Genetic and non-genetic influences on birth weight and type 2 diabetes

Lauren E. Wedekind, MSc

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Biomedical Sciences (NIH-OU)

Hillary Term 2022

Acknowledgements

My grandmother's experiences with diabetes have fuelled my drive to study the disease and its complications. My parents, Jerry and Swan Wedekind; siblings, Karen and Kati Wedekind and Catherine Ngo; and extended family have reminded me of the important things throughout this long journey and stayed close while physically far apart. Others who have made the past few years, supported me, and opened my mind, include: Angela Lulu Zhang, David Kay, Logan Sigua, Jare Fagbemi, Paul Schroeder, Hoang Nguyen, Sneha Saroja Ayyagari, Raga Ayyagari, Evelyn Akuaa Kusi, Akua Nyarko-Odoom, Hannah Lin, Esmael Habtamu, Bezawit Atenafu Beyene, Alexis Kahanu, Vishesh Gupta, Dera Luce, Farzana Kalladi, James Farnhill, Emily Kelly, Farhana Hamid Butt, Toel Koyithara, Jasmine A Mack, Andrew Wills, Coley Lameman, Cassie M Mitchell, Audrey Winkelsas, and Camille Sheets.

My supervisor Robert Hanson is my role model as a future physician-scientist for his curiosity, brilliance, and generosity with time and advice. Mark McCarthy's dedication to clear science communication and global research collaborations has also been inspiring. Anubha Mahajan has enabled me to access data sets from consortia she orchestrated that were instrumental to this thesis research. Other mentors who have greatly impacted my life and the lives of others include: Suzann Pershing, Darvin Scott Smith, Phyllis Tien, Melissa Chambers, Bill Knowler, and Patrick Needham. I would also like to thank the NIH Oxford-Cambridge Scholars Program and NIH/NIDDK Phoenix staff, including Tracy Lovato, Dorota Wasak, and Sharon Johns, as well as the research participants who made these studies possible. The work that went into this thesis was completed between Phoenix, Arizona, US (homelands of the Akimel O'odham, Yavapai, and Hohokam peoples) and Oxford.

Various chapters of this thesis reproduce parts of text of abstracts and articles that are published or currently under consideration for publication in various peer-reviewed journals. Chapter 1 and Chapter 2 reproduces parts of the text published in the article ‘Epidemiology of type 2 diabetes in Indigenous communities in the United States’ (Wedekind et al., 2021).

Chapter 3 of this thesis reproduces parts of the text of a published abstract that was presented at the American Diabetes Association 2019 Scientific Sessions: ‘Genome-wide association study identifies potential associations with birth weight in a Southwestern American Indian population’ (Wedekind et al., 2019). It also reproduces parts of the text of a published abstract that was presented at the American Diabetes Association 2021 Scientific Sessions: ‘Genetic relationships between birth weight and type 2 diabetes’ (Wedekind et al., 2021). Both abstracts are published in *Diabetes*.

Chapter 4 reproduces parts of the text of an abstract that was accepted for an oral presentation at the American Diabetes Association 2022 Scientific Sessions, as of the submission of this thesis: ‘Associations between type 2 diabetes partitioned/process-specific polygenic scores and metabolic traits’ and subsequent publication in abstract form in *Diabetes* (Wedekind et al., 2022).

Chapter 5 reproduces parts of the text from an original article in *Diabetologia*: ‘The utility of a type 2 diabetes polygenic score in addition to clinical variables for the prediction of type 2 diabetes incidence in birth, youth, and adult cohorts in an Indigenous population’ (Wedekind et al., 2023).

Any work that is not wholly my own has been indicated in the relevant sections.

Abstract

Type 2 diabetes (T2D) is a complex condition that is characterised by hyperglycaemia and associated with over 400 independent genetic variants to date, as well as non-genetic factors. Birth weight, an indicator of intrauterine nutrition and predictor of infant survival, is also associated with the risk of T2D in adulthood. In multiple study populations, the association between birth weight and the risk of subsequent T2D forms a U-shaped curve, with lower and higher birth weights associated with greater risk of T2D. This complex relationship between birth weight and subsequent risk of T2D is influenced by an interplay of genetic and non-genetic influences.

To explore the genetics of birth weight and T2D and to investigate the relationship between these traits, I conducted genome-wide association studies (GWAS) of birth weight. Seven different models of GWAS of birth weight were conducted, using various combinations of covariates, imputed, or directly genotyped variants, and maternal and foetal genotypes. None of the GWAS models yielded genome-wide significant ($p < 5.0 \times 10^{-8}$) associations. Certain signals that have high levels of significance across multiple models could still merit further exploration.

I also quantified the relationship between polygenic scores (PS) that were constructed using clusters of T2D-associated genetic variants and T2D and related phenotypes. These analyses were based on data that had been collected in a longitudinal study of diabetes, which was conducted in an Indigenous study population from the Southwestern US between 1965 and 2007 and had data for birth weight, T2D, T2D subphenotypes, population substructure, and genotype. Partitioned/process-specific polygenic scores (pPS) were constructed using genotypes from the longitudinal study population and genetic clusters that were identified by Mahajan et al to be associated with T2D subphenotypes. PS were constructed using genotypes from the longitudinal study population and summary statistics from meta-analyses of GWAS of T2D. Results broadly agree with putative pathophysiological mechanisms that underlie each genetic cluster by Mahajan et al.

I also investigated the prediction of the incidence of T2D using clinical and genetic variables in the longitudinal study. The performances of predictive models including various combinations of T2D-related clinical variables and T2D PS were assessed and compared. All T2D PS were significantly associated with T2D and, when added to models including clinical variables alone, exhibited a significant but small improvement in the prediction of T2D incidence.

