UNIVERSITY OF CALIFORNIA,
IRVINE

Investigation of genetic & environmental factors for glycemic status: a family-based study

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Epidemiology

by

Sepideh Naggash Ferdos

Dissertation Committee:
Professor Karen Edwards, Chair
Professor Baolin Wu, Chair
Professor Zhaoxia Yu

2026

DEDICATION

To

My parents, Ali Ferdos and Mahnaz Chavoshian, in recognition of their love, dedication, and unwavering support.

My brother, Iden Naggash Ferdos. Your support, love, and dedication relentlessly inspire me to believe in myself and reach beyond what I thought was possible. All the world is driftwood emerged from your sea.

ای زندگی و تن و توانم همه تو
جانی و دلی، ای دل و جانم همه تو
تو هستی من شدی، از آنی همه من
من نیست شدم در تو، از آنم همه تو

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would first like to express my deepest gratitude to my advisor, Dr. Karen Edwards, for her guidance, mentorship, and support throughout my doctoral training. Her expertise in genetic epidemiology and commitment to rigorous research have shaped my development as a scientist and have been instrumental in the completion of this dissertation and my training. I am especially grateful for her thoughtful feedback, patience, and encouragement throughout each stage of this work.

I would also like to thank my dissertation committee members, Dr. Wu and Dr. Yu, for their time, insight, and constructive feedback. Their perspectives strengthened the scientific direction of this dissertation and challenged me to think more critically about my research questions, methods, and interpretation of findings. I would like to recognize and thank Jia Wan, who played an important role in helping me develop and address the research questions central to this dissertation.

I am grateful to the faculty, staff, and students in the Department of Epidemiology and Biostatistics at the University of California, Irvine, for creating an environment that supported my academic and professional growth. I would especially like to highlight the mentorship of Dr. Thomas Taylor during my doctoral training and his dedication to teaching the next generation of graduates.

I would also like to thank Dr. Rany Salem, my mentor and principal investigator during my time as a master's student at University of California San Diego. Thank you for training me from the beginning and teaching me how to think like a genetic epidemiologist. Your guidance, support, and advice have stayed with me throughout my doctoral career.

To my colleagues, classmates, and friends, thank you for the encouragement, conversations, support, and the willingness to bounce around ideas at any moment. I am especially grateful to those who reminded me that progress is built through persistence, uncertainty, and effort.

Most importantly, I would like to thank my family for their unconditional love, patience, and belief in me. Their sacrifices, encouragement, and support have carried me through every challenge and milestone of this process, reminding me that it is okay not to have it all figured out. I am deeply grateful for the foundation they have given me and for their reminder of our generational academic and teaching roots from Iran.

Finally, I would like to acknowledge the personal growth that accompanied this academic journey. Completing this dissertation required resilience, discipline, and trust in myself, and I am grateful for the lessons learned along the way. This work represents not only the culmination of my doctoral training, but also the support, mentorship, and love of the many people who helped me reach this point.

VITA

Sepideh Naggash Ferdos

2017	B.A. in Health Studies, University of Washington
2018-2020	M.P.H. in Epidemiology, University of California, San Diego
2021-2025	Ph.D. in Epidemiology, University of California, Irvine

FIELD OF STUDY

Genetic Epidemiology

PUBLICATIONS

Eshraghian EA, Ferdos SN, Mehta SR. The Impact of Human Mobility on Regional and Global Efforts to Control HIV Transmission. *Viruses*. 2020; 12(1):67. https://doi.org/10.3390/v12010067

Ferdos S. A Genome Wide Association Study of Gene by Smoking Interaction on Type II Diabetes [dissertation]. University of California San Diego; 2020.

ABSTRACT OF THE DISSERTATION

Investigation of genetic & environmental factors for Glycemic Status:
a family-based study

by
Sepideh Naggash Ferdos
Doctor of Philosophy in Epidemiology
University of California, Irvine, 2026
Professor Karen Edwards, Chair
Professor Baolin Wu, Chair

Prediabetes is a chronic metabolic condition characterized by impaired glucose regulation and represents an early stage in the progression toward type 2 diabetes. Both genetic susceptibility and lifestyle factors – including diet, physical activity, and smoking – contribute to abnormal glycemia. Although type 2 diabetes has been extensively studied, less is known about the genetic and environmental determinants of prediabetes. Importantly, prediabetes presents a critical window for intervention, as maintenance or reversion to normal glycemia remains possible. This study investigates whether prediabetes has distinct genetic determinants compared to type 2 diabetes and whether smoking modifies these associations, providing opportunities to identify individuals at risk earlier in disease development. Multinomial (polytomous) regression models evaluated genetic associations for 22 candidate SNPs with prediabetes or type 2 diabetes, as well as SNP-smoking interactions, using the multi-ethnic GENNID (Genetics of Non-Insulin Dependent Diabetes) family-based dataset. Overall, this study identified evidence of nominally significant genetic associations with both prediabetes and type 2 diabetes, but these associations varied across ancestry groups and did not support the hypothesis that prediabetes has distinct genetic determinants compared to type 2 diabetes. SNP-smoking interaction analyses provided evidence of effect modification for select SNP-glycemic status associations of nominal

significance. These findings suggest that genetic contributions to dysglycemia in GENNID may be heterogeneous across ancestry and glycemic outcome, underscoring the importance of studying prediabetes directly.

INTRODUCTION

Prediabetes

Prediabetes is a metabolic condition characterized by elevated blood glucose levels that are above the normal range but below the diagnostic threshold for type 2 diabetes.[1] In the United States, an estimated 98 million adults – about 38% of the population – meet the criteria for prediabetes, yet roughly 80% are unaware of their condition. [1] This lack of awareness delays interventions, often resulting in disease progression and poorer long-term health outcomes.[2]

Prediabetes is typically viewed as an early stage in the progression to type 2 diabetes, representing a gradual transition from normoglycemia to hyperglycemia.[3,4] An estimated 70% of individuals with prediabetes will progress to type 2 diabetes; however, many remain in a prediabetic state long-term, and a minority revert to normoglycemia.[5] Individuals with prediabetes are largely understudied, with both clinical and population-level statistics notably lacking. Understanding prediabetes is critical, since it not only serves as a major risk factor for progression to type 2 diabetes but also presents a key opportunity for early intervention and long-term maintenance of metabolic health. This intervention potential has been demonstrated in landmark randomized controlled trials of individuals with impaired glucose regulation. Randomized controlled trials (RCT's), including the Diabetes Prevention Program (DPP) in the United States and the Finnish Diabetes Prevention Study (DPS) in Finland, have demonstrated that lifestyle interventions can substantially reduce progression from prediabetes to type 2 diabetes.[6,7] The DPP and DPS trials both recruited individuals with prediabetes and used the American Diabetes Association (ADA) criteria for establishing prediabetes.[6,7] The DPP trial, sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK),

demonstrated that individuals with prediabetes can prevent or delay type 2 diabetes development by 58% with lifestyle interventions over the three years the trial took place in comparison to the their control groups which did not show a statistically significant risk reduction.[6] The DPS trial found the same level of risk reduction as the DPP trial.[6,7] In the DPP study, regression to normal glucose regulation occurred in both intervention and placebo groups, but was significantly more frequent among those receiving intensive lifestyle intervention. At 4 years, 29.6% of lifestyle intervention participants had both normal fasting and 2-hour glucose values compared with 17.2 of placebo participants.[6] More recent long-term follow-up from the DPP and DPP Outcomes Study further supports the clinical importance of intervening during the prediabetic stage. In a 21-year follow-up of adults with prediabetes, lifestyle intervention was associated with a lower long-term burden of multimorbidity compared with placebo, whereas metformin was not associated with a statistically significant reduction.[6,8] These consistent results highlight the importance of identifying individuals with prediabetes and intervening at this stage to prevent future cardiometabolic disease and reduce the risk of type 2 diabetic comorbidities and mortality. Addressing and understanding prediabetes will aid in risk awareness and prevention before the onset of progression. Despite strong evidence supporting early intervention, important gaps remain in understanding the biological and epidemiological factors of prediabetes. These gaps are also complicated by increasing recognition that prediabetes is a heterogenous condition, with impaired fasting glucose-, impaired glucose tolerance-, and glycosylated hemoglobin-based definitions capturing overlapping but not identical aspects of dysglyemcia.[9] Prediabetes has historically been underrepresented in scientific literature. Most studies focus on type 2 diabetes, and many earlier studies excluded individuals with prediabetes, or occasionally grouped them with those who have diabetes. Only recently have more studies begun to examine prediabetes

separately from type 2 diabetes. Nevertheless, research still centers largely on type 2 diabetes, limiting efforts to characterize the prediabetic state and its role in preventing disease progression.

Type 2 Diabetes

Type 2 Diabetes, or adult-onset diabetes, is a complex chronic condition characterized by abnormal blood glucose levels due to insufficient insulin production.[10-12] The International Diabetes Federation has reported that 590 million individuals, or 11.1%, of the adult population are living with diabetes as of 2025, which is projected to increase to 853 million by 2050.[13] Type 2 Diabetes affects approximately 38.4 million people in the U.S. (11.6% of the U.S. population), and of those individuals, 8.7 million are undiagnosed.[14] Type 2 Diabetes diagnoses are occurring at earlier ages than previously expected.[15,16] The mean age of diagnosis decreased from 52 to 46 years old when comparing two NHANES survey periods of 1988-1994 and 1999-2000, where individuals diagnosed at younger age ranges were found to have higher levels of all-cause mortality than those diagnosed as late-onset diabetes, according to a study by Zhang et al.[15,16] The lifetime treatment and management of type 2 diabetes, along with its rising prevalence poses a threat to the healthcare expenditure and health system burden.[17] The medical costs for individuals living with type 2 diabetes is 2.6 times higher than those without diabetes.[17] The economic cost of type 2 diabetes in the U.S. in 2022 was 412.9 billion USD.[17] Approximately half of this expenditure is attributed to disease progression involving diabetic complications.[17] Detection and prevention of type 2 diabetes is critical for the reduction of disease burden, morbidity, and mortality. Racial and ethnic disparities are especially evident in the severity and progression of the disease.[18] A study published in 2013 found that non-Hispanic Blacks and Hispanic Americans are 2.3 and 1.5 times more likely to die from diabetes or diabetic-related

complications when compared to non-Hispanic Whites, respectively.[19] These racial and ethnic disparities are thought to reflect a complex interplay of genetic, environmental, behavioral, socioeconomic, accessibility, and broader structural determinants of health, which may influence diabetes risk, management, and mortality.[18-21] Nevertheless, these factors remain insufficiently characterized across diverse racial/ethnic populations.[18-21]

Progression of Glycemic Dysregulation

Prediabetes is defined as a state of abnormal glucose homeostasis identified through laboratory measures, most commonly impaired glucose tolerance (IGT) and impaired fasting glucose (IFG).[22-25] Although both represent prediabetic states, they reflect somewhat distinct metabolic abnormalities. IGT is related to post-challenge glucose dysregulation and measures impaired insulin resistance and reduced glucose uptake, whereas IFG is related to fasting glucose dysregulation and measures hepatic insulin resistance in the liver and excess hepatic production of glucose.[26] This distinction is important because while IFG and IGT are both considered prediabetic states and used for classification, they reflect partially distinct pathophysiologic processes, metabolic abnormalities, natural histories, and responses to intervention.[26] For example, the use of fasting glucose measurements to classify prediabetes and type 2 diabetes may represent different levels of fasting glucose-related dysglycemia along the same glycemic continuum where both conditions can capture a fasting glucose-related phenotype.[27] This suggests that the two phenotypes may be more comparable with respect to fasting-glucose related dysglycemia. In comparison, IGT captures abnormalities in glucose tolerance even if the fasting glucose values are below the IFG range.[27] These distinctions are important since the method used to define dysglycemia determines which aspect of the heterogenous prediabetes

phenotype is being captured.[9] As a result, associations with genetic variation and environmental exposures may differ depending on whether they're characterized by fasting glucose, post-challenge glucose, or both.

Prediabetes has commonly been treated as an intermediate stage before type 2 diabetes, reflecting a state of lower disease severity along a non-linear glycemic continuum. However limited attention has been given to whether genetic associations differ between prediabetes and type 2 diabetes, particularly when prediabetes is defined using fasting glucose and therefore primarily reflects and IFG-related phenotype.[22-27] Figure 1 presents the conceptual framework to describe abnormal glycemia as a staged but non-linear continuum from normal glycemia to prediabetes and then to type 2 diabetes. This framework emphasizes that progression across glycemic stages may involve increasing glycemic impairment, cumulative exposure to dysglycemia, beta-cell dysfunction, medication use, and disease burden. However, treating prediabetes only as an intermediate stage may overlook important genetic or environmental mechanisms relevant to prediabetes itself, particularly given evidence that reversion from prediabetes to normal glycemia is possible. The DPP and DPS trials' subsequent analyses showed that regression to normal glucose regulation from prediabetes was associated with a lower risk of future diabetes.[6,7] More recent work further supports regression or remission from prediabetes as a clinically meaningful outcome.[28-30] Studies of prediabetes remission indicate that individuals vary considerably in their risk of progression and their capacity for remission. Regardless, recent prospective evidence further shows that restoration of normoglycemia among individuals with prediabetes is associated with substantially lower subsequent risk of type 2 diabetes compared with persistent prediabetes.[28-30] These findings support viewing prediabetes as a dynamic and potentially reversible stage of dysglycemia rather than an inevitable transition to

type 2 diabetes. This pattern highlights the clinical importance of intervention at an earlier stage of abnormal glycemia as it may offer greater opportunity for prevention and intervention before progression to established type 2 diabetes. Prediabetes is clinically important because it is often asymptomatic and detected only through laboratory measures, which may contribute to underdiagnosis and delayed intervention.[22-30] Moreover, individuals with prediabetes remain at elevated risk for progression to diabetes and for cardiovascular complications relative to those with normal glycemia.[22-30]

Prediabetes and type 2 diabetes share common pathophysiological roots, yet progression and biological expression may differ substantially. Both conditions involve insulin resistance and disruptions in glucose homeostasis; however, in prediabetes, these disturbances are generally less severe, and pancreatic β-cell function is often partially preserved.[24] This partial compensation allows some individuals to maintain stable glycemia control or revert to normoglycemia with appropriate intervention methods.[27] In contrast, progression to type 2 diabetes is marked by progressive β-cell dysfunction and sustained insulin resistance across key metabolic tissues: the liver, the skeletal muscle, and adipose tissue.[10-12,31] These physiological failures give rise to the clinical manifestations and symptoms.[10-12] Despite their placement along the same glycemic continuum, the variability in outcomes among individuals with prediabetes compared to frank diabetes suggests that this state is more than just a transient precursor. Progression to type 2 diabetes is heterogeneous and not inevitable, as some progress rapidly to type 2 diabetes, while others remain metabolically stable for years without worsening.[27-29] This heterogeneity highlights the importance of identifying factors that influence progression across stages of glycemia.

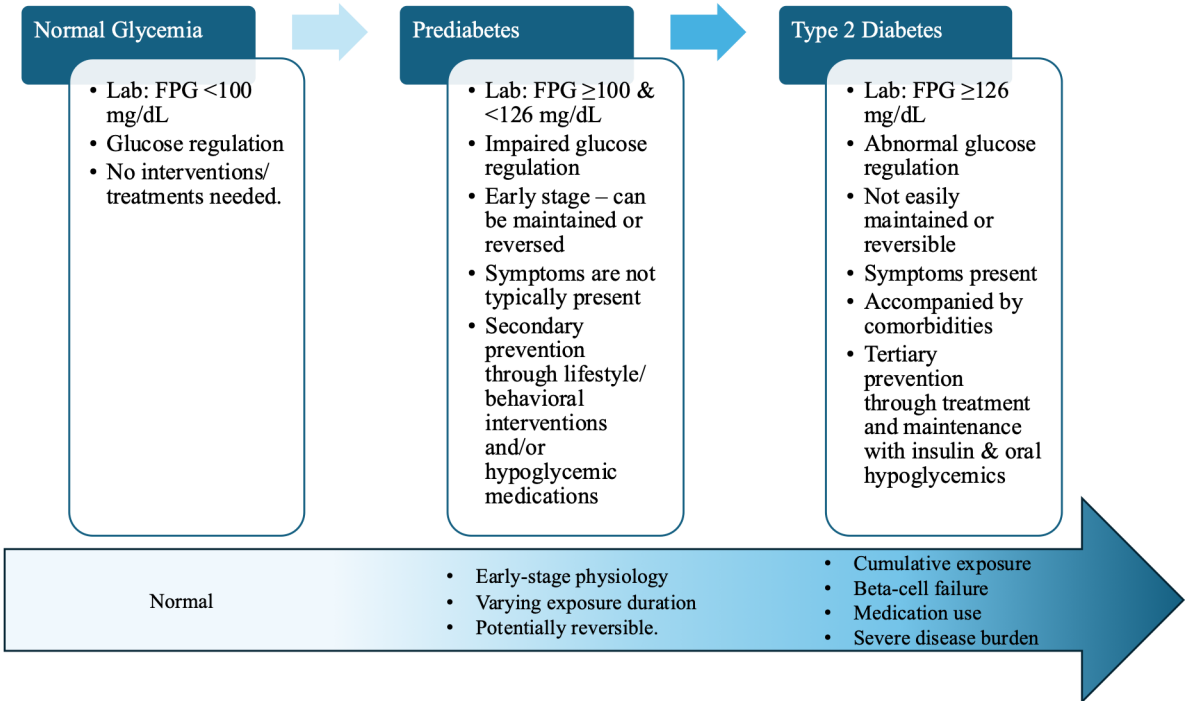

Figure 1. Conceptual Framework of Abnormal Glycemia

Genetic Contributions: Prediabetes vs. Type 2 Diabetes

Although fewer studies have examined genetic associations with prediabetes, emerging evidence suggests that genetic factors may contribute to variation in early glycemic dysregulation.[32-37] Family history studies suggest that early dysglycemia aggregates among individuals with familial diabetes risk.[32,34] In a multicenter analysis from the German Center for Diabetes Research, family history of diabetes was associated with high risk of prediabetes, particularly combined with impaired fasting glucose and impaired glucose tolerance.[33] These findings suggest that susceptibility to early dysglycemia may be influenced by inherited factors and shared familial risk.[32-34] However, direct evidence from twin concordance, segregation analysis, and family-based heritability studies remains limited for prediabetes specifically. Most recent genetic studies provide additional support that prediabetes is genetically influenced. Studies involving polygenic risk scores have suggested that prediabetes is genetically influenced

and may be heterogeneous, with distinct genetic subtypes associated with different risks of progression to type 2 diabetes and variation in underlying metabolic pathways.[34] Similarly, studies evaluating genetic risk for beta-cell dysfunction among individuals with prediabetes indicate that genetic predisposition may influence early pathophysiologic changes before the onset of overt diabetes.[35] Therefore, loci may represent a range of biological pathways involved in impaired glycemia. One study has also examined genetic variation in relation to changes in prediabetes status over time. For example, Liu et al. conducted a genome-wide single SNP-based and gene-based analysis of prediabetes status change among participants with baseline prediabetes in the Atherosclerosis Risk in Communities (ARIC) study, with replication in Framingham Heart Study (FHS).[36] The study evaluated transition from prediabetes to either type 2 diabetes or normoglycemia and identified several gene-based associations with prediabetes status change, including *SGCZ*, *HPSE2*, *ADGRA1*, *GLB1L3*, and *PCSK6*.[36] These genes were associated with either transition to normoglycemia or type 2 diabetes, however, it is unclear specifically which genes were involved with either processes, or the gene associations with prediabetes status itself, which is important for understanding early dysglycemia.[36] Other studies have examined genetic variation in relation to continuous glycemic traits among individuals with prediabetes, including fasting plasma glucose, 2-hour postprandial glucose, HbA1c, and BMI, rather than prediabetes as a categorical outcome.[37] Together, these studies show the heterogeneity in prior prediabetes-related genetic research and lack of consistent phenotyping for prediabetes in studies since they've evaluated different outcomes, including progression to type 2 diabetes, reversion to normoglycemia, and continuous glycemic traits. As a result, SNPs identified from prior studies of prediabetes, progression or related glycemic traits may not all represent the same biological processes captured by a single type of measurement to define

prediabetes. In particular, existing prediabetes genetic studies often lack replication across diverse racial/ethnic ancestry groups, limiting the generalizability of reported findings. This gap supports the need to evaluate prediabetes status and not just transition, or glycemic measures associated with the phenotype. Specifically, if different variants influence prediabetes than type 2 diabetes, it may be possible to identify at-risk individuals earlier when interventions can reduce morbidity and premature mortality.

Conversely, genetic association studies have been instrumental in identifying genetic markers associated with type 2 diabetes. Although type 2 diabetes definitions can also vary across genetic and epidemiologic studies, including the use of glucose thresholds, oral glucose tolerance tests, HbA1C, medication use, clinical diagnosis, and self-report, the genetics of type 2 diabetes has been more extensively studied and replicated than that of prediabetes.[38] The estimated heritability of type 2 diabetes has been found to range from 20-80% based on assessing the proportion of phenotypic variation attributable to genetic factors from a diverse set of population, family, and twin-based studies.[39-50] The lifetime risk of developing the disease depends on the number of affected parents, with the lifetime risk as high as 70% with both parents affected.[39] A range of genetic studies, including segregation analyses, linkage studies, candidate gene studies, and genome-wide association studies (GWAS), have significantly advanced the understanding of the genetic basis of type 2 diabetes.[39-50] These efforts have been instrumental in identifying common and rare variants that contribute to disease susceptibility across diverse ancestry.[39-50] Linkage studies identified variants such as the *CAPN10* (calpain 10) and *TCF7L2* (transcription factor 7-like 2) genes, which were among the earliest genes linked to type 2 diabetes.[39] Candidate gene studies, which involve assessing biologically plausible genes involved in the pathogenesis of insulin secretion, glucose metabolism, post-receptor signaling,

and lipid metabolism emerged and identified notable loci through this approach, which included the *PPARG* (peroxisome proliferator-activated receptor gamma), *IRS1* (insulin receptor substrate 1), *KCNJ11* (potassium inwardly-rectifying channel, subfamily J member 11), *WFS1* (Wolfram syndrome 1), and *HNF1A/B* (HNF1 homeobox A and B).[39] The advent of high-throughput genotyping technologies and access to large-scale datasets like HapMap introduced the field to the development of genome-wide association studies (GWAS).[39] GWAS's allowed researchers to assess hundreds of thousands of SNPs across the genome, ultimately uncovering over 700 genetic variants associated with type 2 diabetes risk by 2024.[39-50] The loci that were confirmed with use of GWAS's include the following genes: *TCF7L2* (transcription factor 7-like 2), *KCNJ11* (potassium inwardly rectifying channel subfamily j member 11), and *PPARG* (peroxisome proliferator-activated receptor gamma).[39] GWAS allowed for identification of novel loci including the following: *HHEX* (hematopoietically expressed homeobox), *SLC30A8* (solute carrier family 30 member 8), *CDKN2A/B* (cyclin-dependent kinase inhibitor 2A/B), *IGF2BP2* (insulin-like growth factor 2 mRNA binding protein 2), *CDC123/CAMK1D* (cell division cycle 123 and calcium/calmodulin-dependent protein kinase ID), *ARAP1* (ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1), *RBMS1* (RNA binding motif single stranded interacting protein 1), *LOC387761*, and *FTO* (fat mass and obesity associated).[39-50] These genes have been replicated in multiple populations and studies and contributed to the understanding of type 2 diabetes risk.

Risk Factors

Assessment of risk factors associated with prediabetes has commonly been combined with that of type 2 diabetes, with only a handful of studies solely focusing on prediabetes. These

studies indicate risk of developing prediabetes involves a combination of genetic, environmental, and lifestyle factors, similar to type 2 diabetes.[51-53] These risk factors can be categorized into two overarching classifications for both prediabetes and type 2 diabetes, which are non-modifiable and modifiable risk factors. Non-modifiable risk factors are age, race/ethnicity, family history of diabetes, and genetics.[51-53] Modifiable risk factors include smoking intake, alcohol intake, sleep quality, stress levels, physical activity level, sedentary habits, diet, obesity, high blood pressure, low high-density lipoprotein (HDL), high triglycerides, and hypertension.[51-53] Smoking is a well-established risk factor for the development of type 2 diabetes and has been mentioned in the 2014 Surgeon General's Report for the Health Consequences of Smoking, which has stated that it is a cause of diabetes, increasing the risk of type 2 diabetes by 30-40% in smokers compared to non-smokers.[54] This was confirmed by the Insulin Resistance Atherosclerosis Study of five years, that found the risk of type 2 diabetes increased among smokers compared to non-smokers when accounting for other risk factors such as anthropometric and behavioral factors, age, and race.[55] Both active and passive smoking are strongly associated with type 2 diabetes, and although the risk declines substantially with time since quitting, it remains modestly elevated among former smokers who quit within five to nine years.[56] Willi et al. found that in a pooled analysis of 25 prospective cohort studies, there was a 44% higher risk of type 2 diabetes among active smokers compared to never smokers.[57] A plausible biological explanation for the relationship between smoking and type 2 diabetes could be due to the chemical components from tobacco that may expose toxic compounds to the pancreas and, as a result, impact β-cell function, which can contribute to a rise in inflammatory markers that increase the risk of type 2 diabetes.[55] In addition, smoking has been linked to reduced insulin sensitivity and increased

insulin resistance, providing a different mechanisms through which smoking may increase diabetes risk.[55]

The relationship between prediabetes and smoking has been evaluated in a handful of studies. A cross-sectional study examining the association of smoking and nicotine dependence with prediabetes in young and healthy adults found that both cumulative smoking exposure and nicotine dependence were associated with prediabetes.[58] Nicotine dependence may partly explain the association between smoking and prediabetes because of greater tobacco exposure that can lead to insulin resistance, reduced insulin sensitivity, inflammation, and impaired beta-cell function.[55,58] These studies are among the few that investigated the relationship between smoking and prediabetes; however, the literature replicating these studies is limited, and even fewer have considered smoking as a modifying factor in the relationship between genetics and prediabetes, especially among a multi-ethnic cohort. An area that has yet to be explored is whether the same set of genetic and environmental factors that jointly influence type 2 diabetes also contribute to prediabetes risk. Consequently, it is important to understand the joint influence of genetic markers and environmental risk factors, such as smoking, on disease susceptibility for prediabetes and type 2 diabetes.

