	LINC01481
	5,322,063	5,326,099	4	8p	RPL23AP54
	97,712,467	97,716,597	4	8q	SDC2, CPQ
MAs	120,759,999	120,773,513	14	3q	CD80, ADPRH
	152,140,530	152,146,140	6	3q	CLRN1
	165,887,308	165,950,952	64	5q	CTB-7E3.1, CTB-63M22.1

Table 38: Genes That Might Be Associated With GFR Value

Region Associated with eGFR values					
Ethnicity	Start (BP)	End (BP)	Length (Kb)	CHR	Gene(s)
AAs	96,297,850	106,253,758	9,956	Xq	NUP62CL, RMB41, DIAPH2
	37,974,972	38,392,858	418	10p	RP11-393J16.4, ZNF33A, ZNF25, ZNF33CP, MTRNR2L7
Female AAs	37,974,972	38,392,858	418	10p	RP11-393J16.4, ZNF33A, ZNF25, ZNF33CP, MTRNR2L7
Male AAs	233,667,683	233,671,494	4	2q	INPP5D
EAs	136,443,519	136,503,814	60	3q	EPHB1, RP11-657O9.1
Male EAs	97,712,467	97,725,056	13	8q	SDC2, CPQ
Mas	120,759,999	120,773,513	14	3q	CD80, ADPRH
	152,140,530	152,146,140	6	3q	CLRN1
	9,658,710	9,659,880	1	5p	SNHG18, TAS2R1
	11,851,961	11,862,674	11	6p	ADTRP
Female MAs	103,337,002	103,345,054	8	7q	RELN
Male MAs	137,970,249	138,001,360	31	3q	STAG1, SLC35G2

Chapter 3

Discussion

In this study, the results of a genome-wide association scan in the FIND collection that have been identified to be associated with the risk of DN and changes in GFR values in type 2 diabetes are reported. Since the condition of interest was DN, statistical analyses first were conducted for categorical outcome variable; in addition, to detect a genetic influence, association with a quantitative outcome variable i.e. GFR value was also performed.

Previous studies revealed that regions on chromosome 7q, 3q, 9q and 20p have been linked with DN in Pima Indians (Imperatore, Hanson et al. 1998). Region on chromosome 18q has also been found to be associated with DN in Turkish families (Vardarli, Baier et al. 2002). In African Americans, regions on chromosomes 3q, 7p and 18q have been reported to be linked with DN (Bowden, Colicigno et al. 2004). In Mexican Americans, a region on chromosome 20q and in European Americans a region on chromosome 15q showed evidence of linkage to eGFR in a genome-wide linkage study (Thameem, Igo et al. 2013).

To my knowledge, potential sex dependency with DN or eGFR variation was not investigated in these studies. In this study the strongest association with DN have been detected on the regions located on the 2p, 4q, 5q, 9q, 12q, 15q and Xq (Table 26), and regions located on the 2q, 3q, 7q, 8q, 14q, 15q and Xq are more probably associated with eGFR values (Table 36).

Since case-control designs are mostly used in the GWA studies, correct selection of cases and controls is a crucial issue for the participant ascertainment. Different ascertainment models include population-based case-control, case-parent trios and affected sib-pairs (ASPs). More significant results and greater odds ratio could be found in unrelated cases and controls and case-parent trios compared to ASP designs; however ASPs could show more power for detecting an effect in most models except for additive and some epistatic models (Howson, Barratt et al. 2005). More association could be achieved in familial cases if multiple genes are involved in the development of the disease while genes with small effect could be identified in case-control studies (Li, Boehnke et al. 2006). Familial cases and/or extreme phenotypes (e.g. selecting cases with early age of onset) can increase phenotypic homogeneity.

In the FIND study, diabetes and nephropathy phenotypes were defined for recruitment of participants and the study is based on a case-control design. Both strict and loose criteria have been used for inclusion of the cases which were selected from FIND probands or siblings (Knowler, Coresh et al. 2005). However, since type 1 diabetes has not been differentiated from type 2 diabetes, I excluded those who received insulin as medication and had the age of onset of diabetes less than 20 years from the original participants before performing statistical analyses; therefore it is quite likely that due to misclassification the results have been biased.

A quantitative biological trait, which is heritable and associated with the disease of interest in the population, is defined as an intermediate phenotype or endophenotype. It should be associated with causes of disease and co-segregated within families

(Gottesman, Gould 2003, Preston, Weinberger 2005, Cannon, Keller 2006). For example, plasma levels of adiponectin have been reported as an endophenotype of obesity and metabolic syndrome. Then, a variance components linkage analysis, by reporting LOD scores, can be used to identify linkage of phenotype to the disease of interest. This method also includes a quantitative trait locus (QTL) component (Comuzzie, Funahashi et al. 2001). If the trait of interest is common, e.g. obesity, misclassification bias and inclusion of the controls which may develop the phenotype of interest later could decrease study power; the more stringent case definition is recommended to control this type of misclassification. Since both diabetes and diabetic nephropathy are long lasting conditions, although controls were selected among members who were long term diabetics (≥10 years) with normal renal function, this kind of misclassification bias still could happen.

Other issues in case-control studies include sufficient sample size, cryptic relatedness and population stratification; statistical methods such as Principal Component Analysis (PCA) are used to detect and control potential population stratification (McCarthy, Abecasis et al. 2008). When mating happens between individuals originated from isolated ancestral populations, admixture occurs. If exogenous risk factors have to be adjusted in association studies, control of admixture fracture at individual level may lead to better results than at group level. Unlinked genotypes have been used to detect population substructure based on a Bayesian approach with Markov Chain Monte Carlo (MCMC) method. Because the convergence findings by these methods might not be reliable, the estimated parameters used in these models may not be sensitive, and the computation is

very intensive, Tang et al developed a maximum likelihood estimate (MLE) of individual admixture (IA) by considering IA and ancestral allele frequency as unknown parameters (Tang, Peng et al. 2005).

Common disease common variant (CD/CV) hypothesis is based on the concept that common genetic variations are responsible for the development of common diseases. Small effect size (low penetrance) of such variants and involvement of multiple such common variants in the inheritance of correlated disorder are consequences of CD/CV hypothesis; therefore, the allele frequency in the population and the risk attributed to such allele are important factors which influence the sample size required to detect the effect and the technical method to be used in the genetic laboratory (Bush, Moore 2012). Either direct or indirect association with a trait can be revealed in genetic association studies by using LD. In the direct association, the SNP associated with a phenotype is involved biologically in the development of the disease and is genotyped directly; in this case the SNP is called the functional SNP. If the disease risk SNP is not directly genotyped, a tag SNP in high LD with this SNP could be genotyped and be associated with the trait; in this case the association is named an indirect association. Therefore, in GWAS an association between a SNP and a phenotype should not be considered as the causal variant.

Microarray technology is used to assay about one million SNPs for each individual. The basic concept that four bases in the DNA structure exclusively bind to each other (i.e. adenine with thymine and cytosine with guanine, known as hybridization) is the basic principle used in microarrays. Since sequencing of the Human Genome Project determined the sequence of each SNP, microarrays use this information to identify each

SNP genotype (Affymetrix 2009a, Affymetrix 2009). The two most commonly used technology developed by Illumina and Affymetrix; chip-based genotyping platforms are used in both methods; ordered array of beads and tagging approach is used in Illumina technology, while printed-array format and spacing markers within whole genome is used in Affymetrix chips. Illumina uses longer DNA sequences than Affymetrix. Illumina technology is more expensive, but more specific than Affymetrix method. Next-generation sequencing methods, which are less expensive, will be used for sequencing the whole genome instead of chip-based genotyping platforms in future (Bush, Moore 2012, DiStefano, Taverna 2011).

Different statistical methods are used to represent the results of a genome-wide association study. Quantile-quantile plot (Q-Q plot) is used to compare observed test statistics distribution with the expected distribution under the null hypothesis; the results might be biased if cryptic relatedness and population stratification exist. Single-point analyses, with one degree of freedom, such as the Cochran-Armitage test, are powerful methods for the analysis of GWAS data. In these methods genotype frequencies for each SNP are compared between cases and controls. Adjustment for covariates e.g. PCA can be included in the analyses. However, due to multiple testing which influence type 1 error rates, adjustment for more than one million independent tests results in a desirable α value to be about 5×10^{-8} for reporting significant results. Haplotype-based methods and imputation methods are examples of multi-marker analyses. These methods have more power to detect associations of variants that have not been genotyped directly. Associations of HapMap SNPs, which are not included in the commercial platforms, can

be detected more powerfully by imputation methods (McCarthy, Abecasis et al. 2008). If multiple hypotheses are tested in one study, classical threshold $\alpha = 0.05$ could result in many spurious significant results. The effect size of an association between a marker allele and the trait of interest could be overestimated in the initial studies compared to the further studies which are conducted to replicate the reported results; this phenomenon is known as "Winner's curse" (Kraft 2008). The novel loci showed association with DN and/or eGFR values in this study could be examples of Winner's curse phenomenon. Small sample size influences the study power, incorrect selection of the significance thresholds, analysis methods, selective reporting, conflicts of interest in interpretation, and the definition of replication have been reported to be the potential reasons which could inflate the association findings and in turn lead to false positive results (Ioannidis 2008).

GWAS tries to identify the correlation between allele frequencies and the frequency of a phenotype by comparing cases and controls; however, the differences in allele frequencies could be related to other factors than only to measured outcome. If the population of study includes subgroups with different genetic background, which result in variation of allele frequencies in these subgroups, and if the disease risk also differs in these subgroups, population stratification must be considered as a confounding factor that could increase type 1 error (Cardon, Palmer 2003). Ethnicity matched design, selecting controls from family members, Transmission Disequilibrium Test (TDT), ethnicity adjustment, structured association method, genomic control and PCA are statistical tools which are used to resolve the existence of population stratification (Cardon, Palmer 2003,

Ziegler, Konig 2010). Different statistical methods are used to detect population stratification. In Pritchard and Rosenberg method a number of null and unlinked markers are genotyped in cases and controls, then a χ^2 test is used to compare the difference in allele frequencies between two groups; the sum of the test statistics from each marker is used to test the null hypothesis (i.e. there is no population stratification), if the difference is significant, population stratification is present (Pritchard, Rosenberg 1999). Devlin and Roeder proposed a Bayesian outlier model to estimate the variance inflation factor λ (i.e. $\hat{\lambda} = [median(X_1, X_2, \dots, X_n)/0.675]^2$ to detect population stratification; $\hat{\lambda}$ values more than 1 means there is population stratification, cryptic relatedness or genotyping error. This method in known as "genomic control" and estimated inflation factor $\hat{\lambda}$ is used for correction of the value of the observed χ^2 test statistic (Devlin, Roeder 1999b). Population stratification, cryptic relatedness, and the genetic association with the trait can also be assessed by quantile-quantile plot (Q-Q plot), which compares the observed values with the expected ones (McKnight 2013). In this study, the genomic inflation factor ranges between 1.0013 and 1.0439 in different subgroups; therefore, in the subgroups, population stratification is less likely to be happened.

Small sample size, which influences the study power to detect real effect size, might be the main reason for not being successful to replicate the results of previous GWA studies. Replication and validation methods are used to identify which of the primary GWAS results are true associations; hence potential errors and biases should also be noticed and addressed. Technical validation could be assessed by re-analyzing the primary samples using a different genotyping technology. Using independent samples is another important

issue to test replication. Similar genetic model, same phenotype, and the same allele/haplotype that have been used in the original studies should be considered in designing replication studies. Successful data combination, which is used in meta-analysis studies, is a potential method to increase the study power.

Finally, it should be considered that reintroducing biological pathways involved in the pathogenesis and development of the disease of interest might explain finding of additional susceptible loci with more modest effect sizes e.g. combination of statistical evidence with some evaluation of performing contention (McCarthy, Abecasis et al. 2008).

This study has several other limitations. Firstly, data used in this study are limited to what is provided by NCBI through authorized access. The data show single time information; therefore, the trend or variation over time is not available. Secondly, it was not possible to test replication of the results found in the FIND collection in another dataset also provided by NCBI, named the chronic renal insufficiency cohort (CRIC) study, because all participants in CRIC had nephropathy when entered into the study and just one time-measured values were available. Therefore, the results of this study are not generalizable. Thirdly, sample size may not be large enough to detect actual association; in addition, after applying quality metrics control the number of individuals encountered in the analyses decreased considerably, specifically in Mexican Americans. Since the condition of interest has been categorical (case/control design), misclassification of phenotype is also likely to happen.

References

ABOUZEID, M., PHILPOT, B., JANUS, E.D., COATES, M.J. and DUNBAR, J.A., 2013. Type 2 diabetes prevalence varies by socio-economic status within and between migrant groups: analysis and implications for Australia. *BMC Public Health,* **13**, pp. 252-2458-13-252.

ACCETTURO, M., CREANZA, T.M., SANTORO, C., TRIA, G., GIORDANO, A., BATTAGLIERO, S., VACCINA, A., SCIOSCIA, G. and LEO, P., 2010. Finding New Genes for Non-Syndromic Hearing Loss through an In Silico Prioritization Study. *PLoS ONE,* **5**(9), e12742 doi:10.1371/journal.pone.0012742.

ACHARJEE, S., GHOSH, B., AL-DHUBIAB, B.E. and NAIR, A.B., 2013. Understanding type 1 diabetes: etiology and models. *Canadian Journal of Diabetes,* **37**(4), pp. 269-276.

AFFYMETRIX, 2009a-last update, Data Sheet, Genome-Wide Human SNP Array 6.0. Available: http://media.affymetrix.com/support/technical/datasheets/genomewide_snp6_datasheet.pdf [July/20, 2014].

AFFYMETRIX, 2009-last update, How Affymetrix Gene Chip DNA Microarrays Work. Available: http://www.autismspeaks.org/docs/Affy_gene_chip.pdf [July/20, 2014].

AISSANI, B., ZHANG, K. and WIENER, H., 2015. Follow-up to genome-wide linkage and admixture mapping studies implicates components of the extracellular matrix in susceptibility to and size of uterine fibroids. *Fertility and Sterility,* **103**(2), pp. 528-534.e13.

ALBERTI, K.G. and ZIMMET, P.Z., 1998. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabetic medicine : a Journal of the British Diabetic Association,* **15**(7), pp. 539-553.

AMERICAN DIABETES ASSOCIATION, 2013. Economic costs of diabetes in the U.S. in 2012. *Diabetes Care,* **36**(4), pp. 1033-1046.

AMERICAN DIABETES ASSOCIATION, 2010. Standards of medical care in diabetes--2010. *Diabetes Care,* **33 Suppl 1**, pp. S11-61.

AMERICAN DIABETES ASSOCIATION, 2004. Diagnosis and classification of diabetes mellitus. *Diabetes Care,* **27 Suppl 1**, pp. S5-S10.

AMERICAN DIABETES ASSOCIATION, 2000. Type 2 diabetes in children and adolescents. American Diabetes Association. *Diabetes Care,* **23**(3), pp. 381-389.

ANDERSON, C.A., PETTERSSON, F.H., CLARKE, G.M., CARDON, L.R., MORRIS, A.P. and ZONDERVAN, K.T., 2010. Data quality control in genetic case-control association studies. *Nature Protocols,* **5**(9), pp. 1564-1573.

ANDERSON, D., CORDELL, H.J., FAKIOLA, M., FRANCIS, R.W., SYN, G., SCAMAN, E.S.H., DAVIS, E., MILES, S.J., MCLEAY, T., JAMIESON, S.E. and BLACKWELL, J.M., 2015. First Genome-Wide Association Study in an Australian Aboriginal Population Provides Insights into Genetic Risk Factors for Body Mass Index and Type 2 Diabetes. *PLoS ONE,* **10**(3), e0119333. doi:10.1371/journal.pone.0119333.

ARDLIE, K.G., KRUGLYAK, L. and SEIELSTAD, M., 2002. Patterns of linkage disequilibrium in the human genome. *Nature Reviews.Genetics,* **3**(4), pp. 299-309.

ARFA, I., ABID, A., MALOUCHE, D., BEN ALAYA, N., AZEGUE, T.R., MANNAI, I., ZORGATI, M.M., BEN RAYANA, M.C., BEN AMMAR, S., BLOUSA-CHABCHOUB, S., BEN ROMDHANE, H., ZOUARI, B., DELLAGI, M.K. and ABDELHAK, S., 2007. Familial aggregation and excess maternal transmission of type 2 diabetes in Tunisia. *Postgraduate Medical Journal,* **83**(979), pp. 348-351.

ASSASSI, S., REVEILLE, J.D., ARNETT, F.C., WEISMAN, M.H., WARD, M.M., AGARWAL, S.K., GOURH, P., BHULA, J., SHARIF, R., SAMPAT, K., MAYES, M.D. and TAN, F.K., 2011. Whole-blood Gene Expression Profiling in Ankylosing Spondylitis Shows Upregulation of Toll-like Receptor 4 and 5. *The Journal of Rheumatology,* **38**(1), pp. 87-98.

BADENHOOP, K., KAHLES, H., SEIDL, C., KORDONOURI, O., LOPEZ, E.R., WALTER, M., ROSINGER, S., ZIEGLER, A., BOHM, B.O. and DIABETES GENETICS CONSORTIUM, 2009. MHC-environment interactions leading to type 1 diabetes: feasibility of an analysis of HLA DR-DQ alleles in relation to manifestation periods and dates of birth. *Diabetes, Obesity & Metabolism,* **11 Suppl 1**, pp. 88-91.

BAGHBANI, F., RAOOFIAN, R., HASANZADEH NAZARABADI, M., HAMZEHLOEI, T., SOUKHTANLOO, M., HEIDARI, M., AFSHARZADEH, S.M., SHEKOUHI, S., MORADI, F., SARLI, A.A., ZAVAR-REZA, J. and MOJARRAD, M., 2013. Identification of Novel Hypoxia Response Genes in Human Glioma Cell Line A172. *Iranian Journal of Basic Medical Sciences,* **16**(5), pp. 675-682.

BALDING, D.J., BISHOP, M. and CANNINGS, C., eds, 2007. *Handbook of Statistical Genetics.* Third edn. West Sussex, England: John Wiley & Sons, Ltd.

BALDING, D.J., 2006. A tutorial on statistical methods for population association studies. *Nature Reviews Genetics,* **7**(10), pp. 781-791.

BAO, M.Z., WANG, J.X., DORMAN, J.S. and TRUCCO, M., 1989. HLA-DQ beta non-ASP-57 allele and incidence of diabetes in China and the USA. *Lancet,* **2**(8661), pp. 497-498.

BARANZINI, S.E., WANG, J., GIBSON, R.A., GALWEY, N., NAEGELIN, Y., BARKHOF, F., RADUE, E.W., LINDBERG, R.L., UITDEHAAG, B.M., JOHNSON, M.R., ANGELAKOPOULOU, A., HALL, L., RICHARDSON, J.C., PRINJHA, R.K., GASS, A., GEURTS, J.J., KRAGT, J., SOMBEKKE, M., VRENKEN, H., QUALLEY, P., LINCOLN, R.R., GOMEZ, R., CAILLIER, S.J., GEORGE, M.F., MOUSAVI, H., GUERRERO, R., OKUDA, D.T., CREE, B.A., GREEN, A.J., WAUBANT, E., GOODIN, D.S., PELLETIER, D., MATTHEWS, P.M., HAUSER, S.L., KAPPOS, L., POLMAN, C.H. and OKSENBERG, J.R., 2009. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Human Molecular Genetics,* **18**(4), pp. 767-778.

BARKER, D.J., HALES, C.N., FALL, C.H., OSMOND, C., PHIPPS, K. and CLARK, P.M., 1993. Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome X): relation to reduced fetal growth. *Diabetologia,* **36**(1), pp. 62-67.

BARNETT, A.H., EFF, C., LESLIE, R.D. and PYKE, D.A., 1981. Diabetes in identical twins. A study of 200 pairs. *Diabetologia,* **20**(2), pp. 87-93.