This suggests that further improvements in understanding the genetics of T2D and related traits can result in improved understanding of diabetes pathophysiology and clinical applications of PS and pPS.

Contents

Table of Abbreviations

Acronym	Terminology
2hPG	2-hour plasma glucose
β-cell	Pancreatic beta cell
λ	Genomic inflation factor
ADA	American Diabetes Association
AF	Atrial fibrillation
AFR	African ancestry population as defined in the 1000 Genomes Project
AGEN	Asian Genetic Epidemiology Network Consortium
AI	American Indian
AIR	Acute insulin response
AN	Alaska Native
AUC	Area under the receiver operating characteristic curve
BBJ	BioBank Japan
BMI	Body mass index
bNMF	Bayesian nonnegative matrix factorisation
Bwt	Birth weight
CAD	Coronary artery disease
CHARGE	Cohorts for Heart and Aging Research in Genomic Epidemiology
Chr	Chromosome
CIR30	30-minute corrected insulin response
CIR120	120-minute corrected insulin response
CKD	Chronic kidney disease
DIAGRAM	DIAbetes Genetics Replication and Meta-Analysis Consortium
DIAMANTE	DIAbetes Meta-Analysis of Trans-Ethnic association studies Consortium
DN	Diabetic nephropathy
DNA	Deoxyribonucleic acid
DOHaD	Developmental Origins of Health and Disease hypothesis
DR	Diabetic retinopathy
EAS	East Asian ancestry population as defined in the 1000 Genomes Project
eGFR	Estimated glomerular filtration rate
EGG	Early Growth Genetics Consortium
ESRD	End-stage renal disease
EUR	European ancestry population as defined in the 1000 Genomes Project
FPG	Fasting plasma glucose
GC	Genetic cluster
GC1-BMI	Genetic cluster 1, with loci with primary effects on body mass index
GC2-lipids	Genetic cluster 2, with loci with primary effects on lipids
GC3-IA	Genetic cluster 3, with loci with primary effects on insulin action
GC4-BCF	Genetic cluster 4, with loci with primary effects on beta-cell function

GC5-BCF	Genetic cluster 5, with loci with primary effects on beta-cell function
GC6-MIX	Genetic cluster 6, with loci with mixed features
GCK-MODY	Maturity-onset diabetes of the young that results from mutations in the glucokinase (*GCK)* gene
GCTA	Genome-wide Complex Trait Analysis
GDM	Gestational diabetes mellitus
gePS	Global extended polygenic score
GRCh37	Genome Reference Consortium Human Build 37
GT	Categorical variable denoting whether a genetic variant was directly genotyped or imputed
GWAS	Genome-wide association studies
HAPO	Hyperglycaemia and Adverse Pregnancy Outcomes
HbA1c	Haemoglobin A1c
HDL	High-density lipoprotein cholesterol
H-E clamp	Hyperinsulinaemic euglycaemic clamp
hg19	Genome Reference Consortium Human Build 37
HHI	Hyperinsulinemic hyperglycaemia of infancy
HLA	Human leukocyte antigen
HOMA-β	Homeostatic model assessment of β cell function
HOMA-IR	Homeostatic model assessment of insulin resistance
HTN	Hypertension
IDDM	Insulin-dependent diabetes mellitus
IDF	International Diabetes Federation
IFG	Impaired fasting glucose
IGT	Impaired glucose tolerance
Imp	Imputed variant
INUTERO	Binary variable indicating whether a mother was diagnosed with diabetes before the birth of the offspring
IVGTT	Intravenous glucose tolerance test
kb	Kilobase
KOGES	Korean Genome and Epidemiology Study
LD	Linkage disequilibrium
LDPred	Linkage disequilibrium prediction program for deriving polygenic scores based on summary statistics
LDSC	Linkage Disequilibrium Score regression
logAIR	The natural logarithm of acute insulin response as measured using an intravenous glucose tolerance test in the PECRB study
logM	The natural logarithm of the M value, which was measured using a hyperglycaemic euglycaemic clamp in the PECRB study
logmaxBMI	The natural logarithm of the maximum BMI recorded for an individual in the PECRB cohort
MAF	Minor allele frequency
MAR	Missing at random
MCAR	Missing completely at random
MEDIA	Meta-analysis of type 2 Diabetes in African Americans
M-GCTA	Maternal effects Genome-wide Complex Trait Analysis

MNAR	Missing not at random
MR	Mendelian randomisation
MODY	Maturity-onset diabetes of the young
MVP	Million Veteran Program
MXL	Mexican/Latin American ancestry population as defined in the 1000 Genomes Project
NHW	Non-Hispanic white
NIDDK	National Institute of Diabetes and Digestive and Kidney Diseases
NIH	National Institutes of Health
NOT_IUE	Binary variable indicating whether the participant's mother had a nondiabetic examination at least one year after the child's birth
OR	Odds ratio
PAGE	Population Architecture Genomics and Epidemiology
PC	Principal component
PECRB	Phoenix Epidemiology and Clinical Research Branch
PFAT	Percentage body fat
pPS	Partitioned/process-specific polygenic score
PRSice	Polygenic Risk Score software for calculating, applying, evaluating, and plotting the results of polygenic risk scores
PS	Polygenic score
p_t	Threshold probability
Qinfo	Imputation quality score
QQ	Quantile-quantile
r	Pearson correlation coefficient
ROC	Receiver operating characteristic curve
rs	Reference single nucleotide polymorphism identification number
rsPS	Restricted-to-significant polygenic score
SAS	South Asian ancestry population as defined in the 1000 Genomes Project
SD	Standard deviation
SE	Standard error
SIGMA	Slim Initiative in Genomic Medicine for the Americas
SNP	Single-nucleotide polymorphism
SOLAR-Eclipse	Extension of Sequential Oligogenetic Linkage Analysis Routines (SOLAR) software
T1D	Type 1 diabetes
T2D	Type 2 diabetes
T2D-GENES	Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples
TC	Total cholesterol
TG	Triglycerides
UKB	UK Biobank
US	United States
WC	Waist circumference
WHO	World Health Organisation