Gene Environment Interaction

Prediabetes and type 2 diabetes arise from the interplay of genetic and environmental factors.[38] Although genetic factors contribute to dysglycemia, they do not explain all of the phenotypic variation in disease risk, underscoring the importance of both genetic and environmental exposures in genetic epidemiology studies.[39,59-60] Environmental and behavioral factors have been associated with type 2 diabetes risk and may contribute to variation in

glycemic regulation.[39,59-60] Gene-environment (GxE) interaction analyses provide a framework for evaluating whether genetic associations with glycemic status differ according to environmental exposures. Prior studies of gene-environment interaction in type 2 diabetes have examined whether lifestyle and behavioral exposures modify genetic susceptibility to type 2 diabetes, but findings have varied by exposure, population, and genetic risk measure.[59,60] Sorensen et al. noted that in gene-environment interaction and precision prevention for type 2 diabetes, lifestyle and interventions can reduce diabetes incidence among high-risk individuals, but existing evidence does not consistently show that prevention effects differ strongly by known genetic predisposition.[60] This suggests that gene-environment interaction remains biologically plausible but difficult to detect consistently, particularly when environmental exposure and genetic susceptibility are measured differently across studies.[60] Nevertheless, a study by Bossowska et al. described personalized nutrition in obesity and prediabetes and emphasized genetic variation may influence individual responses to modifiable lifestyle exposures, supporting continued evaluation of gene-environment relations in dysglycemia.[61] Prior work has evaluated interactions between genetic factors and diet, physical activity, lifestyle, alcohol use, and smoking on type 2 diabetes.[59,60] Smoking is particularly relevant for gene-environment interaction analyses because it is a modifiable diabetes risk factor and may affect metabolic pathways involved in dysglycemia.[55] Since smoking may contribute to diabetes risk through different biological mechanisms, it could modify the association between genetic variants and risk of prediabetes or type 2 diabetes.[55] Furthermore, it may also contribute to earlier abnormalities in glucose regulation.[60,62] In a genome-wide smoking-by-genotype interaction study of 61,164 individuals from 19 cohorts, Wu et al. identified potential SNP-by-smoking interactions for incident type 2 diabetes near several loci, including the *TCF7L2* gene, although

no significant interactions were observed for fasting glucose.[62] This suggests that smoking-related genetic modification may be relevant to diabetes risk, but findings may differ by phenotype and population. Overall, prior gene-environment interaction studies have primarily focused on established type 2 diabetes, with limited evaluation of prediabetes as a distinct outcome.

Research Aims & Significance

The goal of this study is to evaluate associations of selected candidate SNPs and smoking status with prediabetes and type 2 diabetes relative to normal glycemia in the GENNID family cohort, and to determine whether prediabetes shows a pattern of association that differs from type 2 diabetes within this candidate SNP panel. **Aim 1** will evaluate genetic associations of selected candidate SNPs with prediabetes and type 2 diabetes. **Aim 2** will evaluate gene-environment interactions between selected candidate SNPs and smoking status, defined as ever vs. never smoking, on prediabetes and type 2 diabetes. It was hypothesized that within this candidate SNP set, prediabetes would exhibit distinct genetic associations compared to type 2 diabetes. Smoking was hypothesized to modify these genetic associations, with interaction effects differing between prediabetes and type 2 diabetes.

MATERIALS & METHODS

Multinomial, Family-Based, and Ancestry-Aware Analyses

The present study categorized glycemic status into three groups: normal glucose, prediabetes, and type 2 diabetes. A multinomial modeling framework was used to evaluate the genetic associations across disease stages (prediabetes and type 2 diabetes) within a single model. Normal glucose was specified as the reference category, allowing separate estimation of SNP associations for prediabetes versus normal glucose and type 2 diabetes versus normal glucose.

Analyses were conducted using a family-based multinomial modeling approach to account for correlation among related individuals in the GENNID study sample. Specifically, kinship coefficients were incorporated into the working correlation matrix to account for non-independence among family members and to obtain appropriate variance estimates. This approach allowed related individuals to be retained in the analysis rather than excluding one individual per family or treating observations as independent.

All analyses were conducted separately within each ancestry group: European American, Hispanic American, and African American. Ancestry-stratified analyses were used to reduce confounding from population structure and to allow SNP associations to be evaluated within more genetically comparable groups. Principal components were also included as covariates within ancestry-specific models to further adjust for residual population structure.

Collectively, this approach allowed SNP associations to be estimated across glycemic disease stages while accounting for both familial correlation and ancestry-specific genetic structure.

Study Samples

This project leverages data from the Genetics of Non-Insulin Dependent Diabetes Mellitus (NIDDM) (GENNID) study, a rich multi-ethnic family study of type 2 diabetes sponsored by the American Diabetes Association (ADA).[63] The GENNID data source is a cross-sectional study conducted between 1993 and 1997 that contains genetic, laboratory, and questionnaire information on non-Hispanic white, Hispanic, African American, and Japanese American multiplex families collected at eight different sites across the United States.[63] The purpose of this study was to cultivate a family-based resource for identifying genes involved in the etiology of type 2 diabetes.[63] The GENNID study's ascertainment of families occurred through enrollment of families with at least two siblings affected with type 2 diabetes, a minimum of three additional first-degree relatives, and a non-diabetic relative.[63] Index cases are defined using the National Diabetes Data Group criteria during the time of study ascertainment, which involved either a fasting plasma concentration of 7.8 mmol/l (≥140 mg/dL) on more than one occasion or a plasma glucose concentration of 11.1 mmol/l (≥200 mg/dL) in a 2-hour sample and in at least one other sample during an OGTT test.[63] The eligibility criteria of this study required families to have a minimum of two siblings with documented NIDDM, and three first-degree relatives (parents, siblings, or offspring) in addition to the two siblings.[63] The first-degree relatives of additional family members were included as well.[63] All participants had to be at least 18 years old.[63] The GENNID study recruited 1,520 participants from 259 families with the self-reported racial and ethnic breakdown as follows: 600 non-Hispanic white American subjects, 450 Hispanic American subjects, 150 African American subjects, and 150 Japanese American subjects.[63] The original study population included 1,891 individuals, and of those, a sample of 1,520 participants had genetic data. Index cases are those with T2D based on the National

Diabetes Data Group criteria during the time of study ascertainment, which involved either a fasting plasma concentration of 7.8 mmol/l (≥140 mg/dL) on more than one occasion or a plasma glucose concentration of 11.1 mmol/l (≥200 mg/dL) in a 2-hour sample and in at least one other sample during an OGTT test.[63] Participants provided two blood samples collected 15 minutes apart after a 10-hour fast and abstaining from medication and smoking, which were used to assess lipid, glucose, and insulin measures.[63] Questionnaires were administered to determine family eligibility during the study visit, where information regarding age, sex, relationship to the index case, height, weight, medical history, education, occupation, activity level, tobacco and alcohol use, medical conditions, medication use, and diet.[63]

Glycemic Status: Prediabetes and Type 2 Diabetes Phenotype Definitions

The phenotype definitions of prediabetes and type 2 diabetes were based on fasting glucose values and self-reported type 2 diabetes medication use from the questionnaire data. Since diagnostic criteria for type 2 diabetes have changed since the original time of data collection, the current American Diabetes Association (ADA) 2024 classifications were applied to categorize participants as having normal glycemia, prediabetes, and type 2 diabetes.[64] According to the ADA guidelines, normal glycemia was defined as a fasting plasma glucose (FPG) of <100 mg/dL, impaired glycemia or prediabetes as an FPG ≥100 mg/dL to <126 mg/dL, and hyperglycemia or type 2 diabetes as a FPG ≥126 mg/dL.[64] Type 2 diabetes medication use was determined from questionnaire data on medication use and medication type. Participants who reported taking medication for diabetes were categorized as having type 2 diabetes regardless of FPG level. Participants who did not provide a yes/no response to the diabetes medication use question but reported a medication type consistent with treatment for type 2

diabetes, including oral glucose-lowering medications and insulin, were also categorized as having type 2 diabetes.[64,65] This approach accounted for individuals with established diabetes whose FPG values may have been lowered by pharmacologic treatment. At the time of the data collection, pharmacological treatment was reserved for individuals with established or diagnosed type 2 diabetes. This was consistent with the GENNID dataset, in which none of the individuals categorized as having prediabetes reported use of anti-diabetic medication, including oral medications and insulin. Ninety-four individuals were missing fasting plasma glucose measurements. However, after incorporating diabetes medication-use information into the phenotype definition, only 59 participants remained without sufficient phenotype-defining information (missing both fasting plasma glucose and diabetes medication-use) and were coded as missing/NA for glycemic status. Figure 2 presents a flow diagram of the study population used for glycemic status definition, including participants with available genetic data, exclusions, fasting plasma glucose classifications, and reclassification after incorporating anti-diabetic medication use.

Smoking Status Definition

The GENNID study required all study participants to complete a family/medical history questionnaire that included information on behavioral and environmental variables, such as cigarette smoking. Participants self-reported whether they had ever smoked cigarettes during their lifetime and whether they currently smoked cigarettes. Based on these responses, participants were classified as either ever smokers or never smokers. Ever smokers included individuals who reported current or former cigarette smoking, whereas never smokers included individuals who reported no history of cigarette smoking and no current smoking.

Genetic Data Preparation

GENNID samples were genotyped using Illumina's Infinium LCG genotyping assay on the Multiethnic Global Beadchip (Build 37). Quality control was performed separately for each racial/ethnic group and included DNA quantification, gender validation, specifically, exclusion of samples due to low amount, concentration, or integrity of DNA, and sex-typing inconsistencies. Principal components (PC's) were developed for each racial/ethnic group, and eigenvalues were obtained to determine the number of PC's that should be included in each racial/ethnic group analysis. Scree plots for European, Hispanic, and African American ancestry groups indicated that the first principal component explained the highest proportion of variance.

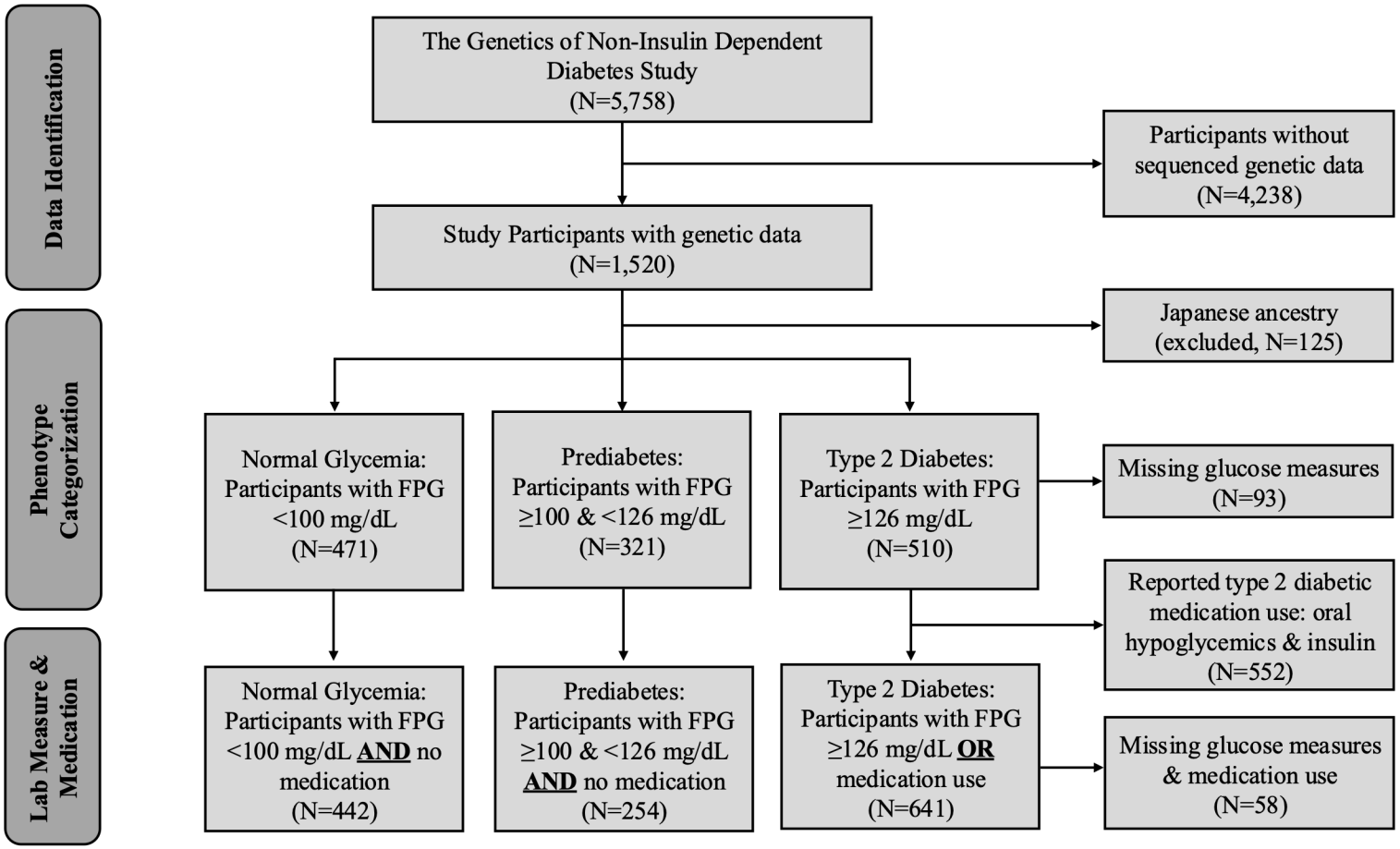

Figure 2. Flow Diagram of Diabetic Phenotype Study Population.

Note: Japanese ancestry was excluded due to limited sample size. After incorporating anti-diabetic medication use, 29 participants with normal glycemia and 67 participants with prediabetes were reclassified to Type 2 Diabetes, and 35 participants were classified with anti-diabetic medication use alone.

Candidate SNP Selection

Candidate SNPs were selected using two separate literature reviews: one to find prediabetes-associated SNPs (Group A) and another to find type 2 diabetes-associated SNPs (Group B). For Group A, SNP selection was further supplemented using the GWAS Catalog to identify additional variants previously associated with prediabetes.

The first literature review was performed to identify studies reporting SNPs associated with prediabetes (Group A). The review was performed in Google Scholar with selected articles cross-referenced in PubMed through the National Center for Biotechnology Information's (NCBI) National Library of Medicine. Search terms were constructed using Boolean operators and included: ("prediabetes") AND ("genome-wide association study" OR GWAS OR genetics). This search identified two studies: (1) A genome-wide association study of prediabetes status change by Liu et al., and (2) Genetic Variants relate to fasting plasma glucose, 2-hour postprandial glucose, glycosylated hemoglobin, and BMI in prediabetes by Lin et al.[36,37] The first study used data from the Atherosclerosis Risk in Communities (ARIC) study, a population-based cohort of 15,792 Black and White participants from four communities and adjusted for age, sex, body mass index, and three principal components.[36] Five SNPs from this study were selected based on suggestive significance ($P<1x10^{-5}$) for association with glycemic status change.[36] The second study is a GWAS of clinical indicators of prediabetes, including fasting plasma glucose (FPG) test, a two-hour oral glucose tolerance test (OGTT), glycosylated hemoglobin (HbA1c) test, and levels of body mass index (BMI) among a cohort of 451 Chinese individuals using the National Diabetes Prevalence Survey of the Chinese Medical Association.[37] Nine SNPs were identified from this study, including four SNPs associated with HbA1c among participants with prediabetes (rs142002616, rs6880621, rs946911, rs11853125), three SNPs associated with 2-

hour plasma glucose test (rs34938732, rs9550371, rs62212118), and two SNPs associated with the fasting plasma glucose (rs4661250, rs3094201).[37] The Genome-Wide Association Study (GWAS) Catalog, which reports SNP-trait associations from published genetic studies, was used as a third resource to search for prediabetes-associated SNPs. [66,67] This search identified one additional SNP not captured in the literature search, which traced to a genome-wide association study of fasting proinsulin, fasting insulin, 2-hour postprandial proinsulin, and 2-hour postprandial insulin in a Chinese cohort of 451 individuals from Hainan Medical University Hospital.[68] In total, 15 prediabetes SNPs were compiled from these three sources and considered for analysis in the present study.

A second literature review was performed to identify review articles reporting type 2 diabetes-associated SNPs. Given the extensive literature on type 2 diabetes, the search was restricted to review articles summarizing replicated and well-established type 2 diabetes-associated loci. This strategy allowed for prioritization of SNPs with prior evidence across multiple studies while maintaining a focused candidate SNP set for analysis. The review was performed in Google Scholar, with selected articles cross-referenced in PubMed through the NCBI's National Library of Medicine. Search terms were constructed using Boolean operators and included: ("type 2 diabetes" OR "type 2 diabetes mellitus" OR T2D) AND (genetics OR genetic OR "genetic basis" OR GWAS OR "genome-wide association study" OR SNP) AND (review). This search identified two review articles summarizing genome-wide significant type 2 diabetes susceptibility loci: (1) Insights into the genetic basis of type 2 diabetes by Kato, N. published in 2013, and (2) Type 2 Diabetes Genetics: Beyond GWAS by Sanghera and Blackett, published in 2012.[41,69] The review by Kato compiled a table of 70 type 2 diabetes susceptibility loci with genome-wide significance.[69] The Sanghera and Blackett review summarized a list of 78

type 2 diabetes susceptibility loci found in genome-wide association studies.[41] Of these, 43 SNPs were reported in both reviews. From these 43 overlapping loci, ten candidate SNPs were selected to provide a comparison set similar in size to the prediabetes candidate SNP group while maintaining representation across established type 2 diabetes loci. SNPs were prioritized using the following criteria: (1) presence in both review articles, (2) evidence of prior association with type 2 diabetes at the genome-wide significant level, (3) evidence of association across ancestry groups, particularly European, African, and Hispanic ancestry groups, and lastly (4) inclusion of loci implicated in different biological pathways related to type 2 diabetes pathophysiology. This final criterion was used to avoid selecting SNPs concentrated within a single biological pathway and to instead reflect the biological heterogeneity of type 2 diabetes, including insulin secretion, beta-cell function, insulin processing, adiposity-related risk, and insulin sensitivity. Five of the ten selected SNPs demonstrated prior evidence in all three ancestry groups examined in this study, while the remaining five SNPs were selected based on strong evidence in European ancestry populations, which comprised the largest ancestry group in GENNID.

Lastly, all candidate SNPs were cross-referenced with GENNID genotype data to determine whether they were available in the study sample. The full list of candidate SNPs and their sources is listed in Table 1, and the SNPs available for analysis by ancestry group is shown in Table 2. A total of 22 SNPs were available in the European and Hispanic ancestry groups, and 21 SNPs were available in the African ancestry group.

Table 1. Candidate SNPs				
Source	**rsID**	**CHR**	**Pos37**	**Gene**
A.1. Liu et al.	rs36087584	3	119097548	ARHGAP31
	rs498414	3	185199267	MAP3K13
	rs10229340	7	51596421	LOC105375277
	rs4886290	13	61490922	LINC01442
	rs41497851	15	40198302	GPR176
A.2. Lin et al.	rs4661250	1	223399328	SUSD4
	rs142002616	1	30589506	MATN1/LINC01648
	rs34938732	4	78345863	CCNG2
	rs6880621	5	34504382	GUSBP18/RAI14-DT
	rs3094201	6	31092763	PSORS1C1
	rs9550371	13	31264917	ALOX5AP
	rs946911	14	38508754	TTC6
	rs11853125	15	92288156	THRAP3P2/SLCO3A1
	rs62212118	21	35465506	MRPS6/SLC5A3
A.3. GWAS Catalog	rs3803084	12	52942627	KRT71
B. Replicated Type 2 Diabetes SNPs	rs1801282	3	12393125	PPARG
	rs13266634	8	118184783	SLC30A8
	rs7903146	10	114758349	TCF7L2
	rs1111875	10	94462882	HHEX
	rs12779790	10	12328010	CDC123/CAMK1D
	rs1552224	11	72433098	ARAP1
	rs5219	11	17409572	KCNJ11
	rs8050136	16	53816275	FTO
	rs564398	9	22029547	CDKN2B-AS1
	rs7593730	2	161171454	RBMS1
Source of SNPs listed as A refer to prediabetes-associated SNPs and B refers to type 2 diabetes-associated SNPs				

Table 2. SNPs in each Racial/Ethnic Group in GENNID					
	Marker		Race/Ethnicity		
Source	CHR	Position	EA	AA	MA
A.1. Liu et al. - Prediabetes	3	119097548	✓	✓	✓
	3	185199267	✓		✓
	7	51596421	✓	✓	
	13	61490922	✓	✓	✓
	15	40198302	✓	✓	✓
A.2. GWAS Catalog SNPs - Prediabetes	1	223399328	✓	✓	✓
	4	78345863	✓	✓	✓
	5	34504382	✓	✓	✓
	12	52942627			✓
	13	31264917	✓	✓	✓
	14	38508754	✓	✓	✓
	15	92288156	✓	✓	✓
	21	35465506	✓	✓	✓
B. Common Type 2 Diabetes SNPs	3	12393125	✓	✓	✓
	8	118184783	✓	✓	✓
	10	114758349	✓	✓	✓
	10	94462882	✓	✓	✓
	10	12328010	✓	✓	✓
	11	72433098	✓	✓	✓
	11	17409572	✓	✓	✓
	16	53816275	✓	✓	✓
	9	22029547	✓	✓	✓
	2	161171454	✓	✓	✓
Total			22	21	22

Table 3. Candidate SNP Genotype Counts and Minor Allele Frequencies by Ancestry-Specific Allele Coding													
CHR:POS	rsID	European American				Hispanic American				African American			
		0	**1**	**2**	**gnomAD MAF**	**0**	**1**	**2**	**ensemble MAF**	**0**	**1**	**2**	**gnomAD MAF**
X1.223399328.A.G_G	rs4661250	400	103	13	0.1393	436	156	6	0.094	265	15	*NA*	0.0251
X2.161171454.T.C_T	rs7593730	340	154	22	0.7997	434	159	4	0.18	113	147	21	0.6238
X3.12393125.C.G_G	rs1801282	363	134	18	0.1245	471	120	6	0.133	263	18	*NA*	0.0215
X3.119097548.T.C_C	rs36087584	229	223	64	0.401	221	275	101	0.477	-	-	-	-
X3.119097548.T.C_T	rs36087584	-	-	-	-	-	-	-	-	78	130	73	0.452
X3.185199267.C.A_A	rs498414	482	34	*NA*	0.0382	568	29	*NA*	0.016	-	-	-	-
X4.78345863.G.A_A	rs34938732	372	133	11	0.1494	437	141	20	0.219	201	78	2	0.1061
X5.34504382.G.A_G	rs6880621	141	243	132	0.5131	231	262	105	0.328	119	134	28	0.328
X7.51596421.A.G_A	rs10229340	140	267	109	0.514	-	-	-	-	-	-	-	-
X7.51596421.A.G_G	rs10229340	-	-	-	-	-	-	-	-	39	134	108	0.6193
X8.118184783.C.T_T	rs13266634	273	209	33	0.3038	339	210	49	0.242	229	48	4	0.0889
X9.22029547.T.C_C	rs564398	176	267	73	0.4169	392	180	26	0.117	232	44	5	0.0658
X10.12328010.A.G_G	rs12779790	322	169	25	0.1732	411	166	18	0.133	208	70	3	0.148
X10.94462882.C.T_T	rs1111875	176	271	68	0.4098	279	259	59	0.391	182	84	15	0.2306
X10.114758349.C.T_T	rs7903146	239	219	57	0.2693	262	280	55	0.219	126	119	36	0.3022
X11.17409572.T.C_T	rs5219	223	218	75	0.6312	242	285	71	0.406	250	31	*NA*	0.9385
X11.72433098.A.C_C	rs1552224	426	80	10	0.1639	491	101	5	0.070	269	12	*NA*	0.0254
X12.52942627.A.G_G	rs3803084	-	-	-	-	406	164	28	0.336	-	-	-	-
X13.31264917.G.A_A	rs9550371	-	-	-	-	-	-	-	-	157	97	27	0.2372
X13.31264917.G.A_G	rs9550371	192	232	86	0.607	162	293	143	0.469	-	-	-	-
X13.61490922.C.T_T	rs4886290	419	96	1	0.1196	462	132	4	0.141	135	118	28	0.2776
X14.38508754.C.T_C	rs946911	171	246	99	0.5743	365	196	36	0.281	215	62	4	0.9136
X15.40198302.G.C_C	rs41497851	246	225	45	0.3019	431	157	10	0.156	199	72	10	0.1326
X15.92288156.G.T_T	rs11853125	436	72	8	0.0760	334	213	51	0.297	120	129	32	0.3371
X16.53816275.C.A_A	rs8050136	163	258	95	0.4206	272	261	63	0.219	87	143	51	0.4432
X21.35465506.G.A_A	rs62212118	421	94	1	0.0661	557	41	*NA*	0.047	250	30	1	0.0449

Note: Table of allele counts (0,1,2) in the GENNID study and the Minor Allele Frequency Percentage based on gnomAD and Ensembl. CHR:POS, Chromosome and Position; 0: homozygous major, 1: heterozygous, 2: homozygous minor. rsID's appear in more than one row due to the coded allele differed across ancestry-specific genotype files. Dashes indicate the SNP may be represented in a separate row with different allele coding.