BATLLE, E., BACANI, J., BEGTHEL, H., JONKHEER, S., GREGORIEFF, A., VAN DE BORN, M., MALATS, N., SANCHO, E., BOON, E., PAWSON, T., GALLINGER, S., PALS, S. and CLEVERS, H., 2005. EphB receptor activity suppresses colorectal cancer progression. *Nature,* **435**(7045), pp. 1126-1130.

BELL, G.I., HORITA, S. and KARAM, J.H., 1984. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes,* **33**(2), pp. 176-183.

BENJAMIN, E.J., DUPUIS, J., LARSON, M.G., LUNETTA, K.L., BOOTH, S.L., GOVINDARAJU, D.R., KATHIRESAN, S., KEANEY, J.F., KEYES, M.J., LIN, J.P., MEIGS, J.B., ROBINS, S.J., RONG, J., SCHNABEL, R., VITA, J.A., WANG, T.J., WILSON, P.W., WOLF, P.A. and VASAN, R.S., 2007. Genome-wide association with select biomarker traits in the Framingham Heart Study. *BMC Medical Genetics,* **8**(Suppl 1), pp. S11-2350-8-S1-S11.

BEN-YOSEF, T., NESS, S.L., MADEO, A.C., BAR-LEV, A., WOLFMAN, J.H., AHMED, Z.M., DESNICK, R.J., WILLNER, J.P., AVRAHAM, K.B., OSTRER, H.,

ODDOUX, C., GRIFFITH, A.J. and FRIEDMAN, T.B., 2003. A mutation of PCDH15 among Ashkenazi Jews with the type 1 Usher syndrome. *The New England Journal of Medicine,* **348**(17), pp. 1664-1670.

BIONE, S., SALA, C., MANZINI, C., ARRIGO, G., ZUFFARDI, O., BANFI, S., BORSANI, G., JONVEAUX, P., PHILIPPE, C., ZUCCOTTI, M., BALLABIO, A. and TONIOLO, D., 1998. A human homologue of the Drosophila melanogaster diaphanous gene is disrupted in a patient with premature ovarian failure: evidence for conserved function in oogenesis and implications for human sterility. *American Journal of Human Genetics,* **62**(3), pp. 533-541.

BLUESTONE, J.A., HEROLD, K. and EISENBARTH, G., 2010. Genetics, pathogenesis and clinical interventions in type 1 diabetes. *Nature,* **464**(7293), pp. 1293-1300.

BO, S., CICCONE, G., BALDI, C., BENINI, L., DUSIO, F., FORASTIERE, G., LUCIA, C., NUTI, C., DURAZZO, M., CASSADER, M., GENTILE, L. and PAGANO, G., 2007. Effectiveness of a lifestyle intervention on metabolic syndrome. A randomized controlled trial. *Journal of General Internal Medicine,* **22**(12), pp. 1695-1703.

BODEN, G., 1997. Role of fatty acids in the pathogenesis of insulin resistance and NIDDM. *Diabetes,* **46**(1), pp. 3-10.

BORCH-JOHNSEN, K., NORGAARD, K., HOMMEL, E., MATHIESEN, E.R., JENSEN, J.S., DECKERT, T. and PARVING, H.H., 1992. Is diabetic nephropathy an inherited complication? *Kidney International,* **41**(4), pp. 719-722.

BOTTINI, N., MUSUMECI, L., ALONSO, A., RAHMOUNI, S., NIKA, K., ROSTAMKHANI, M., MACMURRAY, J., MELONI, G.F., LUCARELLI, P., PELLECCHIA, M., EISENBARTH, G.S., COMINGS, D. and MUSTELIN, T., 2004. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nature Genetics,* **36**(4), pp. 337-338.

BOWDEN, D.W., COLICIGNO, C.J., LANGEFELD, C.D., SALE, M.M., WILLIAMS, A., ANDERSON, P.J., RICH, S.S. and FREEDMAN, B.I., 2004. A genome scan for diabetic nephropathy in African Americans. *Kidney International,* **66**(4), pp. 1517-1526.

BOWDEN, D.W., 2003. Genetics of kidney disease. *Kidney International Supplement,* (83), pp. 8-12.

BRADSHAW, E.M., RADDASSI, K., ELYAMAN, W., ORBAN, T., GOTTLIEB, P.A., KENT, S.C. and HAFLER, D.A., 2009. Monocytes from patients with type 1 diabetes spontaneously secrete proinflammatory cytokines inducing Th17 cells. *Journal of Immunology (Baltimore, Md.: 1950),* **183**(7), pp. 4432-4439.

BROQUÈRES-YOU, D., LERÉ-DÉAN, C., MERKULOVA-RAINON, T., MANTSOUNGA, C.S., ALLANIC, D., HAINAUD, P., CONTRÈRES, J.O., WANG, Y., VILAR, J., VIRALLY, M., MOURAD, J.J., GUILLAUSSEAU, P.J., SILVESTRE, J.S. and LÉVY, B.I., 2012. Ephrin-B2-Activated Peripheral Blood Mononuclear Cells From Diabetic Patients Restore Diabetes-Induced Impairment of Postischemic Neovascularization. *Diabetes,* **61**(10), pp. 2621-2632.

BROWN, A.A., JENSEN, J., NIKOLOVA, Y.S., DJUROVIC, S., AGARTZ, I., SERVER, A., FERRELL, R.E., MANUCK, S.B., MATTINGSDAL, M., MELLE, I., HARIRI, A.R., FRIGESSI, A. and ANDREASSEN, O.A., 2012. Genetic variants affecting the neural processing of human facial expressions: evidence using a genome-wide functional imaging approach. *Translational Psychiatry,* **2**(7), e143. Epub 2012 Jul 24. doi:10.1038/tp.2012.67.

BUSCHARD, K., 2011. What causes type 1 diabetes? Lessons from animal models. *APMIS.Supplementum,* **(132):1-19. doi**(132), pp. 1-19.

BUSH, W.S. and MOORE, J.H., 2012. Chapter 11: Genome-wide association studies. *PLoS Computational Biology,* **8**(12), e1002822.

CANNON, T.D. and KELLER, M.C., 2006. Endophenotypes in the genetic analyses of mental disorders. *Annual Review of Clinical Psychology,* **2**, pp. 267-290.

CARDON, L.R. and PALMER, L.J., 2003. Population stratification and spurious allelic association. *Lancet,* **361**(9357), pp. 598-604.

CAUCHI, S., EL ACHHAB, Y., CHOQUET, H., DINA, C., KREMPLER, F., WEITGASSER, R., NEJJARI, C., PATSCH, W., CHIKRI, M., MEYRE, D. and FROGUEL, P., 2007. TCF7L2 is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis. *Journal of Molecular Medicine (Berlin, Germany),* **85**(7), pp. 777-782.

CHALIGNÉ, R., POPOVA, T., MENDOZA-PARRA, M.A., SALEEM, M.A.M., GENTIEN, D., BAN, K., PIOLOT, T., LEROY, O., MARIANI, O., GRONEMEYER, H., VINCENT-SALOMON, A., STERN, M.H. and HEARD, E., 2015. The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer. *Genome Research,* **25**(4), pp. 488-503.

CHAN, J.C., MALIK, V., JIA, W., KADOWAKI, T., YAJNIK, C.S., YOON, K.H. and HU, F.B., 2009. Diabetes in Asia: epidemiology, risk factors, and pathophysiology. *Jama,* **301**(20), pp. 2129-2140.

CHASMAN, D.I., PARÉ, G., MORA, S., HOPEWELL, J.C., PELOSO, G., CLARKE, R., CUPPLES, L.A., HAMSTEN, A., KATHIRESAN, S., MÄLARSTIG, A.,

ORDOVAS, J.M., RIPATTI, S., PARKER, A.N., MILETICH, J.P. and RIDKER, P.M., 2009. Forty-Three Loci Associated with Plasma Lipoprotein Size, Concentration, and Cholesterol Content in Genome-Wide Analysis. *PLoS Genetics*, **5**(11), e1000730. Epub 2009 Nov 20. doi:10.1371/journal.pgen.1000730.

CHEN, H.L., LI, C.F., GRIGORIAN, A., TIAN, W. and DEMETRIOU, M., 2009. T Cell Receptor Signaling Co-regulates Multiple Golgi Genes to Enhance N-Glycan Branching. *The Journal of Biological Chemistry*, **284**(47), pp. 32454-32461.

CHEN, L., MAGLIANO, D.J. and ZIMMET, P.Z., 2011. The worldwide epidemiology of type 2 diabetes mellitus--present and future perspectives. *Nature Reviews Endocrinology*, **8**(4), pp. 228-236.

CHENG, C.Y., REICH, D., HAIMAN, C.A., TANDON, A., PATTERSON, N., SELVIN, E., AKYLBEKOVA, E.L., BRANCATI, F.L., CORESH, J., BOERWINKLE, E., ALTSHULER, D., TAYLOR, H.A., HENDERSON, B.E., WILSON, J.G. and KAO, W.H., 2012. African ancestry and its correlation to type 2 diabetes in African Americans: a genetic admixture analysis in three U.S. population cohorts. *PloS ONE*, **7**(3), e32840.

COCKCROFT, D.W. and GAULT, M.H., 1976. Prediction of creatinine clearance from serum creatinine. *Nephron*, **16**(1), pp. 31-41.

COMMITTEE ON DIABETIC TWINS, 1988. Diabetes mellitus in twins: a cooperative study in Japan. Committee on Diabetic Twins, Japan Diabetes Society. *Diabetes Research and Clinical Practice*, **5**(4), pp. 271-280.

COMUZZIE, A.G., FUNAHASHI, T., SONNENBERG, G., MARTIN, L.J., JACOB, H.J., BLACK, A.E., MAAS, D., TAKAHASHI, M., KIHARA, S., TANAKA, S., MATSUZAWA, Y., BLANGERO, J., COHEN, D. and KISSEBAH, A., 2001. The genetic basis of plasma variation in adiponectin, a global endophenotype for obesity and the metabolic syndrome. *The Journal of Clinical Endocrinology and Metabolism*, **86**(9), pp. 4321-4325.

CONCANNON, P., ERLICH, H.A., JULIER, C., MORAHAN, G., NERUP, J., POCIOT, F., TODD, J.A., RICH, S.S. and TYPE 1 DIABETES GENETICS CONSORTIUM, 2005. Type 1 diabetes: evidence for susceptibility loci from four genome-wide linkage scans in 1,435 multiplex families. *Diabetes*, **54**(10), pp. 2995-3001.

CONDON, J., SHAW, J.E., LUCIANO, M., KYVIK, K.O., MARTIN, N.G. and DUFFY, D.L., 2008. A study of diabetes mellitus within a large sample of Australian twins. *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies*, **11**(1), pp. 28-40.

COSNES, J., CELLIER, C., VIOLA, S., COLOMBEL, J.F., MICHAUD, L., SARLES, J., HUGOT, J.P., GINIES, J.L., DABADIE, A., MOUTERDE, O., ALLEZ, M., NION-LARMURIER, I. and GROUPE D'ETUDE ET DE RECHERCHE SUR LA MALADIE COELIAQUE, 2008. Incidence of autoimmune diseases in celiac disease: protective effect of the gluten-free diet. *Clinical gastroenterology and hepatology : The Official Clinical Practice Journal of the American Gastroenterological Association,* **6**(7), pp. 753-758.

DAHLQUIST, G.G., IVARSSON, S., LINDBERG, B. and FORSGREN, M., 1995. Maternal enteroviral infection during pregnancy as a risk factor for childhood IDDM. A population-based case-control study. *Diabetes,* **44**(4), pp. 408-413.

DE FRANCISCO, A.L., 1999. Clinical outcomes for patients with diabetic nephropathy. *EDTNA/ERCA Journal (English ed.),* **25**(1), pp. 23-27.

DE LUCA, M., KLIMENTIDIS, Y.C., CASAZZA, K., MOSES CHAMBERS, M., CHO, R., HARBISON, S.T., JUMBO-LUCIONI, P., ZHANG, S., LEIPS, J. and FERNANDEZ, J.R., 2010. A Conserved Role for Syndecan Family Members in the Regulation of Whole-Body Energy Metabolism. *PLoS ONE,* **5**(6), e11286. doi:10.1371/journal.pone.0011286.

DE VRIES, D.R., TER LINDE, J.J.M., VAN HERWAARDEN, M.A., SCHWARTZ, M.P., SHEPHARD, P., GENG, M.M., SMOUT, A.J.P.M. and SAMSOM, M., 2009. In GERD patients, mucosal repair associated genes are upregulated in non-inflamed oesophageal epithelium. *Journal of Cellular and Molecular Medicine,* **13**(5), pp. 936-947.

DENG, T., LYON, C.J., MINZE, L.J., LIN, J., ZOU, J., LIU, J.Z., REN, Y., YIN, Z., HAMILTON, D.J., REARDON, P.R., SHERMAN, V., WANG, H.Y., PHILLIPS, K.J., WEBB, P., WONG, S.T., WANG, R.F. and HSUEH, W.A., 2013. Class II major histocompatibility complex plays an essential role in obesity-induced adipose inflammation. *Cell Metabolism,* **17**(3), pp. 411-422.

DEVLIN, B. and ROEDER, K., 1999a. Genomic control for association studies. *Biometrics,* **55**(4), pp. 997-1004.

DEVLIN, B. and ROEDER, K., 1999b. Genomic control for association studies. *Biometrics,* **55**(4), pp. 997-1004.

DIABETES PREVENTION PROGRAM RESEARCH GROUP, KNOWLER, W.C., FOWLER, S.E., HAMMAN, R.F., CHRISTOPHI, C.A., HOFFMAN, H.J., BRENNEMAN, A.T., BROWN-FRIDAY, J.O., GOLDBERG, R., VENDITTI, E. and NATHAN, D.M., 2009. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet,* **374**(9702), pp. 1677-1686.

DINNEEN, S., GERICH, J. and RIZZA, R., 1992. Carbohydrate metabolism in non-insulin-dependent diabetes mellitus. *The New England Journal of Medicine,* **327**(10), pp. 707-713.

DISTEFANO, J.K. and TAVERNA, D.M., 2011. Technological Issues and Experimental Design of Gene Association Studies. In: J.K. DISTEFANO, ed, *Disease Gene Identification Methods and Protocols.* New York, USA: Humana Press, c/o Springer Science+Business Media, pp. 3-16.

DMITRIENKO, A., WIENS, B. and WESTFALL, P., 2006. Fallback tests in dose-response clinical trials. *Journal of Biopharmaceutical Statistics,* **16**(5), pp. 745-755.

DOTSON, C.D., ZHANG, L., XU, H., SHIN, Y.K., VIGUES, S., OTT, S.H., ELSON, A.E.T., CHOI, H.J., SHAW, H., EGAN, J.M., MITCHELL, B.D., LI, X., STEINLE, N.I. and MUNGER, S.D., 2008. Bitter Taste Receptors Influence Glucose Homeostasis. *PLoS ONE,* **3**(12), e3974. doi:10.1371/journal.pone.0003974.

DUDBRIDGE, F. and GUSNANTO, A., 2008. Estimation of significance thresholds for genomewide association scans. *Genetic Epidemiology,* **32**(3), pp. 227-234.

EL ACHHAB, Y., MEYRE, D., BOUATIA-NAJI, N., BERRAHO, M., DEWEIRDER, M., VATIN, V., DELPLANQUE, J., SERHIER, Z., LYOUSSI, B., NEJJARI, C., FROGUEL, P. and CHIKRI, M., 2009. Association of the ENPP1 K121Q polymorphism with type 2 diabetes and obesity in the Moroccan population. *Diabetes & Metabolism,* **35**(1), pp. 37-42.

ENGLISH, K.M., MANDOUR, O., STEEDS, R.P., DIVER, M.J., JONES, T.H. and CHANNER, K.S., 2000. Men with coronary artery disease have lower levels of androgens than men with normal coronary angiograms. *European Heart Journal,* **21**(11), pp. 890-894.

ENSEMBL, 2016-last update, Ensembl. Available: http://uswest.ensembl.org/Homo_sapiens/Info/Index [July/20, 2014].

ERLICH, H., VALDES, A.M., NOBLE, J., CARLSON, J.A., VARNEY, M., CONCANNON, P., MYCHALECKYJ, J.C., TODD, J.A., BONELLA, P., FEAR, A.L., LAVANT, E., LOUEY, A., MOONSAMY, P. and TYPE 1 DIABETES GENETICS CONSORTIUM, 2008. HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. *Diabetes,* **57**(4), pp. 1084-1092.

ESPINOZA-JIMÉNEZ, A., PEÓN, A.N. and TERRAZAS, L.I., 2012. Alternatively Activated Macrophages in Types 1 and 2 Diabetes. *Mediators of Inflammation,* **2012**, pp. 10.1155/2012/815953.

ESTACIO, R.O. and SCHRIER, R.W., 2001. Diabetic nephropathy: pathogenesis, diagnosis, and prevention of progression. *Advances in Internal Medicine,* **46**, pp. 359-408.

FALLER, N., GAUTSCHI, I. and SCHILD, L., 2014. Functional Analysis of a Missense Mutation in the Serine Protease Inhibitor SPINT2 Associated with Congenital Sodium Diarrhea. *PLoS ONE,* **9**(4), e94267. doi:10.1371/journal.pone.0094267.

FAMILY INVESTIGATION OF NEPHROPATHY AND DIABETES RESEARCH GROUP, 2003. Genetic determinants of diabetic nephropathy: The family investigation of nephropathy and diabetes (FIND). *Journal of the American Society of Nephrology :* **14**(7 Suppl 2), pp. S202-4.

FLOREZ, J.C., PRICE, A.L., CAMPBELL, D., RIBA, L., PARRA, M.V., YU, F., DUQUE, C., SAXENA, R., GALLEGO, N., TELLO-RUIZ, M., FRANCO, L., RODRIGUEZ-TORRES, M., VILLEGAS, A., BEDOYA, G., AGUILAR-SALINAS, C.A., TUSIE-LUNA, M.T., RUIZ-LINARES, A. and REICH, D., 2009. Strong association of socioeconomic status with genetic ancestry in Latinos: implications for admixture studies of type 2 diabetes. *Diabetologia,* **52**(8), pp. 1528-1536.

FOGARTY, D.G., HANNA, L.S., WANTMAN, M., WARRAM, J.H., KROLEWSKI, A.S. and RICH, S.S., 2000. Segregation analysis of urinary albumin excretion in families with type 2 diabetes. *Diabetes,* **49**(6), pp. 1057-1063.

FREEDMAN, B.I., BOSTROM, M., DAEIHAGH, P. and BOWDEN, D.W., 2007. Genetic factors in diabetic nephropathy. *Clinical Journal of the American Society of Nephrology : CJASN,* **2**(6), pp. 1306-1316.

FREEDMAN, B.I., BOWDEN, D.W., RICH, S.S., VALIS, C.J., SALE, M.M., HICKS, P.J. and LANGEFELD, C.D., 2005. A genome scan for all-cause end-stage renal disease in African Americans. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association,* **20**(4), pp. 712-718.

FREEDMAN, B.I., TUTTLE, A.B. and SPRAY, B.J., 1995. Familial predisposition to nephropathy in African-Americans with non-insulin-dependent diabetes mellitus. *American Journal of Kidney Diseases : The Official Journal of the National Kidney Foundation,* **25**(5), pp. 710-713.

FULLERTON, S.M., BARTOSZEWICZ, A., YBAZETA, G., HORIKAWA, Y., BELL, G.I., KIDD, K.K., COX, N.J., HUDSON, R.R. and DI RIENZO, A., 2002. Geographic and haplotype structure of candidate type 2 diabetes susceptibility variants at the calpain-10 locus. *American Journal of Human Genetics,* **70**(5), pp. 1096-1106.

GALE, E.A., 2002. The rise of childhood type 1 diabetes in the 20th century. *Diabetes*, **51**(12), pp. 3353-3361.

GOTTE, M., 2003. Syndecans in inflammation. *FASEB Journal : official publication of the Federation of American Societies for Experimental Biology*, **17**(6), pp. 575-591.

GOTTESMAN, I.I. and GOULD, T.D., 2003. The endophenotype concept in psychiatry: etymology and strategic intentions. *The American Journal of Psychiatry*, **160**(4), pp. 636-645.