1. Introduction

1.1. Clinical presentation of diabetes

Diabetes mellitus is a group of chronic metabolic disorders that are characterized by elevated blood glucose (i.e., hyperglycaemia). Symptoms of hyperglycaemia include polyuria, polydipsia, weight loss, and blurred vision (American Diabetes Association, 2014). Diabetes can be diagnosed based on plasma glucose—fasting or two-hour plasma glucose during a 75-gram oral glucose tolerance test (OGTT)—or haemoglobin A1c (HbA1c). Several pathogenic processes that cause deficient insulin action and impaired insulin secretion are involved in the development of diabetes; these processes frequently coexist in individual patients (American Diabetes Association, 2014). Diabetes is officially classified into multiple categories, including type 1 and type 2 diabetes (T1D and T2D), gestational diabetes mellitus (GDM), monogenic diabetes syndromes, and specific types of diabetes due to other causes (Figure 1.1).

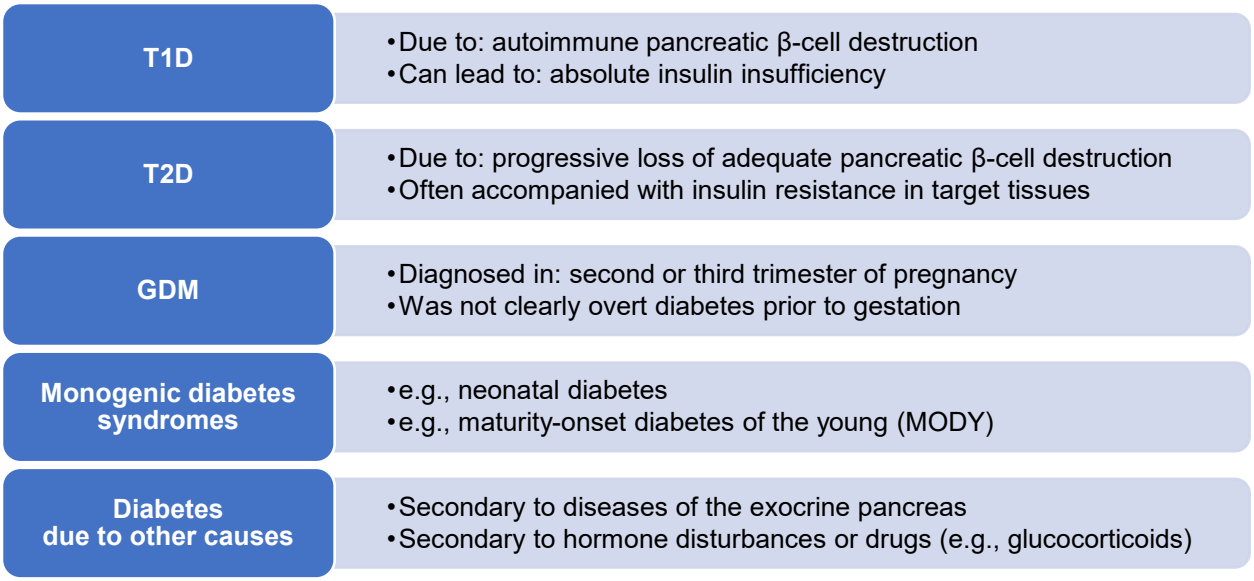


Figure 1.1. A non-exhaustive list of common forms of diabetes, adapted from the ADA *Standards of Medical Care in Diabetes—2021* (American Diabetes Association, 2021).

People with diabetes and subclinical dysglycaemia are at risk for the development of diabetes complications, as well as psychological comorbidities, over the life course (Haines et al., 2007; Pavkov et al., 2008). Such complications can be separated into macrovascular (i.e., involving larger blood vessels) and microvascular (i.e., involving smaller vessels) categories. Macrovascular complications primarily affect arteries and include cardiovascular disease, cerebrovascular disease, and peripheral artery disease; microvascular complications affect microcirculation (i.e., arterioles, capillaries and venules) and include diabetic retinopathy, diabetic neuropathy, and diabetic nephropathy (Fowler, 2011).

T1D, formerly known as insulin-dependent diabetes mellitus (IDDM), is an autoimmune condition that follows the onset of one or more pancreatic beta cell islet autoantibodies and results from the interplay of genetic and environmental influences (Hakonarson & Grant, 2011; Orban et al., 2009). Notably, early genome-wide linkage scans demonstrated that a substantial proportion of familial aggregation of T1D can be attributed to variation in human leukocyte antigen (HLA) loci on chromosome 6p21 (Concannon et al., 2005). HLA plays a key role in the adaptive immune response, and multiple HLA loci are associated with T1D (Krischer et al., 2017).