Table 4. European Ancestry Models					
rsID	**CHR**	**Pos37**	**Gene**	**Aim 1 Model**	**Aim 2 Model**
rs4661250	1	223399328	SUSD4	Additive	Additive
rs7593730	2	161171454	RBMS1	Additive	Additive
rs1801282	3	12393125	PPARG	Additive	Additive
rs36087584	3	119097548	ARHGAP31	Additive	Additive
rs498414	3	185199267	MAP3K13	Additive	Additive
rs34938732	4	78345863	CCNG2	Additive	Additive
rs6880621	5	34504382	GUSBP18/RAI14-DT	Additive	Additive
rs13266634	8	118184783	SLC30A8	Additive	Additive
rs564398	9	22029547	CDKN2B-AS1	Additive	Additive
rs12779790	10	12328010	CDC123/CAMK1D	Additive	Additive
rs1111875	10	94462882	HHEX	Additive	Additive
rs7903146	10	114758349	TCF7L2	Additive	Additive
rs5219	11	17409572	KCNJ11	Additive	Additive
rs1552224	11	72433098	ARAP1	Additive	Additive
rs3803084	12	52942627	KRT71	Additive	Additive
rs9550371	13	31264917	ALOX5AP	Additive	Additive
rs4886290	13	61490922	LINC01442	Dominant	Dominant
rs946911	14	38508754	TTC6	Additive	Additive
rs41497851	15	40198302	GPR176	Additive	Additive
rs11853125	15	92288156	THRAP3P2/SLCO3A1	Additive	Additive
rs8050136	16	53816275	FTO	Additive	Additive
rs62212118	21	35465506	MRPS6/SLC5A3	Dominant	Dominant

Table 5. Hispanic Ancestry Models					
rsID	**CHR**	**Pos37**	**Gene**	**Aim 1 Model**	**Aim 2 Model**
rs4661250	1	223399328	SUSD4	Dominant	Dominant
rs7593730	2	161171454	RBMS1	Dominant	Dominant
rs1801282	3	12393125	PPARG	Dominant	Dominant
rs36087584	3	119097548	ARHGAP31	Additive	Dominant
rs498414	3	185199267	MAP3K13	Additive	Additive
rs34938732	4	78345863	CCNG2	Additive	Additive
rs6880621	5	34504382	GUSBP18/RAI14-DT	Dominant	Dominant
rs13266634	8	118184783	SLC30A8	Additive	Additive
rs564398	9	22029547	CDKN2B-AS1	Additive	Additive
rs12779790	10	12328010	CDC123/CAMK1D	Additive	Additive
rs1111875	10	94462882	HHEX	Additive	Additive
rs7903146	10	114758349	TCF7L2	Additive	Additive
rs5219	11	17409572	KCNJ11	Additive	Additive
rs1552224	11	72433098	ARAP1	Dominant	Dominant
rs3803084	12	52942627	KRT71	Additive	Additive
rs9550371	13	31264917	ALOX5AP	Additive	Additive
rs4886290	13	61490922	LINC01442	Dominant	Dominant
rs946911	14	38508754	TTC6	Additive	Additive
rs41497851	15	40198302	GPR176	Additive	Additive
rs11853125	15	92288156	THRAP3P2/SLCO3A1	Additive	Additive
rs8050136	16	53816275	FTO	Additive	Additive
rs62212118	21	35465506	MRPS6/SLC5A3	Additive	Additive

Table 6. African Ancestry Models					
rsID	**CHR**	**Pos37**	**Gene**	**Aim 1 Model**	**Aim 2 Model**
rs4661250	1	223399328	SUSD4	Additive	-
rs7593730	2	161171454	RBMS1	Additive	Additive
rs1801282	3	12393125	PPARG	Additive	-
rs36087584	3	119097548	ARHGAP31	Additive	Additive
rs34938732	4	78345863	CCNG2	Dominant	Dominant
rs6880621	5	34504382	GUSBP18/RAI14-DT	Additive	Additive
rs10229340	7	51596421	LOC105375277/ ENSG00000285741	Additive	Additive
rs13266634	8	118184783	SLC30A8	Dominant	-
rs564398	9	22029547	CDKN2B-AS1	Dominant	Dominant
rs12779790	10	12328010	CDC123/CAMK1D	Dominant	Dominant
rs1111875	10	94462882	HHEX	Additive	Additive
rs7903146	10	114758349	TCF7L2	Additive	Additive
rs5219	11	17409572	KCNJ11	Additive	Additive
rs1552224	11	72433098	ARAP1	Additive	-
rs9550371	13	31264917	ALOX5AP	Additive	Additive
rs4886290	13	61490922	LINC01442	Additive	Additive
rs946911	14	38508754	TTC6	Dominant	Dominant
rs41497851	15	40198302	GPR176	Additive	Additive
rs11853125	15	92288156	THRAP3P2/SLCO3A1	Additive	Additive
rs8050136	16	53816275	FTO	Additive	Additive
rs62212118	21	35465506	MRPS6/SLC5A3	Dominant	Dominant
Cells with a dash are SNPs that could not be assessed due to low sample size.					

Genetic Model Specification: Dominant and Additive Models

SNP minor allele frequencies were used to determine the model type implemented for each candidate SNP analysis. The additive model was used as the default, since the underlying inheritance mechanism for Prediabetes is unknown and type 2 diabetes is predominantly inherited in an additive fashion. However, in cases where SNPs had ≤10 participants homozygous for the minor allele, a dominant model was used to reduce the number of small

cells. A table of allele frequencies and the model used for each racial/ethnic group is provided in Tables 3-6.

Statistical Methods

The statistical model used in this study was based on an approach developed by Wang et al., who introduced the model and accompanying code for genetic association testing of multi-category phenotypes in family-based samples.[70] Generalized estimated equations (GEE) is a method that can handle correlated and clustered data, making it well-suited for family-based samples in which individuals within the same family are not independent. [71,72] GEE can also accommodate a variety of outcome types, including continuous, binary, and count data, which is especially useful for the statistical structure of this analysis since it uses categorical data.[71,72] The GEE estimates the population average or marginal mean, and uses a working correlation matrix to account for within-family correlation while producing robust standard errors, without requiring specification of the full joint likelihood. [70-72] The extended generalized estimated equation (EGEE) expands this approach by approximating the joint likelihood of correlated observations using a quasi-likelihood function, avoiding the need to fully specify the joint distribution of related individuals within a family; this allows mean and association parameters to be estimated simultaneously.[70-72] This method produces well-controlled type I error rates when simulated in family studies, compared to traditional methods of using a multinomial logistic regression that ignores familial correlation and produces inflated type I error rates.[70] The statistical analysis structure utilizes the EGEE to incorporate a quasi-likelihood to account for correlated observations and derives a Wald Test Statistic with one parameter to test a single coefficient at a time.[70] The Wald Test Statistic was chosen for this analysis over alternative

approaches, such as a likelihood ratio or score test, because constructing these alternatives is difficult in the GEE framework, which does not rely on a fully specified likelihood. The Wald test statistic is comparatively straightforward to implement using the estimated coefficients and their variance-covariance matrix and additionally provides effects sizes and confidence intervals for each SNP. The model assumes that there are N independent families, with a n_i individuals in family i, and a total of n subjects, with a K-category multinomial trait, with the K^{th} level as the reference level of the study's multinomial structure.[70]

The model consists of a matrix used to model the correlation between any two individuals in the same family along with a relationship matrix.[70] The correlation structure models the correlation between individuals within the same family, and the relationship matrix, defined as twice the kinship matrix, reflects genetic relatedness between individuals. In this dataset, kinship coefficients ranged from 0.03125 to 0.25, consistent across all three ancestry groups included in the analysis. Because this formula scales the modeled correlation by each pair's kinship coefficient, individuals with higher genetic relatedness are modeled as having proportionally stronger correlation in diabetes status than more distantly related individuals. The multinomial model and accompanying code were utilized to address the study objectives, evaluating the following glycemic group comparisons: prediabetes versus normal glycemia and type 2 diabetes versus normal glycemia.

Multiple Testing Correction: False Discovery Rate (FDR)

To account for multiple testing, the Benjamini-Hochberg false discovery (FDR) method was applied within each ancestry group to control the expected proportion of false discoveries among rejected hypotheses (significant findings). In the multinomial analysis, each SNP

contributed two hypothesis tests corresponding to the number of comparisons: prediabetes vs. normal glycemia and type 2 diabetes vs. normal glycemia. Therefore, the number of tests used for FDR correction was defined as the number of SNPs multiplied by the two outcome contrasts within each ancestry group. This resulted in 44 tests for the European and Hispanic ancestry groups, corresponding to 22 SNPs x 2 comparisons, and 42 tests for the African ancestry group, corresponding to 21 SNPs x 2 comparisons. The FDR method of correction was selected because the study involved a moderate number of correlated tests in a candidate SNP framework, for which controlling the expected proportion of false discoveries among the detected ones is more appropriate than more conservative methods, such as Bonferroni correction, which instead controls the family-wise error rate or the probability of any single false discoveries across all tests. The Benjamini and Hochberg (BH) false discovery rate FDR is an approach that was introduced in 1995 to control for multiple testing and essentially tests for the expected proportion of false discoveries (incorrect rejection of the null hypothesis among all discoveries).[73-76] This method still controls for the probability of making at least one false positive when performing multiple testing.[73-76]

Aim 1 Methods & Statistical Analysis

The objective of aim 1 is to assess the association between a set of candidate SNPs and glycemic status through a multinomial (polytomous) logistic regression model using an extended version of generalized estimating equations (EGEE) to conduct a family-based candidate SNP association study. Analyses will be run separately for each ancestry group to minimize the effects of population stratification. The models will include kinship coefficients to account for familial relatedness, principal components (PC's) for each ancestry group to further account for

any residual population stratification. The logit model below provides an overview of the statistical model structure that shows for a K-category multinomial outcome, the Kth category was specified as the reference category, resulting in $K-1$ equations. In this study, normal glycemia was specified as the reference category, and separate model parameters were estimated for prediabetes versus normal glycemia and type 2 diabetes versus norma glycemia.

$$log\left(\frac{P(Y_{ij}=k|G_{ij},X_{ij})}{P(Y_{ij}=K\,|G_{ij},\,X_{ij})}\right)=\hat{\alpha}_k+\,\hat{\beta}_k G_{ij}+X_{ij}^T\hat{\gamma}_k\;,\qquad k=1,\ldots,K-1$$

$$K=3\text{: }Normal\ Glycemia$$
$$k\ \in\{1\text{: }Prediabetes,2\text{: }Type\ 2\ Diabetes\}$$

$Y_{ij}=glycemic\ status\ for\ individual\ j\ in\ family\ i$
$K=reference\ category\ (Normal\ Glycemia)$
$k=non-reference\ category\ (either\ prediabetes\ or\ type\ 2\ diabetes)$
$G_{ij}=SNP\ genotype\ for\ individual\ j\ in\ family\ i$
$X_{ij}=vector\ of\ ancestry-specific\ principal\ components\ for\ individual\ j\ in\ family\ i$
$European\ \&\ African\ Ancestry\text{: }PC1$
$Hispanic\ Ancestry\text{: }PC1\ \&\ PC2$
$\hat{\alpha}_k=intercept\ for\ category\ k$
$\hat{\beta}_k=SNP\ effect\ for\ category\ k,relative\ to\ normal\ glycemia$
$\hat{\gamma}_k=vector\ of\ regression\ coefficients\ for\ the\ principal\ components\ for\ category\ k$

where Y_{ij} represents glycemic status for individual j in the family i, K is the reference outcome category, and k is the non-reference outcome category. G_{ij} represents the SNP genotype for individual *j* in family *i*. The term X_{ij} represents the vector of ancestry-specific principal components included as covariates to adjust for residual population structure. Models included up to two principal components depending on the ancestry group. The parameter $\hat{\alpha}_k$ is the intercept for the outcome category k, $\hat{\beta}_k$ is the SNP effect for outcome category k relative to normal glycemia, and $\hat{\gamma}_k$ is the vector of regression coefficients for the principal components for outcome category k.

Data cleaning, preparation, and analyses were conducted using R to handle the multinomial function and obtain Wald Chi-Square test results for comparison of the two

phenotypes of interest (prediabetes, diabetes) with the comparison group (normal glycemia). Familial correlation among related individuals was accounted for using kinship coefficients in the working correlation structure. Kinship coefficients were generated using the makekinship function from the kinship2 package, and the resulting sparse kinship matrix was converted to standard matrix format using the Matrix package. An FDR p-value cut-off of 0.05 (5%) will be considered statistically significant and suggestive significance will be determined based on a p-value of 0.05 prior to FDR correction. Forest plots and scatter plots were generated using the packages forestplot and ggplot2.

Cross-Ancestry Comparison

As a secondary descriptive step to evaluate consistency of findings across populations, results from the ancestry-specific analyses are summarized to identify SNPs that demonstrate evidence of association in more than one ancestry group. Cross-ancestry overlap was defined based on the 95% confidence intervals that excluded the null value across the three comparison groups.

Aim 2 Methods & Statistical Analysis

Aim 2 evaluated whether smoking modified the association between each candidate SNP and glycemic status. This was assessed by including SNP by smoking statistical interaction terms in the multinomial family-based model. Smoking was treated as the environmental modifier, and the interaction effects were estimated separately for prediabetes versus normal glycemia and type 2 diabetes versus normal glycemia. All candidate SNPs are assessed in Aim 2. Analyses are

stratified by ancestry group and include kinship coefficients and principal components (PC's). The models include both the smoking term ($\widehat{B}_{2k}S_{ij}$) and the interaction ($\widehat{B}_{3k}(G_{ij} * S_{ij})$) term.

$$log\left(\frac{P(Y_{ij}=k|G_{ij},S_{ij},X_{ij})}{P(Y_{ij}=K\ |G_{ij},S_{ij},X_{ij})}\right)=\hat{\alpha}_k+\ \hat{\beta}_{1k}G_{ij}+\hat{\beta}_{2k}S_{ij}+\hat{\beta}_{3k}\left(G_{ij}\ \times S_{ij}\right)+X_{ij}^{T}\hat{\gamma}_k\ ,$$
$$k=1,\ldots,K-1$$

$$K=3\text{: }Normal\ Glycemia$$
$$k\ \in\{1\text{: }Prediabetes,2\text{: }Type\ 2\ Diabetes\}$$

$Y_{ij}=glycemic\ status\ for\ individual\ j\ in\ family\ i$
$K=reference\ category\ (Normal\ Glycemia)$
$k=non-reference\ category\ (either\ prediabetes\ or\ type\ 2\ diabetes)$
$G_{ij}=SNP\ genotype\ for\ individual\ j\ in\ family\ i$
$S_{ij}=smoking\ status\ for\ individual\ j\ in\ family\ i$
$G_{ij}\times S_{ij}=SNP\ by\ smoking\ interaction\ term$
$X_{ij}=vector\ of\ ancestry-specific\ principal\ components\ for\ individual\ j\ in\ family\ i$
$European\ \&\ African\ Ancestry\text{: }PC1$
$Hispanic\ Ancestry\text{: }PC1\ \&\ PC2$
$\hat{a}_k=intercept\ for\ category\ k$
$\hat{\beta}_{1k}=SNP\ effect\ for\ category\ k,relative\ to\ normal\ glycemia$
$\hat{\beta}_{2k}=smoking\ effect\ for\ category\ k\ relative\ to\ normal\ glycemia$
$\hat{\beta}_{3k}=SNP\ by\ smoking\ interaction\ effect\ for\ category\ k\ relative\ to\ normal\ glycemia$
$\hat{\gamma}_k=vector\ of\ regression\ coefficients\ for\ the\ principal\ components\ for\ category\ k$

Data cleaning, preparation, and analyses were conducted using R to handle the multinomial function and obtain Wald Chi-Square test results for comparison of the two phenotypes of interest (prediabetes, diabetes) with the comparison group (normal glycemia). Familial correlation among related individuals was accounted for using kinship coefficients in the working correlation structure. Kinship coefficients were generated using the makekinship function from the kinship2 package, and the resulting sparse kinship matrix was converted to standard matrix format using the Matrix package. An FDR p-value cut-off of 0.05 (5%) will be considered statistically significant and suggestive significance will be determined based on a p-

value of 0.05 prior to FDR correction. Forest plots and scatter plots were generated using the packages forestplot and ggplot2.

To interpret SNP-by-smoking interaction findings, smoking stratum-specific SNP to outcome odds ratios and 95% confidence intervals were obtained for SNPs showing nominal evidence of interaction prior to FDR-correction. These estimates were derived from the fitted multinomial SNP-by-smoking interaction models. For each outcome comparison, the SNP odds ratio among never smokers was calculated using the genotype coefficient, while the SNP odds ratio among ever smokers was calculated using the sum of the genotype coefficient and the corresponding SNP-by-smoking interaction coefficient. The 95% confidence intervals were calculated using the variance-covariance matrix from the fitted model. The calculation for never smokers involves exponentiating the genotype beta coefficient,

$$OR_{Never} = \exp(\hat{\beta}_G)$$

and obtaining the standard error using the variance-covariance matrix,

$$SE_{Never} = \sqrt{Var(\hat{\beta}_G)}$$

and calculating the 95% confidence interval.

$$\exp\left(\hat{\beta}_G \pm 1.96 \times SE_{Never}\right)$$

The calculation for the ever smokers involves exponentiating the sum of the genotype and interaction coefficients,

$$OR_{Ever} = \exp(\hat{\beta}_G + \hat{\beta}_{GS})$$

and obtaining the standard error using the variance-covariance matrix,

$$SE_{Ever} = \sqrt{Var(\hat{\beta}_G) + Var(\hat{\beta}_{GS}) + 2Cov(\hat{\beta}_G, \hat{\beta}_{GS})}$$

And calculating the 95% confidence interval.

$$\exp\left[(\hat{\beta}_G + \hat{\beta}_{GS}) \pm 1.96 \times SE_{Ever})\right]$$

For dominant models, SNP odds ratios were compared individuals with 1 or more coded allele copies to individuals with 0 coded allele copies. For additive models, SNP odds ratios were interpreted per additional coded allele. Observed counts by genotype group, smoking status, and glycemic category were presented alongside these model-derived estimates. These stratum-specific analyses were descriptive and were used to evaluate whether the SNP-outcome association differed across smoking groups. The SNP-by-smoking interaction p-value was used to evaluate whether the SNP-outcome association differed by smoking status.

RESULTS

Study Population

The sample included 1,395 individuals across three ancestry groups. Glycemic status was available for 1,337 participants, including 442 with normal fasting glucose, 254 with prediabetes, and 641 with type 2 diabetes; 58 participants were missing data on their glycemic status. Among the 1,337 participants, 505 were of European ancestry, 569 were of Hispanic ancestry, and 263 were of African ancestry. The sex distribution of the sample includes 514 males and 823 females, and 608 ever smokers, 725 never smokers. The average age of those with normal glycemia is 45 years (SD ~17 years), prediabetes is 49 years (SD ~16 years), and those with type 2 diabetes is 57 years (SD ~12 years). Table 7 presents the study demographics across three glycemic categories and includes the demographic characteristics. The sex distribution was similar among men and women with prediabetes, however, women represented a larger portion of participants with type 2 diabetes compared to men.

Table 7. Study Characteristics of Full Sample			
N (%)			
Glycemic Categories	Normal FG	Prediabetes	Type 2 Diabetes
	442	254	641
Body Mass Index (BMI)			
Underweight	6 (1.4%)	-	2 (0.3%)
Healthy Weight	144 (32.6%)	41 (16.1%)	81 (12.6%)
Overweight	157 (35.5%)	89 (35.0%)	206 (32.1%)
Obese	123 (27.8%)	117 (46.1%)	315 (49.1%)
NA	12 (2.7%)	7 (2.8%)	37 (5.8%)
Age, yr (SD)			
	45.00 (16.98)	49.07 (15.94)	57.66 (12.75)
Race/Ethnicity			
European Ancestry	209 (47.3%)	106 (41.7%)	190 (29.6%)
Hispanic Ancestry	160 (36.2%)	115 (45.3%)	294 (45.9%)
African Ancestry	73 (16.5%)	33 (13.0%)	157 (24.5%)
Sex			
Men	154 (34.8%)	123 (48.4%)	237 (37.0%)
Women	288 (65.2%)	131 (51.6%)	404 (63.0%)
Smoking Status			
Never	259 (58.6%)	145 (57.1%)	321 (50.1%)
Ever (*Current & Former*)	183 (41.4%)	106 (41.7%)	319 (49.8%)
NA	-	3 (1.2%)	1 (0.1%)

Note: BMI categorizations (kg/m^2): underweight <18.5, healthy weight ≥18.5 to <25, overweight ≥25 to <30, obesity ≥30. FG: Fasting glucose. Table includes data on those with glycemic status.

Table 8. European Ancestry Study Characteristics			
N (%)			
Glycemic Categories	Normal FG	Prediabetes	Type 2 Diabetes
	209	106	190
Body Mass Index (BMI)			
Underweight	3 (1.43%)	-	-
Healthy Weight	73 (34.9%)	21 (19.8%)	24 (12.6%)
Overweight	76 (36.4%)	36 (34.0%)	64 (33.7%)
Obese	47 (22.5%)	44 (41.5%)	86 (45.3%)
Age, yr (SD)			
	43.2 (16.1)	48 (14.4)	62.1 (12.2)
Sex			
Men	76 (36.4%)	55 (52.0%)	90 (47.4%)
Women	133 (63.6%)	51 (48.0%)	100 (52.6%)
Smoking Status			
Never	135 (64.6%)	65 (61.3%)	101 (53.1%)
Current	31 (14.8%)	12 (11.3%)	15 (7.9%)
Former	43 (20.6%)	28 (26.4%)	74 (39.0%)

Note: BMI categorizations (kg/m^2): underweight <18.5, healthy weight ≥18.5 to <25, overweight ≥25 to <30, obesity ≥30.

Table 9. Hispanic Ancestry Study Characteristics			
N (%)			
Glycemic Categories	Normal FG	Prediabetes	Type 2 Diabetes
	160	115	294
Body Mass Index (BMI)			
Underweight	2 (1.3%)	0	0
Healthy Weight	53 (33.1%)	14 (12.2%)	41 (13.9%)
Overweight	54 (33.8%)	48 (41.7%)	94 (32.0%)
Obese	50 (31.3%)	53 (46.0%)	149 (50.7%)
Age, yr (SD)			
	46.7 (17.6)	50.0 (17.5)	55.9 (12.7)
Sex			
Men	48 (30.0%)	54 (47.0%)	108 (36.7%)
Women	112 (70.0%)	61 (53.0%)	186 (63.3%)
Smoking Status			
Never	93 (58.1%)	63 (54.8%)	158 (53.7%)
Current	40 (25.0%)	23 (20.0%)	53 (18.0%)
Former	27 (16.8%)	27 (23.5%)	83 (28.2%)

Note: BMI categorizations (kg/m^2): underweight <18.5, healthy weight ≥18.5 to <25, overweight ≥25 to <30, obesity ≥30.

Table 10. African Ancestry Study Characteristics			
N (%)			
Glycemic Categories	Normal FG	Prediabetes	Type 2 Diabetes
	73	33	157
Body Mass Index (BMI)			
Underweight	1 (1.4%)	-	2 (1.3%)
Healthy Weight	18 (24.7%)	6 (18.2%)	16 (10.2%)
Overweight	27 (37.0%)	5 (15.2%)	48 (30.6%)
Obese	26 (35.6%)	20 (60.6%)	80 (51.0%)
Age, yr (SD)			
	46.3 (17.8)	49.3 (15.0)	55.4 (12.2)
Sex			
Men	30 (41.1%)	14 (42.4%)	39 (24.8%)
Women	43 (58.9%)	19 (57.6%)	118 (75.2%)
Smoking Status			
Never	31 (42.5%)	17 (51.5%)	62 (39.5%)
Current	31 (42.5%)	8 (24.2%)	35 (22.3%)
Former	11 (15.1%)	8 (24.2%)	59 (37.6%)

Note: BMI categorizations (kg/m^2): underweight <18.5, healthy weight ≥18.5 to <25, overweight ≥25 to <30, obesity ≥30.