GUÉNARD, F., LAMONTAGNE, M., BOSSÉ, Y., DESHAIES, Y., CIANFLONE, K., KRAL, J.G., MARCEAU, P. and VOHL, M.C., 2015. Influences of Gestational Obesity on Associations between Genotypes and Gene Expression Levels in Offspring following Maternal Gastrointestinal Bypass Surgery for Obesity. *PLoS ONE*, **10**(1), e0117011. doi:10.1371/journal.pone.0117011.

GUHA, S., ROSENFELD, J.A., MALHOTRA, A.K., LEE, A.T., GREGERSEN, P.K., KANE, J.M., PE'ER, I., DARVASI, A. and LENCZ, T., 2012. Implications for health and disease in the genetic signature of the Ashkenazi Jewish population. *Genome Biology*, **13**(1), pp. R2-2012-13-1-r2.

GUO, C.Y., GU, Y., LI, L., JIA, E.Z., LI, C.J., WANG, L.S., YANG, Z.J., CAO, K.J. and MA, W.Z., 2012. Association of SNP rs6903956 on chromosome 6p24.1 with angiographical characteristics of coronary atherosclerosis in a Chinese population. *PloS ONE*, **7**(8), e43732.

GUO, X., GAO, L., WEI, C., YANG, X., ZHAO, Y. and DONG, A., 2011. A Computational Method Based on the Integration of Heterogeneous Networks for Predicting Disease-Gene Associations. *PLoS ONE*, **6**(9), e24171. doi:10.1371/journal.pone.0024171.

HE, K., XUN, P., LIU, K., MORRIS, S., REIS, J. and GUALLAR, E., 2013. Mercury exposure in young adulthood and incidence of diabetes later in life: the CARDIA Trace Element Study. *Diabetes Care*, **36**(6), pp. 1584-1589.

HE, X., FULLER, C., SONG, Y., MENG, Q., ZHANG, B., YANG, X. and LI, H., 2013. Sherlock: Detecting Gene-Disease Associations by Matching Patterns of Expression QTL and GWAS. *American Journal of Human Genetics*, **92**(5), pp. 667-680.

HEMMINKI, K., LI, X., SUNDQUIST, K. and SUNDQUIST, J., 2010. Familial risks for type 2 diabetes in Sweden. *Diabetes Care*, **33**(2), pp. 293-297.

HEROLD, K.C., VIGNALI, D.A., COOKE, A. and BLUESTONE, J.A., 2013. Type 1 diabetes: translating mechanistic observations into effective clinical outcomes. *Nature Reviews Immunology,* **13**(4), pp. 243-256.

HOFMAN, P.L., REGAN, F., JACKSON, W.E., JEFFERIES, C., KNIGHT, D.B., ROBINSON, E.M. and CUTFIELD, W.S., 2004. Premature birth and later insulin resistance. *The New England Journal of Medicine,* **351**(21), pp. 2179-2186.

HOLM, S., 1979. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.,* **6**, pp. 65-70.

HOLSTEIN, A., HAHN, M., KORNER, A., STUMVOLL, M. and KOVACS, P., 2011. TCF7L2 and therapeutic response to sulfonylureas in patients with type 2 diabetes. *BMC Medical Genetics,* **12**, pp. 12:30. doi:10.1186/1471-2350-12-30.

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. *Nature Genetics,* **26**(2), pp. 163-175.

HOWSON, J.M., BARRATT, B.J., TODD, J.A. and CORDELL, H.J., 2005. Comparison of population- and family-based methods for genetic association analysis in the presence of interacting loci. *Genetic Epidemiology,* **29**(1), pp. 51-67.

HU, F.B., MANSON, J.E., STAMPFER, M.J., COLDITZ, G., LIU, S., SOLOMON, C.G. and WILLETT, W.C., 2001. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *The New England Journal of Medicine,* **345**(11), pp. 790-797.

HUANG, E.W., PENG, L.Y., ZHENG, J.X., WANG, D., XU, Q.Y., HUANG, L., WU, Q.P., TANG, S.B., LUO, B., LIU, S.P., LIU, X.S., LI, Z.H., QUAN, L., LI, Y., SHI, H., LV, G.L., ZHAO, J., CHENG, J.D. and LIU, C., 2015. Common Variants in Promoter of ADTRP Associate with Early-Onset Coronary Artery Disease in a Southern Han Chinese Population. *PLoS ONE,* **10**(9), e0137547. doi:10.1371/journal.pone.0137547.

HYTTINEN, V., KAPRIO, J., KINNUNEN, L., KOSKENVUO, M. and TUOMILEHTO, J., 2003. Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. *Diabetes,* **52**(4), pp. 1052-1055.

IGO, R.P.,Jr, IYENGAR, S.K., NICHOLAS, S.B., GODDARD, K.A., LANGEFELD, C.D., HANSON, R.L., DUGGIRALA, R., DIVERS, J., ABBOUD, H., ADLER, S.G.,

ARAR, N.H., HORVATH, A., ELSTON, R.C., BOWDEN, D.W., GUO, X., IPP, E., KAO, W.H., KIMMEL, P.L., KNOWLER, W.C., MEONI, L.A., MOLINEROS, J., NELSON, R.G., PAHL, M.V., PAREKH, R.S., RASOOLY, R.S., SCHELLING, J.R., SHAH, V.O., SMITH, M.W., WINKLER, C.A., ZAGER, P.G., SEDOR, J.R., FREEDMAN, B.I. and FAMILY INVESTIGATION OF NEPHROPATHY AND DIABETES RESEARCH GROUP, 2011. Genomewide linkage scan for diabetic renal failure and albuminuria: the FIND study. *American Journal of Nephrology,* **33**(5), pp. 381-389.

IMAM, K., 2012. Clinical Features, diagnostic criteria and pathogenesis of diabetes mellitus. In: S.I. AHMAD, *Diabetes An Old Disease, a New Insight.* Texas, USA: Landes Bioscience I Springer Science+ Business Media, pp. 340-355.

IMPERATORE, G., HANSON, R.L., PETTITT, D.J., KOBES, S., BENNETT, W. and KNOWLER, W.C., 1997. Segregation and linkage analyses of diabetic nephropathy in Pima Indians. The Pima Diabetes Genes Group, NIDDK, NIH, Phoenix AZ. Poster presentations. *American Journal of Human Genetics,* **61**, pp. A280.

IMPERATORE, G., HANSON, R.L., PETTITT, D.J., KOBES, S., BENNETT, P.H. and KNOWLER, W.C., 1998. Sib-pair linkage analysis for susceptibility genes for microvascular complications among Pima Indians with type 2 diabetes. Pima Diabetes Genes Group. *Diabetes,* **47**(5), pp. 821-830.

INTERNATIONAL HAPMAP CONSORTIUM, 2005. A haplotype map of the human genome. *Nature,* **437**(7063), pp. 1299-1320.

IOANNIDIS, J.P., 2008. Why most discovered true associations are inflated. *Epidemiology (Cambridge, Mass.),* **19**(5), pp. 640-648.

IVLIEV, A.E., 'T HOEN, P.A.C., VAN ROON-MOM, W.M.C., PETERS, D.J.M. and SERGEEVA, M.G., 2012. Exploring the Transcriptome of Ciliated Cells Using In Silico Dissection of Human Tissues. *PLoS ONE,* **7**(4), e35618. doi:10.1371/journal.pone.0035618.

IYENGAR, S.K., ABBOUD, H.E., GODDARD, K.A., SAAD, M.F., ADLER, S.G., ARAR, N.H., BOWDEN, D.W., DUGGIRALA, R., ELSTON, R.C., HANSON, R.L., IPP, E., KAO, W.H., KIMMEL, P.L., KLAG, M.J., KNOWLER, W.C., MEONI, L.A., NELSON, R.G., NICHOLAS, S.B., PAHL, M.V., PAREKH, R.S., QUADE, S.R., RICH, S.S., ROTTER, J.I., SCAVINI, M., SCHELLING, J.R., SEDOR, J.R., SEHGAL, A.R., SHAH, V.O., SMITH, M.W., TAYLOR, K.D., WINKLER, C.A., ZAGER, P.G., FREEDMAN, B.I. and FAMILY INVESTIGATION OF NEPHROPATHY AND DIABETES RESEARCH GROUP, 2007. Genome-wide scans for diabetic nephropathy and albuminuria in multiethnic populations: the family investigation of nephropathy and diabetes (FIND). *Diabetes,* **56**(6), pp. 1577-1585.

JANSSEN, S.F., GORGELS, T.G.M.F., BOSSERS, K., TEN BRINK, J.B., ESSING, A.H.W., NAGTEGAAL, M., VAN DER SPEK, P.J., JANSONIUS, N.M. and BERGEN, A.A.B., 2012. Gene Expression and Functional Annotation of the Human Ciliary Body Epithelia. *PLoS ONE,* **7**(9), e44973. doi:10.1371/journal.pone.0044973.

JENUM, A.K., HOLME, I., GRAFF-IVERSEN, S. and BIRKELAND, K.I., 2005. Ethnicity and sex are strong determinants of diabetes in an urban Western society: implications for prevention. *Diabetologia,* **48**(3), pp. 435-439.

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**(11), pp. 1060-1067.

KELDER, T., VERSCHUREN, L., VAN OMMEN, B., VAN GOOL, A.J. and RADONJIC, M., 2014. Network signatures link hepatic effects of anti-diabetic interventions with systemic disease parameters. *BMC Systems Biology,* **8**, 108. doi: 10.1186/s12918-014-0108-0.

KHAN, M.I., KERSTEN, F.F., AZAM, M., COLLIN, R.W., HUSSAIN, A., SHAH, S.T., KEUNEN, J.E., KREMER, H., CREMERS, F.P., QAMAR, R. and DEN HOLLANDER, A.I., 2011. CLRN1 mutations cause nonsyndromic retinitis pigmentosa. *Ophthalmology,* **118**(7), pp. 1444-1448.

KNOWLER, W.C., BARRETT-CONNOR, E., FOWLER, S.E., HAMMAN, R.F., LACHIN, J.M., WALKER, E.A., NATHAN, D.M. and DIABETES PREVENTION PROGRAM RESEARCH GROUP, 2002. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *The New England Journal of Medicine,* **346**(6), pp. 393-403.

KNOWLER, W.C., CORESH, J., ELSTON, R.C., FREEDMAN, B.I., IYENGAR, S.K., KIMMEL, P.L., OLSON, J.M., PLAETKE, R., SEDOR, J.R., SELDIN, M.F. and FAMILY INVESTIGATION OF NEPHROPATHY AND DIABETES RESEARCH GROUP, 2005. The Family Investigation of Nephropathy and Diabetes (FIND): design and methods. *Journal of Diabetes and Its Complications,* **19**(1), pp. 1-9.

KNOWLER, W.C., WILLIAMS, R.C., PETTITT, D.J. and STEINBERG, A.G., 1988. Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *American Journal of Human Genetics,* **43**(4), pp. 520-526.

KONTOGIANNI, M.D., LIATIS, S., GRAMMATIKOU, S., PERREA, D., KATSILAMBROS, N. and MAKRILAKIS, K., 2012. Changes in dietary habits and their association with metabolic markers after a non-intensive, community-based lifestyle

intervention to prevent type 2 diabetes, in Greece. The DEPLAN study. *Diabetes Research and Clinical Practice,* **95**(2), pp. 207-214.

KOSAKA, K., NODA, M. and KUZUYA, T., 2005. Prevention of type 2 diabetes by lifestyle intervention: a Japanese trial in IGT males. *Diabetes Research and Clinical Practice,* **67**(2), pp. 152-162.

KRAFT, P., 2008. Curses--winner's and otherwise--in genetic epidemiology. *Epidemiology (Cambridge, Mass.),* **19**(5), pp. 649-51; discussion 657-8.

KROLEWSKI, A.S., WARRAM, J.H., RAND, L.I. and KAHN, C.R., 1987. Epidemiologic approach to the etiology of type I diabetes mellitus and its complications. *The New England Journal of Medicine,* **317**(22), pp. 1390-1398.

KUSHIMA, I., NAKAMURA, Y., ALEKSIC, B., IKEDA, M., ITO, Y., SHIINO, T., OKOCHI, T., FUKUO, Y., UJIKE, H., SUZUKI, M., INADA, T., HASHIMOTO, R., TAKEDA, M., KAIBUCHI, K., IWATA, N. and OZAKI, N., 2012. Resequencing and Association Analysis of the KALRN and EPHB1 Genes And Their Contribution to Schizophrenia Susceptibility. *Schizophrenia Bulletin,* **38**(3), pp. 552-560.

LAING, S., UNGER, M., KOCH-NOLTE, F. and HAAG, F., 2011. ADP-ribosylation of arginine. *Amino Acids,* **41**(2), pp. 257-269.

LAPORTE, R.E., TAJIMA, N., AKERBLOM, H.K., BERLIN, N., BROSSEAU, J., CHRISTY, M., DRASH, A.L., FISHBEIN, H., GREEN, A. and HAMMAN, R., 1985. Geographic differences in the risk of insulin-dependent diabetes mellitus: the importance of registries. *Diabetes Care,* **8 Suppl 1**, pp. 101-107.

LAYMAN, L.C., 2013. The genetic basis of female reproductive disorders: Etiology and clinical testing. *Molecular and Cellular Endocrinology,* **370**(0), pp. 138-148.

LEAHY, J.L., 2005. Pathogenesis of type 2 diabetes mellitus. *Archives of Medical Research,* **36**(3), pp. 197-209.

LEAK, T.S., PERLEGAS, P.S., SMITH, S.G., KEENE, K.L., HICKS, P.J., LANGEFELD, C.D., MYCHALECKYJ, J.C., RICH, S.S., KIRK, J.K., FREEDMAN, B.I., BOWDEN, D.W. and SALE, M.M., 2009. Variants in intron 13 of the ELMO1 gene are associated with diabetic nephropathy in African Americans. *Annals of Human Genetics,* **73**(2), pp. 152-159.

LEE, S.H., LEE, T.W., IHM, C.G., KIM, M.J., WOO, J.T. and CHUNG, J.H., 2005. Genetics of diabetic nephropathy in type 2 DM: candidate gene analysis for the pathogenic role of inflammation. *Nephrology (Carlton, Vic.),* **10 Suppl**, pp. S32-6.

LEE, S.J., CHOI, M.G., KIM, D.S. and KIM, T.W., 2006. Manganese superoxide dismutase gene polymorphism (V16A) is associated with stages of albuminuria in Korean type 2 diabetic patients. *Metabolism: Clinical and Experimental,* **55**(1), pp. 1-7.

LEONETTI, D.L. and FUJIMOTO, W.Y., 1989. Type 2 diabetes, impaired glucose tolerance, and hypertension in offspring of migrants and the structure of the population of origin. *Human Biology,* **61**(3), pp. 369-386.

LEVEY, A.S., BOSCH, J.P., LEWIS, J.B., GREENE, T., ROGERS, N. and ROTH, D., 1999. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Annals of Internal Medicine,* **130**(6), pp. 461-470.

LEVEY, A.S., CORESH, J., GREENE, T., MARSH, J., STEVENS, L.A., KUSEK, J.W., VAN LENTE, F. and CHRONIC KIDNEY DISEASE EPIDEMIOLOGY COLLABORATION, 2007. Expressing the Modification of Diet in Renal Disease Study equation for estimating glomerular filtration rate with standardized serum creatinine values. *Clinical Chemistry,* **53**(4), pp. 766-772.

LI, D.C., ZHU, X.G., ZHANG, Y. and FU, J.X., 2006. Role of CD80 in stimulating T lymphocyte activation. *World Journal of Gastroenterology,* **12**(6), pp. 880-884.

LI, J.K., NG, M.C., SO, W.Y., CHIU, C.K., OZAKI, R., TONG, P.C., COCKRAM, C.S. and CHAN, J.C., 2006. Phenotypic and genetic clustering of diabetes and metabolic syndrome in Chinese families with type 2 diabetes mellitus. *Diabetes/Metabolism Research and Reviews,* **22**(1), pp. 46-52.

LI, M., BOEHNKE, M. and ABECASIS, G.R., 2006. Efficient study designs for test of genetic association using sibship data and unrelated cases and controls. *American Journal of Human Genetics,* **78**(5), pp. 778-792.

LI, Y., HE, Y., QI, L., JADDOE, V.W., FESKENS, E.J., YANG, X., MA, G. and HU, F.B., 2010. Exposure to the Chinese famine in early life and the risk of hyperglycemia and type 2 diabetes in adulthood. *Diabetes,* **59**(10), pp. 2400-2406.

LINTAS, C. and PERSICO, A.M., 2010. Neocortical RELN promoter methylation increases significantly after puberty. *Neuroreport,* **21**(2), pp. 114-118.

LIU, X.F., BERA, T.K., KAHUE, C., ESCOBAR, T., FEI, Z., RACITI, G.A. and PASTAN, I., 2012. ANKRD26 and Its Interacting Partners TRIO, GPS2, HMMR and DIPA Regulate Adipogenesis in 3T3-L1 Cells. *PLoS ONE,* **7**(5), e38130. doi:10.1371/journal.pone.0038130.

LIU, Y., LEE, Y.F. and NG, 2011. SNP and gene networks construction and analysis from classification of copy number variations data. *BMC Bioinformatics,* **12**(Suppl 5), pp. S4-2105-12-S5-S4.

LIVINGSTON, W.S., RUSCH, H.L., NERSESIAN, P.V., BAXTER, T., MYSLIWIEC, V. and GILL, J.M., 2015. Improved Sleep in Military Personnel is Associated with Changes in the Expression of Inflammatory Genes and Improvement in Depression Symptoms. *Frontiers in Psychiatry*, **6**, pp. 10.3389/fpsyt.2015.00059.

LOWE, C.E., COOPER, J.D., BRUSKO, T., WALKER, N.M., SMYTH, D.J., BAILEY, R., BOURGET, K., PLAGNOL, V., FIELD, S., ATKINSON, M., CLAYTON, D.G., WICKER, L.S. and TODD, J.A., 2007. Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nature Genetics,* **39**(9), pp. 1074-1082.

MACDONALD, M.J., 1980. The frequencies of juvenile diabetes in American Blacks and Caucasians are consistent with dominant inheritance. *Diabetes,* **29**(2), pp. 110-114.

MAEDA, S., IMAMURA, M., KURASHIGE, M., ARAKI, S., SUZUKI, D., BABAZONO, T., UZU, T., UMEZONO, T., TOYODA, M., KAWAI, K., IMANISHI, M., HANAOKA, K., MAEGAWA, H., UCHIGATA, Y. and HOSOYA, T., 2013. Replication study for the association of 3 SNP loci identified in a genome-wide association study for diabetic nephropathy in European type 1 diabetes with diabetic nephropathy in Japanese patients with type 2 diabetes. *Clinical and Experimental Nephrology,* **17**(6), pp. 866-871.

MAESO, I., IRIMIA, M., TENA, J.J., GONZÁLEZ-PÉREZ, E., TRAN, D., RAVI, V., VENKATESH, B., CAMPUZANO, S., GÓMEZ-SKARMETA, J.L. and GARCIA-FERNÀNDEZ, J., 2012. An ancient genomic regulatory block conserved across bilaterians and its dismantling in tetrapods by retrogene replacement. *Genome Research,* **22**(4), pp. 642-655.

MAHER, A.C., FU, M.H., ISFORT, R.J., VARBANOV, A.R., QU, X.A. and TARNOPOLSKY, M.A., 2009. Sex Differences in Global mRNA Content of Human Skeletal Muscle. *PLoS ONE,* **4**(7), e6335. doi:10.1371/journal.pone.0006335.

MAHURKAR, S., MOLDOVAN, M., SUPPIAH, V. and O'DOHERTY, C., 2013. Identification of Shared Genes and Pathways: A Comparative Study of Multiple Sclerosis Susceptibility, Severity and Response to Interferon Beta Treatment. *PLoS ONE,* **8**(2), e57655. doi:10.1371/journal.pone.0057655.