T2D, formerly known as non-insulin-dependent diabetes mellitus (NIDDM), is characterized by hyperglycaemia--which is due to reduced sensitivity to insulin action (i.e., insulin resistance)--as well as pancreatic β-cell dysfunction that impairs the ability of these cells to secrete sufficient insulin to overcome this insulin resistance (Fonseca, 2009; Cerf, 2013; Campbell, 2000). Pancreatic β-cell dysfunction can be exhibited by a substantial decrease in β-cell number and mass and corresponding reduced insulin secretion (Butler et al., 2003; Del Guerra et al., 2005). In the case of insulin resistance, insulin may still be present but target tissues (e.g., adipose) exhibit insensitivity to such insulin. T2D has substantial variability in clinical presentation and response to treatment (Schwartz et al., 2016). The diagnosis of T2D

is often cited as one of ‘exclusion’: relatively reliable but imperfect diagnostic markers for T1D (pancreatic autoantibodies in the blood) and MODY (single gene defects in known monogenic diabetes genes) for other types of diabetes exist; however, there is no similar diagnostic test for T2D (Pearson, 2019). Multiple pathophysiological processes often contribute to the clinical presentation of complex conditions such as T2D (McCarthy, 2017).

The heterogeneity of T2D has prompted studies that employ cluster analyses and data for various related biomarkers to identify structure within the condition. If such T2D subtypes are properly validated and require substantially different disease management, then classifying research participants or patients according to their subtype could facilitate the optimisation of their treatment. Using 73 variables from linked electronic medical records and linked genotypic data from Mount Sinai Medical Centre for 2551 individuals with T2D, Li et al developed a data-driven, topology-based approach that identified three T2D subtypes (Li et al., 2015). Each of these proposed subtypes was associated with increased risk of various disease comorbidities and over 300 unique genetic associations. However, external validation was not conducted, this study has not been replicated, and the clinical utility of these proposed T2D subtypes is unclear. In the Swedish All New Diabetics in Scandia cohort comprised of individuals with adult-onset diabetes, Ahlqvist et al employed data-driven (k-means) clustering and identified five replicable subgroups of patients using six biomarkers that were measured at the time of their diagnosis: glutamate decarboxylase antibodies, age at diagnosis, body mass index (BMI), HbA1c, and homeostatic model assessment 2 estimates of β-cell function (HOMA-β) and insulin resistance (HOMA-IR) (Ahlqvist et al., 2018). A validation study that included individuals with newly diagnosed and existing diabetes replicated those five diabetes subtypes: severe auto-immune diabetes, severe insulin-deficient diabetes, severe insulin-resistant diabetes, mild obesity-related diabetes, and mild age-related diabetes (Pigeyre et al., 2021).

Differences in underlying pathophysiology of the T2D subtypes proposed by Ahlqvist et al—and the clinical utility of these distinctions—remain dubious, without extensive replication in other study populations. However, Dennis et al employed the same data-driven cluster analysis as Ahlqvist et al within the ADOPT trial (N = 4,351 participants) and found similarities in clusters generated and differences in glycaemic progression and treatment response (Dennis et al., 2019). Dennis et al highlighted that clustering can result in loss of information as compared with using 'simple' prediction models comprised of commonly measured, continuous clinical variables alone (Dennis et al., 2019). Further, the technique of k-means clustering forces individuals into only one cluster, even if other clusters fit marginally less well. This prompts further investigation into two issues: the stability of these cluster assignments over time must be addressed in longitudinal study data, and further validation studies must be performed.

Previous studies have shown that hyperglycaemia often precedes the diagnosis of T2D by multiple years (Færch et al., 2013). Prediabetes, a state of intermediate glycaemia, has different definitions according to different health organisations. For example, the World Health Organisation (WHO) currently defines prediabetes as having impaired fasting glucose (IFG; fasting plasma glucose (FPG) of 6.1-6.9 mmol/L = 110-125 mg/dL) and/or impaired glucose tolerance (IGT; 2-hour plasma glucose (2hPG) 7.8-11.0 mmol/L = 140-200 mg/dL) after a 75-gram oral glucose load (World Health Organization & International Diabetes Federation, 2006). By contrast, the respective definitions for IFG and IGT for the diagnosis of prediabetes as currently set by the American Diabetes Association (ADA) include the same 7.8-11.0 mmol/L = 140-200 mg/dL range for IGT and a lower cut-off value for IFG (5.6-6.9 mmol/L = 100-125 mg/dL) (American Diabetes Association, 2020). This lower cut-off value for IFG in the ADA criteria for prediabetes affects the calculated incidence rate of T2D among those with prediabetes, with lower incidence in study participants that are diagnosed with

prediabetes on the basis of the ADA criteria, as compared with the WHO criteria (Forouhi et al., 2006). An early meta-analysis evaluated the progression of prediabetes to T2D and reported that the annual incidence rate of T2D was 6-9% for participants with IFG alone, 4-6% for those with IGT alone, and 15-19% for those with both IFG and IGT (Gerstein et al., 2007). A notable limitation of the meta-analysis was that the criteria for IGT and IFG differed among included studies; for instance, they employed the WHO 1980, 1985, 1998, and 1999 definitions; as well as the ADA 1997 definition (Gerstein et al., 2007).

The previous longstanding definition for gestational diabetes mellitus (GDM) was any degree of glucose intolerance that was first recognised during pregnancy, irrespective of the degree of hyperglycaemia (Zhang et al., 2010). A recent study within the RADIEL intervention trial, among pregnant patients who were at elevated risk for developing GDM with three follow-up visits within a 12-month period, indicated that many cases of GDM are pre-existing hyperglycaemia that is detected by routine screenings during pregnancy (Huvinen et al., 2018). Such routine screenings are not widely performed in non-pregnant people of reproductive age (Huvinen et al., 2018). More generally, there are a substantial number of pregnant people with undiagnosed T2D, often due to underlying β-cell dysfunction (American Diabetes Association, 2020). Therefore, testing those with risk factors for T2D early in pregnancy could lead to earlier management of the condition throughout and after pregnancy (American Diabetes Association, 2020; Mission et al., 2017). Monogenic diabetes syndromes encompass single gene defects that cause diabetes, and include neonatal diabetes (1 in 100,000 births) and maturity-onset diabetes of the young (MODY; approximately 1 to 4% of all cases of diabetes in those who are diagnosed before the age of 30) (Johansson et al., 2017; Shields et al., 2017). The most common of these single gene defects occur in the genes *HNF1A*, *HNF4A*, *HNF1B*, and *GCK* (Johansson et al., 2017; Shields et al., 2017).