Genetic Association Results

Results of the genetic association analyses are presented separately by ancestry group. Within each ancestry group, primary category-specific comparisons are presented for prediabetes versus normal glycemia and type 2 diabetes versus normal glycemia, followed by the post-hoc comparison of prediabetes versus type 2 diabetes. Results from the three comparison groups are summarized using Venn diagrams to show SNPs with nominal evidence of association in more than one comparison group. Scatter plots are then used within each ancestry group to compare log-odds estimates for prediabetes versus normal glycemia and type 2 diabetes versus normal glycemia. Finally, cross-ancestry comparisons are presented using descriptive tables to identify

SNPs showing evidence of association in more than one ancestry group and figures to display SNPs with similar directions of effect across ancestries.

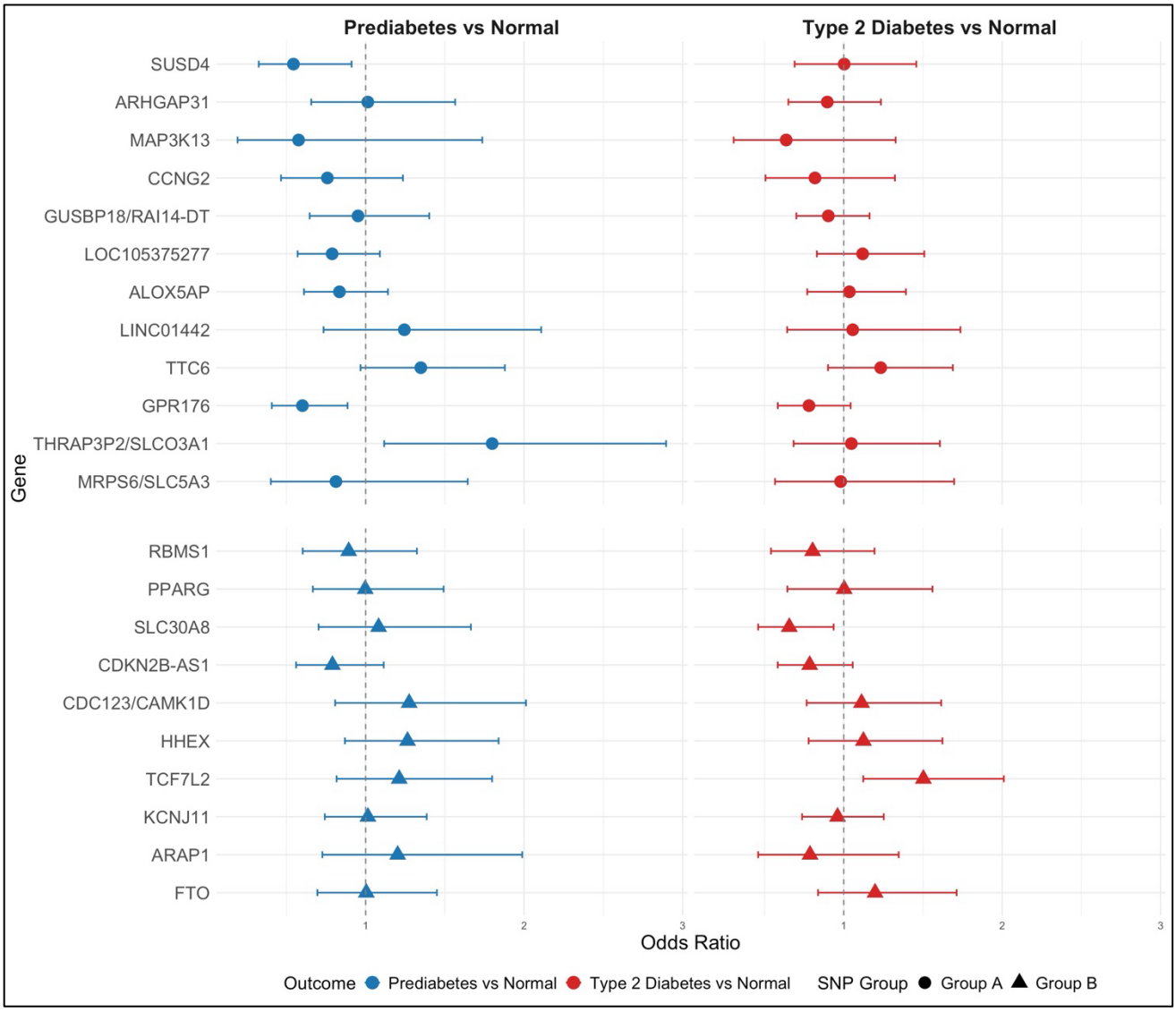

Figure 3. European Ancestry Candidate SNPs Forest Plot.

Note: The forest plot shows the comparison groups of prediabetes vs. normal glycemia and type 2 diabetes vs. normal glycemia side by side for European Ancestry. This plot displays the genes on the y-axis and their corresponding odds ratios and 95% confidence intervals on the x-axis.

Table 11. European Ancestry Candidate SNP Results							
Trait	**CHR**	**rsID**	**Gene**	**Prediabetes vs. Normal**		**Type 2 Diabetes vs. Normal**	
				OR (95% CI)	Unadj. P-Value	OR (95% CI)	Unadj. P-Value
Group A (Prediabetes SNPs)	1	rs4661250	SUSD4	**0.54 (0.33-0.91)**	**0.02**	1.00 (0.69-1.46)	0.99
	3	rs36087584	ARHGAP31	1.01 (0.66-1.56)	0.95	0.90 (0.65-1.23)	0.50
	3	rs498414	MAP3K13	0.58 (0.19-1.74)	0.33	0.64 (0.31-1.33)	0.23
	4	rs34938732	CCNG2	0.76 (0.47-1.24)	0.27	0.82 (0.51-1.32)	0.41
	5	rs6880621	GUSBP18/RAI14-DT	0.95 (0.65-1.40)	0.80	0.90 (0.70-1.16)	0.43
	7	rs10229340	LOC105375277/ENSG00000285741	0.79 (0.57-1.09)	0.15	1.12 (0.83-1.51)	0.46
	13	rs9550371	ALOX5AP	0.83 (0.61-1.14)	0.26	1.04 (0.77-1.39)	0.82
	13	rs4886290	LINC01442	1.24 (0.73-2.11)	0.42	1.06 (0.64-1.74)	0.83
	14	rs946911	TTC6	1.35 (0.97-1.88)	0.08	1.23 (0.90-1.69)	0.19
	15	rs41497851	GPR176	**0.60 (0.41-0.89)**	**0.01**	0.78 (0.58-1.04)	0.09
	15	rs11853125	THRAP3P2/SLCO3A1	**1.80 (1.12-2.90)**	**0.02**	1.05 (0.68-1.61)	0.83
	21	rs62212118	MRPS6/SLC5A3	0.81 (0.40-1.64)	0.56	0.98 (0.57-1.70)	0.94
Group B (Type 2 Diabetes SNPs)	2	rs7593730	RBMS1	0.89 (0.60-1.32)	0.57	0.80 (0.54-1.19)	0.28
	3	rs1801282	PPARG	1.00 (0.67-1.49)	0.99	1.00 (0.64-1.56)	0.99
	8	rs13266634	SLC30A8	1.08 (0.70-1.66)	0.72	**0.66 (0.46-0.94)**	**0.02**
	9	rs564398	CDKN2B-AS1	0.79 (0.56-1.11)	0.18	0.79 (0.58-1.06)	0.11
	10	rs12779790	CDC123/CAMK1D	1.27 (0.81-2.01)	0.30	1.11 (0.77-1.62)	0.58
	10	rs1111875	HHEX	1.26 (0.87-1.84)	0.22	1.12 (0.78-1.62)	0.53
	10	rs7903146	TCF7L2	1.21 (0.82-1.80)	0.34	**1.50 (1.12-2.01)**	**0.006**
	11	rs5219	KCNJ11	1.01 (0.74-1.39)	0.93	0.96 (0.74-1.25)	0.77
	11	rs1552224	ARAP1	1.20 (0.73-1.99)	0.47	0.79 (0.46-1.35)	0.38
	16	rs8050136	FTO	1.00 (0.70-1.45)	0.98	1.20 (0.84-1.71)	0.32

Note: Bolded values are significant prior to FDR correction; Purple outlined SNPs: Group A SNPs analyzed in the prediabetes group in GENNID; Orange outlined SNPs: Group A SNPs analyzed in T2D group; Blue outlined SNPs: Group B SNPs analyzed in prediabetes group; Green outlined SNPs: Group B SNPs analyzed in T2D group.

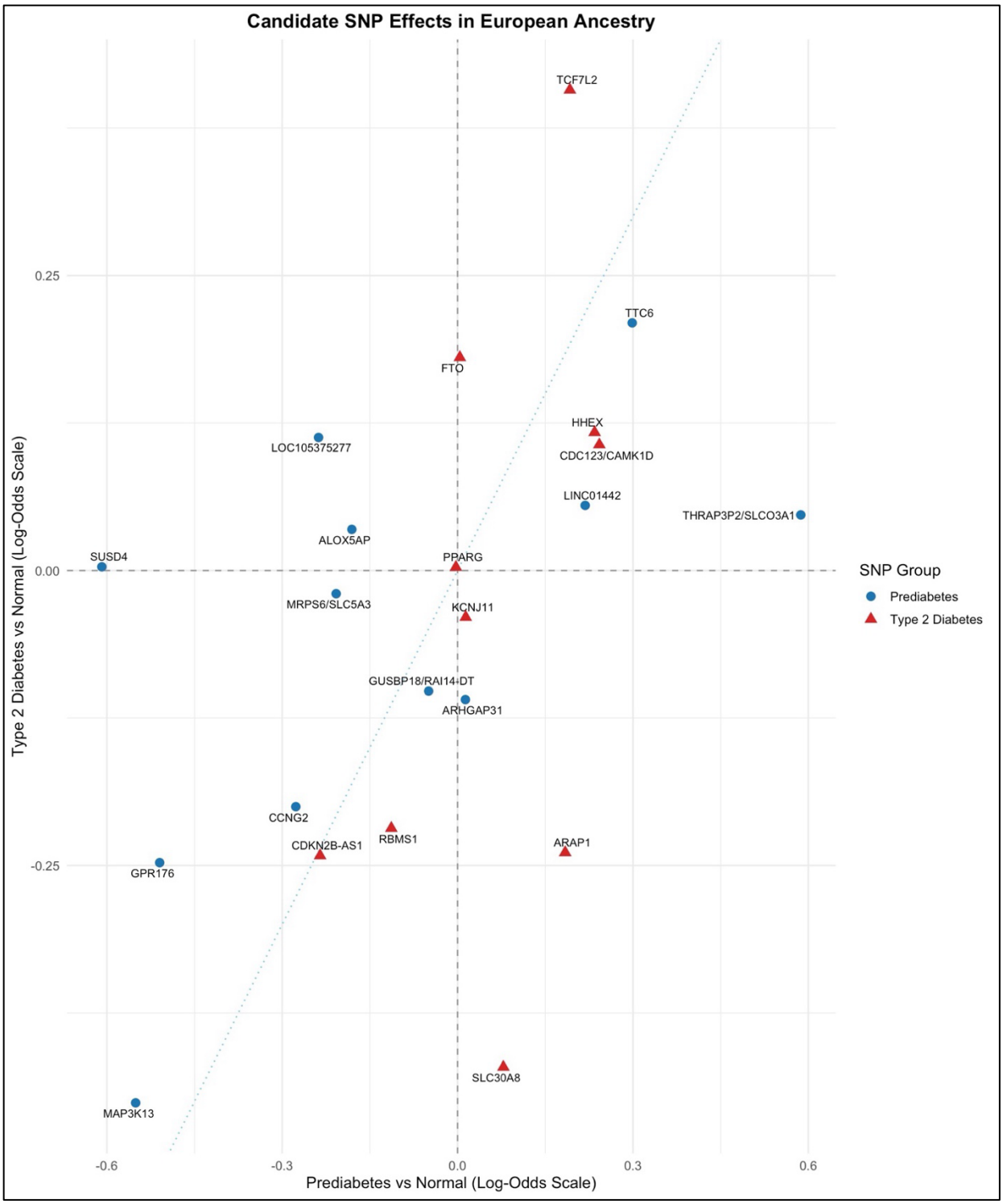

Figure 4. European Ancestry Scatter Plot of SNP Effects on a Log-Odds Scale

Note: The scatter plot displays SNP effects on a Log-Odds Scale for European Ancestry. SNPs in the upper right quadrant to be positive for both conditions (prediabetes & type 2 diabetes), the bottom right quadrant to be negative for both conditions, and the top left/bottom right to be in opposite directions.

European Ancestry Results

The multinomial candidate SNP analysis results for the European ancestry sample are shown in Table 11. In the comparison of prediabetes to normal glycemia, three Group A SNPs were found to be significant prior to FDR correction. The first SNP is rs4661250 on chromosome 1 near the *SUSD4* gene, which was associated with a 46% lower odds of prediabetes (P=0.02, OR: 0.54 [95% CI: 0.33, 0.91]). Prior work by Lin et al. reported the *SUSD4* (sushi domain containing 4) gene in relation to fasting plasma glucose among individuals with prediabetes.[37,77] Although the functional consequence of the rs4661250 SNP remains unknown, the *SUSD4* gene is involved in regulation of the complement system and innate immune response; however, the relevance of these functions to glycemic status has not been established.[78] The second SNP is rs41497851 on chromosome 15 near the *GPR176* gene, which was associated with a 40% lower odds of prediabetes (P=0.01, OR: 0.60 [95% CI: 0.41, 0.89]). The *GPR176* (g protein-coupled receptor 176) gene was identified only in one of the studies from which our SNP panel was derived, where it was associated with HbA1c and have not otherwise been associated with glycemic traits or related phenotypes in the literature.[37,79] The third SNP, rs11853125, is on chromosome 15 near the *THRAP3P2* and *SLCO3A1* genes and was associated with an 80% increase in the odds of prediabetes (P=0.02, OR: 1.80 [95% CI: 1.12, 2.90]). The *THRAP3P2* (thyroid hormone receptor associated protein 3 pseudogene) and *SLCO3A1* (solute carrier organic anion transporter) genes were identified in one of the studies from which our SNP panel was derived, where they showed association with HbA1c among prediabetics, and has not otherwise been associated with glycemic traits or related phenotypes in the literature.[37,80-81] No SNPs remained significant after FDR-based multiple testing correction.

In the comparison of type 2 diabetes to normal glycemia, two Group B SNPs were found to be significant prior to FDR adjustment. The first SNP is rs13266634 on chromosome 8 near the *SLC30A8* gene and was significantly associated with a 34% decrease in odds of type 2 diabetes (P=0.02, OR: 0.66 [95% CI: 0.46, 0.94]). The *SLC30A8* (solute carrier family 30 member 8) transports zinc molecules into insulin granules for insulin crystallization and storage; disruption of this process decreases insulin secretion.[82,83] The second SNP, rs7903146, is on chromosome 10 near the *TCF7L2* gene, which was significantly associated with a 50% increase in odds of type 2 diabetes (P=0.006, OR: 1.50 [95% CI: 1.12, 2.01]). The *TCF7L2* (transcription factor 7 like 2) is the most consistently replicated genetic variant for type 2 diabetes.[39-42,84] No SNPs remained significant in this category after FDR-based multiple testing correction.

The scatter plot of candidate SNP effects among European ancestry participants in Figure 4 provides a visual comparison of log-odds estimates for the prediabetes versus normal glycemia and type 2 diabetes versus normal glycemia groups. Overall, SNP association patterns varied across the two glycemic outcomes. Several SNPs showed estimates in the same direction for both outcomes, including *TCF7L2*, *HHEX*, *CDC123/CAMK1D*, *LINC01442*, *THRAP3P2/SLCO3A1,* and *TTC6*, which showed positive estimates, and *RBMS1*, *CDKN2B-AS1*, *GPR176*, *GUSBP18/RAI14-DT*, and *MAP3K13*, which showed negative estimates. However, certain SNPs showed estimates in the opposite directions for prediabetes and type 2 diabetes, including *SLC30A8* and *ARAP1*. These patterns suggest heterogeneity in candidate SNP association estimates across the two glycemic outcomes within European ancestry participants.

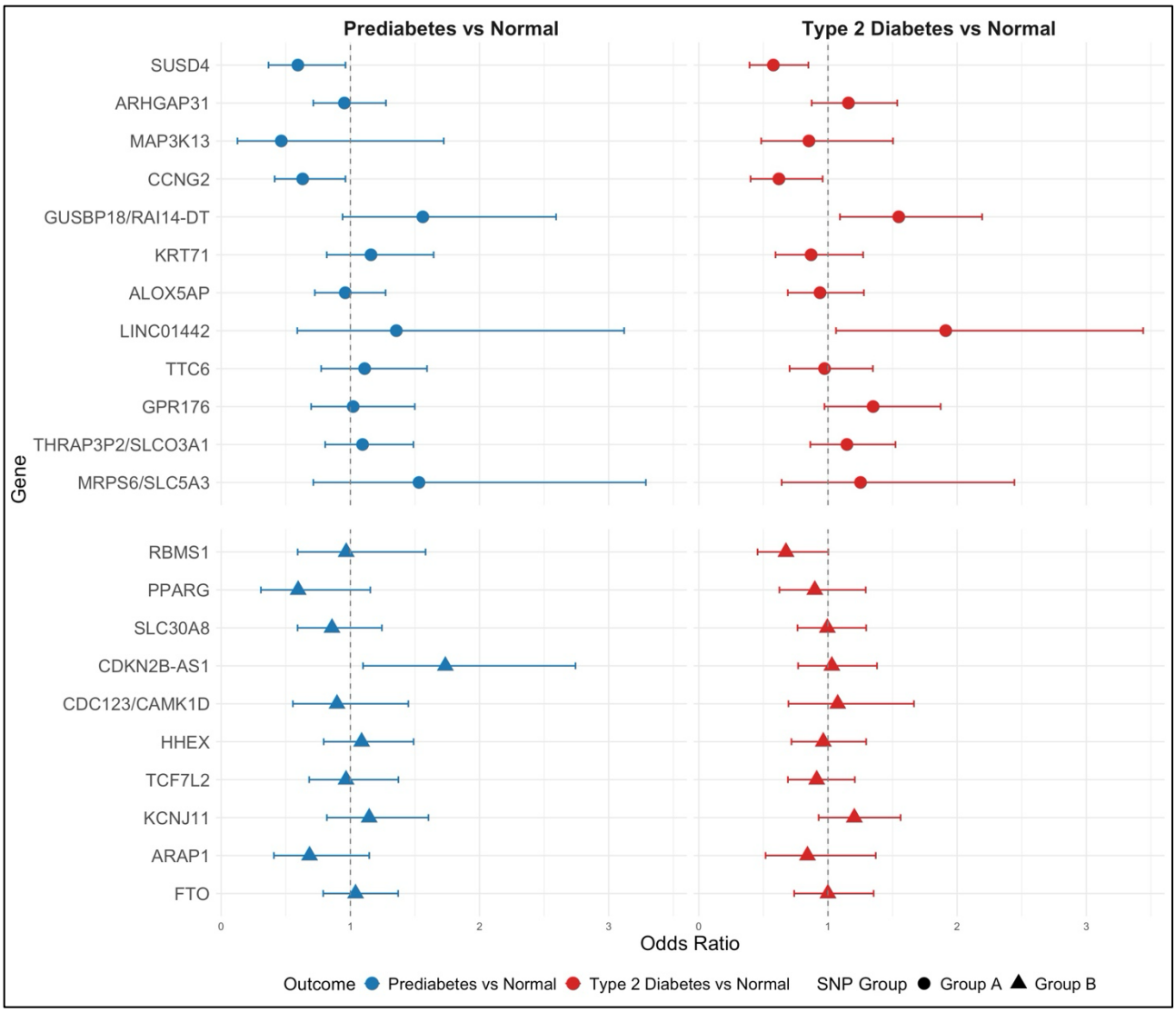

Figure 5. Hispanic Ancestry Candidate SNPs Forest Plot.

Note: The forest plot shows the comparison groups of prediabetes vs. normal glycemia and type 2 diabetes vs. normal glycemia side by side for Hispanic Ancestry. This plot displays the genes on the y-axis and their corresponding odds ratios and 95% confidence intervals on the x-axis.

Table 12. Hispanic Ancestry Candidate SNP Results							
Trait	**CHR**	**rsID**	**Gene**	**Prediabetes vs. Normal**		**Type 2 Diabetes vs. Normal**	
				OR (95% CI)	Unadj. P-Value	OR (95% CI)	Unadj. P-Value
Group A (Prediabetes SNPs)	1	rs4661250	SUSD4	**0.59 (0.37-0.96)**	**0.03**	**0.58 (0.39-0.85)**	**0.005**
	3	rs36087584	ARHGAP31	0.95 (0.71-1.28)	0.75	1.16 (0.87-1.54)	0.31
	3	rs498414	MAP3K13	0.47 (0.13-1.72)	0.25	0.85 (0.48-1.50)	0.58
	4	rs34938732	CCNG2	**0.63 (0.41-0.96)**	**0.03**	**0.62 (0.40-0.96)**	**0.03**
	5	rs6880621	GUSBP18/RAI14-DT	1.56 (0.94-2.59)	0.09	**1.55 (1.09-2.19)**	**0.01**
	12	rs3803084	KRT71	1.16 (0.82-1.64)	0.41	0.87 (0.59-1.27)	0.47
	13	rs9550371	ALOX5AP	0.96 (0.72-1.27)	0.78	0.94 (0.69-1.28)	0.69
	13	rs4886290	LINC01442	1.36 (0.59-3.12)	0.47	**1.91 (1.06-3.44)**	**0.03**
	14	rs946911	TTC6	1.11 (0.77-1.59)	0.57	0.97 (0.70-1.35)	0.87
	15	rs41497851	GPR176	1.02 (0.70-1.50)	0.91	1.35 (0.97-1.87)	0.07
	15	rs11853125	THRAP3P2/SLCO3A1	1.09 (0.81-1.49)	0.56	1.15 (0.86-1.52)	0.34
	21	rs62212118	MRPS6/SLC5A3	1.53 (0.71-3.29)	0.27	1.25 (0.64-2.44)	0.51
Group B (Type 2 Diabetes SNPs)	2	rs7593730	RBMS1	0.97 (0.59-1.58)	0.90	0.68 (0.45-1.00)	0.05
	3	rs1801282	PPARG	0.60 (0.31-1.15)	0.13	0.90 (0.62-1.29)	0.56
	8	rs13266634	SLC30A8	0.86 (0.59-1.24)	0.42	1.00 (0.76-1.30)	0.97
	9	rs564398	CDKN2B-AS1	**1.74 (1.10-2.74)**	**0.02**	1.03 (0.77-1.38)	0.84
	10	rs12779790	CDC123/CAMK1D	0.90 (0.56-1.45)	0.66	1.08 (0.69-1.67)	0.74
	10	rs1111875	HHEX	1.09 (0.79-1.49)	0.60	0.96 (0.72-1.30)	0.81
	10	rs7903146	TCF7L2	0.97 (0.68-1.37)	0.85	0.91 (0.69-1.21)	0.52
	11	rs5219	KCNJ11	1.15 (0.82-1.60)	0.43	1.20 (0.93-1.56)	0.16
	11	rs1552224	ARAP1	0.68 (0.41-1.15)	0.15	0.84 (0.52-1.37)	0.49
	16	rs8050136	FTO	1.04 (0.79-1.37)	0.78	1.00 (0.74-1.35)	0.99

Note: Bolded values are significant prior to FDR correction; Purple outlined SNPs: Group A SNPs analyzed in the Prediabetes group in GENNID; Orange outlined SNPs: Group A SNPs analyzed in T2D group; Blue outlined SNPs: Group B SNPs analyzed in Prediabetes group; Green outlined SNPs: Group B SNPs analyzed in T2D group.