MARASCHIN, J.F., 2012. Classification of Diabetes, In: S.I. AHMAD, ed, *Diabetes An Old Disease, a New Insight.* Texas, USA: Landes Bioscience I Springer Science+ Business Media, pp. 12-19.

MARCHINI, J., CARDON, L.R., PHILLIPS, M.S. and DONNELLY, P., 2004. The effects of human population structure on large genetic association studies. *Nature Genetics,* **36**(5), pp. 512-517.

MARSELLI, L., THORNE, J., DAHIYA, S., SGROI, D.C., SHARMA, A., BONNER-WEIR, S., MARCHETTI, P. and WEIR, G.C., 2010. Gene Expression Profiles of Beta-Cell Enriched Tissue Obtained by Laser Capture Microdissection from Subjects with Type 2 Diabetes. *PLoS ONE,* **5**(7), e11499. doi:10.1371/journal.pone.0011499.

MARTINEZ-MARIGNAC, V.L., VALLADARES, A., CAMERON, E., CHAN, A., PERERA, A., GLOBUS-GOLDBERG, R., WACHER, N., KUMATE, J., MCKEIGUE, P., O'DONNELL, D., SHRIVER, M.D., CRUZ, M. and PARRA, E.J., 2007. Admixture in Mexico City: implications for admixture mapping of type 2 diabetes genetic risk factors. *Human Genetics,* **120**(6), pp. 807-819.

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. *Ethnicity & Disease,* **19**(1), pp. 49-55.

MAULL, E.A., AHSAN, H., EDWARDS, J., LONGNECKER, M.P., NAVAS-ACIEN, A., PI, J., SILBERGELD, E.K., STYBLO, M., TSENG, C.H., THAYER, K.A. and LOOMIS, D., 2012. Evaluation of the association between arsenic and diabetes: a National Toxicology Program workshop review. *Environmental Health Perspectives,* **120**(12), pp. 1658-1670.

MCCANCE, D.R., HANSON, R.L., PETTITT, D.J., JACOBSSON, L.T., BENNETT, P.H., BISHOP, D.T. and KNOWLER, W.C., 1995. Diabetic nephropathy: a risk factor for diabetes mellitus in offspring. *Diabetologia,* **38**(2), pp. 221-226.

MCCARTHY, M.I., ABECASIS, G.R., CARDON, L.R., GOLDSTEIN, D.B., LITTLE, J., IOANNIDIS, J.P. and HIRSCHHORN, J.N., 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews.Genetics,* **9**(5), pp. 356-369.

MCDONOUGH, C.W., PALMER, N.D., HICKS, P.J., ROH, B.H., AN, S.S., COOKE, J.N., HESTER, J.M., WING, M.R., BOSTROM, M.A., RUDOCK, M.E., LEWIS, J.P., TALBERT, M.E., BLEVINS, R.A., LU, L., NG, M.C., SALE, M.M., DIVERS, J., LANGEFELD, C.D., FREEDMAN, B.I. and BOWDEN, D.W., 2011. A genome-wide association study for diabetic nephropathy genes in African Americans. *Kidney International,* **79**(5), pp. 563-572.

MCDONOUGH, C.W., 2010. *The genetics of type 2 diabetic end-stage renal disease in African Americans*, Wake Forest University.

MCKNIGHT, B., 2013. Population Stratification in Genetic Association Studies. In: M.A. AUSTIN, ed, *Genetic Epidemiology Methods and Applications.* Oxfordshire, UK: CABI (CAB International), pp. 97-114.

MEIGS, J.B., CUPPLES, L.A. and WILSON, P.W., 2000. Parental transmission of type 2 diabetes: the Framingham Offspring Study. *Diabetes,* **49**(12), pp. 2201-2207.

MICHAELIS, M., BARTH, S., BREITLING, R., BRUCH, J., STEINBERGER, D., ROTHWEILER, F., HACKMANN, K., SCHRÖCK, E., DOERR, H.W., GRIFFIN, D.K., CINATL, J. and CINATL, J., 2012. Selection of a highly invasive neuroblastoma cell population through long-term human cytomegalovirus infection. *Oncogenesis,* **1**(4), pp. e10-. Epub 2012 Apr 30 doi:10.1038/oncsis.2012.10.

MIKESKA, T. and CRAIG, J.M., 2014. DNA Methylation Biomarkers: Cancer and Beyond. *Genes,* **5**(3), pp. 821-864.

MKHIKIAN, H., GRIGORIAN, A., LI, C.F., CHEN, H.L., NEWTON, B., ZHOU, R.W., BEETON, C., TOROSSIAN, S., TATARIAN, G.G., LEE, S.U., LAU, K., WALKER, E., SIMINOVITCH, K.A., CHANDY, K.G., YU, Z., DENNIS, J.W. and DEMETRIOU, M., 2011. Genetics and the environment converge to dysregulate N-glycosylation in multiple sclerosis. *Nature Communications,* **2**, pp. 334.

MOSLEHI, R., MILLS, J.L., SIGNORE, C., KUMAR, A., AMBROGGIO, X. and DZUTSEV, A., 2013. Integrative transcriptome analysis reveals dysregulation of canonical cancer molecular pathways in placenta leading to preeclampsia. *Scientific Reports,* **3**, pp. 10.1038/srep02407.

MULLER, M., VAN DEN BELD, A.W., BOTS, M.L., GROBBEE, D.E., LAMBERTS, S.W. and VAN DER SCHOUW, Y.T., 2004. Endogenous sex hormones and progression of carotid atherosclerosis in elderly men. *Circulation,* **109**(17), pp. 2074-2079.

NAGEL, J.D., 2014. Abstracts from the NIH Office of Research on Women's HealthEleventh Annual Interdisciplinary Women's Health Research Symposium November 6, 2014: Eleventh Annual NIH Interdisciplinary Women's Health Research Symposium November 6, 2014. *Journal of Women's Health,* **23**(10), pp. 849-877.

NCBI, 2016-last update, National Center for Biotechnology Information. Available: http://www.ncbi.nlm.nih.gov.qe2a-proxy.mun.ca/ [July/20, 2014].

NDIC, 2011-last update, Centers for Disease Control and Prevention. National Diabetes Fact Sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2011. Available: http://www.cdc.gov/DIABETES/pubs/pdf/ndfs_2011.pdf [July/20, 2014].

NEMOTO, M., SASAKI, T., DEEB, S.S., FUJIMOTO, W.Y. and TAJIMA, N., 2002. Differential effect of PPARgamma2 variants in the development of type 2 diabetes between native Japanese and Japanese Americans. *Diabetes Research and Clinical Practice,* **57**(2), pp. 131-137.

NESS, S.L., BEN-YOSEF, T., BAR-LEV, A., MADEO, A.C., BREWER, C.C., AVRAHAM, K.B., KORNREICH, R., DESNICK, R.J., WILLNER, J.P., FRIEDMAN, T.B. and GRIFFITH, A.J., 2003. Genetic homogeneity and phenotypic variability among Ashkenazi Jews with Usher syndrome type III. *Journal of Medical Genetics,* **40**(10), pp. 767-772.

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**(10), pp. 763-768.

NG, D.P., PLACHA, G., CHOO, S., CHIA, K.S., WARRAM, J.H. and KROLEWSKI, A.S., 2006. A disease haplotype for advanced nephropathy in type 2 diabetes at the ACE locus. *Diabetes,* **55**(9), pp. 2660-2663.

NG, D.P., TAI, B.C., KOH, D., TAN, K.W. and CHIA, K.S., 2005. Angiotensin-I converting enzyme insertion/deletion polymorphism and its association with diabetic nephropathy: a meta-analysis of studies reported between 1994 and 2004 and comprising 14,727 subjects. *Diabetologia,* **48**(5), pp. 1008-1016.

NG, M.C., HESTER, J.M., WING, M.R., LI, J., XU, J., HICKS, P.J., ROH, B.H., LU, L., DIVERS, J., LANGEFELD, C.D., FREEDMAN, B.I., PALMER, N.D. and BOWDEN, D.W., 2012. Genome-wide association of BMI in African Americans. *Obesity (Silver Spring, Md.),* **20**(3), pp. 622-627.

O'DEA, K., 1991. Westernisation, insulin resistance and diabetes in Australian aborigines. *The Medical Journal of Australia,* **155**(4), pp. 258-264.

O'DEA, K., 1984. Marked improvement in carbohydrate and lipid metabolism in diabetic Australian aborigines after temporary reversion to traditional lifestyle. *Diabetes,* **33**(6), pp. 596-603.

OHINMAA, A., JACOBS, P., SIMPSON, S. and JOHNSON, J.A., 2004. The projection of prevalence and cost of diabetes in Canada: 2000 to 2016. *Canadian Journal of Diabetes,* **28**(2), pp. 1-8.

OLSSON, A.H., VOLKOV, P., BACOS, K., DAYEH, T., HALL, E., NILSSON, E.A., LADENVALL, C., RÖNN, T. and LING, C., 2014. Genome-Wide Associations between Genetic and Epigenetic Variation Influence mRNA Expression and Insulin Secretion in

Human Pancreatic Islets. *PLoS Genetics,* **10**(11), e1004735. doi:10.1371/journal.pgen.1004735.

OOMS, L.M., HORAN, K.A., RAHMAN, P., SEATON, G., GURUNG, R., KETHESPARAN, D.S. and MITCHELL, C.A., 2009. The role of the inositol polyphosphate 5-phosphatases in cellular function and human disease. *The Biochemical Journal,* **419**(1), pp. 29-49.

ORG, E., EYHERAMENDY, S., JUHANSON, P., GIEGER, C., LICHTNER, P., KLOPP, N., VELDRE, G., DORING, A., VIIGIMAA, M., SOBER, S., TOMBERG, K., ECKSTEIN, G., KORA, KELGO, P., REBANE, T., SHAW-HAWKINS, S., HOWARD, P., ONIPINLA, A., DOBSON, R.J., NEWHOUSE, S.J., BROWN, M., DOMINICZAK, A., CONNELL, J., SAMANI, N., FARRALL, M., BRIGHT, CAULFIELD, M.J., MUNROE, P.B., ILLIG, T., WICHMANN, H.E., MEITINGER, T. and LAAN, M., 2009. Genome-wide scan identifies CDH13 as a novel susceptibility locus contributing to blood pressure determination in two European populations. *Human Molecular Genetics,* **18**(12), pp. 2288-2296.

ORTIZ, R.A. and BARNES, K.C., 2015. Genetics of Allergic Diseases. *Immunology and Allergy Clinics of North America,* **35**(1), pp. 19-44.

OSTBYE, T., WELBY, T.J., PRIOR, I.A., SALMOND, C.E. and STOKES, Y.M., 1989. Type 2 (non-insulin-dependent) diabetes mellitus, migration and westernisation: the Tokelau Island Migrant Study. *Diabetologia,* **32**(8), pp. 585-590.

PAN, X.R., LI, G.W., HU, Y.H., WANG, J.X., YANG, W.Y., AN, Z.X., HU, Z.X., LIN, J., XIAO, J.Z., CAO, H.B., LIU, P.A., JIANG, X.G., JIANG, Y.Y., WANG, J.P., ZHENG, H., ZHANG, H., BENNETT, P.H. and HOWARD, B.V., 1997. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. *Diabetes Care,* **20**(4), pp. 537-544.

PARRA, E.J., HOGGART, C.J., BONILLA, C., DIOS, S., NORRIS, J.M., MARSHALL, J.A., HAMMAN, R.F., FERRELL, R.E., MCKEIGUE, P.M. and SHRIVER, M.D., 2004. Relation of type 2 diabetes to individual admixture and candidate gene polymorphisms in the Hispanic American population of San Luis Valley, Colorado. *Journal of Medical Genetics,* **41**(11), pp. e116.

PATEL, J.V., VYAS, A., CRUICKSHANK, J.K., PRABHAKARAN, D., HUGHES, E., REDDY, K.S., MACKNESS, M.I., BHATNAGAR, D. and DURRINGTON, P.N., 2006. Impact of migration on coronary heart disease risk factors: comparison of Gujaratis in Britain and their contemporaries in villages of origin in India. *Atherosclerosis,* **185**(2), pp. 297-306.

PETTITT, D.J., SAAD, M.F., BENNETT, P.H., NELSON, R.G. and KNOWLER, W.C., 1990. Familial predisposition to renal disease in two generations of Pima Indians with type 2 (non-insulin-dependent) diabetes mellitus. *Diabetologia*, **33**(7), pp. 438-443.

PEZZOLESI, M.G., KATAVETIN, P., KURE, M., POZNIK, G.D., SKUPIEN, J., MYCHALECKYJ, J.C., RICH, S.S., WARRAM, J.H. and KROLEWSKI, A.S., 2009. Confirmation of genetic associations at ELMO1 in the GoKinD collection supports its role as a susceptibility gene in diabetic nephropathy. *Diabetes*, **58**(11), pp. 2698-2702.

PEZZOLESI, M.G., SKUPIEN, J., MYCHALECKYJ, J.C., WARRAM, J.H. and KROLEWSKI, A.S., 2010. Insights to the genetics of diabetic nephropathy through a genome-wide association study of the GoKinD collection. *Seminars in Nephrology*, **30**(2), pp. 126-140.

PINHAS-HAMIEL, O. and ZEITLER, P., 2005. The global spread of type 2 diabetes mellitus in children and adolescents. *The Journal of Pediatrics*, **146**(5), pp. 693-700.

PLACHA, G., POZNIK, G.D., DUNN, J., SMILES, A., KROLEWSKI, B., GLEW, T., PUPPALA, S., SCHNEIDER, J., ROGUS, J.J., RICH, S.S., DUGGIRALA, R., WARRAM, J.H. and KROLEWSKI, A.S., 2006. A genome-wide linkage scan for genes controlling variation in renal function estimated by serum cystatin C levels in extended families with type 2 diabetes. *Diabetes*, **55**(12), pp. 3358-3365.

POULSEN, P., KYVIK, K.O., VAAG, A. and BECK-NIELSEN, H., 1999. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. *Diabetologia*, **42**(2), pp. 139-145.

POURNOURMOHAMMADI, S., FARZAMI, B., OSTAD, S.N., AZIZI, E. and ABDOLLAHI, M., 2005. Effects of malathion subchronic exposure on rat skeletal muscle glucose metabolism. *Environmental Toxicology and Pharmacology*, **19**(1), pp. 191-196.

PRESTON, G.A. and WEINBERGER, D.R., 2005. Intermediate phenotypes in schizophrenia: a selective review. *Dialogues in Clinical Neuroscience*, **7**(2), pp. 165-179.

PRITCHARD, J.K. and ROSENBERG, N.A., 1999. Use of unlinked genetic markers to detect population stratification in association studies. *American Journal of Human Genetics*, **65**(1), pp. 220-228.

PURCELL, S., 2010-last update, PLINK (1.07) Documentation. Available: http://pngu.mgh.harvard.edu/purcell/plink/ [July/20, 2014].

PURCELL, S., 2009-last update, Package: PLINK (v 1.07, 2009). Available: http://pngu.mgh.harvard.edu/purcell/plink/ [July/20, 2014].

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. *American Journal of Human Genetics,* **81**(3), pp. 559-575.

QUINN, M., ANGELICO, M.C., WARRAM, J.H. and KROLEWSKI, A.S., 1996. Familial factors determine the development of diabetic nephropathy in patients with IDDM. *Diabetologia*, **39**(8), pp. 940-945.

R CORE TEAM, 2013-last update, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. A BibTeX entry for LaTeX users is @Manual{title = {R: A Language and Environment for Statistical Computing}, author = {{R Core Team}}, organization = {R Foundation for Statistical Computing}, address = {Vienna, Austria}, year = {2013}, url = {http://www.R-project.org/}, }. Available: http://www.R-project.org/ [July/20, 2014].

RAJALA, M.W. and SCHERER, P.E., 2003. Minireview: The adipocyte--at the crossroads of energy homeostasis, inflammation, and atherosclerosis. *Endocrinology,* **144**(9), pp. 3765-3773.

RAMACHANDRAN, A., SNEHALATHA, C., RAM, J., SELVAM, S., SIMON, M., NANDITHA, A., SHETTY, A.S., GODSLAND, I.F., CHATURVEDI, N., MAJEED, A., OLIVER, N., TOUMAZOU, C., ALBERTI, K.G. and JOHNSTON, D.G., 2013. Effectiveness of mobile phone messaging in prevention of type 2 diabetes by lifestyle modification in men in India: a prospective, parallel-group, randomised controlled trial. *The Lance Diabetes & Endocrinology,* **1**(3), pp. 191-198.

RAVUSSIN, E. and SMITH, S.R., 2002. Increased fat intake, impaired fat oxidation, and failure of fat cell proliferation result in ectopic fat storage, insulin resistance, and type 2 diabetes mellitus. *Annals of the New York Academy of Sciences,* **967**, pp. 363-378.

ROSS, K.A., 2011. Evidence for somatic gene conversion and deletion in bipolar disorder, Crohn's disease, coronary artery disease, hypertension, rheumatoid arthritis, type-1 diabetes, and type-2 diabetes. *BMC Medicine,* **9**, pp. 12-7015-9-12.

ROTTER, J.I., 1981. The modes of inheritance of insulin-dependent diabetes mellitus or the genetics of IDDM, no longer a nightmare but still a headache. *American Journal of Human Genetics,* **33**(6), pp. 835-851.

RUBINSTEIN, P., GINSBERG-FELLNER, F. and FALK, C., 1981. Genetics of Type I diabetes mellitus: a single, recessive predisposition gene mapping between HLA-B and GLO. With an appendix on the estimation of selection bias. *American Journal of Human Genetics,* **33**(6), pp. 865-882.

SABBAH, E., SAVOLA, K., EBELING, T., KULMALA, P., VAHASALO, P., ILONEN, J., SALMELA, P.I. and KNIP, M., 2000. Genetic, autoimmune, and clinical characteristics of childhood- and adult-onset type 1 diabetes. *Diabetes Care*, **23**(9), pp. 1326-1332.

SAKURA, H., BOND, C., WARREN-PERRY, M., HORSLEY, S., KEARNEY, L., TUCKER, S., ADELMAN, J., TURNER, R. and ASHCROFT, F.M., 1995. Characterization and variation of a human inwardly-rectifying-K-channel gene (KCNJ6): a putative ATP-sensitive K-channel subunit. *FEBS Letters*, **367**(2), pp. 193-197.

SCHELLING, J.R., ABBOUD, H.E., NICHOLAS, S.B., PAHL, M.V., SEDOR, J.R., ADLER, S.G., ARAR, N.H., BOWDEN, D.W., ELSTON, R.C., FREEDMAN, B.I., GODDARD, K.A., GUO, X., HANSON, R.L., IPP, E., IYENGAR, S.K., JUN, G., KAO, W.H., KASINATH, B.S., KIMMEL, P.L., KLAG, M.J., KNOWLER, W.C., NELSON, R.G., PAREKH, R.S., QUADE, S.R., RICH, S.S., SAAD, M.F., SCAVINI, M., SMITH, M.W., TAYLOR, K., WINKLER, C.A., ZAGER, P.G., SHAH, V.O. and FAMILY INVESTIGATION OF NEPHROPATHY AND DIABETES RESEARCH GROUP, 2008. Genome-wide scan for estimated glomerular filtration rate in multi-ethnic diabetic populations: the Family Investigation of Nephropathy and Diabetes (FIND). *Diabetes*, **57**(1), pp. 235-243.

SEAQUIST, E.R., GOETZ, F.C., RICH, S. and BARBOSA, J., 1989. Familial clustering of diabetic kidney disease. Evidence for genetic susceptibility to diabetic nephropathy. *The New England Journal of Medicine*, **320**(18), pp. 1161-1165.

SHAW, J.E., SICREE, R.A. and ZIMMET, P.Z., 2010. Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Research and Clinical Practice*, **87**(1), pp. 4-14.

SHAW, J.T., PURDIE, D.M., NEIL, H.A., LEVY, J.C. and TURNER, R.C., 1999. The relative risks of hyperglycaemia, obesity and dyslipidaemia in the relatives of patients with Type II diabetes mellitus. *Diabetologia*, **42**(1), pp. 24-27.