1.2. Epidemiology of diabetes

The global prevalence of diabetes has increased substantially in recent decades and is projected by multiple international organizations to further increase in the coming decades (Table 1.1). For instance, as of 2019, the International Diabetes Federation (IDF) estimated that the global diabetes burden among adults aged 20-79 years was 463 million (9.3%); by 2045, this figure is projected to increase by 51% to 700 million (10.9%) (Saeedi et al., 2019).

Table 1.1. World Health Organisation (WHO) and International Diabetes Federation (IDF) estimates or projections of the global burden of diabetes in adults in years 2000 to 2045.

Source (citation)	Year for figure	Estimated or projected global count of diabetes in adults
WHO (Wild et al., 2004)	2000	171 million
IDF (International Diabetes Federation, 2000)	2000	151 million
IDF (International Diabetes Federation, 2003)	2003	194 million
IDF (International Diabetes Federation, 2006)	2006	246 million
IDF (International Diabetes Federation, 2009)	2009	285 million
IDF (International Diabetes Federation, 2011)	2011	366 million
IDF (International Diabetes Federation, 2013)	2013	382 million
IDF (Ogurtsova et al., 2017)	2015	415 million
IDF (Cho et al., 2018)	2017	451 million
IDF (Saeedi et al., 2019)	2019	463 million
WHO (Wild et al., 2004)	2030	366 million
IDF (Saeedi et al., 2019)	2030	578 million
IDF (Ogurtsova et al., 2017)	2040	642 million
IDF (Saeedi et al., 2019)	2045	700 million

There are substantial country-level and regional disparities in undiagnosed diabetes and the total burden of diabetes. In 2019, 50.1% of diabetes worldwide was estimated to be undiagnosed; the regions with the greatest proportion of undiagnosed diabetes were Africa

(59.7%) and Southeast Asia (56.7%) and those with the least were Europe (40.7%) and North America and the Caribbean (37.8%) (Saeedi et al., 2019). These disparities were largely attributed to differences in health care services across country-level income classifications.

According to a nationwide, population-based, cross-sectional survey among adults in the United States (US) in 2016 and 2017, approximately 5% of prevalent diabetes comprised type 1 diabetes (T1D), 92% comprised T2D, and 3% comprised other subtypes of diabetes (Xu et al., 2018). T1D accounts for the majority of diabetes cases in people under 20 years of age in most populations (Ang, 2020). Investigations within the SEARCH for Diabetes in Youth study have shown that T1D is more prevalent than T2D in youth in most racial/ethnic groups in the United States (US), except Indigenous (American Indian and Alaska Native, AI/AN) peoples (Hamman et al., 2014).

T2D constitutes the vast majority (over 90%) of the total global burden of diabetes. Multiple genetic and non-genetic factors influence an individual's predisposition to T2D (McCarthy, 2017). Risk factors for type 2 diabetes include increased adiposity, decreased physical activity, age, hypertension, and gestational diabetes (Bellou et al., 2018). The increased counts of overall diabetes published by such institutions as the WHO and IDF in recent decades largely reflect the increased burden of type 2 diabetes. Between 1990 and 2017, the global prevalence of T2D rose substantially, with greater increases in lower-income countries; in 2017, an estimated 462 million individuals had type 2 diabetes (6.3% of the global population) (Khan et al., 2020). Early-onset T2D is increasingly prevalent, particularly among adolescents who are overweight or have obesity, with estimated incidence rates currently increasing by multiple percentage points per year (Pyle & Kelsey, 2021). Younger age of onset of T2D is associated with more severe disease, as measured by total mortality, cardiovascular-related mortality and non-cardiovascular mortality (Sattar et al., 2019).

Between 1990 and 2013, the global burden of diabetes, as assessed by mortality and years of life lost due to diabetes, increased by over 50%; however, this estimated increase in mortality is likely an underestimate, due to the study's reliance on death certificate data that listed diabetes as primary cause of death, rather than associated macrovascular complications (Naghavi et al., 2015). Diabetes-related chronic kidney disease (CKD) is the leading cause of end-stage renal disease (ESRD), which contributes to approximately 10% of deaths in people with T2D; increases in the prevalence of diabetic nephropathy (DN) such as these are expected with the increase of the global burden of T2D, particularly in lower and middle income countries (Saran et al., 2019; Tuttle et al., 2014).

1.3. Epidemiology of T2D in Indigenous communities in the US

This section provides an overview of epidemiology of T2D in Indigenous communities (i.e., first peoples) in the US to provide context for Section 1.6, which describes the longitudinal study in the Southwestern US in which the analyses outlined in this thesis have been undertaken. Indigenous peoples in the present-day US, including members of 574 federally recognized tribes and other first peoples of the US and territories, have experienced historical and ongoing challenges to health and its determinants. Due to the tumultuous and complicated relationships between Indigenous nations and the rest of the US, increasing participation in research studies among these populations is complex. The dynamic between Indigenous populations and the federal government in the US is often described by 'colonization,' a term that encompasses the physical and theoretical framework of interactions among the entities, beginning at first contact between Indigenous peoples and European settlers (Alfred, 2009). Post-contact, European settlers and their descendants who largely took ownership of lands inhabited by Indigenous peoples and power in colonial, federal, and state governments in what became the US used Eurocentric arguments to support forced

assimilation and cultural genocide of Indigenous peoples (Braun et al., 2014). At the most basic level, careful consideration of engagement with Indigenous individuals and communities begins with terminology.