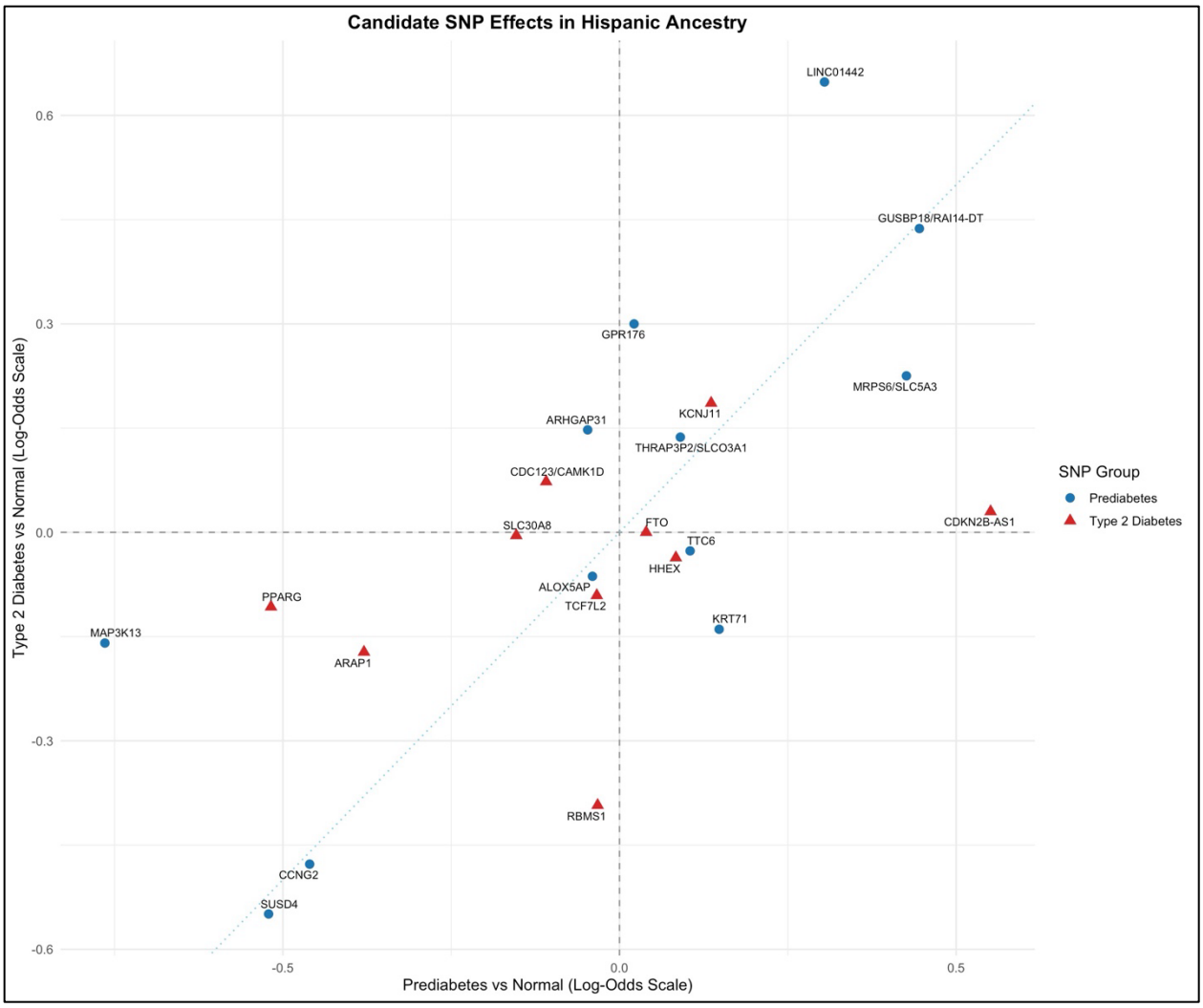

Figure 6. Hispanic Ancestry Scatter Plot of SNP Effects on a Log-Odds Scale.

Note: The scatter plot displays SNP effects on a Log-Odds Scale for Hispanic Ancestry. SNPs in the upper right quadrant to be positive for both conditions (prediabetes & type 2 diabetes), the bottom right quadrant to be negative for both conditions, and the top left/bottom right to be in opposite directions.

Hispanic Ancestry Results

The multinomial candidate SNP analysis results for the Hispanic ancestry are shown in Table 12. In the comparison of prediabetes to normal glycemia, two Group A SNPs (rs4661250 and rs34938732) and one Group B SNP (rs564398) were found to be significant prior to FDR correction. The first SNP, rs4661250 is on chromosome 1, near the *SUSD4* gene and is associated with a 41% decrease in odds of prediabetes (P=0.03, OR: 0.59 [95% CI: 0.37, 0.96]). The second SNP, rs34938732, is on chromosome 4 near the *CCNG2* gene and is associated with a 37% decrease in odds of prediabetes (P=0.03, OR: 0.63 [95% CI: 0.41, 0.96]). The *CCNG2* (cyclin g2) gene was identified in one of the studies from which our SNP panel was derived, which showed association with 2-hour postprandial glucose (2hPG).[37,85] This gene has otherwise not been associated with glycemic traits or related phenotypes in the literature. The third SNP, rs564398, is on chromosome 9 near the *CDKN2B-AS1* gene and is associated with a 74% increase in odds of prediabetes (P=0.02, OR: 1.74 [95% CI: 1.10, 2.74]). The *CDKN2B-AS1* gene is involved with type 2 diabetes risk and reduced beta-cell function.[39-41] However, these results did not remain significant after FDR-based multiple testing correction.

In the comparison of type 2 diabetes compared to normal glycemia, four Group A SNPs were significant prior to FDR adjustment. The first SNP is rs4661250 on chromosome 1 near the *SUSD4* gene and is associated with a 42% lower odds of type 2 diabetes (P=0.005, OR: 0.58 [95% CI: 0.39, 0.85]). The second SNP is rs34938732 on chromosome 4 near the *CCNG2* gene and is significantly associated with a 38% decrease in odds of type 2 diabetes (P=0.03, OR: 0.62 [95% CI: 0.40, 0.96]). The third SNP is rs6880621 on chromosome 5 near the *GUSBP18* and *RAI14-DT* genes and is associated with a 55% increase in odds of type 2 diabetes (P=0.01, OR: 1.55 [95% CI: 1.09, 2.19]). The *GUSBP18* (glucuronidase beta pseudogene 18) and *RAI14-DT*

(retinoic acid induced 14 divergent transcript) genes were identified in one of the studies from which our SNP panel was derived, which showed association with HbA1c.[37,86-87] These genes have otherwise not been associated with glycemic traits or related phenotypes in the literature. The fourth SNP is rs4886290 on chromosome 13 near the *LINC01442* gene and is associated with a 91% increase in odds of type 2 diabetes (P=0.03, OR: 1.91 [95% CI: 1.06, 3.44]). The *LINC01442* (long intergenic non-protein coding) gene was identified only in one of the studies from which our SNP panel was derived, which was associated with prediabetes status change, and has not otherwise been associated with glycemic traits or related phenotypes in the literature.[36,88] These results did not remain significant after FDR-based multiple testing correction.

The scatter plot of candidate SNP effects among Hispanic ancestry participants in Figure 6 shows overall variability in candidate SNP association patterns across prediabetes and type 2 diabetes. Several SNPs showed estimates in the same direction for both outcomes. The SNPs that showed to be positive for both outcomes include *LINC01442*, *GUSBP18/RAI14-DT*, *MRPS6/SLC5A3*, *THRAP3P2/SLCO3A1*, and *KCNJ11*; while the SNPs that showed to be negative estimates for both outcomes include *SUSD4*, *CCNG2*, *ARAP1*, and *MAP3K13*. Several SNPs with opposite-direction estimates were close to the null for one or more of the outcomes.

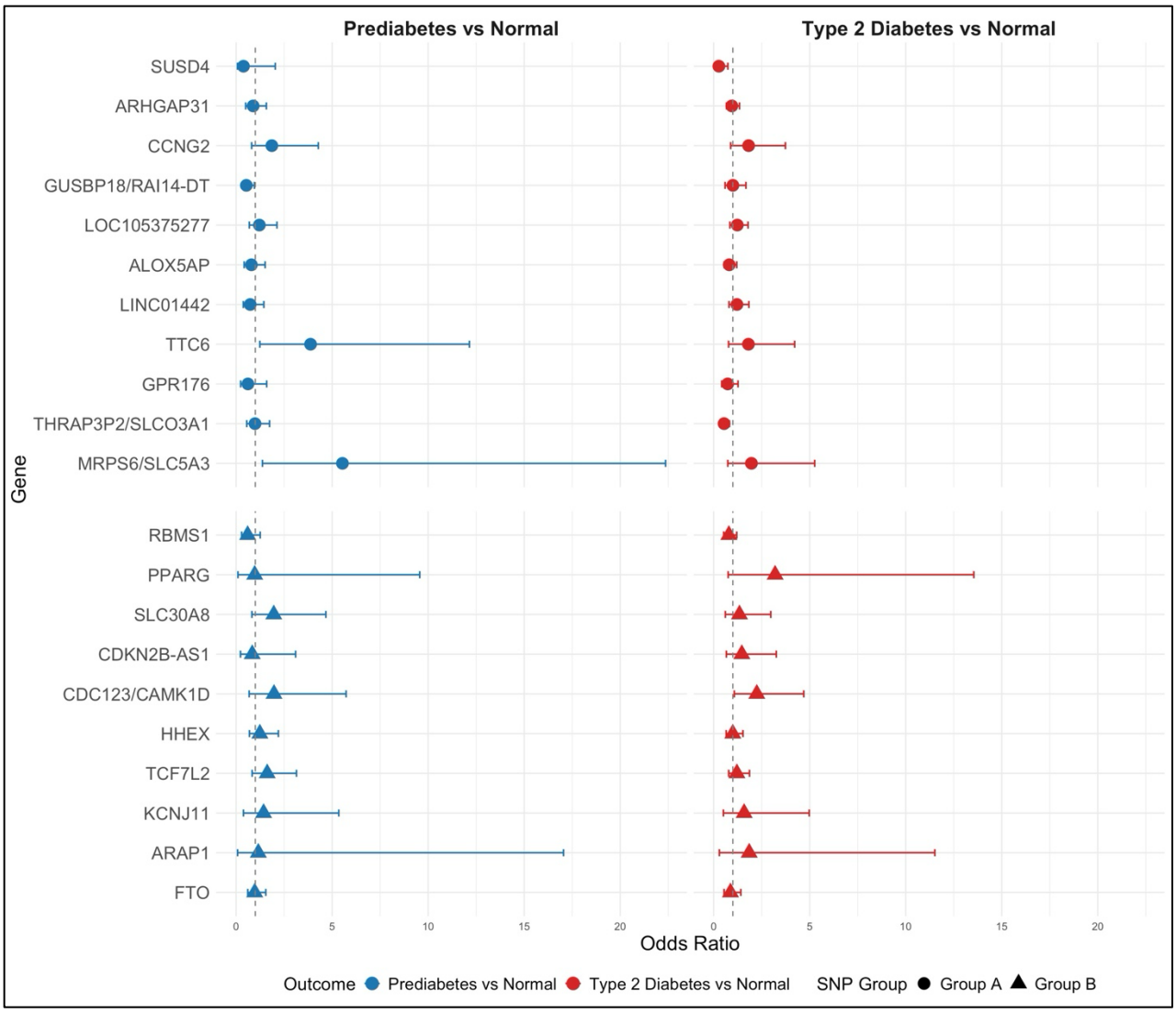

Figure 7. African Ancestry Candidate SNPs Forest Plot.

Note: The forest plot shows the comparison groups of prediabetes vs. normal glycemia and type 2 diabetes vs. normal glycemia side by side for African Ancestry. This plot displays the genes on the y-axis and their corresponding odds ratios and 95% confidence intervals on the x-axis.

Table 13. African Ancestry Candidate SNP Results							
Trait	**CHR**	**rsID**	**Gene**	**Prediabetes vs. Normal**		**Type 2 Diabetes vs. Normal**	
				OR (95% CI)	Unadj. P-value	OR (95% CI)	Unadj. P-value
Group A (Prediabetes SNPs)	1	rs4661250	SUSD4	0.38 (0.07-2.04)	0.26	**0.27 (0.10-0.74)**	**0.01**
	3	rs36087584	ARHGAP31	0.89 (0.50-1.58)	0.68	0.95 (0.67-1.35)	0.76
	4	rs34938732	CCNG2	1.86 (0.81-4.28)	0.14	1.82 (0.89-3.74)	0.10
	5	rs6880621	GUSBP18/RAI14-DT	**0.53 (0.30-0.95)**	**0.03**	1.00 (0.60-1.68)	1.00
	7	rs10229340	LOC105375277/ENSG00000285741	1.21 (0.69-2.13)	0.51	1.23 (0.85-1.79)	0.27
	13	rs9550371	ALOX5AP	0.80 (0.42-1.51)	0.49	0.81 (0.55-1.20)	0.30
	13	rs4886290	LINC01442	0.74 (0.38-1.44)	0.37	1.21 (0.80-1.83)	0.36
	14	rs946911	TTC6	**3.88 (1.24-12.15)**	**0.02**	1.81 (0.78-4.22)	0.17
	15	rs41497851	GPR176	0.62 (0.24-1.59)	0.32	0.73 (0.42-1.27)	0.27
	15	rs11853125	THRAP3P2/SLCO3A1	0.99 (0.56-1.74)	0.96	**0.54 (0.36-0.82)**	**0.004**
	21	rs62212118	MRPS6/SLC5A3	**5.54 (1.37-22.36)**	**0.02**	1.97 (0.74-5.26)	0.18
Group B (Type 2 Diabetes SNPs)	2	rs7593730	RBMS1	0.60 (0.29-1.26)	0.18	0.79 (0.53-1.20)	0.27
	3	rs1801282	PPARG	0.97 (0.10-9.56)	0.98	3.20 (0.76-13.55)	0.11
	8	rs13266634	SLC30A8	1.96 (0.82-4.68)	0.13	1.34 (0.61-2.97)	0.47
	9	rs564398	CDKN2B-AS1	0.84 (0.23-3.10)	0.79	1.47 (0.66-3.26)	0.35
	10	rs12779790	CDC123/CAMK1D	1.97 (0.68-5.73)	0.21	**2.25 (1.08-4.69)**	**0.03**
	10	rs1111875	HHEX	1.24 (0.70-2.20)	0.47	1.00 (0.65-1.53)	0.99
	10	rs7903146	TCF7L2	1.62 (0.84-3.14)	0.15	1.21 (0.79-1.87)	0.38
	11	rs5219	KCNJ11	1.43 (0.38-5.35)	0.59	1.59 (0.51-4.98)	0.43
	11	rs1552224	ARAP1	1.16 (0.08-17.05)	0.92	1.85 (0.30-11.52)	0.51
	16	rs8050136	FTO	0.97 (0.61-1.55)	0.90	0.88 (0.54-1.42)	0.59

Bolded values are significant prior to FDR correction; Purple outlined SNPs: Group A SNPs analyzed in the Prediabetes group in GENNID; Orange outlined SNPs: Group A SNPs analyzed in T2D group; Blue outlined SNPs: Group B SNPs analyzed in Prediabetes group; Green outlined SNPs: Group B SNPs analyzed in T2D group.

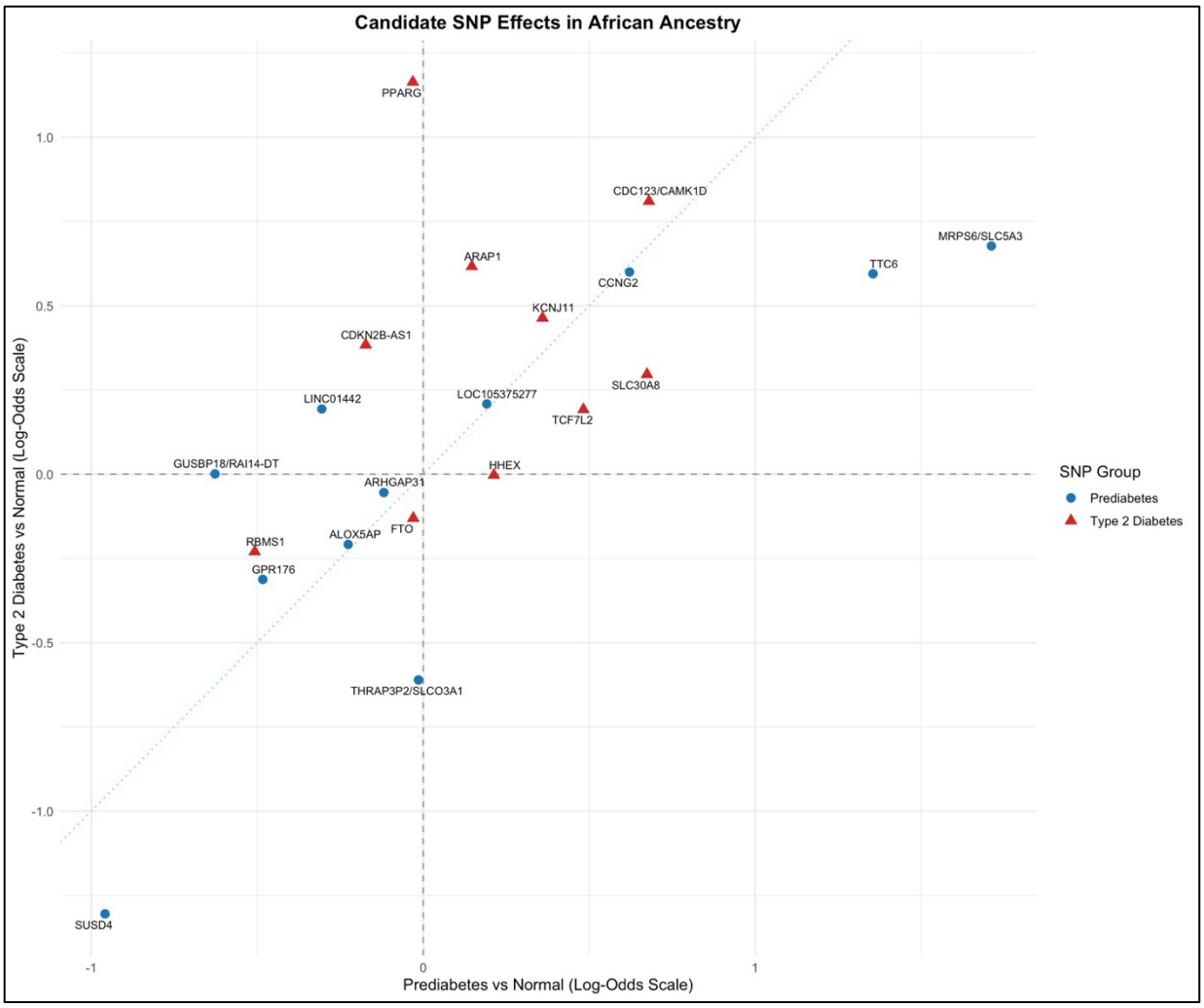

Figure 8. African Ancestry Scatter Plot of SNP Effects on a Log-Odds Scale.

Note: The scatter plot displays SNP effects on a Log-Odds Scale for African Ancestry. SNPs in the upper right quadrant to be positive for both conditions (prediabetes & type 2 diabetes), the bottom right quadrant to be negative for both conditions, and the top left/bottom right to be in opposite directions.

African Ancestry Results

The multinomial candidate SNP analysis results for African ancestry are shown in Table 13. In the comparison of prediabetes to normal glycemia, three Group A SNPs were found to be significant prior to FDR correction. The first SNP is rs6880621 on chromosome 5 and located within the *GUSBP18* and *RAI14-DT* genes and is significantly associated with a 47% decrease in odds of prediabetes (P=0.03, OR: 0.53 [95% CI: 0.30, 0.95]). The second SNP is rs946911 on chromosome 14 and located near the *TTC6* gene and is significantly associated with a 288% increase in odds of prediabetes (P=0.02, OR: 3.88 [95% CI: 1.24, 12.15]). The TTC6 (tetratricopeptide repeat domain 6) gene was identified in one of the studies from which our SNP panel was derived, which showed an association with HbA1c.[37,89] The third SNP is rs62212118 on chromosome 21 and located near the *MRPS6* and *SLC5A3* genes and is significantly associated with a 454% increase in odds of prediabetes (P=0.02, OR: 5.54 [95% CI: 1.37, 22.36]). The *MRPS6* (mitochondrial ribosomal protein s6) and *SLC5A3* (solute carrier family 5 member 3) marker rs62212118 was associated with 2-hour postprandial glucose in discovery GWAS, and a follow-up functional study subsequently confirmed that MRPS6 (not SLC5A3) regulates glucose-stimulated insulin secretion in β-cells via modulation of the mitochondrial unfolded protein response.[37,90-92] No SNPs remained significant in this category after FDR-based multiple testing correction.

In the comparison of type 2 diabetes to normal glycemia, two Group A SNPs (rs4661250 and rs11853125) and one Group B SNP (rs12779790) were found to be significant prior to FDR adjustment. The first SNP is the rs4661250 on chromosome 1 near the *SUSD4* gene, which was significantly associated with a 73% lower odds of type 2 diabetes (P=0.01, OR: 0.27 [95% CI: 0.10, 0.74]). The second SNP is the rs11853125 on chromosome 15 near the *THRAP3P2* and

SLCO3A1 genes, which was significantly associated with a 46% lower odds of type 2 diabetes (P=0.004, OR: 0.54 [95% CI: 0.36, 0.82]). The third SNP is the rs12779790 SNP on chromosome 10 near the *CDC123/CAMK1D* genes, which was significantly associated with a 125% increase in odds of type 2 diabetes (P=0.03, OR: 2.25 [95% CI: 1.08, 4.69]). The *CDC123* (cell division cycle 123) and *CAMK1D* (calcium/calmodulin dependent protein kinase 1D) genes are associated with type 2 diabetes risk and are involved in reduced glucose-stimulated insulin secretion.[39,93-94] However, no SNPs remained significant in this category after FDR-based multiple testing correction.

Among African ancestry participants, the scatter plot of prediabetes versus normal and type 2 diabetes versus normal on a log-odds scale in Figure 8 was interpreted cautiously due to low sample size and large confidence intervals. Several SNPs showed estimates in the same direction for prediabetes and type 2 diabetes. SNPs that showed positive estimates for both outcomes included *MRPS6/SLC5A3*, *TTC6*, *CDC123/CAMK1D*, *KCNJ11*, *CCNG2*, *TCF7L2*, and *SLC30A8*; while the SNPs that showed negative estimates for both outcomes included *SUSD4*, *RBMS1*, *GPR176*, and *ALOX5AP*. Several SNPs with opposite-direction estimates were close to the null for one or more of the outcomes.

Cross-Ancestry Results

Findings from the ancestry-specific analyses were compared across the racial/ethnic groups to identify SNPs showing evidence of association in more than one ancestry-specific analysis among the three outcome comparison groups. Table 14 presents three SNPs from Group A that showed evidence of association prior to FDR adjustment in two or more ancestry groups

among the three comparison groups based on the 95% CI. This pattern was not detected in Group B SNPs.

Table 14. Cross-Ancestry Results							
Trait	**CHR**	**Pos37**	**rsID**	**Gene**	**ANC**	**Prediabetes vs. Normal**	**Type 2 Diabetes vs. Normal**
Group A (Prediabetes SNPs)	1	223399328	rs4661250	SUSD4	EA	**0.54 (0.33-0.91)**	1.00 (0.69-1.46)
					HA	**0.59 (0.37-0.96)**	**0.58 (0.39-0.85)**
					AA	0.38 (0.07-2.04)	**0.27 (0.10-0.74)**
	5	34504382	rs6880621	GUSBP18/ RAI14-DT	HA	1.56 (0.94-2.59)	**1.55 (1.09-2.19)**
					AA	**0.53 (0.30-0.95)**	1.00 (0.60-1.68)
	15	92288156	rs11853125	THRAP3P2/ SLCO3A1	EA	**1.80 (1.12-2.90)**	1.05 (0.68-1.61)
					AA	0.99 (0.56-1.74)	**0.54 (0.36-0.82)**
CHR: Chromosome, Pos37: Genome Reference Human Build 37, ANC: ancestry, Bolded OR (95% CI) are significant prior to FDR-adj.							

Gene by Smoking Interaction Results

Results of the gene-by-smoking environment test are presented separately by ancestry group for the primary category-specific comparisons: prediabetes vs. normal glycemia and type 2 diabetes vs. normal glycemia. The interaction term evaluates whether the association between each SNP and glycemic status differs by smoking status (ever smokers and never smokers).