SHERWIN, R.S., 2004. Diabetes Mellitus. In: L. GOLDMAN and D. AUSIELLO, eds, *Cecil Textbook of Medicine.* 22nd edn. Pennsylvania, USA: Suanders (An Imprint of Elsevier), pp. 1425.

SHIFMAN, S., JOHANNESSON, M., BRONSTEIN, M., CHEN, S.X., COLLIER, D.A., CRADDOCK, N.J., KENDLER, K.S., LI, T., O'DONOVAN, M., O'NEILL, F.A., OWEN, M.J., WALSH, D., WEINBERGER, D.R., SUN, C., FLINT, J. and DARVASI, A., 2008. Genome-Wide Association Identifies a Common Variant in the Reelin Gene That Increases the Risk of Schizophrenia Only in Women. *PLoS Genetics*, **4**(2), e28. Epub 2008 Feb 15 doi:10.1371/journal.pgen.0040028.

SHIMAZAKI, A., KAWAMURA, Y., KANAZAWA, A., SEKINE, A., SAITO, S., TSUNODA, T., KOYA, D., BABAZONO, T., TANAKA, Y., MATSUDA, M., KAWAI, K., IIIZUMI, T., IMANISHI, M., SHINOSAKI, T., YANAGIMOTO, T., IKEDA, M., OMACHI, S., KASHIWAGI, A., KAKU, K., IWAMOTO, Y., KAWAMORI, R., KIKKAWA, R., NAKAJIMA, M., NAKAMURA, Y. and MAEDA, S., 2005. Genetic variations in the gene encoding ELMO1 are associated with susceptibility to diabetic nephropathy. *Diabetes,* **54**(4), pp. 1171-1178.

SHIMAZAKI, A., TANAKA, Y., SHINOSAKI, T., IKEDA, M., WATADA, H., HIROSE, T., KAWAMORI, R. and MAEDA, S., 2006. ELMO1 increases expression of extracellular matrix proteins and inhibits cell adhesion to ECMs. *Kidney International,* **70**(10), pp. 1769-1776.

SINGH, P.P., NAZ, I., GILMOUR, A., SINGH, M. and MASTANA, S., 2006. Association of APOE (Hha1) and ACE (I/D) gene polymorphisms with type 2 diabetes mellitus in North West India. *Diabetes Research and Clinical Practice,* **74**(1), pp. 95-102.

SINHA, S., GRUBER, S.B., MUKHERJEE, B. and RENNERT, G., 2008. Inference of the Haplotype Effect in a Matched Case-Control Study Using Unphased Genotype Data*. *The International Journal of Biostatistics,* **4**(1), Article 6. doi:10.2202/1557-4679.1079.

SMITH, M.W. and O'BRIEN, S.J., 2005. Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. *Nature Reviews Genetics,* **6**(8), pp. 623-632.

SO, H.C. and SHAM, P.C., 2011. Multiple testing and power calculations in genetic association studies. *Cold Spring Harbor Protocols,* **2011**(1), pp. pdb.top95.

SPIELMAN, R.S., BAKER, L. and ZMIJEWSKI, C.M., 1980. Gene dosage and suceptibility to insulin-dependent diabetes. *Annals of Human Genetics,* **44**(Pt 2), pp. 135-150.

SPRAY, B.J., ATASSI, N.G., TUTTLE, A.B. and FREEDMAN, B.I., 1995. Familial risk, age at onset, and cause of end-stage renal disease in white Americans. *Journal of the American Society of Nephrology : JASN,* **5**(10), pp. 1806-1810.

STUMVOLL, M., GOLDSTEIN, B.J. and VAN HAEFTEN, T.W., 2007. Pathogenesis of type 2 diabetes. *Endocrine Research,* **32**(1-2), pp. 19-37.

SUN, N.K., HUANG, S.L., LU, H.P., CHANG, T.C. and CHAO, C.C.K., 2015. Integrative transcriptomics-based identification of cryptic drivers of taxol-resistance genes in ovarian carcinoma cells: Analysis of the androgen receptor. *Oncotarget,* **6**(29), pp. 27065-27082.

SUZUKI, H., NAGASE, S., KIKUCHI, S., WANG, Y. and KOYAMA, A., 2000. Association of a missense Glu298Asp mutation of the endothelial nitric oxide synthase gene with end stage renal disease. *Clinical Chemistry,* **46**(11), pp. 1858-1860.

SZATMÁRI, T., MUNDT, F., HEIDARI-HAMEDANI, G., ZONG, F., FEROLLA, E., ALEXEYENKO, A., HJERPE, A. and DOBRA, K., 2012. Novel Genes and Pathways Modulated by Syndecan-1: Implications for the Proliferation and Cell-Cycle Regulation of Malignant Mesothelioma Cells. *PLoS ONE,* **7**(10), e48091. doi:10.1371/journal.pone.0048091.

TAN, E.D., DAVIS, W.A. and DAVIS, T.M., 2013. Characteristics and prognosis of Asian patients with type 2 diabetes from a multi-racial Australian community: the Fremantle Diabetes Study. *Internal Medicine Journal,* **43**(10), pp. 1125-1132.

TANG, H., PENG, J., WANG, P. and RISCH, N.J., 2005. Estimation of individual admixture: analytical and study design considerations. *Genetic Epidemiology,* **28**(4), pp. 289-301.

TAYLOR, K.W., NOVAK, R.F., ANDERSON, H.A., BIRNBAUM, L.S., BLYSTONE, C., DEVITO, M., JACOBS, D., KOHRLE, J., LEE, D.H., RYLANDER, L., RIGNELL-HYDBOM, A., TORNERO-VELEZ, R., TURYK, M.E., BOYLES, A.L., THAYER, K.A. and LIND, L., 2013. Evaluation of the association between persistent organic pollutants (POPs) and diabetes in epidemiological studies: a national toxicology program workshop review. *Environmental Health Perspectives,* **121**(7), pp. 774-783.

TECHASINTANA, P., DAVIS, J.W., GUBIN, M.M., MAGEE, J.D. and ATASOY, U., 2015. Transcriptomic-Wide Discovery of Direct and Indirect HuR RNA Targets in Activated CD4+ T Cells. *PLoS ONE,* **10**(7), e0129321. doi:10.1371/journal.pone.0129321.

THAMEEM, F., IGO, R.P.,Jr, FREEDMAN, B.I., LANGEFELD, C., HANSON, R.L., SCHELLING, J.R., ELSTON, R.C., DUGGIRALA, R., NICHOLAS, S.B., GODDARD, K.A., DIVERS, J., GUO, X., IPP, E., KIMMEL, P.L., MEONI, L.A., SHAH, V.O., SMITH, M.W., WINKLER, C.A., ZAGER, P.G., KNOWLER, W.C., NELSON, R.G., PAHL, M.V., PAREKH, R.S., KAO, W.H., RASOOLY, R.S., ADLER, S.G., ABBOUD, H.E., IYENGAR, S.K., SEDOR, J.R. and FAMILY INVESTIGATION OF NEPHROPATHY AND DIABETES RESEARCH GROUP, 2013. A genome-wide search for linkage of estimated glomerular filtration rate (eGFR) in the Family Investigation of Nephropathy and Diabetes (FIND). *PloS one,* **8**(12), e81888. doi: 10.1371/journal.pone.0081888. eCollection 2013.

THE EVIAN ANNUAL REPRODUCTION (EVAR) WORKSHOP GROUP 2010, THE EVIAN ANNUAL REPRODUCTION (EVAR) WORKSHOP GROUP 2010, FAUSER, B., DIEDRICH, K., BOUCHARD, P., DOMÍNGUEZ, F., MATZUK, M., FRANKS, S.,

HAMAMAH, S., SIMÓN, C., DEVROEY, P., EZCURRA, D. and HOWLES, C., 2011. Contemporary genetic technologies and female reproduction. *Human Reproduction Update,* **17**(6), pp. 829-847.

THOMAS, D.C., ed, 2004. *Statistical Methods in Genetic Epidemiology.* New York, USA: Oxford University Press.

TING, C., BANSAL, V., BATAL, I., MOUNAYAR, M., CHABTINI, L., EL AKIKI, G. and AZZI, J., 2012. Impairment of Immune Systems in Diabetes. In: S.I. AHMAD, ed, *Diabetes An Old Disease, a New Insight.* Texas, USA: Landes Bioscience I Springer Science+ Business Media, pp. 62-75.

TOLPPANEN, A.M., NEVALAINEN, T., KOLEHMAINEN, M., SEITSONEN, S., IMMONEN, I., UUSITUPA, M., KAARNIRANTA, K. and PULKKINEN, L., 2009. Single nucleotide polymorphisms of the tenomodulin gene (TNMD) in age-related macular degeneration. *Molecular Vision,* **15**, pp. 762-770.

TONG, J., TAYLOR, P., JOVCEVA, E., ST-GERMAIN, J.R., JIN, L.L., NIKOLIC, A., GU, X., LI, Z.H., TRUDEL, S. and MORAN, M.F., 2008. Tandem immunoprecipitation of phosphotyrosine-mass spectrometry (TIPY-MS) indicates C19ORF19 becomes tyrosine-phosphorylated and associated with activated epidermal growth factor receptor. *Journal of Proteome Research,* **7**(3), pp. 1067-1077.

TUNNACLIFFE, A., LIU, L., MOORE, J.K., LEVERSHA, M.A., JACKSON, M.S., PAPI, L., FERGUSON-SMITH, M.A., THIESEN, H.J. and PONDER, B.A., 1993. Duplicated KOX zinc finger gene clusters flank the centromere of human chromosome 10: evidence for a pericentric inversion during primate evolution. *Nucleic Acids Research,* **21**(6), pp. 1409-1417.

TUOMILEHTO, J., LINDSTROM, J., ERIKSSON, J.G., VALLE, T.T., HAMALAINEN, H., ILANNE-PARIKKA, P., KEINANEN-KIUKAANNIEMI, S., LAAKSO, M., LOUHERANTA, A., RASTAS, M., SALMINEN, V., UUSITUPA, M. and FINNISH DIABETES PREVENTION STUDY GROUP, 2001. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *The New England Journal of Medicine,* **344**(18), pp. 1343-1350.

TURNER, S.T., BAILEY, K.R., SCHWARTZ, G.L., CHAPMAN, A.B., CHAI, H.S. and BOERWINKLE, E., 2012. Genomic association analysis identifies multiple loci influencing antihypertensive response to an angiotensin II receptor blocker. *Hypertension,* **59**(6), pp. 1204-1211.

UEDA, H., HOWSON, J.M., ESPOSITO, L., HEWARD, J., SNOOK, H., CHAMBERLAIN, G., RAINBOW, D.B., HUNTER, K.M., SMITH, A.N., DI GENOVA, G., HERR, M.H., DAHLMAN, I., PAYNE, F., SMYTH, D., LOWE, C., TWELLS, R.C.,

HOWLETT, S., HEALY, B., NUTLAND, S., RANCE, H.E., EVERETT, V., SMINK, L.J., LAM, A.C., CORDELL, H.J., WALKER, N.M., BORDIN, C., HULME, J., MOTZO, C., CUCCA, F., HESS, J.F., METZKER, M.L., ROGERS, J., GREGORY, S., ALLAHABADIA, A., NITHIYANANTHAN, R., TUOMILEHTO-WOLF, E., TUOMILEHTO, J., BINGLEY, P., GILLESPIE, K.M., UNDLIEN, D.E., RONNINGEN, K.S., GUJA, C., IONESCU-TIRGOVISTE, C., SAVAGE, D.A., MAXWELL, A.P., CARSON, D.J., PATTERSON, C.C., FRANKLYN, J.A., CLAYTON, D.G., PETERSON, L.B., WICKER, L.S., TODD, J.A. and GOUGH, S.C., 2003. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature,* **423**(6939), pp. 506-511.

USRDS, 2013-last update, U.S. Renal Data System, USRDS 2013 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2013. *Notice: The data reported here have been supplied by the United States Renal Data System (USRDS). The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy or interpretation of the U.S. government.*. Available: http://www.usrds.org/atlas.aspx [July/20, 2014].

VAARALA, O., ATKINSON, M.A. and NEU, J., 2008. The "perfect storm" for type 1 diabetes: the complex interplay between intestinal microbiota, gut permeability, and mucosal immunity. *Diabetes,* **57**(10), pp. 2555-2562.

VARDARLI, I., BAIER, L.J., HANSON, R.L., AKKOYUN, I., FISCHER, C., ROHMEISS, P., BASCI, A., BARTRAM, C.R., VAN DER WOUDE, F.J. and JANSSEN, B., 2002. Gene for susceptibility to diabetic nephropathy in type 2 diabetes maps to 18q22.3-23. *Kidney International,* **62**(6), pp. 2176-2183.

WAGENKNECHT, L.E., ROSEMAN, J.M. and HERMAN, W.H., 1991. Increased incidence of insulin-dependent diabetes mellitus following an epidemic of Coxsackievirus B5. *American Journal of Epidemiology,* **133**(10), pp. 1024-1031.

WAGNER, M., SOBCZYNSKI, M., KARABON, L., BILINSKA, M., POKRYSZKO-DRAGAN, A., PAWLAK-ADAMSKA, E., CYRUL, M., KUSNIERCZYK, P. and JASEK, M., 2015. Polymorphisms in CD28, CTLA-4, CD80 and CD86 genes may influence the risk of multiple sclerosis and its age of onset. *Journal of Neuroimmunology,* **288**, pp. 79-86.

WAKEM, P., BURNS, R.P.,Jr, RAMIREZ, F., ZLOTNICK, D., FERBEL, B., HAIDARIS, C.G. and GASPARI, A.A., 2000. Allergens and irritants transcriptionally upregulate CD80 gene expression in human keratinocytes. *The Journal of Investigative Dermatology,* **114**(6), pp. 1085-1092.

WANDELL, P.E., JOHANSSON, S.E., GAFVELS, C., HELLENIUS, M.L., DE FAIRE, U. and SUNDQUIST, J., 2008. Estimation of diabetes prevalence among immigrants from the Middle East in Sweden by using three different data sources. *Diabetes & Metabolism,* **34**(4 Pt 1), pp. 328-333.

WANG, F., XU, C.Q., HE, Q., CAI, J.P., LI, X.C., WANG, D., XIONG, X., LIAO, Y.H., ZENG, Q.T., YANG, Y.Z., CHENG, X., LI, C., YANG, R., WANG, C.C., WU, G., LU, Q.L., BAI, Y., HUANG, Y.F., YIN, D., YANG, Q., WANG, X.J., DAI, D.P., ZHANG, R.F., WAN, J., REN, J.H., LI, S.S., ZHAO, Y.Y., FU, F.F., HUANG, Y., LI, Q.X., SHI, S.W., LIN, N., PAN, Z.W., LI, Y., YU, B., WU, Y.X., KE, Y.H., LEI, J., WANG, N., LUO, C.Y., JI, L.Y., GAO, L.J., LI, L., LIU, H., HUANG, E.W., CUI, J., JIA, N., REN, X., LI, H., KE, T., ZHANG, X.Q., LIU, J.Y., LIU, M.G., XIA, H., YANG, B., SHI, L.S., XIA, Y.L., TU, X. and WANG, Q.K., 2011. Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. *Nature Genetics,* **43**(4), pp. 345-349.

WANIC, K., KROLEWSKI, B., JU, W., PLACHA, G., NIEWCZAS, M.A., WALKER, W., WARRAM, J.H., KRETZLER, M. and KROLEWSKI, A.S., 2013. Transcriptome Analysis of Proximal Tubular Cells (HK-2) Exposed to Urines of Type 1 Diabetes Patients at Risk of Early Progressive Renal Function Decline. *PLoS ONE,* **8**(3), e57751. doi:10.1371/journal.pone.0057751.

WATERS, K.M., STRAM, D.O., HASSANEIN, M.T., LE MARCHAND, L., WILKENS, L.R., MASKARINEC, G., MONROE, K.R., KOLONEL, L.N., ALTSHULER, D., HENDERSON, B.E. and HAIMAN, C.A., 2010. Consistent association of type 2 diabetes risk variants found in europeans in diverse racial and ethnic groups. *PLoS Genetics,* **6**(8), e1001078. doi:10.1371/journal.pgen.1001078.

WEISS, R., DZIURA, J., BURGERT, T.S., TAMBORLANE, W.V., TAKSALI, S.E., YECKEL, C.W., ALLEN, K., LOPES, M., SAVOYE, M., MORRISON, J., SHERWIN, R.S. and CAPRIO, S., 2004. Obesity and the metabolic syndrome in children and adolescents. *The New England Journal of Medicine,* **350**(23), pp. 2362-2374.

WELLCOME TRUST CASE CONTROL CONSORTIUM, 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature,* **447**(7145), pp. 661-678.

WHINCUP, P.H., KAYE, S.J., OWEN, C.G., HUXLEY, R., COOK, D.G., ANAZAWA, S., BARRETT-CONNOR, E., BHARGAVA, S.K., BIRGISDOTTIR, B.E., CARLSSON, S., DE ROOIJ, S.R., DYCK, R.F., ERIKSSON, J.G., FALKNER, B., FALL, C., FORSEN, T., GRILL, V., GUDNASON, V., HULMAN, S., HYPPONEN, E., JEFFREYS, M., LAWLOR, D.A., LEON, D.A., MINAMI, J., MISHRA, G., OSMOND, C., POWER, C., RICH-EDWARDS, J.W., ROSEBOOM, T.J., SACHDEV, H.S., SYDDALL, H., THORSDOTTIR, I., VANHALA, M., WADSWORTH, M. and

YARBROUGH, D.E., 2008. Birth weight and risk of type 2 diabetes: a systematic review. *Jama*, **300**(24), pp. 2886-2897.

WHITING, D.R., GUARIGUATA, L., WEIL, C. and SHAW, J., 2011. IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Research and Clinical Practice*, **94**(3), pp. 311-321.

WHO, 2011-last update, World Health Organization. Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus: abbreviated report of a WHO consultation. Available: http://www.who.int/diabetes/publications/report-hba1c_2011.pdf [July/20, 2014].

WILLARD, S.S. and KOOCHEKPOUR, S., 2012. Regulators of gene expression as biomarkers for prostate cancer. *American Journal of Cancer Research*, **2**(6), pp. 620-657.

WU, C., DEWAN, A., HOH, J. and WANG, Z., 2011. A comparison of association methods correcting for population stratification in case-control studies. *Annals of Human Genetics*, **75**(3), pp. 418-427.

YANG, Y., GURUNG, B., WU, T., WANG, H., STOFFERS, D.A. and HUA, X., 2010. Reversal of preexisting hyperglycemia in diabetic mice by acute deletion of the Men1 gene. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(47), pp. 20358-20363.

ZHANG, J., DING, L., HOLMFELDT, L., WU, G., HEATLEY, S.L., PAYNE-TURNER, D., EASTON, J., CHEN, X., WANG, J., RUSCH, M., LU, C., CHEN, S.C., WEI, L., COLLINS-UNDERWOOD, J.R., MA, J., ROBERTS, K.G., POUNDS, S.B., ULYANOV, A., BECKSFORT, J., GUPTA, P., HUETHER, R., KRIWACKI, R.W., PARKER, M., MCGOLDRICK, D.J., ZHAO, D., ALFORD, D., ESPY, S., BOBBA, K.C., SONG, G., PEI, D., CHENG, C., ROBERTS, S., BARBATO, M.I., CAMPANA, D., COUSTAN-SMITH, E., SHURTLEFF, S.A., RAIMONDI, S.C., KLEPPE, M., COOLS, J., SHIMANO, K.A., HERMISTON, M.L., DOULATOV, S., EPPERT, K., LAURENTI, E., NOTTA, F., DICK, J.E., BASSO, G., HUNGER, S.P., LOH, M.L., DEVIDAS, M., WOOD, B., WINTER, S., DUNSMORE, K.P., FULTON, R.S., FULTON, L.L., HONG, X., HARRIS, C.C., DOOLING, D.J., OCHOA, K., JOHNSON, K.J., OBENAUER, J.C., EVANS, W.E., PUI, C.H., NAEVE, C.W., LEY, T.J., MARDIS, E.R., WILSON, R.K., DOWNING, J.R. and MULLIGHAN, C.G., 2012. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature*, **481**(7380), pp. 157-163.