There has been considerable debate regarding appropriate terminology in discussing and engaging with Indigenous peoples; historically in the US, such terminology has included 'Indian,' 'American Indian,' and 'Native American.' Name designation with respect to Indigenous communities is especially convoluted (Davis-Delano et al., 2020). Name preferences vary depending on the individual and community, although the consensus among communities is to ask each their preference, which may be in their native tongue (e.g., 'Diné' versus the colonial name 'Navajo'). This thesis uses a range of terms based on the context. The terms 'Indian,' 'American Indian (AI),' and 'Native American' are based on inaccurate declarations and applied inconsistently (e.g., toward Indigenous peoples of Alaska, Hawaii, Guam, Puerto Rico, US Virgin Islands, etc.); however, these terms are often used in bureaucratic settings. For clarity, the broadly applicable term 'Indigenous' has been used throughout this thesis where applicable, to recognize the first peoples in the US and other countries. Subgroups that are included in relevant studies will be specified.

Various classification systems encompassing race, ethnicity, and ancestry groups have been used to categorise humans in epidemiological studies and in clinical settings with limited consistency across countries and over time (Carter-Pokras et al., 2012). Racial and ethnic categorisations that have been developed and employed within the US have been particularly imprecise, and changes in census categorisations for race and ethnicity have been extensive across decades (Strmic-Pawl et al., 2017). For example, the US government continues to use blood quantum—which was originally formulated as a tool to facilitate genocide—to measure the degree of Indigenous ancestry that Indigenous peoples possess; this is often printed on individuals' tribal identification cards (Rodriguez-Lonebear, 2021). Within research settings,

certain investigators who have conducted research within Indigenous communities employ this measure of blood quantum as a covariate in analyses to approximate admixture (TallBear, 2013). While addressing population stratification and admixture is necessary for methodologically sound genome-wide association study (GWAS) in diverse study populations, historical context and the pitfalls of using such covariates as blood quantum in research studies must be considered (Garrison et al., 2019).

Certain imprecise categorisation systems and communication about them in scientific literature have also been rooted in and perpetuated historical racial stereotypes. For example, a race correction factor that differentiates Black and non-Black patients and research participants in the US has been employed in calculating estimated glomerular filtration rate (eGFR), a measure of kidney function (Eneanya et al., 2019). The race correction factor emerged from methodologically flawed research studies that showed that Black participants had physiological differences in muscle mass as compared with non-Black people (Ge et al., 2021)—which had been historically used to justify the enslavement of Black peoples in the US (Guillory, 1968). This has led to systematically lower access to life-saving care such as kidney transplantation for Black patients with ESRD (determined by eGFR thresholds) than for their non-Black counterparts (Epstein et al., 2009).

While disparities across such groups are often cited in articles on the epidemiology of complex traits such as T2D, the groups alone provide an inadequate proxy for the many genetic and non-genetic influences on T2D risk, such as access to health care (Campbell et al., 2012; Kirk et al., 2008). While race and ethnicity correlate with certain genetic variants, ancestry is often defined on a global scale that is less convoluted and dynamic than those within the US and is more precise than self-reported race and ethnicity (Ali-Khan et al., 2011). In studies of genetic associations with T2D, meta-analyses of GWAS that include participants of different ancestry groups have played an increasingly important role in detecting variants

that are lower in frequency or monomorphic in European-ancestry study populations, who are overly represented in the literature (Peterson et al., 2019). More precisely defined descriptors are needed to ensure that policies and clinical standards that are intended to address disparities in the risk of T2D and other complex diseases are based on sound evidence that employs precise, consistent research methodologies.

Previous studies have reported the relatively high prevalence of diabetes, especially T2D, among Indigenous peoples in the US According to the 2020 National Diabetes Statistics Report published by the Centers for Disease Control and Prevention, 34.2 million people in the US are estimated to have diabetes (10.5% of the adult population); of this figure, 7.3 million people (21.4%) are estimated to have undiagnosed diabetes (Centers for Disease Control and Prevention, 2020). In the National Health Interview Survey (NHIS), in 2011-2015, the age-standardised prevalence of self-reported diagnosed diabetes was 2.2 times higher in Indigenous (American Indian/Alaska Native, AI/AN) than in non-Hispanic white (NHW) men, and 2.5 times higher in Indigenous than in NHW women (Table 1.2) (Cowie et al., 2018).

Table 1.2. Age-standardised prevalence of diagnosed diabetes in adults in the US in the NHIS, 2011-2015 (Cowie et al., 2018). This table was reproduced from Cowie et al., 2018; Diabetes in America, published by the National Institute of Diabetes and Digestive and Kidney Diseases, is in the public domain of the United States.

Ethnic group	Diabetes prevalence in men (95% CI)	Diabetes prevalence in women (95% CI)
Non-Hispanic Indigenous	19.7% (15.9-24.1)	18.8% (13.7-25.2)
Non-Hispanic Black	14.0% (13.7-14.8)	14.3% (13.6-15.0)
Hispanic	13.9% (13.1-14.8)	13.2% (12.4-14.0)
Non-Hispanic Asian	10.4% (9.3-11.7)	8.3% (7.5-9.1)
Non-Hispanic White	9.0% (8.7-9.3)	7.6% (7.3-7.9)

In the National Health Interview Survey, age-sex standardised prevalence of diagnosed diabetes in adults aged at least 20 years was 19.1% in Indigenous populations, as compared with 9.5% in the general population (Cowie et al., 2018).