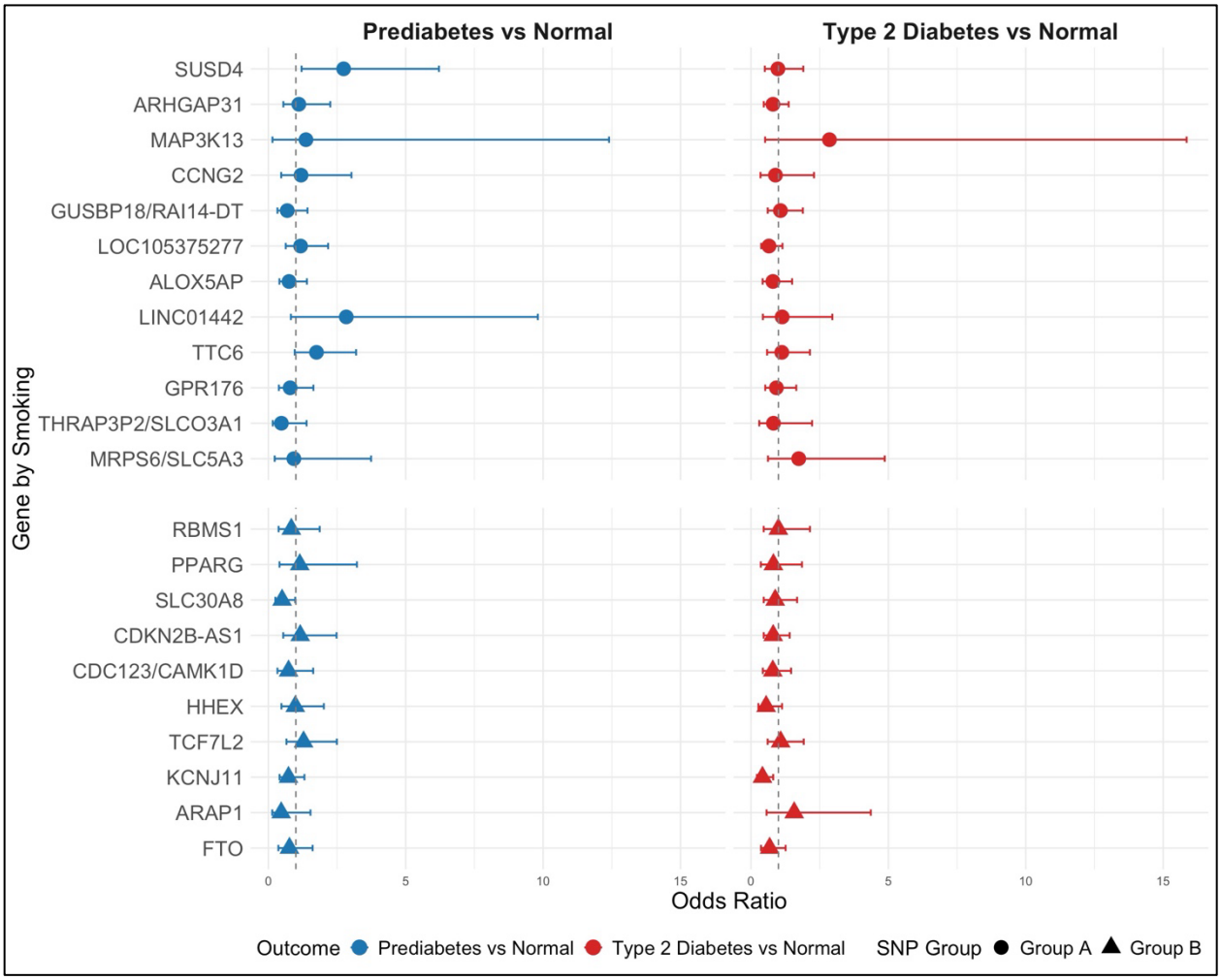

Figure 9. European Ancestry Candidate SNPs Interaction Forest Plot.

Note: The forest plot illustrates the comparison groups of prediabetes vs. normal glycemia and type 2 diabetes vs. normal glycemia side by side for European Ancestry. This plot displays the gene by smoking interaction on the y-axis and their corresponding odds ratios and 95% confidence intervals on the x-axis.

Table 15. European Ancestry Candidate SNP by Smoking Interaction Results							
Trait	**CHR**	**rsID**	**Gene**	**Prediabetes vs. Normal**		**Type 2 Diabetes vs. Normal**	
SNP by Smoking				OR (95% CI)	Unadj. P-value	OR (95% CI)	Unadj. P-value
Group A (Prediabetes SNPs)	1	rs4661250	SUSD4	**2.74 (1.21-6.21)**	**0.02**	0.98 (0.50-1.90)	0.95
	3	rs36087584	ARHGAP31	1.11 (0.55-2.25)	0.77	0.80 (0.47-1.37)	0.42
	3	rs498414	MAP3K13	1.37 (0.15-12.40)	0.78	2.86 (0.51-15.86)	0.23
	4	rs34938732	CCNG2	1.19 (0.47-3.02)	0.72	0.89 (0.35-2.29)	0.81
	5	rs6880621	GUSBP18/RAI14-DT	0.69 (0.33-1.42)	0.31	1.07 (0.61-1.89)	0.81
	7	rs10229340	LOC105375277/ENSG00000285741	1.17 (0.63-2.17)	0.61	0.65 (0.37-1.15)	0.14
	13	rs9550371	ALOX5AP	0.75 (0.40-1.40)	0.37	0.80 (0.43-1.50)	0.48
	13	rs4886290	LINC01442	2.84 (0.82-9.81)	0.10	1.13 (0.43-2.96)	0.80
	14	rs946911	TTC6	1.75 (0.96-3.19)	0.07	1.12 (0.59-2.14)	0.73
	15	rs41497851	GPR176	0.79 (0.38-1.64)	0.53	0.92 (0.52-1.65)	0.79
	15	rs11853125	THRAP3P2/SLCO3A1	0.47 (0.16-1.39)	0.17	0.82 (0.30-2.22)	0.69
	21	rs62212118	MRPS6/SLC5A3	0.93 (0.23-3.74)	0.92	1.74 (0.62-4.87)	0.29
Group B (Type 2 Diabetes SNPs)	2	rs7593730	RBMS1	0.84 (0.37-1.87)	0.66	1.00 (0.46-2.15)	0.99
	3	rs1801282	PPARG	1.14 (0.41-3.22)	0.80	0.82 (0.36-1.86)	0.63
	8	rs13266634	SLC30A8	**0.50 (0.26-0.98)**	**0.04**	0.88 (0.46-1.68)	0.70
	9	rs564398	CDKN2B-AS1	1.16 (0.54-2.48)	0.70	0.81 (0.46-1.41)	0.45
	10	rs12779790	CDC123/CAMK1D	0.74 (0.33-1.63)	0.45	0.79 (0.43-1.46)	0.46
	10	rs1111875	HHEX	0.98 (0.48-2.02)	0.96	0.55 (0.27-1.13)	0.10
	10	rs7903146	TCF7L2	1.28 (0.66-2.49)	0.46	1.08 (0.61-1.92)	0.79
	11	rs5219	KCNJ11	0.73 (0.41-1.31)	0.30	**0.42 (0.22-0.81)**	**0.01**
	11	rs1552224	ARAP1	0.47 (0.14-1.53)	0.21	1.57 (0.57-4.36)	0.38
	16	rs8050136	FTO	0.77 (0.37-1.61)	0.48	0.68 (0.37-1.26)	0.22

Bolded values are significant prior to FDR correction; Purple outlined SNPs: Group A SNPs analyzed in the prediabetes group in GENNID; Orange outlined SNPs: Group A SNPs analyzed in type 2 diabetes group; Blue outlined SNPs: Group B SNPs analyzed in prediabetes group; Green outlined SNPs: Group B SNPs analyzed in type 2 diabetes group.

Table 16. Stratum-Specific SNP Odds Ratios by Smoking Status for Dominant Models in European Ancestry							
SNP/GENE	**Smoking Status**	**0 G Copies: Normal**	**0 G Copies: Prediabetes**	**1+ G Copies: Normal**	**1+ G Copies: Prediabetes**	**SNP OR: Prediabetes vs Normal**	**Interaction P-Value**
rs4661250/ SUSD4	Never Smokers	102	59	33	6	0.32 (0.14-0.76)	0.02
	Ever Smokers	54	30	20	10	0.95 (0.46-1.95)	
SNP/GENE	**Smoking Status**	**0 T Copies: Normal**	**0 T Copies: Prediabetes**	**1+ T Copies: Normal**	**1+ T Copies: Prediabetes**	**SNP OR: Prediabetes vs Normal**	**Interaction P-Value**
rs13266634/ SLC30A8	Never Smokers	70	30	65	35	1.39 (0.83-2.33)	0.04
	Ever Smokers	31	21	43	19	0.69 (0.40-1.21)	
Note: Counts represent observed frequencies by genotype group, smoking status, and glycemic category. Odds ratios, 95% confidence intervals, and the SNP-by-smoking interaction p-values presented derive from the original GxE interaction model.							

Table 17. Stratum-Specific SNP Odds Ratios by Smoking Status for Additive Models in European Ancestry									
SNP/GENE	**Smoking Status**	**0 T Copies: Normal**	**0 T Copies: Type 2 Diabetes**	**1 T Copy: Normal**	**1 T Copy: Type 2 Diabetes**	**2 T Copies: Normal**	**2 T Copies: Type 2 Diabetes**	**SNP OR: Type 2 Diabetes vs Normal**	**Interaction P-Value**
rs5219/ KCNJ11	Never Smokers	58	38	64	43	13	20	1.42 (0.96-2.12)	0.01
	Ever Smokers	29	53	29	24	16	12	0.59 (0.37-0.96)	
Note: Counts represent observed frequencies by genotype group, smoking status, and glycemic category. Odds ratios, 95% confidence intervals, and the SNP-by-smoking interaction p-values presented derive from the original GxE interaction model.									

Results for European Ancestry Interaction

The multinomial candidate SNP interaction analysis results for European ancestry assessed 22 SNPs. In the comparison of prediabetes to normal glycemia, one Group A SNP (rs4661250) by smoking interaction and one Group B SNP (rs13266634) by smoking interaction were found to be significant prior to FDR-based multiple testing correction. The first SNP by smoking interaction found to be significant is rs4661250 on chromosome 1 near the *SUSD4* gene (P=0.02, OR: 2.74 [95% CI: 1.21, 6.21]). The second SNP by smoking interaction is rs13266634 on chromosome 8 near the *SLC30A8* gene (P=0.04, OR: 0.50 [95% CI: 0.26, 0.98]). No SNPs remained significant after FDR-based multiple testing correction. Results for the 22 SNPs are outlined in Table 15.

The multinomial candidate SNP interaction analysis results for European Ancestry assessing type 2 diabetes compared to normal glycemia identified one Group B SNP by smoking interaction to be significant prior to FDR correction. The SNP is rs5219 on chromosome 11, which is located within the *KCNJ11* gene (P=0.01, OR: 0.42 [95% CI: 0.22, 0.81]). The *KCNJ11* (potassium inwardly rectifying channel subfamily J member 11) gene plays a role in pancreatic β-cell function and encodes a subunit of the ATP-sensitive potassium channel that triggers insulin release.[39,95] No SNPs remained significant after FDR-based multiple testing correction.

To interpret the SNP-by-smoking interaction model results, stratum-specific SNP odds ratios derived from the original multinomial models were used to describe whether the association between SNPs and glycemic status differed by smoking status. These estimates derived from the variance-covariance matrix and p-value estimates of the original model compare the association between the SNP and glycemic status within each smoking stratum

(ever smokers and never smokers) and are presented for those SNPs that showed nominal evidence of SNP-by-smoking interaction prior to FDR correction, as shown in Table 16 and 17. For rs4661250 near *SUSD4*, the SNP OR for prediabetes versus normal glycemia was 0.32 among never smokers (95% CI: 0.14, 0.76) and 0.95 among ever smokers (95% CI: 0.46, 1.95), with an interaction p-value of 0.02. The interaction p-value suggests that the SNP-prediabetes association differed by smoking status. The stratum-specific 95% CIs suggests that the SNP-prediabetes association was suggestively significant among never smokers, but not among ever smokers. For rs13266634 in *SLC30A8*, the SNP OR for prediabetes versus normal glycemia was 1.29 among never smokers (95% CI: 0.83, 2.33) and 0.69 among ever smokers (95% CI: 0.40, 1.21), with an interaction p-value of 0.04. The interaction p-value suggests that the SNP-prediabetes association differed by smoking status. The stratum-specific 95% confidence intervals suggest that there is no association between SNP-prediabetes within ever smokers and never smokers. For rs5219 in *KCNJ11*, the SNP OR for type 2 diabetes versus normal glycemia was 1.42 among never smokers (95% CI: 0.96, 2.12) and 0.59 among ever smokers (95% CI: 0.37, 0.96), with an interaction p-value of 0.01. The interaction p-value suggests that the SNP-type 2 diabetes association differed by smoking status. The stratum-specific 95% confidence interval suggest there is an association between SNP-type 2 diabetes among ever smokers, but not among never smokers.

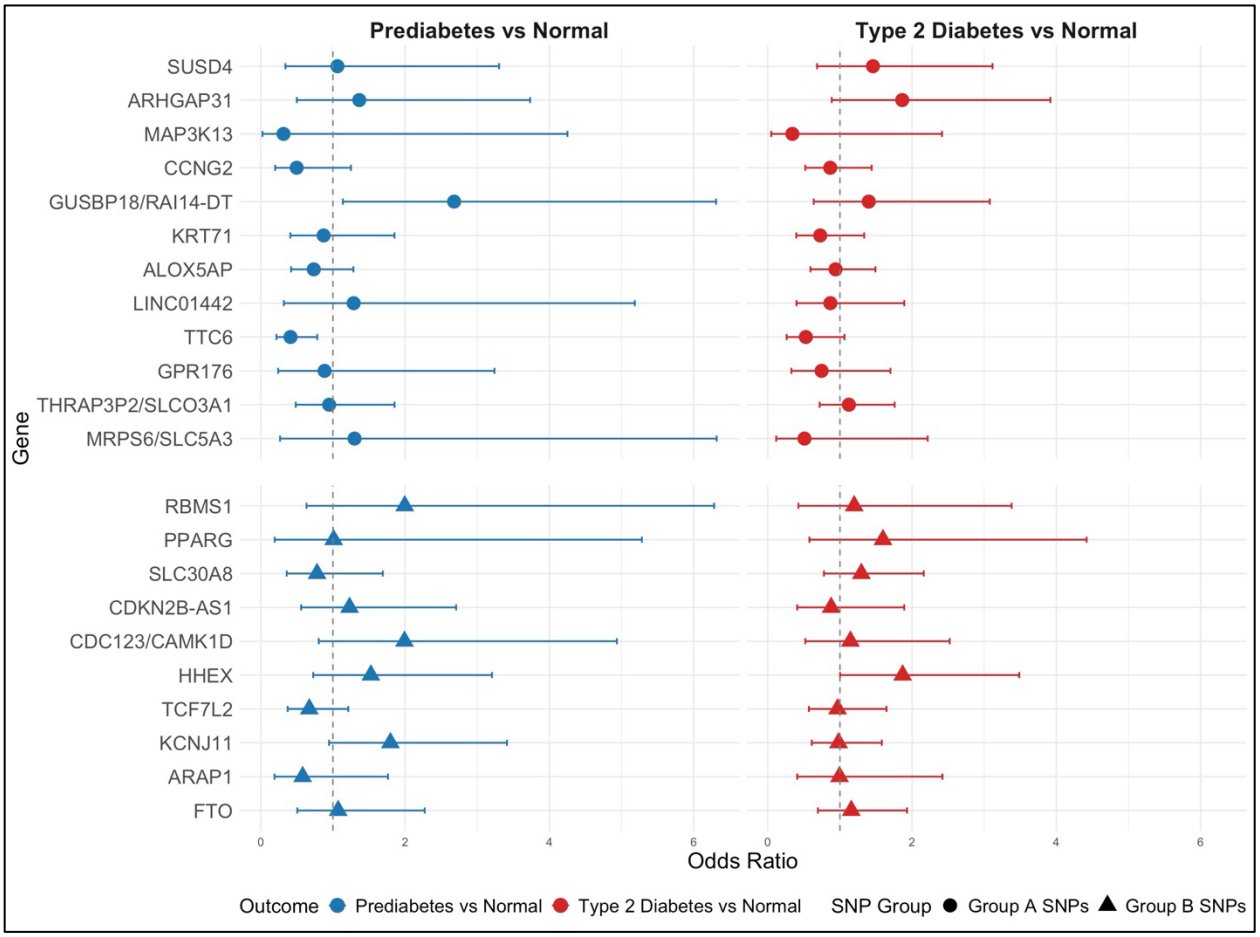

Figure 10. Hispanic Ancestry Candidate SNPs Interaction Forest Plot.

Note: The forest plot illustrates the comparison groups of prediabetes vs. normal glycemia and type 2 diabetes vs. normal glycemia side by side for Hispanic Ancestry. This plot displays the gene-by-smoking interaction on the y-axis and their corresponding odds ratios and 95% confidence intervals on the x-axis.

Table 18. Hispanic Ancestry Candidate SNP by Smoking Interaction Results							
Trait	**CHR**	**rsID**	**Gene**	**Prediabetes vs. Normal**		**Type 2 Diabetes vs. Normal**	
SNP by Smoking				OR (95% CI)	Unadj. P-Value	OR (95% CI)	Unadj. P-Value
Group A (Prediabetes SNPs)	1	rs4661250	SUSD4	1.06 (0.34-3.30)	0.92	1.46 (0.68-3.12)	0.33
	3	rs36087584	ARHGAP31	1.37 (0.50-3.73)	0.54	1.86 (0.89-3.92)	0.10
	3	rs498414	MAP3K13	0.32 (0.02-4.25)	0.38	0.34 (0.05-2.41)	0.28
	4	rs34938732	CCNG2	0.50 (0.20-1.25)	0.14	0.87 (0.52-1.44)	0.58
	5	rs6880621	GUSBP18/RAI14-DT	**2.68 (1.14-6.31)**	**0.02**	1.40 (0.64-3.08)	0.40
	12	rs3803084	KRT71	0.87 (0.41-1.85)	0.72	0.73 (0.40-1.34)	0.30
	13	rs9550371	ALOX5AP	0.73 (0.42-1.28)	0.28	0.94 (0.59-1.49)	0.79
	13	rs4886290	LINC01442	1.29 (0.32-5.19)	0.72	0.87 (0.40-1.89)	0.72
	14	rs946911	TTC6	**0.41 (0.22-0.78)**	**0.007**	0.53 (0.26-1.06)	0.07
	15	rs41497851	GPR176	0.88 (0.24-3.24)	0.85	0.75 (0.33-1.70)	0.49
	15	rs11853125	THRAP3P2/SLCO3A1	0.95 (0.48-1.85)	0.88	1.12 (0.72-1.76)	0.61
	21	rs62212118	MRPS6/SLC5A3	1.30 (0.27-6.32)	0.75	0.51 (0.12-2.22)	0.37
Group B (Type 2 Diabetes SNPs)	2	rs7593730	RBMS1	2.00 (0.63-6.28)	0.24	1.20 (0.42-3.38)	0.73
	3	rs1801282	PPARG	1.01 (0.19-5.28)	0.99	1.60 (0.58-4.42)	0.37
	8	rs13266634	SLC30A8	0.78 (0.36-1.69)	0.53	1.30 (0.78-2.16)	0.32
	9	rs564398	CDKN2B-AS1	1.23 (0.56-2.71)	0.61	0.88 (0.41-1.89)	0.74
	10	rs12779790	CDC123/CAMK1D	1.99 (0.80-4.94)	0.14	1.15 (0.52-2.52)	0.74
	10	rs1111875	HHEX	1.53 (0.73-3.21)	0.26	1.87 (1.00-3.49)	0.05
	10	rs7903146	TCF7L2	0.67 (0.37-1.21)	0.19	0.97 (0.57-1.65)	0.90
	11	rs5219	KCNJ11	1.80 (0.95-3.41)	0.07	0.98 (0.61-1.58)	0.94
	11	rs1552224	ARAP1	0.58 (0.19-1.76)	0.34	0.99 (0.41-2.42)	0.99
	16	rs8050136	FTO	1.07 (0.51-2.27)	0.85	1.16 (0.69-1.93)	0.57

Bolded values are significant prior to FDR correction; Purple outlined SNPs: Group A SNPs analyzed in the prediabetes group in GENNID; Orange outlined SNPs: Group A SNPs analyzed in type 2 diabetes group; Blue outlined SNPs: Group B SNPs analyzed in prediabetes group; Green outlined SNPs: Group B SNPs analyzed in type 2 diabetes group.

Table 19. Stratum-Specific SNP Odds Ratios by Smoking Status for Dominant Models in Hispanic Ancestry							
SNP/GENE	**Smoking Status**	**0 G Copies: Normal**	**0 G Copies: Prediabetes**	**1+ G Copies: Normal**	**1+ G Copies: Prediabetes**	**SNP OR: Prediabetes vs Normal**	**Interaction P-Value**
rs6880621/ GUSBP18/ RAI14-DT	Never Smokers	38	25	55	38	1.04 (0.52-2.08)	0.02
	Ever Smokers	37	15	30	35	2.79 (1.49-5.25)	
Note: Counts represent observed frequencies by genotype group, smoking status, and glycemic category. Odds ratios, 95% confidence intervals, and the SNP-by-smoking interaction p-values presented derive from the original GxE interaction model.							

Table 20. Stratum-Specific SNP Odds Ratios by Smoking Status for Additive Models in Hispanic Ancestry									
SNP/GENE	**Smoking Status**	**0 C Copies: Normal**	**0 C Copies: Prediabetes**	**1 C Copy: Normal**	**1 C Copy: Prediabetes**	**2 C Copies: Normal**	**2 C Copies: Prediabetes**	**SNP OR: Prediabetes vs Normal**	**Interaction P-Value**
rs946911/ TTC6	Never Smokers	65	34	21	24	7	5	1.52 (1.00-2.32)	0.006
	Ever Smokers	35	33	27	14	5	3	0.63 (0.37-1.06)	
Note: Counts represent observed frequencies by genotype group, smoking status, and glycemic category. Odds ratios, 95% confidence intervals, and the SNP-by-smoking interaction p-values presented derive from the original GxE interaction model.									

Results for Hispanic Ancestry Interaction

The multinomial candidate SNP interaction analysis results for Hispanic ancestry assessed 22 SNPs. In the comparison of prediabetes to normal glycemia, two Group A SNP by smoking interactions were found to be significant prior to FDR-based multiple testing correction. The first SNP by smoking interaction was rs6880621 on chromosome 5 near the *GUSBP18/RAI14-DT* gene (P=0.02, OR: 2.68 [95% CI: 1.14, 6.31]). The second SNP by smoking interaction is rs946911 on chromosome 14 near the *TTC6* gene (P=0.007, OR: 0.41 [95% CI: 0.22, 0.78]). No SNPs remained significant after FDR-based multiple testing correction. Results for the 22 SNPs analyzed are outlined in Table 18.

The multinomial candidate SNP interaction analysis results for Hispanic ancestry assessing type 2 diabetes to normal glycemia did not identify any significant SNP by smoking interactions.

The SNP-by-smoking stratum-specific odds ratio, confidence interval, and p-value from the interaction model results are shown in Tables 19 and 20. For rs6880621 near *GUSBP18/RAI14-DT*, the SNP OR for prediabetes versus normal glycemia was 1.04 among never smokers (95% CI:0.52, 2.08) and 2.79 among ever smokers (95% CI: 1.49, 5.25), with an interaction p-value of 0.02. The interaction p-value suggests that the SNP-prediabetes association differed by smoking status. The stratum specific 95% CIs suggest that the SNP-prediabetes association was not significant among ever smokers and never smokers. For rs946911 in *TTC6*, the SNP OR for prediabetes versus normal glycemia was 1.52 among never smokers (95% CI:1.00, 2.32) and 0.63 among ever smokers (95% CI: 0.37, 1.06), with an interaction p-value of 0.006. The interaction p-values suggest that the SNP-prediabetes association differed by smoking

status. The stratum-specific 95% CIs suggests that the SNP-prediabetes association was not significant among ever smokers and never smokers.

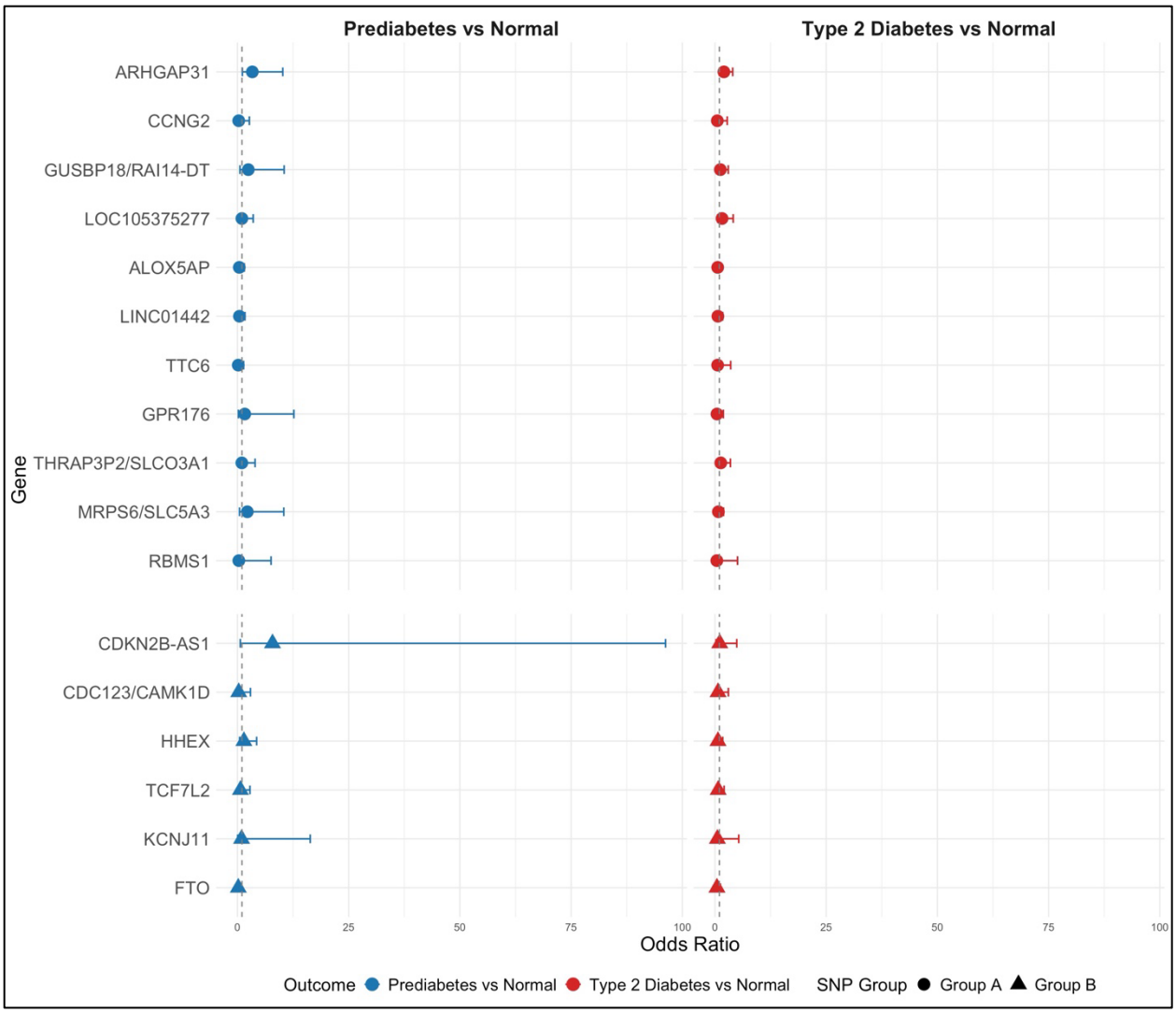

Figure 11. African Ancestry Candidate SNPs Interaction Forest Plot.