ZHANG, S.X., SANDERS, E., FLIESLER, S.J. and WANG, J.J., 2014. Endoplasmic Reticulum Stress and the Unfolded Protein Responses in Retinal Degeneration. *Experimental Eye Research*, **0**, pp. 30-40.

ZHENG, G., JOO, J., ZHANG, C. and GELLER, N.L., 2007. Testing association for markers on the X chromosome. *Genetic Epidemiology*, **31**(8), pp. 834-843.

ZIEGLER, A. and KONIG, I.R., eds, 2010. *A Statistical Approach to Genetic Epidemiology*. 2nd edn. Lubeck, Germany: Wiley-VCH Verlag GmbH & Co. KGaA.

ZIMMET, P., ALBERTI, K.G. and SHAW, J., 2001. Global and societal implications of the diabetes epidemic. *Nature*, **414**(6865), pp. 782-787.

Appendices

Appendix 1: Study Approval

Received: from ella.ucs.mun.ca (134.153.232.56) by exchange.med.mun.ca
(134.153.88.87) with Microsoft SMTP Server (TLS) id 14.1.289.1; Wed, 22 Jan
2014 11:19:38 -0330
Received: from nihrelayxway.hub.nih.gov (nihrelayxway.hub.nih.gov
[128.231.90.106]) by ella.ucs.mun.ca (8.13.8/8.13.8) with ESMTP id
s0MEnYVn016901 for <saeed.samet@med.mun.ca>; Wed, 22 Jan 2014 11:19:35 -0330
X-IronPortListener: NIH_Relay
X-SBRS: None
X-IronPort-AV: E=Sophos;i="4.95,700,1384318800";
d="scan'208,217";a="149008260"
Received: from cesedge01.nih.gov (HELO mailfwd.nih.gov) ([128.231.90.107]) by
nihrelayxway.hub.nih.gov with ESMTP/TLS/AES128-SHA; 22 Jan 2014 09:49:01
-0500
Received: from mail2.ncbi.nlm.nih.gov (130.14.26.42) by edge.hub.nih.gov
(128.231.90.76) with Microsoft SMTP Server id 14.3.158.1; Wed, 22 Jan 2014
09:49:01 -0500
Received: from mssql75 (mssql75.ncbi.nlm.nih.gov [130.14.18.140]) by
mail2.ncbi.nlm.nih.gov (Postfix) with ESMTP id 6B5F040003; Wed, 22 Jan 2014
09:49:01 -0500 (EST)
From: <dbgap-reply@ncbi.nlm.nih.gov>
To: <saeed.samet@med.mun.ca>
CC: <diroor@mun.ca>
Date: Wed, 22 Jan 2014 09:49:01 -0500
Subject: APPROVAL of your request [#24469-4] for access phs000333/DNAR
Message-ID: <20140122144901.6B5F040003@mail2.ncbi.nlm.nih.gov>
X-Spam-Flag: NO
X-Spam-Checker-Version: SpamAssassin 3.003001 (2010-03-16)
X-Spam-Status: No, hits=-3.2 required=4.0
tests=HTML_MESSAGE,MIME_HTML_ONLY,PHISH_HOOKINFO,
RCVD_IN_DNSWL_HI,T_RP_MATCHES_RCVD
version=3.3.1
X-Spam-Report: -5.0 RCVD_IN_DNSWL_HI RBL: Sender listed at http://www.dnswl.org/, high
* trust
* [128.231.90.106 listed in list.dnswl.org]
* -0.0 T_RP_MATCHES_RCVD Envelope sender domain matches handover relay
* domain
* 0.0 HTML_MESSAGE BODY: HTML included in message
* 1.0 MIME_HTML_ONLY BODY: Message only has text/html MIME parts
* 0.8 PHISH_HOOKINFO Email appears to be attempting to hook user
* information but is not from a trusted source.
X-Scanned-By: MIMEDefang 2.73 on 134.153.232.56
Return-Path: dbgap-reply@ncbi.nlm.nih.gov
X-MS-Exchange-Organization-AuthSource: exchange.med.mun.ca
X-MS-Exchange-Organization-AuthAs: Anonymous
X-MS-Exchange-Organization-AVStamp-Mailbox: NAI;56074317;0;novirus
Content-Type: multipart/alternative; boundary="B_3473234618_17816891"
MIME-Version: 1.0

--B_3473234618_17816891
Content-Type: text/plain; charset="US-ASCII"
Content-Transfer-Encoding: 7bit

Dear Saeed Samet,

This email was generated by the National Center for Biotechnology Information Genotypes and Phenotypes Database (NCBI dbGaP) Data Access Request system at the National Institutes of Health.

NIH has APPROVED your request [#24469-4]
for Diabetes, Kidney Disease, Retinopathy and/or Related Diseases access of Family Investigation of Nephropathy and Diabetes (FIND) Study for project #5841: "Detecting the possible impact of polymorphisms in the engulfment and cell motility 1 (ELMO1) and silent information regulator 1 (SIRT1) gene regions on diabetic nephropathy in different ethnic groups." .

Before accessing the data, please review the terms of access of the Data Use Agreement
<http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?view_pdf&wlid=5841&tlsid=634>
that you and your institution have signed.

Approved Users are strongly encouraged to publish their results in peer-reviewed journals and to present research findings at scientific meetings. However, under the terms of the NIH GWAS Policy and the Data Use Certification agreed to by you and your institution, the NIH expects that Approved Users will not submit findings for publication OR presentation for a defined period of exclusivity provided to Contributing Investigators. This exclusivity period concludes with the Embargo Date identified for each dataset within the NIH GWAS Data Repository (dbGaP). The specific Embargo Date(s) for the dataset(s) in which you are being granted access can be located at the Downloadable Manifest
<http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?view_pdf&permid=1110&tlsid=634>
(the latest embargo date listed is: "2011-10-25"). During this embargo period, only Contributing Investigators, or those who have been granted permission by the Contributing Investigator, may submit analyses or findings for publication, presentation, or any other form of public dissemination of the work.

To access the data, log into the dbGaP Authorized Access
<https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login> website, then click the "My Requests" link in the menu at the top of the page. You will be presented with a list of approved requests. Select the requests you wish to access by clicking on 'Request Files' (last column of the table). You can also use the following direct link to a list of files:
http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=request_details_pi&filter=arid&arid=57176

If you have questions related to your access request for the 'Family Investigation of Nephropathy and Diabetes (FIND) Study', please contact the 'NIDDK Central Repository GWAS Data Access Committee' Data Access Committee at niddk-dac@mail.nih.gov.

If you have any questions regarding Controlled Access Portal please contact NCBI dbGaP help desk at dbgap-help@ncbi.nlm.nih.gov. Please do not reply to this message.

--B_3473234618_17816891
Content-Type: text/html; charset="US-ASCII"
Content-Transfer-Encoding: quoted-printable

<html><head>
<meta http-equiv=3D"Content-Type" content=3D"text/html; charset=3Dus-ascii"=
>
</head>
<body>
Dear Saeed Samet,
This email was generated by the National Center for Biotechnology Inform=
ation Genotypes and Phenotypes Database (NCBI dbGaP) Data Access Request sy=
stem at the National Institutes of Health.

NIH has APPROVED your request [#24469-4]
for Diabetes, Kidney Disease, Retinopathy and/or Related Diseases ac=
cess
of Family Investigation of Nephropathy and Diabetes (FIND) Study
for project #5841: "Detecting the possible impact of polymorphism=
s in the engulfment and cell motility 1 (ELMO1) and silent information regu=
lator 1 (SIRT1) gene regions on diabetic nephropathy in different ethnic gr=
oups." .

Before accessing the data, please review the terms of access of the <a h=
ref=3D"http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?view_pdf&wlid=3D5841&am=
p;tlsid=3D634">
Data Use Agreement that you and your institution have signed.
Approved Users are strongly encouraged to publish their results in peer-=
reviewed journals and to present research findings at scientific meetings. =
However, under the terms of the NIH GWAS Policy and the Data Use Certificat=
ion agreed to by you and your institution,
the NIH expects that Approved Users will not submit findings for publicati=
on OR presentation for a defined period of exclusivity provided to Contribu=
ting Investigators. This exclusivity period concludes with the Embargo Date=
identified for each dataset within
the NIH GWAS Data Repository (dbGaP). The specific Embargo Date(s) for the=
dataset(s) in which you are being granted access can be located at the

<a href=3D"http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?view_pdf&permid=3D1=
110&tlsid=3D634">
Downloadable Manifest (the latest embargo date listed is: "2011-10=
-25"). During this embargo period, only Contributing Investigators, or=
 those who have been granted permission by the Contributing Investigator, m=
ay submit analyses or findings for publication,
 presentation, or any other form of public dissemination of the work.
To access the data, log into the <a href=3D"https://dbgap.ncbi.nlm.nih.g=
ov/aa/wga.cgi?page=3Dlogin">
dbGaP Authorized Access website, then click the "My Requests"=
 link in the menu at the top of the page. You will be presented with a list=
 of approved requests. Select the requests you wish to access by clicking o=
n 'Request Files' (last column of the table).
 You can also use the following direct link to a list of files: <a href=3D"=
http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=3Drequest_details_pi&filt=
er=3Darid&arid=3D57176">
http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=3Drequest_details_pi&filt=
er=3Darid&arid=3D57176
If you have questions related to your access request for the 'Family Inv=
estigation of Nephropathy and Diabetes (FIND) Study', please contact the 'N=
IDDK Central Repository GWAS Data Access Committee' Data Access Committee a=
t
niddk-dac@mail.nih.gov.
If you have any questions regarding Controlled Access Portal please cont=
act NCBI dbGaP help desk at
<a href=3D"mailto:dbgap-help@ncbi.nlm.nih.gov">dbgap-help@ncbi.nlm.nih.gov<=
/a>. Please do not reply to this message.

</body>
</html>

--B_3473234618_17816891--

Appendix 2: Study Variables

Variable accession	Variable name	Variable description
phv00161411.v1.p1	SubjectId	De-identified subject ID
phv00161412.v1.p1	Sex	Sex
phv00161413.v1.p1	Race	Race
phv00161414.v1.p1	Age	Age at enrollment
phv00161415.v1.p1	DiabeticNephropathy	Diagnosed with diabetic nephropathy
phv00161416.v1.p1	Diabetes	Diagnosed with diabetes
phv00161417.v1.p1	Diabetes_Duration	Age of diabetes diagnosis
phv00161418.v1.p1	Diabetes_AgeOnset	Duration of diabetes
phv00161419.v1.p1	Retinopathy	Retinopathy in either eye?
phv00161420.v1.p1	KidneyHistology	Histological findings in kidneys
phv00161421.v1.p1	KidneyFailure	Kidney failure requiring dialysis or transplant?
phv00161422.v1.p1	KidneyFailure_AgeOnset	Age of kidney failure
phv00161423.v1.p1	KidneyFailure_Duration	Duration of kidney failure
phv00161424.v1.p1	Historic_UrineProteinExcretion	Historic protein excretion in urine
phv00161425.v1.p1	Historic_SerumCreatinine	Historic serum creatinine
phv00161426.v1.p1	Historic_UrineProtein	Historic urine protein excretion
phv00161427.v1.p1	Historic_UrineAlbuminSign	For values that fall below the

		threshold of detection
phv00161428.v1.p1	Historic_UrineAlbumin	Historic urine albumin excretion
phv00161429.v1.p1	Historic_UrineCreatinine	Historic urine creatinine excretion
phv00161430.v1.p1	Historic_PCR	Historic urine protein: creatinine ratio
phv00161431.v1.p1	Historic_ACR	Historic urine albumin: creatinine ratio
phv00161432.v1.p1	Historic_GFR	Historic (estimated) glomerular filtration rate
phv00161433.v1.p1	Lab_SerumCreatinine	Lab serum creatinine
phv00161434.v1.p1	Lab_UrineProtein	Lab urine protein excretion
phv00161435.v1.p1	Lab_UrineAlbumin	Lab urine albumin excretion
phv00161436.v1.p1	Lab_UrineCreatinine	Lab urine creatinine excretion
phv00161437.v1.p1	Lab_PCR	Lab protein-to-creatinine ratio
phv00161438.v1.p1	Lab_ACR	Lab albumin-to-creatinine ratio
phv00161439.v1.p1	Lab_GFR	Lab estimated glomerular filtration rate
phv00161440.v1.p1	Lab_HbA1c	Lab hemoglobin A1c, measured as part of study protocol
phv00161441.v1.p1	CurrentHeight	Height at enrollment
phv00161442.v1.p1	CurrentWeight	Weight at enrollment

phv00161443.v1.p1	AgeMaxWeight	Age when at maximum weight
phv00161444.v1.p1	HeightMaxWeight	Height when at maximum weight
phv00161445.v1.p1	MaxWeight	Maximum weight
phv00161446.v1.p1	Amputation	Did either leg need amputation?
phv00161447.v1.p1	LegAmputatedLeft	Left leg amputated?
phv00161448.v1.p1	LegAmputatedRight	Right leg amputated?
phv00161449.v1.p1	HBP	High blood pressure
phv00161450.v1.p1	HBP_Duration	Duration of high blood pressure
phv00161451.v1.p1	ACE_ARBs	Taking ACE inhibitors or ARBS?
phv00161452.v1.p1	Insulin	Taking insulin?

Appendix 3: African Americans log files

1) sex_check.log

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@

Web-based version check ( --noweb to skip )
Connecting to web...  OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [ sex_check.log ]
Analysis started: Wed May 20 01:22:58 2015

Options in effect:
        --bfile FIND_AA_5plus
        --check-sex
        --out sex_check

Reading map (extended format) from [ FIND_AA_5plus.bim ]
932534 markers to be included from [ FIND_AA_5plus.bim ]
Reading pedigree information from [ FIND_AA_5plus.fam ]
864 individuals read from [ FIND_AA_5plus.fam ]
859 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
581 cases, 278 controls and 5 missing
291 males, 573 females, and 0 of unspecified sex
Reading genotype bitfile from [ FIND_AA_5plus.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 932534 SNPs
864 founders and 0 non-founders found
36273 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ sex_check.hh ]
24247 SNPs with no founder genotypes observed
Warning, MAF set to 0 for these SNPs (see --nonfounders)
Writing list of these SNPs to [ sex_check.nof ]
Total genotyping rate in remaining individuals is 0.964565
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 932534 SNPs
After filtering, 581 cases, 278 controls and 5 missing
After filtering, 291 males, 573 females, and 0 of unspecified sex
Converting data to Individual-major format
Writing X-chromosome sex check results to [ sex_check.sexcheck ]
```

Analysis finished: Wed May 20 01:25:27 2015
2) genomissing.log

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@

Web-based version check ( --noweb to skip )
Recent cached web-check found... OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [ genomissing.log ]
Analysis started: Wed May 20 01:29:52 2015

Options in effect:
        --bfile FIND_AA_5plus
        --missing
        --out genomissing

Reading map (extended format) from [ FIND_AA_5plus.bim ]
932534 markers to be included from [ FIND_AA_5plus.bim ]
Reading pedigree information from [ FIND_AA_5plus.fam ]
864 individuals read from [ FIND_AA_5plus.fam ]
859 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
581 cases, 278 controls and 5 missing
291 males, 573 females, and 0 of unspecified sex
Reading genotype bitfile from [ FIND_AA_5plus.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 932534 SNPs
864 founders and 0 non-founders found
36273 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ genomissing.hh ]
24247 SNPs with no founder genotypes observed
Warning, MAF set to 0 for these SNPs (see --nonfounders)
Writing list of these SNPs to [ genomissing.nof ]
Writing individual missingness information to [ genomissing.imiss ]
Writing locus missingness information to [ genomissing.lmiss ]
Total genotyping rate in remaining individuals is 0.96452
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 932534 SNPs
After filtering, 581 cases, 278 controls and 5 missing
After filtering, 291 males, 573 females, and 0 of unspecified sex

Analysis finished: Wed May 20 01:32:19 2015
```

3) FIND_AA_5plus_ibd.log

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@

Web-based version check ( --noweb to skip )
Recent cached web-check found... OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [ FIND_AA_5plus_ibd.log ]
Analysis started: Wed May 20 03:06:21 2015

Options in effect:
        --bfile FIND_AA_5plus_sex_imputed_independentSNPs
        --genome
        --out FIND_AA_5plus_ibd

Reading map (extended format) from [ FIND_AA_5plus_sex_imputed_independentSNPs.bim ]
312743 markers to be included from [ FIND_AA_5plus_sex_imputed_independentSNPs.bim ]
Reading pedigree information from [ FIND_AA_5plus_sex_imputed_independentSNPs.fam ]
853 individuals read from [ FIND_AA_5plus_sex_imputed_independentSNPs.fam ]
853 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
575 cases, 278 controls and 0 missing
286 males, 567 females, and 0 of unspecified sex
Reading genotype bitfile from [ FIND_AA_5plus_sex_imputed_independentSNPs.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 312743 SNPs
853 founders and 0 non-founders found
22335 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ FIND_AA_5plus_ibd.hh ]
1 SNPs with no founder genotypes observed
Warning, MAF set to 0 for these SNPs (see --nonfounders)
Writing list of these SNPs to [ FIND_AA_5plus_ibd.nof ]
Total genotyping rate in remaining individuals is 0.994493
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 312743 SNPs
After filtering, 575 cases, 278 controls and 0 missing
After filtering, 286 males, 567 females, and 0 of unspecified sex
Converting data to Individual-major format
Writing whole genome IBS/IBD information to [ FIND_AA_5plus_ibd.genome ]
Filtering output to include pairs with ( 0 <= PI-HAT <= 1 )
```

Analysis finished: Wed May 20 05:08:36 2015

4) FIND_AA_5plus_sex_problem_removed.log

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@

Web-based version check ( --noweb to skip )
Recent cached web-check found... OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [ FIND_AA_5plus_sex_problem_removed.log ]
Analysis started: Sun Jul 19 13:59:00 2015

Options in effect:
        --bfile FIND_AA_5plus
        --remove sex_problem.txt
        --make-bed
        --out FIND_AA_5plus_sex_problem_removed

Reading map (extended format) from [ FIND_AA_5plus.bim ]
932534 markers to be included from [ FIND_AA_5plus.bim ]
Reading pedigree information from [ FIND_AA_5plus.fam ]
864 individuals read from [ FIND_AA_5plus.fam ]
859 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
581 cases, 278 controls and 5 missing
291 males, 573 females, and 0 of unspecified sex
Reading genotype bitfile from [ FIND_AA_5plus.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Reading individuals to remove [ sex_problem.txt ] ... 11 read
11 individuals removed with --remove option
Before frequency and genotyping pruning, there are 932534 SNPs
853 founders and 0 non-founders found
36273 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ FIND_AA_5plus_sex_problem_removed.hh ]
24247 SNPs with no founder genotypes observed
Warning, MAF set to 0 for these SNPs (see --nonfounders)
Writing list of these SNPs to [ FIND_AA_5plus_sex_problem_removed.nof ]
Total genotyping rate in remaining individuals is 0.970195
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 932534 SNPs
```

After filtering, 575 cases, 278 controls and 0 missing
After filtering, 286 males, 567 females, and 0 of unspecified sex
Writing pedigree information to [FIND_AA_5plus_sex_problem_removed.fam]
Writing map (extended format) information to [FIND_AA_5plus_sex_problem_removed.bim]
Writing genotype bitfile to [FIND_AA_5plus_sex_problem_removed.bed]
Using (default) SNP-major mode