Among Indigenous communities in the US—a demographic that currently constitutes approximately 5.2 million individuals (1.6% of the US population)—documented cases of diabetes began to increase drastically in the 20th century (Will et al., 1997). Previous studies on cardiometabolic health outcomes in certain Indigenous communities in the US have reported that nearly all cases of diabetes in participants were clinically characterized as T2D (Dabelea et al., 1999). While few such investigations exist, along similar lines, studies in Indigenous communities in the Southwestern US and Alaska report that nearly all diabetes in study populations was T2D, based on the absence of T1D-related autoantibodies and clinical characteristics (Knowler et al., 1979; Mohatt et al., 2002). A recent observational study on the incidence and prevalence of diagnosed diabetes (T1D and T2D combined) was conducted on Indian Health Service health records of Indigenous (defined in that study as American Indian/Alaska Native, AI/AN) adults from 2006-2017 (Bullock et al., 2020). Therein, a trend analysis found that the prevalence of diagnosed diabetes increased from 2006 (14.4%) to 2013 (15.4%) and decreased from 2013 to 2017 (14.6%) (Bullock et al., 2020). These trends were consistent across all age groups (18-44 years, 45-64 years, 65-74 years and ≥75 years); however, these trends have not been replicated and could originate in part from sampling or other technical reasons (Bullock et al., 2020).

There are limited data on prevalence of prediabetes, but several studies have indicated that prevalence of prediabetes—defined as fasting plasma glucose > 100-125 mg/dL or HbA1c 5.7-6.4%, inclusive—in Indigenous populations is similar to, or slightly lower than, in the general population (Centers for Disease Control and Prevention, 2020). For example, in a consortium of US health care systems including nearly five million individuals, over 25,000

individuals who identified themselves in health records, or who were identified on their birth certificate, as AI/AN, had lower prevalence of prediabetes (31.1%; 95% CI 30.6-31.7%) as compared with the overall study population (33.4%; 95% CI 33.3-33.5%) (Zhu et al., 2019). In this study, the age-adjusted prevalence of diabetes was 19.6% (95% CI 19.1-20.0%) in AI/AN individuals, versus 15.9% (95% CI 15.8-16.0%) in the overall population (Zhu et al., 2019). Non-white groups in the study had higher rates of diabetes and prediabetes at lower BMIs than whites and the overall population (Zhu et al., 2019); thus, factors other than obesity (as measured by BMI) may have contributed to contribute to both diabetes and prediabetes risk in these groups.

According to the National Health Interview Survey, AI/AN adults are 1.5 times as likely to have obesity relative to NHW adults, and Indigenous adolescents are 30% more likely to have obesity than NHW adolescents (Centers for Disease Control and Prevention, 2018). While further research on fat distribution and cardiometabolic risk is needed and BMI is not necessarily the optimal predictor of disease risk, there is strong evidence for the association between obesity and hypertension, dyslipidaemia, hypercholesterolemia, insulin resistance, and glucose intolerance (Saxton et al., 2019). Longitudinal studies in Indigenous populations have shown that obesity, as measured by BMI, is strongly and linearly associated with incidence of diabetes and that diabetes incidence is low in those with BMI < 25 kg/m^2 (Figure 1.2) (Knowler et al., 1981).

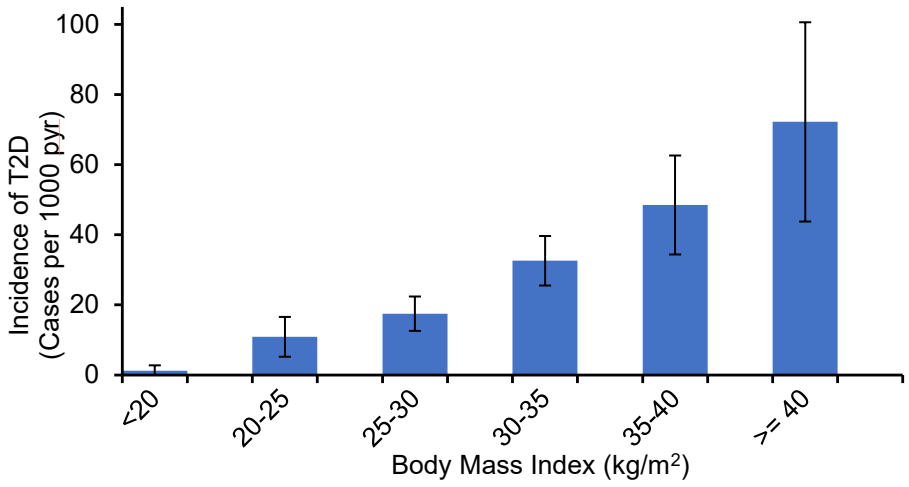


Figure 1.2. Age-sex standardised incidence of type 2 diabetes by body mass index, with 95% confidence intervals, in a longitudinal study of Indigenous adults from a community in the Southwestern US (Knowler et al., 1981). This figure was reproduced with Copyright Clearance Centre licence number 5263240028580.

The relatively high prevalence of T2D in Indigenous populations is particularly pronounced in youth. An analysis of time trends in diagnosed diabetes (including T1D and T2D) prevalence among AI and AN young adults (aged 18-34 years) between 1994 and 2007 found increasing trends in annual prevalence estimates of diagnosed diabetes among Indigenous (AI/AN, and those from within the 50 US, Puerto Rico, Guam, and the Virgin Islands) and NHW young adults; however, these findings have not been replicated (Hamman et al., 2014). Indigenous young adults were, on average, 1.7 (95% CI 1.1-2.6) times more likely than NHW young adults to be diagnosed with diabetes during 1994-2000 and 2.5 (95% CI 1.9-3.3) times more likely during 2001-2007 (Roberts et al., 2009). In the SEARCH for Diabetes in Youth study, a population-based assessment of prevalence in youth aged 10-19 years across 5 different American centres as well as several tribal reservation communities, the prevalence of diagnosed T2D was 1.20 per 1000 Indigenous participants, compared with 0.17 per 1000 NHW participants, 0.34 per 1000 Asian and Pacific Islander participants, 0.79 per 1000 Hispanic participants, and 1.06 per 1000 Black participants (Dabelea et al., 2014).