Note: The forest plot illustrates the comparison groups of prediabetes vs. normal glycemia and type 2 diabetes vs. normal glycemia side by side for African Ancestry. This plot displays the gene-by-smoking interaction on the y-axis and their corresponding odds ratios and 95% confidence intervals on the x-axis.

Table 21. African Ancestry Candidate SNP by Smoking Interaction Results							
Trait	**CHR**	**rsID**	**Gene**	**Prediabetes vs. Normal**		**Type 2 Diabetes vs. Normal**	
SNP by Smoking				OR (95% CI)	Unadj. P-value	OR (95% CI)	Unadj. P-value
Group A (Prediabetes SNPs)	3	rs36087584	ARHGAP31	**3.34 (1.09-10.19)**	**0.03**	**2.02 (1.02-3.99)**	**0.04**
	4	rs34938732	CCNG2	0.28 (0.03-2.67)	0.27	0.53 (0.10-2.76)	0.45
	5	rs6880621	GUSBP18/RAI14-DT	2.44 (0.57-10.49)	0.23	1.21 (0.49-2.99)	0.68
	7	rs10229340	LOC105375277/ENSG00000285741	1.00 (0.28-3.54)	0.99	1.59 (0.62-4.10)	0.33
	13	rs9550371	ALOX5AP	0.41 (0.11-1.53)	0.18	0.61 (0.29-1.29)	0.19
	13	rs4886290	LINC01442	0.44 (0.12-1.69)	0.23	0.65 (0.25-1.73)	0.39
	14	rs946911	TTC6	0.18 (0.02-1.37)	0.10	0.60 (0.10-3.54)	0.57
	15	rs41497851	GPR176	1.64 (0.21-12.68)	0.64	0.45 (0.11-1.88)	0.27
	15	rs11853125	THRAP3P2/SLCO3A1	0.98 (0.24-3.96)	0.98	1.31 (0.49-3.48)	0.59
	21	rs62212118	MRPS6/SLC5A3	0.31 (0.01-7.55)	0.47	0.42 (0.03-5.07)	0.49
Group B (Type 2 Diabetes SNPs)	2	rs7593730	RBMS1	2.22 (0.47-10.42)	0.31	0.78 (0.31-1.93)	0.58
	9	rs564398	CDKN2B-AS1	7.88 (0.65-96.24)	0.11	1.07 (0.23-4.91)	0.93
	10	rs12779790	CDC123/CAMK1D	0.28 (0.03-2.92)	0.29	0.69 (0.16-3.02)	0.63
	10	rs1111875	HHEX	1.45 (0.49-4.31)	0.50	0.68 (0.27-1.75)	0.42
	10	rs7903146	TCF7L2	0.67 (0.16-2.81)	0.59	0.72 (0.25-2.06)	0.54
	11	rs5219	KCNJ11	0.92 (0.05-16.37)	0.96	0.55 (0.06-5.34)	0.60
	16	rs8050136	FTO	**0.19 (0.06-0.62)**	**0.006**	0.45 (0.17-1.21)	0.11

Bolded values are significant prior to FDR correction; Purple outlined SNPs: Group A SNPs analyzed in the prediabetes group in GENNID; Orange outlined SNPs: Group A SNPs analyzed in type 2 diabetes group; Blue outlined SNPs: Group B SNPs analyzed in prediabetes group; Green outlined SNPs: Group B SNPs analyzed in type 2 diabetes group.

Table 22. Stratum-Specific SNP Odds Ratios by Smoking Status for African Ancestry									
SNP/GENE	**Smoking Status**	**0 T Copies: Normal**	**0 T Copies: Prediabetes**	**1 T Copy: Normal**	**1 T Copy: Prediabetes**	**2 T Copies: Normal**	**2 T Copies: Prediabetes**	**SNP OR: Prediabetes vs Normal**	**Interaction P-Value**
rs36087584/ ARHGAP31	Never Smokers	7	6	12	8	12	3	0.46 (0.21-1.04)	0.03
	Ever Smokers	13	4	20	6	9	6	1.54 (0.72-3.29)	
SNP/GENE	**Smoking Status**	**0 T Copies: Normal**	**0 T Copies: Type 2 Diabetes**	**1 T Copy: Normal**	**1 T Copy: Type 2 Diabetes**	**2 T Copies: Normal**	**2 T Copies: Type 2 Diabetes**	**SNP OR: Type 2 Diabetes vs Normal**	**Interaction P-Value**
rs36087584/ ARHGAP31	Never Smokers	7	16	12	32	12	14	0.63 (0.35-1.16)	0.04
	Ever Smokers	13	25	20	41	9	28	1.28 (0.86-1.89)	
SNP/GENE	**Smoking Status**	**0 A Copies: Normal**	**0 A Copies: Type 2 Diabetes**	**1 A Copy: Normal**	**1 A Copy: Type 2 Diabetes**	**2 A Copies: Normal**	**2 A Copies: Type 2 Diabetes**	**SNP OR: Type 2 Diabetes vs Normal**	**Interaction P-Value**
rs8050136/ FTO	Never Smokers	12	19	16	32	3	11	2.45 (1.16-5.16)	0.006
	Ever Smokers	11	31	18	46	13	17	0.46 (0.21-0.97)	
Note: Counts represent observed frequencies by genotype group, smoking status, and glycemic category. Odds ratios, 95% confidence intervals, and the SNP-by-smoking interaction p-values presented derive from the original GxE interaction model.									

Results for African Ancestry Interaction

The multinomial candidate SNP interaction analysis results for African ancestry assessed 17 SNPs. In the comparison of prediabetes to normal glycemia, one Group A SNP (rs36087584) by smoking interaction and one Group B SNP (rs8050136) by smoking interaction were found to be significant prior to FDR correction. The first SNP was rs36087584 on chromosome 3 near the *ARHGAP31* gene (P=0.03, OR: 3.34 [95% CI: 1.09, 10.19]). The *ARHGAP31* (rho GTPase-activating protein 31) gene was identified in one of the studies from which our SNP panel was derived, which was associated with prediabetes status change, and has not otherwise been associated with glycemic traits or related phenotypes in the literature.[36,96] The second SNP was rs8050136 on chromosome 16 near the *FTO* gene (P=0.006, OR: 0.19 [95% CI: 0.06, 0.62]). The FTO (fat mass and obesity associated) gene is a well-replicated T2D risk variant and acts primarily through adiposity rather than direct effects on insulin secretion.[39,97] Results did not remain significant after FDR-based multiple testing correction. Results for the 17 SNPs analyzed are outlined in Table 21.

In the comparison of type 2 diabetes to normal glycemia one Group A SNP by smoking interaction was found to be significant prior to FDR-based multiple testing correction. The SNP is rs36087584 on chromosome 3 near the *ARHGAP31* gene (P=0.04, OR: 2.02 [95% CI: 1.02, 3.99]).

The SNP-by-smoking stratum-specific odds ratios, confidence interval, and p-value from the interaction model results are shown in Table 22. For rs36087584 in *ARHGAP31*, the SNP OR for prediabetes versus normal glycemia was 0.46 among never smokers (95% CI: 0.21, 1.04) and 1.54 among ever smokers (95% CI: 0.72, 3.29), with an interaction p-value of 0.03. The interaction p-value suggests that the SNP-prediabetes association differed by smoking status. The

stratum-specific 95% CIs suggest that the SNP-prediabetes association was not significant among ever smokers and never smokers. For rs36087584 in *ARHGAP31*, the SNP OR for type 2 diabetes versus normal glycemia was 0.63 among never smokers (95% CI: 0.35, 1.16) and 1.28 among ever smokers (95% CI: 0.86, 1.89), with an interaction p-value of 0.04. The interaction p-value suggests that the SNP-type 2 diabetes association differed by smoking status. The stratum-specific 95% CIs suggest that the SNP-type 2 diabetes association was not significant among ever smokers and never smokers. For rs8050136 in *FTO*, the SNP OR for type 2 diabetes versus normal glycemia was 2.45 among never smokers (95% CI: 1.16, 5.16) and 0.46 among ever smokers (95% CI: 0.21, 0.97), with an interaction p-value of 0.006. The interaction p-value suggests that the SNP-type 2 diabetes association differed by smoking status. The stratum-specific 95% CIs suggest that the SNP-type 2 diabetes association was suggestively significant among ever smokers, but not among never smokers.

DISCUSSION

The purpose of this study was to assess genetic influences on glycemic status and potential interactions with smoking. We assessed whether candidate SNPs previously associated with either prediabetes or type 2 diabetes showed associations with glycemic status in a multi-ancestry family-based sample and whether these associations varied by smoking status. We observed candidate SNP associations and SNP-by-smoking interactions for both prediabetes and type 2 diabetes prior to FDR-based multiple testing correction. Three main findings were detected. First, we observed Group A SNPs to be associated with prediabetes and Group B SNPs to be associated with type 2 diabetes; however, the results were not consistent across ancestry groups. In European ancestry, a distinct pattern was observed between the Prediabetes-associated SNPs and the type 2 diabetes-associated SNPs. However, this pattern was not observed within Hispanic and African American ancestry groups – Group A SNPs were not consistently associated with Prediabetes across all three ancestries, similarly, Group B SNPs were not consistently associated with type 2 diabetes. Second, despite the overall inconsistency across ancestry groups and outcome comparisons, the rs4661250 SNP, located near the *SUSD4* gene, showed suggestive evidence of association across all ancestry groups and was consistently found in more than one outcome comparison group among both European and Hispanic ancestry groups. This marker was associated with prediabetes among European ancestry, with both prediabetes and type 2 diabetes in Hispanic ancestry, and with type 2 diabetes in African American ancestry, suggesting that the two stages of disease could potentially involve a similar marker; however, these findings did not remain significant after FDR-based multiple testing correction. Third, SNP findings for prediabetes and type 2 diabetes showed fewer associations

than expected for both groups. Collectively, the findings were not consistent with our hypothesis of differences between prediabetes and type 2 diabetes within this study's set of candidate SNPs.

The absence of a clear or consistent distinction between prediabetes-associated and type 2 diabetes-associated candidate SNPs may reflect factors including selection of candidate SNPs and the phenotypic heterogeneity of prediabetes categorization. Categorical definitions of glycemic status do not fully capture the underlying heterogeneity of dysglycemia. Prediabetes defined by impaired fasting glucose, impaired glucose tolerance, or elevated HbA1c identifies overlapping but not identical populations and may reflect different underlying aspects of dysglycemia.[27] Because these categories reflect different aspects of glycemic dysfunction, SNPs identified using one measure may not associate with outcomes defined by another, which may partly explain the mixed pattern of associations observed in this study. The majority of prediabetes- and type 2 diabetes-associated SNPs selected for this study had been closely linked to aspects of dysglycemia. The prediabetes traits were linked to fasting plasma glucose, HbA1c, 2-hour oral glucose tolerance test levels, and prediabetes status change, whereas the type 2 diabetes traits were strongly linked to insulin resistance, insulin sensitivity, insulin secretion, and beta-cell function.[36,37] Further, prediabetes may not be a single uniform intermediate stage before type 2 diabetes.[27] Recent work has identified prediabetes subtypes that differ biologically by insulin sensitivity, insulin secretion, visceral and hepatic fat, and genetic risk.[32] Therefore, individuals classified as having prediabetes may not share a single underlying pathophysiologic or genetic profile. This heterogeneity could obscure associations with candidate SNPs, particularly if some variants are relevant only to specific prediabetes subtypes or metabolic pathways. This could help explain the lack of consistent Group A SNP associations based on the heterogeneity within the prediabetes phenotype rather than absence of genetic contribution to

early dysglycemia. Another factor is that individuals with prediabetes may follow different clinical pathways, including progression to type 2 diabetes, persistence of prediabetes, or regression to normoglycemia. Evidence that remission or regression from prediabetes is possible and that restoration of normoglycemia is associated with lower subsequent type 2 diabetes risk, suggests that prediabetes includes biologically diverse and potentially modifiable states rather than a single fixed precursor phenotype.[28-29]

Candidate SNPs were selected based on their relevance to prediabetes (Group A) or type 2 diabetes (Group B). However, consistent findings across Groups A and B SNPs were not observed beyond European ancestry, particularly in Hispanic and African American ancestry. Seven distinct Group A SNPs, selected because prior studies associated them with prediabetes or related measures of early dysglycemia, showed suggestive association with prediabetes in the present study across European, Hispanic, and African American ancestries (one of which showed suggestive association in two ancestries, yielding eight total SNP-ancestry associations). These SNPs included rs4661250, rs41497851, rs11853125, rs6880621, rs946911, rs62212118, and rs34938732. However, only a subset of Group A SNPs showed suggestive association with prediabetes for each ancestry and were not consistent across ancestry groups in the present study. This may partly reflect differences between the phenotypes evaluated in the original studies and the fasting plasma glucose-based definition used in our analysis. Several Group A SNPs were previously associated with impaired glucose tolerance, hemoglobin A1C (HbA1c), and changes in prediabetes status. These traits capture related but biologically distinct aspects of dysglycemia. Impaired fasting glucose is more closely related to fasting glucose regulation and hepatic insulin resistance, whereas impaired glucose tolerance reflects post-challenge glucose handling, peripheral insulin sensitivity, and beta-cell response following a glucose load.[98] HbA1c reflects

cumulative glucose exposure over time rather than insulin function directly.[99] Therefore, some Group A SNPs may be more closely related to specific aspects of glycemic regulation than to prediabetes defined by fasting plasma glucose. Consistent with this possibility, the rs4661250 SNP near *SUSD4* was observed in European and Hispanic ancestries in the present study and was previously associated with fasting plasma glucose.[37] Additionally, several Group A SNPs were associated with type 2 diabetes rather than prediabetes, or with both outcomes. Five markers found to be suggestively significant in only the Hispanic and African American ancestry groups indicate that the prediabetes candidate SNPs were associated with type 2 diabetes in this study, suggesting that these loci may be associated with both phenotypes. These variants could therefore influence biological processes that contribute to worsening glycemic regulation across the disease continuum rather than being specific to prediabetes alone. Because fasting plasma glucose contributed to the classification of both prediabetes and type 2 diabetes, some Group A SNP associations may also reflect genetic influences on fasting glucose related dysglycemia across both categories, rather than exclusively to prediabetes. The lack of clear separation between Group A and Group B SNPs may reflect shared genetic influences across stages of dysglycemia. These markers influence biological processes that precede clinical diagnosis, including beta-cell function, insulin secretion, proinsulin processing, fasting glucose regulation, and adiposity. Therefore, some Group B SNPs may also be relevant to earlier dysglycemia as well. Consistent with this interpretation, variants rs4661250 and rs34938732 near the *SUSD4* and *CCNG2* genes, respectively, showed nominal evidence of association with both prediabetes and type 2 diabetes among participants of Hispanic ancestry.

In contrast, we did not find many associations with type 2 diabetes and Group B SNPs, despite the fact that these SNPs have been replicated in multiple studies and across ancestry

groups. Only three of the Group B SNPs (rs13266634, rs7903146, rs12779790) showed suggestive association with type 2 diabetes across ancestries. This finding is unexpected given that these Group B SNPs were chosen for having been replicated in several kinds of studies for their association with type 2 diabetes across ancestry groups, including those reflective of the GENNID sample. Two factors may explain this finding. First, the ancestry-specific analyses may have had limited power to detect the modest effects typically associated with common susceptibility variants. Since large-scale GWAS identify variants with modest effect sizes, detection of these variants requires substantially larger sample sizes than were available within each ancestry group of the GENNID cohort. Second, as in many genetic studies of complex disease, reduced penetrance – a phenomenon in which individuals may carry a risk allele but not express the associated phenotype – can affect population-based and family-based studies alike.[100,101] Selection of controls from relatives of cases does not bias estimates of genetic effect, provided disease risk remains constant over time; however, that assumption could be violated since type 2 diabetes risk increases with age.[102] This is not unique to the type 2 diabetes comparisons and can affect prediabetes since the same family-based ascertainment applies throughout the sample. Normoglycemic relatives who have not yet developed prediabetes could, in principle, dilute the prediabetes comparisons in the same way. However, age distribution suggests this issue disproportionately affected the type 2 diabetes comparisons when compared to prediabetes. The normoglycemic group (mean age 45 years, SD 16) was on average younger than the type 2 diabetes group (mean age 57 years, SD 12), whereas the age difference between the normoglycemic and prediabetes groups (mean age 49 years, SD 15) was comparatively small. Because type 2 diabetes risk increases with age, a substantial portion of younger normoglycemic individuals have not reached the age of risk for disease onset, despite carrying

susceptibility alleles, whereas the narrower age gap relevant to the prediabetes comparison limits this effect even though prediabetes risk also increases with age. Because the GENNID families were ascertained through a proband affected by type 2 diabetes, the normoglycemic relatives are more likely than unrelated individuals to carry the same susceptibility alleles. Combined with the younger age of the group, a substantial portion of normoglycemic relatives may carry these alleles but have not reached the age at which they would be phenotypically expressed. The narrow allele frequency differences between the normoglycemic and type 2 diabetes group may reflect limited power, reduced penetrance, or absence of true association. Reduced penetrance could potentially explain why the genotype and phenotype may not correspond perfectly, since individuals carrying susceptibility alleles may not show indication of disease, which may be applicable to both prediabetics and type 2 diabetics. Separately, there is also potential for non-differential outcome misclassification in this study. If present, this kind of bias can arise when incorrect classification of the outcome (glycemic status) is not dependent on the exposure (genotype). In this study, glycemic status was established based on a single fasting plasma glucose measurement subject to day-to-day physiological variability, which is expected to be independent of genotype and can contribute to this type of misclassification.

The ancestry-specific findings did not identify a consistent pattern of separation between Group A and Group B SNPs. Further, different SNPs reached suggestive significance across ancestry, with little overlap. A few factors may have contributed to these differences. First, they may reflect the ancestry composition of the studies from which the candidate SNPs were selected. Many prediabetes-associated SNPs were originally identified in predominantly European and East Asian ancestry populations, which may explain the clearer separation observed in the European ancestry group and less consistent patterns observed in the Hispanic

and African American groups. This discrepancy arises from two important factors: differences in allele frequencies and linkage disequilibrium (LD) across ancestries. Allele frequency differences may vary by ancestry, resulting in a situation where a variant may be common in one ancestry group and rare or less common in another. For example, in the GENNID sample, allele frequencies varied across ancestries for the rs41497851 SNP (*GPR176*). The marker was associated with prediabetes in the European ancestry group (MAF 0.3), but the minor allele frequency was lower in Hispanic (MAF 0.14) and African American (MAF 0.16) ancestry groups, where no associations were detected. Further, GWAS index SNPs often tag nearby causal variation rather than representing the causal variant itself. If linkage disequilibrium patterns differ across ancestry groups, or if LD patterns differ between the discovery population and the GENNID ancestry groups, the index SNP may not tag the same underlying causal variant with the same level of accuracy. For example, in the GENNID sample, the SNP mentioned previously, rs41497851 (*GPR176*) was found to be associated with prediabetes in the European group. The LDproxy results from 1000 Genomes populations show nine tagged variants with perfect-LD block ($r^2 = 1$) around the SNP, compared to the African American ancestry panel that showed only four, with the r^2 dropping sharply after. The LD structure in the Hispanic American panel closely resembled that of the European panel, suggesting the reduced tagging efficiency is not the explanation for the null association, rather, the lower minor allele frequency may have limited power to detect an association of similar effect size. Collectively, this may explain why consistent patterns did not emerge for Group A SNPs with prediabetes or Group B SNPs with type 2 diabetes across ancestry groups.

Across all three ancestries, suggestive SNP-by-smoking interactions were observed for both the prediabetes and the type 2 diabetes assessments, and the corresponding stratified results

provided the subgroup-specific estimates that contextualize these interactions. Although most interaction terms did not remain significant after FDR-based multiple testing correction, SNP-by-smoking results suggest that smoking status modifies the relationship between some of the candidate SNPs and dysglycemia. Biologically, this is plausible because smoking has been associated with insulin resistance, inflammation, oxidative stress, altered adiposity distribution, and beta-cell dysfunction, all of which are relevant to prediabetes and type 2 diabetes.[103,104] Since the *SLC30A8* and *KCNJ11* genes both play established roles in pancreatic beta-cell function and insulin release, smoking-related oxidative stress could compound beta-cell dysfunction in carriers of risk alleles for these loci.[105-108] *FTO*, a well-established regulator of adiposity, may interact with smoking-related changes in fat distribution to influence insulin resistance.[109] *SUSD4*'s role in innate immune regulation suggests a plausible interaction with smoking-induced inflammation.[77] However, these findings should be interpreted cautiously for two reasons. First, the interaction results did not remain significant after FDR-based multiple testing correction and second, the family-based structure of the GENNID study may also contribute to ancestry-specific association patterns beyond genetics alone. Although analyses stratified by ancestry account for differences between racial and ethnic populations, participants within each ancestry group were members of families who shared both inherited genetic background and environmental context. Shared familial factors – including diet, physical activity patterns, smoking norms, socioeconomic conditions, healthcare access – may influence glycemic status and shape how genetic susceptibility is expressed. While kinship coefficients model correlation attributable to genetic relatedness in the variance structure of the model, they do not adjust for potential confounding by shared environment. This is especially relevant for gene-by-smoking interaction analyses in this study since smoking status may reflect family- or community- level patterns in

addition to individual-level exposure. This may result in inflated type I error, consistent with prior literature for gene-environment interaction in family studies.[110] Regardless, the value of these findings is that they indicate potential opportunities for intervention since smoking is a modifiable exposure.

This study involved a few limitations, one of which involved limited availability of candidate variants for prediabetes across diverse racial and ethnic groups. Although the genetic literature is substantially more developed for type 2 diabetes, many Group A (prediabetes) variants were identified in populations of European and East Asian ancestry. Differences in allele frequencies and linkage disequilibrium patterns may have contributed to differences in association patterns across ancestry groups. A second limitation is the cross-sectional establishment of the glycemic status. Because glycemic status was measured at a single point in time, the study could not distinguish individuals with stable normoglycemia or prediabetes from those who may later progress to another glycemic category. This form of misclassification is likely non-differential since individuals with the effect allele could be placed into the normoglycemia group based on their fasting plasma glucose at the time of study enrollment, however, may not remain in that category. This would diminish differences between the outcome groups and attenuate the results. Third, smoking exposure was categorized as ever versus never smoking and therefore did not capture smoking duration, intensity, or time since cessation. This limited exposure characterization may have obscured differences among ever smokers and reduced the ability to detect SNP by smoking interactions. Lastly, nominal associations did not remain significant after correction for multiple testing. This may reflect limited statistical power to detect modest genetic effects rather than the absence of a true biological signal. Similarly,

power to detect interactions was limited given the relatively modest genetic effects and the need to account for multiple testing.

This study included several key strengths. One of those strengths is the use of a multinomial study design that allowed glycemic status to be evaluated as a multicategory phenotype rather than separate binary models. This helped preserve sample size, reduce multiple comparisons, and improve power while accounting for family structure. Second, while family datasets can be susceptible to biases such as reduced power from related, non0independent observations, they offer methodological advantages in genetic epidemiology due to enrichment of genetic susceptibility, improved control of population stratification, and the ability to model shared familial correlation structures. The presence of family structure still enhances the power to detect genetic associations and gene-environment interactions. Further, ancestry-stratified analyses provided racial/ethnic-specific insights into genetic factors involved in disease development, given known differences in allele frequencies, and linkage disequilibrium patterns across populations. The candidate SNP approach is both a strength and a limitation since it enabled a focused, biologically interpretable analysis with reduced multiple-testing burden, but restricted inference to a limited set of previously reported variants that may not capture relevant genetic variation equally across ancestry groups.