Analysis finished: Sun Jul 19 14:00:59 2015

5) FIND_AA_5plus_sex_imputed.log

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@
```

Web-based version check (--noweb to skip)
Recent cached web-check found... OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [FIND_AA_5plus_sex_imputed.log]
Analysis started: Sun Jul 19 14:11:17 2015

Options in effect:
--bfile FIND_AA_5plus_sex_problem_removed
--impute-sex
--make-bed
--out FIND_AA_5plus_sex_imputed

Reading map (extended format) from [FIND_AA_5plus_sex_problem_removed.bim]
932534 markers to be included from [FIND_AA_5plus_sex_problem_removed.bim]
Reading pedigree information from [FIND_AA_5plus_sex_problem_removed.fam]
853 individuals read from [FIND_AA_5plus_sex_problem_removed.fam]
853 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
575 cases, 278 controls and 0 missing
286 males, 567 females, and 0 of unspecified sex
Reading genotype bitfile from [FIND_AA_5plus_sex_problem_removed.bed]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 932534 SNPs
853 founders and 0 non-founders found
36273 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [FIND_AA_5plus_sex_imputed.hh]
24247 SNPs with no founder genotypes observed
Warning, MAF set to 0 for these SNPs (see --nonfounders)
Writing list of these SNPs to [FIND_AA_5plus_sex_imputed.nof]

Total genotyping rate in remaining individuals is 0.970195
0 SNPs failed missingness test (GENO > 1)
0 SNPs failed frequency test (MAF < 0)
After frequency and genotyping pruning, there are 932534 SNPs
After filtering, 575 cases, 278 controls and 0 missing
After filtering, 286 males, 567 females, and 0 of unspecified sex
Converting data to Individual-major format
Writing X-chromosome sex check results to [FIND_AA_5plus_sex_imputed.sexcheck]
Writing pedigree information to [FIND_AA_5plus_sex_imputed.fam]
Writing map (extended format) information to [FIND_AA_5plus_sex_imputed.bim]
Writing genotype bitfile to [FIND_AA_5plus_sex_imputed.bed]
Using (default) SNP-major mode
Converting data to SNP-major format

Analysis finished: Sun Jul 19 14:16:40 2015

6) FIND_AA_5plus_pruning.log

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@
```

Web-based version check (--noweb to skip)
Recent cached web-check found... OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [FIND_AA_5plus_pruning.log]
Analysis started: Sun Jul 19 14:53:48 2015

Options in effect:
--bfile FIND_AA_5plus
--indep 50 5 2
--out FIND_AA_5plus_pruning

Reading map (extended format) from [FIND_AA_5plus.bim]
932534 markers to be included from [FIND_AA_5plus.bim]
Reading pedigree information from [FIND_AA_5plus.fam]
864 individuals read from [FIND_AA_5plus.fam]
859 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
581 cases, 278 controls and 5 missing
291 males, 573 females, and 0 of unspecified sex
Reading genotype bitfile from [FIND_AA_5plus.bed]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 932534 SNPs

```
864 founders and 0 non-founders found
36273 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ FIND_AA_5plus_pruning.hh ]
24247 SNPs with no founder genotypes observed
Warning, MAF set to 0 for these SNPs (see --nonfounders)
Writing list of these SNPs to [ FIND_AA_5plus_pruning.nof ]
Total genotyping rate in remaining individuals is 0.96452
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 932534 SNPs
After filtering, 581 cases, 278 controls and 5 missing
After filtering, 291 males, 573 females, and 0 of unspecified sex
Performing LD-based pruning...
Writing pruned-in SNPs to [ FIND_AA_5plus_pruning.prune.in ]
Writing pruned-out SNPs to [ FIND_AA_5plus_pruning.prune.out ]
Scanning from chromosome 1 to Y

Scan region on chromosome 1 from [ rs10458597 ] to [ rs7521920 ]
For chromosome 1, 48947 SNPs pruned out, 24399 remaining
Scan region on chromosome 2 from [ rs10181821 ] to [ rs12478296 ]
For chromosome 2, 51460 SNPs pruned out, 24404 remaining
Scan region on chromosome 3 from [ rs13089679 ] to [ rs2313099 ]
For chromosome 3, 41972 SNPs pruned out, 20295 remaining
Scan region on chromosome 4 from [ rs1988520 ] to [ rs1474523 ]
For chromosome 4, 38716 SNPs pruned out, 18781 remaining
Scan region on chromosome 5 from [ rs7704488 ] to [ rs6894609 ]
For chromosome 5, 38816 SNPs pruned out, 19092 remaining
Scan region on chromosome 6 from [ rs4959515 ] to [ rs6931065 ]
For chromosome 6, 38946 SNPs pruned out, 18839 remaining
Scan region on chromosome 7 from [ rs12666071 ] to [ rs1985369 ]
For chromosome 7, 32157 SNPs pruned out, 16165 remaining
Scan region on chromosome 8 from [ rs7462951 ] to [ rs6599566 ]
For chromosome 8, 34013 SNPs pruned out, 15977 remaining
Scan region on chromosome 9 from [ rs2811026 ] to [ rs7863719 ]
For chromosome 9, 28576 SNPs pruned out, 14077 remaining
Scan region on chromosome 10 from [ rs11252546 ] to [ rs12218790 ]
For chromosome 10, 33222 SNPs pruned out, 16323 remaining
Scan region on chromosome 11 from [ rs3802985 ] to [ rs11224232 ]
For chromosome 11, 30980 SNPs pruned out, 14874 remaining
Scan region on chromosome 12 from [ rs1500095 ] to [ rs11147298 ]
For chromosome 12, 28732 SNPs pruned out, 15050 remaining
Scan region on chromosome 13 from [ rs1747380 ] to [ rs9525207 ]
For chromosome 13, 23412 SNPs pruned out, 11707 remaining
Scan region on chromosome 14 from [ rs4983173 ] to [ rs7144412 ]
For chromosome 14, 18955 SNPs pruned out, 9969 remaining
Scan region on chromosome 15 from [ rs12905389 ] to [ rs1533309 ]
For chromosome 15, 17124 SNPs pruned out, 9742 remaining
Scan region on chromosome 16 from [ rs41340949 ] to [ rs4785775 ]
For chromosome 16, 18387 SNPs pruned out, 10238 remaining
Scan region on chromosome 17 from [ rs6565733 ] to [ rs4986109 ]
For chromosome 17, 12761 SNPs pruned out, 8540 remaining
Scan region on chromosome 18 from [ rs7235612 ] to [ rs7236400 ]
For chromosome 18, 17591 SNPs pruned out, 9612 remaining
```

```
Scan region on chromosome 19 from [ rs8100066 ] to [ rs8110595 ]
For chromosome 19, 6998 SNPs pruned out, 5358 remaining
Scan region on chromosome 20 from [ rs4814683 ] to [ rs6062363 ]
For chromosome 20, 15037 SNPs pruned out, 8430 remaining
Scan region on chromosome 21 from [ rs169757 ] to [ rs7116 ]
For chromosome 21, 8182 SNPs pruned out, 4748 remaining
Scan region on chromosome 22 from [ rs4389403 ] to [ rs28729663 ]
For chromosome 22, 7129 SNPs pruned out, 4901 remaining
Scan region on chromosome 23 from [ rs6423165 ] to [ rs2981835 ]
For chromosome 23, 26820 SNPs pruned out, 11179 remaining
Scan region on chromosome 24 from [ rs2075640 ] to [ rs9724556 ]
For chromosome 24, 858 SNPs pruned out, 43 remaining

Analysis finished: Sun Jul 19 16:04:18 2015
```

7) FIND_AA_5plus_sex_imputed_independentSNPs

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@

Web-based version check ( --noweb to skip )
Recent cached web-check found... OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [ FIND_AA_5plus_sex_imputed_independentSNPs.log ]
Analysis started: Sun Jul 19 16:09:25 2015

Options in effect:
        --bfile FIND_AA_5plus_sex_imputed
        --extract FIND_AA_5plus_pruning.prune.in
        --make-bed
        --out FIND_AA_5plus_sex_imputed_independentSNPs

Reading map (extended format) from [ FIND_AA_5plus_sex_imputed.bim ]
932534 markers to be included from [ FIND_AA_5plus_sex_imputed.bim ]
Reading pedigree information from [ FIND_AA_5plus_sex_imputed.fam ]
853 individuals read from [ FIND_AA_5plus_sex_imputed.fam ]
853 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
575 cases, 278 controls and 0 missing
286 males, 567 females, and 0 of unspecified sex
Reading genotype bitfile from [ FIND_AA_5plus_sex_imputed.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Reading list of SNPs to extract [ FIND_AA_5plus_pruning.prune.in ] ... 312743 read
```

Before frequency and genotyping pruning, there are 312743 SNPs
853 founders and 0 non-founders found
22335 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [FIND_AA_5plus_sex_imputed_independentSNPs.hh]
1 SNPs with no founder genotypes observed
Warning, MAF set to 0 for these SNPs (see --nonfounders)
Writing list of these SNPs to [FIND_AA_5plus_sex_imputed_independentSNPs.nof]
Total genotyping rate in remaining individuals is 0.994577
0 SNPs failed missingness test (GENO > 1)
0 SNPs failed frequency test (MAF < 0)
After frequency and genotyping pruning, there are 312743 SNPs
After filtering, 575 cases, 278 controls and 0 missing
After filtering, 286 males, 567 females, and 0 of unspecified sex
Writing pedigree information to [FIND_AA_5plus_sex_imputed_independentSNPs.fam]
Writing map (extended format) information to [FIND_AA_5plus_sex_imputed_independentSNPs.bim]
Writing genotype bitfile to [FIND_AA_5plus_sex_imputed_independentSNPs.bed]
Using (default) SNP-major mode

Analysis finished: Sun Jul 19 16:10:35 2015

8) FIND_AA_5plus_sex_imputed_people_snp_removed

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@
```

Web-based version check (--noweb to skip)
Recent cached web-check found... OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [FIND_AA_5plus_sex_imputed_people_snp_removed.log]
Analysis started: Sun Jul 19 16:18:55 2015

Options in effect:
--bfile FIND_AA_5plus_sex_imputed
--remove people_missing0.03.txt
--exclude snplist_missing0.05.txt
--make-bed
--out FIND_AA_5plus_sex_imputed_people_snp_removed

Reading map (extended format) from [FIND_AA_5plus_sex_imputed.bim]
932534 markers to be included from [FIND_AA_5plus_sex_imputed.bim]
Reading pedigree information from [FIND_AA_5plus_sex_imputed.fam]
853 individuals read from [FIND_AA_5plus_sex_imputed.fam]
853 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)

Missing phenotype value is also -9
575 cases, 278 controls and 0 missing
286 males, 567 females, and 0 of unspecified sex
Reading genotype bitfile from [FIND_AA_5plus_sex_imputed.bed]
Detected that binary PED file is v1.00 SNP-major mode
Reading list of SNPs to exclude [snplist_missing0.05.txt] ... 35692 read
Reading individuals to remove [people_missing0.03.txt] ... 177 read
177 individuals removed with --remove option
Before frequency and genotyping pruning, there are 896842 SNPs
676 founders and 0 non-founders found
224 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [
FIND_AA_5plus_sex_imputed_people_snp_removed.hh]
Total genotyping rate in remaining individuals is 0.998683
0 SNPs failed missingness test (GENO > 1)
0 SNPs failed frequency test (MAF < 0)
After frequency and genotyping pruning, there are 896842 SNPs
After filtering, 451 cases, 225 controls and 0 missing
After filtering, 218 males, 458 females, and 0 of unspecified sex
Writing pedigree information to [FIND_AA_5plus_sex_imputed_people_snp_removed.fam]
Writing map (extended format) information to [FIND_AA_5plus_sex_imputed_people_snp_removed.bim
]
Writing genotype bitfile to [FIND_AA_5plus_sex_imputed_people_snp_removed.bed]
Using (default) SNP-major mode

Analysis finished: Sun Jul 19 16:20:39 2015

9) FIND_AA_5plus_sex_imputed_people_snp_t1d_removed

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@
```

Web-based version check (--noweb to skip)
Recent cached web-check found... OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [FIND_AA_5plus_sex_imputed_people_snp_t1d_removed.log]
Analysis started: Sun Jul 19 16:57:50 2015

Options in effect:
--bfile FIND_AA_5plus_sex_imputed_people_snp_removed
--remove AA_Age19andless_insulin.txt
--make-bed
--out FIND_AA_5plus_sex_imputed_people_snp_t1d_removed

Reading map (extended format) from [FIND_AA_5plus_sex_imputed_people_snp_removed.bim]
896842 markers to be included from [FIND_AA_5plus_sex_imputed_people_snp_removed.bim]
Reading pedigree information from [FIND_AA_5plus_sex_imputed_people_snp_removed.fam]
676 individuals read from [FIND_AA_5plus_sex_imputed_people_snp_removed.fam]
676 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
451 cases, 225 controls and 0 missing
218 males, 458 females, and 0 of unspecified sex
Reading genotype bitfile from [FIND_AA_5plus_sex_imputed_people_snp_removed.bed]
Detected that binary PED file is v1.00 SNP-major mode
Reading individuals to remove [AA_Age19andless_insulin.txt] ... 14 read
14 individuals removed with --remove option
Before frequency and genotyping pruning, there are 896842 SNPs
662 founders and 0 non-founders found
214 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_removed.hh]
Total genotyping rate in remaining individuals is 0.998681
0 SNPs failed missingness test (GENO > 1)
0 SNPs failed frequency test (MAF < 0)
After frequency and genotyping pruning, there are 896842 SNPs
After filtering, 441 cases, 221 controls and 0 missing
After filtering, 212 males, 450 females, and 0 of unspecified sex
Writing pedigree information to [FIND_AA_5plus_sex_imputed_people_snp_t1d_removed.fam]
Writing map (extended format) information to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_removed.bim]
Writing genotype bitfile to [FIND_AA_5plus_sex_imputed_people_snp_t1d_removed.bed]
Using (default) SNP-major mode

Analysis finished: Sun Jul 19 16:59:22 2015

10) FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@
```

Web-based version check (--noweb to skip)
Recent cached web-check found...Problem connecting to web

Writing this text to log file [FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.log]
Analysis started: Sat Jul 25 14:00:37 2015

Options in effect:
--bfile FIND_AA_5plus_sex_imputed_people_snp_t1d_removed
--remove PI_HAT0.1875.txt

--make-bed
--out FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed

Reading map (extended format) from [FIND_AA_5plus_sex_imputed_people_snp_t1d_removed.bim]
896842 markers to be included from [FIND_AA_5plus_sex_imputed_people_snp_t1d_removed.bim]
Reading pedigree information from [FIND_AA_5plus_sex_imputed_people_snp_t1d_removed.fam]
662 individuals read from [FIND_AA_5plus_sex_imputed_people_snp_t1d_removed.fam]
662 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
441 cases, 221 controls and 0 missing
212 males, 450 females, and 0 of unspecified sex
Reading genotype bitfile from [FIND_AA_5plus_sex_imputed_people_snp_t1d_removed.bed]
Detected that binary PED file is v1.00 SNP-major mode
Reading individuals to remove [PI_HAT0.1875.txt] ... 29 read
26 individuals removed with --remove option
Before frequency and genotyping pruning, there are 896842 SNPs
636 founders and 0 non-founders found
203 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.hh]
Total genotyping rate in remaining individuals is 0.998677
0 SNPs failed missingness test (GENO > 1)
0 SNPs failed frequency test (MAF < 0)
After frequency and genotyping pruning, there are 896842 SNPs
After filtering, 426 cases, 210 controls and 0 missing
After filtering, 199 males, 437 females, and 0 of unspecified sex
Writing pedigree information to [FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.fam
]
Writing map (extended format) information to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.bim]
Writing genotype bitfile to [FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.bed]
Using (default) SNP-major mode

Analysis finished: Sat Jul 25 14:02:07 2015

11) hwe.log

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@
```

Web-based version check (--noweb to skip)
Recent cached web-check found...Problem connecting to web

Writing this text to log file [hwe.log]
Analysis started: Sat Jul 25 14:27:05 2015

Options in effect:
	--bfile FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed
	--hardy
	--nonfounders
	--out hwe

Reading map (extended format) from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.bim]
896842 markers to be included from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.bim]
Reading pedigree information from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.fam]
636 individuals read from [FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.fam]
636 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
426 cases, 210 controls and 0 missing
199 males, 437 females, and 0 of unspecified sex
Reading genotype bitfile from [FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.bed]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 896842 SNPs
636 founders and 0 non-founders found
203 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [hwe.hh]
Writing Hardy-Weinberg tests (all individuals) to [hwe.hwe]
1690 markers to be excluded based on HWE test (p <= 0.001)
	3152 markers failed HWE test in cases
	1690 markers failed HWE test in controls
Total genotyping rate in remaining individuals is 0.998677
0 SNPs failed missingness test (GENO > 1)
0 SNPs failed frequency test (MAF < 0)
After frequency and genotyping pruning, there are 895152 SNPs
After filtering, 426 cases, 210 controls and 0 missing
After filtering, 199 males, 437 females, and 0 of unspecified sex

Analysis finished: Sat Jul 25 14:30:02 2015

12) freq.log

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|       http://pngu.mgh.harvard.edu/purcell/plink/         |
@----------------------------------------------------------@
```

Web-based version check (--noweb to skip)
Recent cached web-check found...Problem connecting to web

Writing this text to log file [freq.log]
Analysis started: Sat Jul 25 14:36:11 2015

Options in effect:
--bfile FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed
--freq
--out freq

Reading map (extended format) from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.bim]
896842 markers to be included from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.bim]
Reading pedigree information from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.fam]
636 individuals read from [FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.fam]
636 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
426 cases, 210 controls and 0 missing
199 males, 437 females, and 0 of unspecified sex
Reading genotype bitfile from [FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.bed]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 896842 SNPs
636 founders and 0 non-founders found
203 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [freq.hh]
Writing allele frequencies (founders-only) to [freq.frq]

Analysis finished: Sat Jul 25 14:37:20 2015

13) FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_removed.log

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@
```

Web-based version check (--noweb to skip)
Recent cached web-check found...Problem connecting to web

Writing this text to log file [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_removed.log]
Analysis started: Sat Jul 25 15:31:51 2015

Options in effect:
--bfile FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed
--exclude freq0.05.txt
--make-bed

--out FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_removed

** For gPLINK compatibility, do not use '.' in --out **
Reading map (extended format) from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.bim]
896842 markers to be included from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.bim]
Reading pedigree information from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.fam]
636 individuals read from [FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.fam]
636 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
426 cases, 210 controls and 0 missing
199 males, 437 females, and 0 of unspecified sex
Reading genotype bitfile from [FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_removed.bed]
Detected that binary PED file is v1.00 SNP-major mode
Reading list of SNPs to exclude [freq0.05.txt] ... 111642 read
Before frequency and genotyping pruning, there are 785200 SNPs
636 founders and 0 non-founders found
54 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_removed.hh]
Total genotyping rate in remaining individuals is 0.998696
0 SNPs failed missingness test (GENO > 1)
0 SNPs failed frequency test (MAF < 0)
After frequency and genotyping pruning, there are 785200 SNPs
After filtering, 426 cases, 210 controls and 0 missing
After filtering, 199 males, 437 females, and 0 of unspecified sex
Writing pedigree information to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_removed.fam]
Writing map (extended format) information to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_removed.bim]
Writing genotype bitfile to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_removed.bed]
Using (default) SNP-major mode

Analysis finished: Sat Jul 25 15:33:13 2015

14) FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.log

@--@
PLINK!	v1.07	10/Aug/2009
(C) 2009 Shaun Purcell, GNU General Public License, v2
For documentation, citation & bug-report instructions:
http://pngu.mgh.harvard.edu/purcell/plink/
@--@

Web-based version check (--noweb to skip)
Recent cached web-check found...Problem connecting to web

Writing this text to log file [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.log]
Analysis started: Sat Jul 25 15:36:33 2015

Options in effect:
--bfile FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_removed
--exclude hwe0.0001.txt
--make-bed
--out FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed

** For gPLINK compatibility, do not use '.' in --out **
Reading map (extended format) from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_removed.bim]
785200 markers to be included from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_removed.bim]
Reading pedigree information from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_removed.fam]
636 individuals read from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_removed.fam]
636 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
426 cases, 210 controls and 0 missing
199 males, 437 females, and 0 of unspecified sex
Reading genotype bitfile from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_removed.bed]
Detected that binary PED file is v1.00 SNP-major mode
Reading list of SNPs to exclude [hwe0.0001.txt] ... 3988 read
Before frequency and genotyping pruning, there are 782810 SNPs
636 founders and 0 non-founders found
54 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.hh]
Total genotyping rate in remaining individuals is 0.998719
0 SNPs failed missingness test (GENO > 1)
0 SNPs failed frequency test (MAF < 0)
After frequency and genotyping pruning, there are 782810 SNPs
After filtering, 426 cases, 210 controls and 0 missing
After filtering, 199 males, 437 females, and 0 of unspecified sex
Writing pedigree information to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.fam]
Writing map (extended format) information to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bim]
Writing genotype bitfile to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bed]
Using (default) SNP-major mode

Analysis finished: Sat Jul 25 15:37:52 2015

15) FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed_male.log

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@

Web-based version check ( --noweb to skip )
Recent cached web-check found...Problem connecting to web

Writing this text to log file [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed_male.log ]
Analysis started: Sat Jul 25 15:47:43 2015

Options in effect:
        --bfile FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_hwe0.0001_removed
        --remove FemaleAA.txt
        --make-bed
        --out
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed_male

** For gPLINK compatibility, do not use '.' in --out **
Reading map (extended format) from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_hwe0.0001_removed.bim ]
782810 markers to be included from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_hwe0.0001_removed.bim ]
Reading pedigree information from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_hwe0.0001_removed.fam ]
636 individuals read from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_hwe0.0001_removed.fam ]
636 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
426 cases, 210 controls and 0 missing
199 males, 437 females, and 0 of unspecified sex
Reading genotype bitfile from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_hwe0.0001_removed.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Reading individuals to remove [ FemaleAA.txt ] ... 437 read
437 individuals removed with --remove option
Before frequency and genotyping pruning, there are 782810 SNPs
199 founders and 0 non-founders found
54 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed_male.hh ]
Total genotyping rate in remaining individuals is 0.99889
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 782810 SNPs
After filtering, 155 cases, 44 controls and 0 missing
After filtering, 199 males, 0 females, and 0 of unspecified sex
```