Understanding the genetic and environmental factors associated with prediabetes is important for clarifying the early stages of dysglycemia before the development of type 2 diabetes. Although type 2 diabetes has been extensively studied, prediabetes remains comparatively underrepresented in genetic epidemiology research. This dissertation evaluated candidate SNP associations with prediabetes and type 2 diabetes as separate outcomes within a family-based, multi-ancestry framework, while also examining whether associations differed in

the presence of smoking. Overall, the results did not demonstrate a clear separation in candidate SNP association patterns between prediabetes and type 2 diabetes. Instead, association patterns varied across SNPs, outcomes, and ancestry groups. These findings suggest that, within this candidate SNP framework, there is no clear evidence that the genetic associations for prediabetes were distinct from those for type 2 diabetes. Future research should expand investigation of the prediabetic state, focusing on large-scale genome-wide association studies with larger sample sizes, and replication datasets with inclusion of diverse populations to further evaluate genetic associations for prediabetes among the glycemic continuum. Developing a better understanding of prediabetes may help inform future prevention strategies aimed at reducing the public health burden of disease.

REFERENCES

1. Bansal N. Prediabetes diagnosis and treatment: A review. *World J Diabetes*. 2015;6(2):296-303. doi:10.4239/wjd.v6.i2.296
2. Prediabetes: Could it be you? infographic. Centers for Disease Control and Prevention. Accessed April 3, 2025. https://www.cdc.gov/diabetes/communication-resources/prediabetes-statistics.html.
3. Hostalek U. Global epidemiology of prediabetes - present and future perspectives. *Clinical Diabetes and Endocrinology*. 2019;5(1):5. doi:10.1186/s40842-019-0080-0
4. Khan RMM, Chua ZJY, Tan JC, Yang Y, Liao Z, Zhao Y. From Pre-Diabetes to Diabetes: Diagnosis, Treatments and Translational Research. *Medicina (Kaunas)*. 2019;55(9):546. Published 2019 Aug 29. doi:10.3390/medicina55090546
5. Sallar A, Dagogo-Jack S. Regression from prediabetes to normal glucose regulation: State of the science. *Exp Biol Med (Maywood)*. 2020;245(10):889-896. doi:10.1177/1535370220915644
6. Diabetes Prevention Program (DPP) Research Group. The Diabetes Prevention Program (DPP): description of lifestyle intervention. *Diabetes Care*. 2002;25(12):2165-2171. doi:10.2337/diacare.25.12.2165
7. Lindström J, Louheranta A, Mannelin M, et al. The Finnish Diabetes Prevention Study (DPS): Lifestyle intervention and 3-year results on diet and physical activity. *Diabetes Care*. 2003;26(12):3230-3236. doi:10.2337/diacare.26.12.3230
8. Salive ME, Tjaden AH, Ames JR, et al. Lifestyle and metformin interventions and risk of multimorbidity in adults with Prediabetes. *JAMA*. Published online June 15, 2026. doi:10.1001/jama.2026.8492
9. Wagner R, Selvin E, Sehgal R, et al. Beyond glucose—rethinking prediabetes for precision prevention. *Diabetes Care*. 2025;49(2):226-235. doi:10.2337/dci25-0054
10. Galicia-Garcia U, Benito-Vicente A, Jebari S, et al. Pathophysiology of Type 2 Diabetes Mellitus. *Int J Mol Sci*. 2020;21(17):6275. Published 2020 Aug 30. doi:10.3390/ijms21176275
11. Lu X, Xie Q, Pan X, et al. Type 2 diabetes mellitus in adults: pathogenesis, prevention and therapy. *Signal Transduction and Targeted Therapy*. 2024;9(1):262. doi:10.1038/s41392-024-01951-9
12. American Diabetes Association; Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 1 January 2014; 37 (Suppl 1): S81–S90. doi:10.2337/dc14-S081
13. Facts & figures. International Diabetes Federation. April 7, 2025. Accessed April 22, 2025. https://idf.org/about-diabetes/diabetes-facts-figures/.
14. National Diabetes Statistics Report. Centers for Disease Control and Prevention. Accessed April 3, 2025. https://www.cdc.gov/diabetes/php/data-research/index.html.
15. Koopman RJ. Changes in Age at Diagnosis of Type 2 Diabetes Mellitus in the United States, 1988 to 2000. *The Annals of Family Medicine*. 2005;3(1):60-63. doi:10.1370/afm.214
16. Zhang HJ, Feng J, Zhang XT, Zhang HZ. Age at type 2 diabetes diagnosis and the risk of mortality among US population. *Scientific Reports*. 2024;14(1):29155. doi:10.1038/s41598-024-80790-8
17. Parker ED, Lin J, Mahoney T, et al. Economic Costs of Diabetes in the U.S. in 2022. *Diabetes Care*. 2024;47(1):26-43. doi:10.2337/dci23-0085

18. Golden SH, Yajnik C, Phatak S, Hanson RL, Knowler WC. Racial/ethnic differences in the burden of type 2 diabetes over the life course: a focus on the USA and India. *Diabetologia*. 2019;62(10):1751-1760. doi:10.1007/s00125-019-4968-0
19. Spanakis EK, Golden SH. Race/ethnic difference in diabetes and diabetic complications. *Curr Diab Rep*. 2013;13(6):814-823. doi:10.1007/s11892-013-0421-9
20. Walker RJ, Strom Williams J, Egede LE. Influence of race, ethnicity and social determinants of health on diabetes outcomes. *Am J Med Sci.* 2016;351(4):366-373. doi:10.1016/j.amjms.2016.01.008
21. Canedo JR, Miller ST, Schlundt D, Fadden MK, Sanderson M. Racial/ethnic disparities in diabetes quality of care: the role of healthcare access and socioeconomic status. *J Racial Ethn Health Disparities.* 2018;5(1):7-14. doi:10.1007/s40615-016-0335-8
22. Ferrannini E, Gastaldelli A, Iozzo P. Pathophysiology of Prediabetes. *Medical Clinics of North America*. 2011;95(2):327-339. doi:10.1016/j.mcna.2010.11.005
23. Echouffo-Tcheugui JB, Selvin E. Prediabetes and What It Means: The Epidemiological Evidence. *Annual Review of Public Health*. 2021;42(1):59-77. doi:10.1146/annurev-publhealth-090419-102644
24. Alvarez S, Coffey R, Mathias PM, Algotar AM. Prediabetes. In: *StatPearls*. 2023. Treasure Island (FL): StatPearls Publishing; July 17, 2023.
25. Prediabetes: Could it be you? infographic. Centers for Disease Control and Prevention. Accessed April 3, 2025. https://www.cdc.gov/diabetes/communication-resources/prediabetes-statistics.html.
26. Rooney MR, Fang M, Ogurtsova K, et al. Global Prevalence of Prediabetes. *Diabetes Care*. 2023;46(7):1388-1394. doi:10.2337/dc22-2376
27. Unnikrishnan R, Shaw JE, Chan JC, et al. Prediabetes. *Nature Reviews Disease Primers*. 2025;11(1). doi:10.1038/s41572-025-00635-0
28. Sandforth L, Kullmann S, Sandforth A, et al. Prediabetes remission to reduce the global burden of type 2 diabetes. *Trends in Endocrinology & Metabolism*. 2025;36(10):899-916. doi:10.1016/j.tem.2025.01.004
29. Davoodian N, Lotfaliany M, Huxley RR, et al. Regression from prediabetes to normoglycaemia and the role of cardiometabolic risk factors on the subsequent risk of developing type 2 diabetes. *Diabetologia*. 2025;69(1):69-80. doi:10.1007/s00125-025-06555-8
30. Lawal Y, Bello F, Kaoje YS. Prediabetes Deserves More Attention: A Review. *Clinical Diabetes*. 2020;38(4):328-338. doi:10.2337/cd19-0101
31. Freeman AM, Acevedo LA, Pennings N. Insulin Resistance. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; August 17, 2023.
32. Wagner R, Thorand B, Osterhoff MA, et al. Family history of diabetes is associated with higher risk for prediabetes: a multicentre analysis from the German Center for Diabetes Research. *Diabetologia*. 2013;56(10):2176-2180. doi:10.1007/s00125-013-3002-1
33. Yang Li, Guo-Chong Chen, Jee-Young Moon, Rhonda Arthur, Daniela Sotres-Alvarez, Martha L. Daviglus, Amber Pirzada, Josiemer Mattei, Krista M. Perreira, Jerome I. Rotter, Kent D. Taylor, Yii-Der Ida Chen, Sylvia Wassertheil-Smoller, Tao Wang, Thomas E. Rohan, Joel D. Kaufman, Robert Kaplan, Qibin Qi; Genetic Subtypes of Prediabetes, Healthy Lifestyle, and Risk of Type 2 Diabetes. *Diabetes* 20 June 2024; 73 (7): 1178–1187.
34. Liu X, Collister JA, Clifton L, Hunter DJ, Littlejohns TJ. Polygenic risk of prediabetes, undiagnosed diabetes, and incident type 2 diabetes stratified by diabetes risk factors. *J Endocr Soc.* 2023;7(4):bvad020. doi:10.1210/jendso/bvad020

35. Billings LK, Jablonski KA, Pan Q, et al. Increased Genetic Risk for β-Cell Failure Is Associated With β-Cell Function Decline in People With Prediabetes. *Diabetes*. 2024;73(8):1352-1360. doi:10.2337/db23-0761
36. Liu T, Li H, Conley YP, et al. A Genome-Wide Association Study of Prediabetes Status Change. *Front Endocrinol (Lausanne)*. 2022;13:881633. Published 2022 Jun 13. doi:10.3389/fendo.2022.881633
37. Lin L, Fang T, Lin L, et al. Genetic Variants Relate to Fasting Plasma Glucose, 2-Hour Postprandial Glucose, Glycosylated Hemoglobin, and BMI in Prediabetes. *Front Endocrinol (Lausanne)*. 2022;13:778069. Published 2022 Mar 1. doi:10.3389/fendo.2022.778069
38. Suzuki K, Hatzikotoulas K, Southam L, et al. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature*. 2024;627(8003):347-357. doi:10.1038/s41586-024-07019-6
39. Ali, O. Genetics of type 2 diabetes. *World J Diabetes*. 2013;4(4):114-123. doi:10.4239/wjd.v4.i4.114
40. Kreienkamp RJ, Voight BF, Gloyn AL, Udler MS. Genetics of Type 2 Diabetes. In: Lawrence JM, Casagrande SS, Herman WH, Wexler DJ, Cefalu WT, eds. *Diabetes in America*. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK); December 20, 2023.
41. Sanghera DK, Blackett PR. Type 2 Diabetes Genetics: Beyond GWAS. *J Diabetes Metab*. 2012;3(198):6948. doi:10.4172/2155-6156.1000198
42. Tabara Y, Osawa H, Kawamoto R, et al. Replication Study of Candidate Genes Associated With Type 2 Diabetes Based On Genome-Wide Screening. *Diabetes*. 2009;58(2):493-498. doi:10.2337/db07-1785
43. Voight BF, Scott LJ, Steinthorsdottir V, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genetics*. 2010;42(7):579-589. doi:10.1038/ng.609
44. Willer CJ, Bonnycastle LL, Conneely KN, et al. Screening of 134 single nucleotide polymorphisms (snps) previously associated with type 2 diabetes replicates association with 12 snps in nine genes. *Diabetes*. 2007;56(1):256-264. doi:10.2337/db06-0461
45. Barroso I. Genetics of Type 2 diabetes. *Diabetic Medicine*. 2005;22(5):517-535. doi:10.1111/j.1464-5491.2005.01550.x
46. Herder C, Roden M. Genetics of type 2 diabetes: pathophysiologic and clinical relevance. *European Journal of Clinical Investigation*. 2011;41(6):679-692. doi:10.1111/j.1365-2362.2010.02454.x
47. Laakso M, Fernandes Silva L. Genetics of Type 2 Diabetes: Past, Present, and Future. *Nutrients*. 2022; 14(15):3201. https://doi.org/10.3390/nu14153201
48. Prasad RB, Groop L. Genetics of Type 2 Diabetes—Pitfalls and Possibilities. *Genes*. 2015; 6(1):87-123. https://doi.org/10.3390/genes6010087
49. Owen KR, McCarthy MI. Genetics of type 2 diabetes. *Current Opinion in Genetics & Development*. 2007;17(3):239-244. doi:10.1016/j.gde.2007.04.003
50. Imamura M, Maeda S. Perspectives on genetic studies of type 2 diabetes from the genome-wide association studies era to precision medicine. *J Diabetes Investig*. 2024;15(4):410-422. doi:10.1111/jdi.14149
51. Ismail L, Materwala H, al Kaabi J. Association of risk factors with type 2 diabetes: A systematic review. *Computational and Structural Biotechnology Journal*. 2021;19:1759-1785. doi:10.1016/j.csbj.2021.03.003

52. Diabetes risk factors. www.heart.org. January 8, 2025. Accessed April 4, 2025. https://www.heart.org/en/health-topics/diabetes/understand-your-risk-for-diabetes.
53. Salzberg L. Risk Factors and Lifestyle Interventions. *Prim Care*. 2022;49(2):201-212. doi:10.1016/j.pop.2021.11.001
54. U.S. Department of Health and Human Services. *The Health Consequences of Smoking: 50 Years of Progress. A Report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2014.
55. Śliwińska-Mossoń M, Milnerowicz H. The impact of smoking on the development of diabetes and its complications. Diabetes & Vascular Disease Research. 2017;14(4):265-276. doi:10.1177/1479164117701876
56. Pan A, Wang Y, Talaei M, Hu FB, Wu T. Relation of active, passive, and quitting smoking with incident type 2 diabetes: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol.* 2015;3(12):958-967.
57. Willi C, Bodenmann P, Ghali WA, Faris PD, Cornuz J. Active smoking and the risk of type 2 diabetes: a systematic review and meta-analysis. *JAMA*. 2007;298(22):2654-2664. doi:10.1001/jama.298.22.2654
58. Aeschbacher S, Schoen T, Clair C, et al. Association of smoking and nicotine dependence with pre-diabetes in young and healthy adults. *Swiss Med Wkly*. 2014;144:w14019. Published 2014 Oct 8. doi:10.4414/smw.2014.14019
59. Virolainen, S.J., VonHandorf, A., Viel, K.C.M.F. *et al.* Gene–environment interactions and their impact on human health. *Genes Immun* **24**, 1–11 (2023). https://doi.org/10.1038/s41435-022-00192-6
60. Sørensen, T.I.A., Metz, S. & Kilpeläinen, T.O. Do gene–environment interactions have implications for the precision prevention of type 2 diabetes?. *Diabetologia* **65**, 1804–1813 (2022). https://doi.org/10.1007/s00125-021-05639-5
61. Bossowska M, Bossowski F, Adamska-Patruno E, Maliszewska K, Krętowski A. Personalised nutrition in obesity and prediabetes: Do genotypes matter? *Nutrients*. 2026;18(5):815. doi:10.3390/nu18050815
62. Wu P, Tajes JF, Qi Q, et al. Smoking-by-genotype interaction in type 2 diabetes risk and fasting glucose levels: A genome-wide interaction study in 61,164 individuals. *PLoS One*. 2020;15(3):e0230815. doi:10.1371/journal.pone.0230815
63. Raffel LJ, Robbins DC, Norris JM, et al. The GENNID Study: A resource for mapping the genes that cause NIDDM. *Diabetes Care*. 1996;19(8):864-872. doi:10.2337/diacare.19.8.864
64. American Diabetes Association Professional Practice Committee; 2. Diagnosis and Classification of Diabetes: *Standards of Care in Diabetes—2024*. *Diabetes Care* 1 January 2024; 47 (Suppl 1): S20–S42. https://doi.org/10.2337/dc24-S002
65. American Diabetes Association. Oral & injectable medications for type 2 diabetes. Accessed May 19, 2026. https://diabetes.org/health-wellness/medication/oral-other-injectable-diabetes-medications
66. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics. *Nucleic Acids Res.* 2019;47(D1):D1005-D1012. doi:10.1093/nar/gky1120
67. Cerezo M, Sollis E, Ji Y, et al. The NHGRI-EBI GWAS Catalog: standards for reusability, sustainability and diversity. *Nucleic Acids Res.* 2025;53(D1):D998–D1005. doi:10.1093/nar/gkae1070

68. Lin L, Quan H, Fang T, et al. Genome-wide association study of fasting proinsulin, fasting insulin, 2-hour postprandial proinsulin, and 2-hour postprandial insulin in Chinese Han people. *Endokrynol Pol*. 2022;73(5):856-862. doi:10.5603/EP.a2022.0054
69. Kato N. Insights into the genetic basis of type 2 diabetes. *J Diabetes Investig*. 2013;4(3):233-244. doi:10.1111/jdi.12067
70. Wang S, Meigs JB, Dupuis J. Genetic association tests in family samples for multi-category phenotypes. *BMC Genomics*. 2021;22(1):873. doi:10.1186/s12864-021-08107-x
71. Hall DB, Severini TA. Extended Generalized Estimating Equations for Clustered Data. *Journal of the American Statistical Association*. 1998;93(444):1365-1375. doi:10.1080/01621459.1998.10473798
72. Zeger SL, Liang KY, Albert PS. Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics*. 1988;44(4):1049. doi:10.2307/2531734
73. Benjamini Y, Hochberg U. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological).* 1995; *57*(1), 289-300. http://www.jstor.org/stable/2346101
74. Bender R, Lange S. Adjusting for multiple testing—when and how? *Journal of Clinical Epidemiology*. 2001;54(4):343-349. doi:10.1016/S0895-4356(00)00314-0
75. Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *Journal of Clinical Epidemiology*. 2014;67(8):850-857. doi:10.1016/j.jclinepi.2014.03.012
76. Rouam S. False Discovery Rate (FDR). In: *Encyclopedia of Systems Biology*. Springer New York; 2013:731-732. doi:10.1007/978-1-4419-9863-7_223
77. SUSD4 - sushi domain containing 4 [homo sapiens (human)] - Gene - NCBI. National Center for Biotechnology Information. Accessed April 4, 2025. https://www.ncbi.nlm.nih.gov/gene/55061.
78. Holmquist E, Okroj M, Nodin B, Jirström K, Blom AM. Sushi domain-containing protein 4 (SUSD4) inhibits complement by disrupting the formation of the classical C3 convertase. *FASEB J*. 2013;27(6):2355-2366. doi:10.1096/fj.12-222042
79. GPR176 G protein-coupled receptor 176 [Homo sapiens (human)]. National Center for Biotechnology Information. Gene database. Accessed August 6, 2026. https://www.ncbi.nlm.nih.gov/gene/11245
80. Alliance of Genome Resources. THRAP3P2: THRAP3 pseudogene 2. Alliance of Genome Resources. Accessed August 6, 2026. https://www.alliancegenome.org/gene/HGNC:54830
81. SLCO3A1 gene. GeneCards: The Human Gene Database. Weizmann Institute of Science. Accessed August 6, 2026. https://www.genecards.org/card/SLCO3A1
82. SLC30A8 solute carrier family 30 member 8 [Homo sapiens (human)]. National Center for Biotechnology Information. Gene database. Accessed August 6, 2026. https://www.ncbi.nlm.nih.gov/gene/169026
83. Wijesekara N, Dai FF, Hardy AB, et al. Beta cell-specific Znt8 deletion in mice causes marked defects in insulin processing, crystallisation and secretion. *Diabetologia*. 2010;53(8):1656-1668. doi:10.1007/s00125-010-1733-9
84. TCF7L2 transcription factor 7 like 2 [Homo sapiens (human)]. National Center for Biotechnology Information. Gene database. Accessed August 6, 2026. https://www.ncbi.nlm.nih.gov/gene/6934
85. CCNG2 gene. GeneCards: The Human Gene Database. Weizmann Institute of Science. Accessed August 6, 2026. https://www.genecards.org/card/CCNG2

86. GUSBP18 gene. GeneCards: The Human Gene Database. Weizmann Institute of Science. Accessed August 6, 2026. https://www.genecards.org/card/GUSBP18
87. RAI14 gene. GeneCards: The Human Gene Database. Weizmann Institute of Science. Accessed August 6, 2026. https://www.genecards.org/card/RAI14
88. LINC01442 gene. GeneCards: The Human Gene Database. Weizmann Institute of Science. Accessed August 6, 2026. https://www.genecards.org/card/LINC01442
89. TTC6 gene. GeneCards: The Human Gene Database. Weizmann Institute of Science. Accessed August 6, 2026. https://www.genecards.org/card/TTC6
90. MRPS6 gene. GeneCards: The Human Gene Database. Weizmann Institute of Science. Accessed August 6, 2026. https://www.genecards.org/card/MRPS6
91. SLC5A3 gene. GeneCards: The Human Gene Database. Weizmann Institute of Science. Accessed August 6, 2026. https://www.genecards.org/card/SLC5A3
92. Lin D, Yu J, Lin L, Ou Q, Quan H. MRPS6 modulates glucose-stimulated insulin secretion in mouse islet cells through mitochondrial unfolded protein response. *Sci Rep*. 2023;13:16149. doi:10.1038/s41598-023-43438-7
93. CDC123 gene. GeneCards: The Human Gene Database. Weizmann Institute of Science. Accessed August 6, 2026. https://www.genecards.org/card/CDC123
94. CAMK1D gene. GeneCards: The Human Gene Database. Weizmann Institute of Science. Accessed August 6, 2026. https://www.genecards.org/card/CAMK1D
95. KCNJ11 gene. GeneCards: The Human Gene Database. Weizmann Institute of Science. Accessed August 6, 2026. https://www.genecards.org/card/KCNJ11
96. ARHGAP31 gene. GeneCards: The Human Gene Database. Weizmann Institute of Science. Accessed August 6, 2026. https://www.genecards.org/card/ARHGAP31
97. Huang C, Chen W, Wang X. Studies on the fat mass and obesity-associated (FTO) gene and its impact on obesity-associated diseases. *Genes Dis*. 2022;10(6):2351-2365. doi:10.1016/j.gendis.2022.04.014
98. Abdul-Ghani MA, Tripathy D, DeFronzo RA. Contributions of β-Cell Dysfunction and Insulin Resistance to the Pathogenesis of Impaired Glucose Tolerance and Impaired Fasting Glucose. *Diabetes Care*. 2006;29(5):1130-1139. doi:10.2337/diacare.2951130
99. Kilpatrick ES. Haemoglobin A1c in the diagnosis and monitoring of diabetes mellitus. *J Clin Pathol*. 2008;61(9):977-982. doi:10.1136/jcp.2007.050625
100. Reduced penetrance. NCI Dictionary of Cancer Terms. National Cancer Institute. Accessed August 6, 2026. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/reduced-penetrance
101. Kingdom R, Wright CF. Incomplete penetrance and variable expressivity: from clinical studies to population cohorts. *Front Genet.* 2022;13:920390. doi:10.3389/fgene.2022.920390.
102. Goldstein AM, Hodge SE, Haile RW. Selection bias in case-control studies using relatives as the controls. *International Journal of Epidemiology*. 1989;18(4):985-989. doi:10.1093/ije/18.4.985
103. Rouland A, Thuillier P, Al-Salameh A, et al. Smoking and diabetes. *Annales d'Endocrinologie*. 2024;85(6):614-622. doi:10.1016/j.ando.2024.08.001
104. Tong X, Chaudhry Z, Lee C-C, et al. Cigarette smoke exposure impairs β-cell function through activation of oxidative stress and ceramide accumulation. *Molecular Metabolism*. 2020;37:100975. doi:10.1016/j.molmet.2020.100975

105. Gu H. Genetic, epigenetic and biological effects of zinc transporter (SLC30A8) in type 1 and type 2 diabetes. *Current Diabetes Reviews*. 2017;13(2):132-140. doi:10.2174/1573399812666151123104540
106. Moazzam-Jazi M, Najd-Hassan-Bonab L, Masjoudi S, et al. Risk of type 2 diabetes and KCNJ11 gene polymorphisms: A nested case–Control Study and meta-analysis. *Scientific Reports*. 2022;12(1). doi:10.1038/s41598-022-24931-x
107. Haghvirdizadeh P, Mohamed Z, Abdullah NA, Haghvirdizadeh P, Haerian MS, Haerian BS. KCNJ11: Genetic polymorphisms and risk of diabetes mellitus. *Journal of Diabetes Research*. 2015;2015:1-9. doi:10.1155/2015/908152
108. Keane KN, Cruzat VF, Carlessi R, de Bittencourt PI, Newsholme P. Molecular events linking oxidative stress and inflammation to insulin resistance and β-cell dysfunction. *Oxidative Medicine and Cellular Longevity*. 2015;2015:1-15. doi:10.1155/2015/181643
109. de Soysa AK, Langaas M, Jakic A, et al. The fat mass and obesity-associated (FTO) gene allele RS9939609 and glucose tolerance, hepatic and total insulin sensitivity, in adults with obesity. *PLOS ONE*. 2021;16(3). doi:10.1371/journal.pone.0248247
110. Shi M, Umbach DM, Weinberg CR. Family-based gene-by-environment interaction studies. *Epidemiology*. 2011;22(3):400-407. doi:10.1097/ede.0b013e318212fec6