Writing pedigree information to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed_male.fam]
Writing map (extended format) information to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed_male.bim]
Writing genotype bitfile to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed_male.bed]
Using (default) SNP-major mode

Analysis finished: Sat Jul 25 15:48:41 2015

16)
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed_female.log

@--@
PLINK!	v1.07	10/Aug/2009
(C) 2009 Shaun Purcell, GNU General Public License, v2
For documentation, citation & bug-report instructions:
http://pngu.mgh.harvard.edu/purcell/plink/
@--@

Web-based version check (--noweb to skip)
Recent cached web-check found...Problem connecting to web

Writing this text to log file [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed_female.log]
Analysis started: Sat Jul 25 15:49:13 2015

Options in effect:
	--bfile FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_hwe0.0001_removed
	--remove maleAA.txt
	--make-bed
	--out
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed_female

** For gPLINK compatibility, do not use '.' in --out **
Reading map (extended format) from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_hwe0.0001_removed.bim]
782810 markers to be included from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_hwe0.0001_removed.bim]
Reading pedigree information from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_hwe0.0001_removed.fam]
636 individuals read from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_hwe0.0001_removed.fam]
636 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
426 cases, 210 controls and 0 missing
199 males, 437 females, and 0 of unspecified sex
Reading genotype bitfile from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_hwe0.0001_removed.bed]

Detected that binary PED file is v1.00 SNP-major mode
Reading individuals to remove [maleAA.txt] ... 199 read
199 individuals removed with --remove option
Before frequency and genotyping pruning, there are 782810 SNPs
437 founders and 0 non-founders found
Total genotyping rate in remaining individuals is nan
0 SNPs failed missingness test (GENO > 1)
0 SNPs failed frequency test (MAF < 0)
After frequency and genotyping pruning, there are 782810 SNPs
After filtering, 271 cases, 166 controls and 0 missing
After filtering, 0 males, 437 females, and 0 of unspecified sex
Writing pedigree information to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed_female.fam]
Writing map (extended format) information to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed_female.bim]
Writing genotype bitfile to [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed_female.bed]
Using (default) SNP-major mode

Analysis finished: Sat Jul 25 15:50:27 2015

17) AA_quantitative_all.log

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@
```

Web-based version check (--noweb to skip)
Connecting to web... OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [AA_q_all.log]
Analysis started: Sat Aug 08 15:01:49 2015

Options in effect:
	--bfile
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed
	--pheno GFR_Phen.txt
	--make-bed
	--out AA_q_all

Reading map (extended format) from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bim]
782810 markers to be included from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bim]

Reading pedigree information from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.fam]
636 individuals read from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.fam]
636 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
426 cases, 210 controls and 0 missing
199 males, 437 females, and 0 of unspecified sex
Reading genotype bitfile from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bed]
Detected that binary PED file is v1.00 SNP-major mode
Reading alternate phenotype from [GFR_Phen.txt]
633 individuals with non-missing alternate phenotype
Assuming a quantitative trait
Missing phenotype value is -9
Before frequency and genotyping pruning, there are 782810 SNPs
636 founders and 0 non-founders found
54 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [AA_q_all.hh]
Total genotyping rate in remaining individuals is 0.998719
0 SNPs failed missingness test (GENO > 1)
0 SNPs failed frequency test (MAF < 0)
After frequency and genotyping pruning, there are 782810 SNPs
After filtering, 633 individuals with non-missing status
After filtering, 199 males, 437 females, and 0 of unspecified sex
Writing pedigree information to [AA_q_all.fam]
Writing map (extended format) information to [AA_q_all.bim]
Writing genotype bitfile to [AA_q_all.bed]
Using (default) SNP-major mode

Analysis finished: Sat Aug 08 15:03:03 2015

18) AA_ibs_all.log

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|       http://pngu.mgh.harvard.edu/purcell/plink/         |
@----------------------------------------------------------@
```

Web-based version check (--noweb to skip)
Recent cached web-check found... OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [plink.log]
Analysis started: Sat Aug 08 16:11:51 2015

Options in effect:
--bfile
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed
--read-genome FIND_AA_5plus_ibd.genome
--ibs-test

Reading map (extended format) from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bim]
782810 markers to be included from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bim]
Reading pedigree information from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.fam]
636 individuals read from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.fam]
636 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
426 cases, 210 controls and 0 missing
199 males, 437 females, and 0 of unspecified sex
Reading genotype bitfile from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bed]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 782810 SNPs
636 founders and 0 non-founders found
54 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [plink.hh]
Total genotyping rate in remaining individuals is 0.998719
0 SNPs failed missingness test (GENO > 1)
0 SNPs failed frequency test (MAF < 0)
After frequency and genotyping pruning, there are 782810 SNPs
After filtering, 426 cases, 210 controls and 0 missing
After filtering, 199 males, 437 females, and 0 of unspecified sex
Converting data to Individual-major format
Clustering individuals based on genome-wide IBS
Merge distance p-value constraint = 0
Reading genome-wide IBS estimates from [FIND_AA_5plus_ibd.genome]
Set to permute within 1 cluster(s)

Between-group IBS (mean, SD) = 0.817981, 0.00336465
In-group (2) IBS (mean, SD) = 0.81839, 0.00293358
In-group (1) IBS (mean, SD) = 0.817598, 0.00367504
Approximate proportion of variance between group = 0.00155461
IBS group-difference empirical p-values:

T1: Case/control less similar p = 0.0137299
T2: Case/control more similar p = 0.98628

T3: Case/case less similar than control/control p = 0.98632
T4: Case/case more similar than control/control p = 0.0136899

T5: Case/case less similar p = 0.98635
T6: Case/case more similar p = 0.0136599

T7: Control/control less similar p = 0.0136299
T8: Control/control more similar p = 0.98638

T9: Case/case less similar than case/control p = 0.9863
T10: Case/case more similar than case/control p = 0.0137099

T11: Control/control less similar than case/control p = 0.9816
T12: Control/control more similar than case/control p = 0.0184098

Analysis finished: Sat Aug 08 16:23:29 2015

19) AA_ibd_all.log

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@

Web-based version check ( --noweb to skip )
Recent cached web-check found... OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [ plink.log ]
Analysis started: Sat Aug 08 22:52:34 2015

Options in effect:
          --bfile
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed
          --read-genome FIND_AA_5plus_ibd.genome
          --cluster
          --neighbour 1 5

Reading map (extended format) from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bim ]
782810 markers to be included from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bim ]
Reading pedigree information from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.fam ]
636 individuals read from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.fam ]
636 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
426 cases, 210 controls and 0 missing
199 males, 437 females, and 0 of unspecified sex
Reading genotype bitfile from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bed ]
```

Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 782810 SNPs
636 founders and 0 non-founders found
54 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [plink.hh]
Total genotyping rate in remaining individuals is 0.998719
0 SNPs failed missingness test (GENO > 1)
0 SNPs failed frequency test (MAF < 0)
After frequency and genotyping pruning, there are 782810 SNPs
After filtering, 426 cases, 210 controls and 0 missing
After filtering, 199 males, 437 females, and 0 of unspecified sex
Converting data to Individual-major format
Clustering individuals based on genome-wide IBS
Merge distance p-value constraint = 0
Outlier detection based on neighbours 1 to 5
Reading genome-wide IBS estimates from [FIND_AA_5plus_ibd.genome]
Of these, 201930 are pairable based on constraints
Writing individual neighbour/outlier statatistics to [plink.nearest]
Writing cluster progress to [plink.cluster0]
Writing cluster solution (1) [plink.cluster1]
Writing cluster solution (2) [plink.cluster2]
Writing cluster solution (3) [plink.cluster3]

Analysis finished: Sat Aug 08 22:54:16 2015

20) AA_mds_plot_all.log

```
@----------------------------------------------------------@
|       PLINK!      |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@
```

Web-based version check (--noweb to skip)
Recent cached web-check found... OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [plink.log]
Analysis started: Sun Aug 09 21:37:32 2015

Options in effect:
--bfile
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed
--read-genome FIND_AA_5plus_ibd.genome
--cluster
--mds-plot 4

Reading map (extended format) from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bim]
782810 markers to be included from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bim]
Reading pedigree information from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.fam]
636 individuals read from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.fam]
636 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
426 cases, 210 controls and 0 missing
199 males, 437 females, and 0 of unspecified sex
Reading genotype bitfile from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bed]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 782810 SNPs
636 founders and 0 non-founders found
54 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [plink.hh]
Total genotyping rate in remaining individuals is 0.998719
0 SNPs failed missingness test (GENO > 1)
0 SNPs failed frequency test (MAF < 0)
After frequency and genotyping pruning, there are 782810 SNPs
After filtering, 426 cases, 210 controls and 0 missing
After filtering, 199 males, 437 females, and 0 of unspecified sex
Converting data to Individual-major format
Clustering individuals based on genome-wide IBS
Merge distance p-value constraint = 0
Reading genome-wide IBS estimates from [FIND_AA_5plus_ibd.genome]
Of these, 201930 are pairable based on constraints
Writing cluster progress to [plink.cluster0]
Writing cluster solution (1) [plink.cluster1]
Writing cluster solution (2) [plink.cluster2]
Writing cluster solution (3) [plink.cluster3]
Writing MDS solution to [plink.mds]
MDS plot of individuals (not clusters)

Analysis finished: Sun Aug 09 21:39:37 2015

21) All_AA_Z4_removed.log

```
@----------------------------------------------------------@
|       PLINK!      |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@
```

Web-based version check (--noweb to skip)

Connecting to web... OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [All_AA_Z4_removed.log]
Analysis started: Mon Oct 19 21:15:39 2015

Options in effect:
--bfile
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed
--remove Z_4_ol.txt
--make-bed
--out All_AA_Z4_removed

Reading map (extended format) from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bim]
782810 markers to be included from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bim]
Reading pedigree information from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.fam]
636 individuals read from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.fam]
636 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
426 cases, 210 controls and 0 missing
199 males, 437 females, and 0 of unspecified sex
Reading genotype bitfile from [
FIND_AA_5plus_sex_imputed_people_snp_t1d_PI_HAT_MAF0.05_HWE0.0001_removed.bed]
Detected that binary PED file is v1.00 SNP-major mode
Reading individuals to remove [Z_4_ol.txt] ... 1 read
1 individuals removed with --remove option
Before frequency and genotyping pruning, there are 782810 SNPs
635 founders and 0 non-founders found
54 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [All_AA_Z4_removed.hh]
Total genotyping rate in remaining individuals is 0.998718
0 SNPs failed missingness test (GENO > 1)
0 SNPs failed frequency test (MAF < 0)
After frequency and genotyping pruning, there are 782810 SNPs
After filtering, 426 cases, 209 controls and 0 missing
After filtering, 199 males, 436 females, and 0 of unspecified sex
Writing pedigree information to [All_AA_Z4_removed.fam]
Writing map (extended format) information to [All_AA_Z4_removed.bim]
Writing genotype bitfile to [All_AA_Z4_removed.bed]
Using (default) SNP-major mode

Analysis finished: Mon Oct 19 21:16:58 2015

22) All_AA_Z4_assoc.log

@----------@

```
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|       http://pngu.mgh.harvard.edu/purcell/plink/         |
@----------------------------------------------------------@

Web-based version check ( --noweb to skip )
Connecting to web...  failed connection

Problem connecting to web

Writing this text to log file [ All_AA_Z4_assoc.log ]
Analysis started: Tue Oct 20 14:12:01 2015

Options in effect:
        --bfile All_AA_Z4_removed
        --assoc
        --ci 0.95
        --out All_AA_Z4_assoc

Reading map (extended format) from [ All_AA_Z4_removed.bim ]
782810 markers to be included from [ All_AA_Z4_removed.bim ]
Reading pedigree information from [ All_AA_Z4_removed.fam ]
635 individuals read from [ All_AA_Z4_removed.fam ]
635 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
426 cases, 209 controls and 0 missing
199 males, 436 females, and 0 of unspecified sex
Reading genotype bitfile from [ All_AA_Z4_removed.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 782810 SNPs
635 founders and 0 non-founders found
54 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ All_AA_Z4_assoc.hh ]
Total genotyping rate in remaining individuals is 0.998718
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 782810 SNPs
After filtering, 426 cases, 209 controls and 0 missing
After filtering, 199 males, 436 females, and 0 of unspecified sex
Writing main association results to [ All_AA_Z4_assoc.assoc ]

Analysis finished: Tue Oct 20 14:13:18 2015
```

23) All_AA_Z4_assoc_adjust.log

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
```

```
| (C) 2009 Shaun Purcell, GNU General Public License, v2 |
|----------------------------------------------------------|
| For documentation, citation & bug-report instructions: |
|       http://pngu.mgh.harvard.edu/purcell/plink/       |
@----------------------------------------------------------@

Web-based version check ( --noweb to skip )
Recent cached web-check found...Problem connecting to web

Writing this text to log file [ All_AA_Z4_assoc_adjust.log ]
Analysis started: Tue Oct 20 14:14:45 2015

Options in effect:
        --bfile All_AA_Z4_removed
        --assoc
        --ci 0.95
        --adjust
        --out All_AA_Z4_assoc_adjust

Reading map (extended format) from [ All_AA_Z4_removed.bim ]
782810 markers to be included from [ All_AA_Z4_removed.bim ]
Reading pedigree information from [ All_AA_Z4_removed.fam ]
635 individuals read from [ All_AA_Z4_removed.fam ]
635 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
426 cases, 209 controls and 0 missing
199 males, 436 females, and 0 of unspecified sex
Reading genotype bitfile from [ All_AA_Z4_removed.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 782810 SNPs
635 founders and 0 non-founders found
54 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ All_AA_Z4_assoc_adjust.hh ]
Total genotyping rate in remaining individuals is 0.998718
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 782810 SNPs
After filtering, 426 cases, 209 controls and 0 missing
After filtering, 199 males, 436 females, and 0 of unspecified sex
Writing main association results to [ All_AA_Z4_assoc_adjust.assoc ]
Computing corrected significance values (FDR, Sidak, etc)
Genomic inflation factor (based on median chi-squared) is 1.03567
Mean chi-squared statistic is 1.03908
Correcting for 782810 tests
Writing multiple-test corrected significance values to [ All_AA_Z4_assoc_adjust.assoc.adjusted ]

Analysis finished: Tue Oct 20 14:16:19 2015
```

24) Quantitative_All_AA_Z4_qassoc.log

```
@----------------------------------------------------------@
```

```
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@

Web-based version check ( --noweb to skip )
Recent cached web-check found... OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [ Q_All_AA_Z4_qassoc.log ]
Analysis started: Fri Oct 23 19:13:43 2015

Options in effect:
        --bfile All_AA_Z4_removed
        --pheno GFR_Phen.txt
        --assoc
        --ci 0.95
        --qt-means
        --out Q_All_AA_Z4_qassoc

Reading map (extended format) from [ All_AA_Z4_removed.bim ]
782810 markers to be included from [ All_AA_Z4_removed.bim ]
Reading pedigree information from [ All_AA_Z4_removed.fam ]
635 individuals read from [ All_AA_Z4_removed.fam ]
635 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
426 cases, 209 controls and 0 missing
199 males, 436 females, and 0 of unspecified sex
Reading genotype bitfile from [ All_AA_Z4_removed.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Reading alternate phenotype from [ GFR_Phen.txt ]
632 individuals with non-missing alternate phenotype
Assuming a quantitative trait
Missing phenotype value is -9
Before frequency and genotyping pruning, there are 782810 SNPs
635 founders and 0 non-founders found
54 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ Q_All_AA_Z4_qassoc.hh ]
Total genotyping rate in remaining individuals is 0.998718
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 782810 SNPs
After filtering, 632 individuals with non-missing status
After filtering, 199 males, 436 females, and 0 of unspecified sex
Writing QT association results to [ Q_All_AA_Z4_qassoc.qassoc ]
Writing QT genotypic means to [ Q_All_AA_Z4_qassoc.qassoc.means ]

Analysis finished: Fri Oct 23 19:15:49 2015
```

25) Quantitative_All_AA_Z4_qassoc_adjust.log

```
@----------------------------------------------------------@
|       PLINK!       |     v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@

Web-based version check ( --noweb to skip )
Recent cached web-check found... OK, v1.07 is current

+++ PLINK 1.9 is now available! See above website for details +++

Writing this text to log file [ Q_All_AA_Z4_qassoc_adjust.log ]
Analysis started: Fri Oct 23 19:18:21 2015

Options in effect:
        --bfile All_AA_Z4_removed
        --pheno GFR_Phen.txt
        --assoc
        --ci 0.95
        --adjust
        --qt-means
        --out Q_All_AA_Z4_qassoc_adjust

Reading map (extended format) from [ All_AA_Z4_removed.bim ]
782810 markers to be included from [ All_AA_Z4_removed.bim ]
Reading pedigree information from [ All_AA_Z4_removed.fam ]
635 individuals read from [ All_AA_Z4_removed.fam ]
635 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
426 cases, 209 controls and 0 missing
199 males, 436 females, and 0 of unspecified sex
Reading genotype bitfile from [ All_AA_Z4_removed.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Reading alternate phenotype from [ GFR_Phen.txt ]
632 individuals with non-missing alternate phenotype
Assuming a quantitative trait
Missing phenotype value is -9
Before frequency and genotyping pruning, there are 782810 SNPs
635 founders and 0 non-founders found
54 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ Q_All_AA_Z4_qassoc_adjust.hh ]
Total genotyping rate in remaining individuals is 0.998718
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 782810 SNPs
```

After filtering, 632 individuals with non-missing status
After filtering, 199 males, 436 females, and 0 of unspecified sex
Writing QT association results to [Q_All_AA_Z4_qassoc_adjust.qassoc]
Writing QT genotypic means to [Q_All_AA_Z4_qassoc_adjust.qassoc.means]
Computing corrected significance values (FDR, Sidak, etc)
Genomic inflation factor (based on median chi-squared) is 1.00698
Mean chi-squared statistic is 1.01439
Correcting for 782810 tests
Writing multiple-test corrected significance values to [Q_All_AA_Z4_qassoc_adjust.qassoc.adjusted]

Analysis finished: Fri Oct 23 19:21:04 2015