Thus far, this section has focused upon T2D prevalence, which is essentially the cumulative incidence of T2D in a given population, not including deaths that may occur in that population over time. Shifting to discuss the incidence of T2D, previous studies have found that the incidence of young-onset T2D in Indigenous children, adolescents and young adults was greater than that of the overall US population in any given period. A longitudinal study of Indigenous young adults within a larger general US study population found that Indigenous young adults had higher HbA1c levels, higher self-reported blood glucose, self-reported diabetes, and obesity (Marley & Metzger, 2015). They also had higher rates of structural risk factors (e.g., residing in poorer and more transient neighbourhoods and having greater levels of stress) than NHW young adults (Marley & Metzger, 2015). The relatively high prevalence and incidence of early-onset T2D in Indigenous youth relative to that of all youth in the US poses challenges to the prevention of complications of diabetes in Indigenous communities.

While the prevalence of DN—including elevated microalbuminuria, macroalbuminuria, and end-stage kidney disease—is relatively high in Indigenous communities in the US, its incidence has decreased within this demographic group in recent years (Pavkov et al., 2008). The public health impact of diabetes on kidney disease is particularly severe in Indigenous communities; diabetes is the primary cause of 67% of the ESRD that occurs in Indigenous (AI/AN) populations, as compared with 39% in the general US population (Saran et al., 2020). In a longitudinal population-based study among members of an Indigenous community from the Southwestern US, early-onset T2D was associated with substantially increased incidence of ESRD and mortality in middle age (25-55 years); the longer duration of T2D experienced by individuals diagnosed with T2D at under 20 years of age predominantly accounted for this burden (Pavkov et al., 2006). Early-onset T2D, and the increased prevalence of childhood obesity among Indigenous communities as well as the broader US population that contribute

to individuals' risk of developing T2D, are strongly associated with DN and pose major challenges to its prevention (Pavkov et al., 2006, 2007).

Among Indigenous peoples with diabetes in the US, prevalence estimates for diabetic retinopathy (DR) are often reported for specific tribal communities. For example, among 418 Lakota participants in the Strong Heart Study from the Cheyenne River Sioux Tribe and Oglala Sioux Tribe who had T2D (aged 45-75 years), 45.3% had DR: 40.2% had proliferative DR (mean duration of T2D 12.3 years, SD 7.5 years) and 5.1% had non-proliferative DR (mean duration of T2D 14.2 years, SD 7.5 years) (Berinstein et al., 1997). In the longitudinal study in which this thesis research was conducted, among study participants from an Indigenous community from the Southwestern US, among 171 participants who had T2D (aged 30-70 years), 40.9% had DR (aged mean 51.6 ± 8.8 years; median duration of T2D 15.0 years, range 0.2-26.1 years) and 2.3% had proliferative DR (Nagi et al., 1997). According to a larger multi-site, cross-sectional study including 45,482 AI/AN participants with diabetes recruited at Indian Health Service-Joslin Vision Network Teleophthalmology Program clinics, 20.0% of patients had DR: 17.7% had non-proliferative DR and 2.3% had proliferative DR (Bursell et al., 2018).

1.4. Genetics of T2D

Estimates of the heritability of T2D, derived from twin and family studies, range from 20 to 80%, capturing combinations of genetic and nongenetic (e.g., shared familial, prenatal, and postnatal environmental exposures) factors (Ali, 2013; Almgren et al., 2011; K. Wang et al., 2017; Willemsen et al., 2015). Studies in Indigenous populations in the US have reported broad ranges of heritability for T2D—ranging from 22 to 62% (Baier & Hanson, 2004; North et al., 2003). These heritability estimates reflect phenotypic resemblance among family

members and illustrate potential genetic effects on T2D, although they can also reflect clustering of environmental risk factors within families, depending on study design. The advent of GWAS has in the past 15 years provided a powerful approach for identifying genetic variants that contribute to T2D risk. GWAS have also enabled estimates of narrow-sense heritability (h^2) that are able to distinguish between genetic and shared environmental effects. Researchers have developed methods to estimate a lower bound on heritability directly from GWAS data, using genetic similarities between unrelated individuals (Lee & Chow, 2014). Further information on heritability estimates is provided in Section 2.3.

Early GWAS mostly captured common variants associated with T2D; GWAS and GWAS meta-analyses with increasingly large sample sizes, improved quality control measures, and density of imputation have yielded increased statistical power to detect rarer variants (McCarthy et al., 2016).

A 2016 study on the genetic architecture of T2D reported that modest effect sizes of T2D-associated variants limit the collective contribution of low-frequency and rare variants to heritability (Fuchsberger et al., 2016). While simulated rare variant models had predicted more rare and low-frequency variants to be associated with T2D, Fuchsberger et al identified only one low-frequency variant whose association with T2D achieved genome-wide significance (Fuchsberger et al., 2016). In the years since, other studies involving large-scale whole-genome sequencing, exome sequencing, and GWAS meta-analyses have further limited the role of low-frequency and rare variants in T2D susceptibility. A 2017 study by Jun et al that involved whole-genome sequencing of over 1000 individuals across large pedigrees found no evidence of rare variants that associated with T2D with large effect sizes (Jun et al., 2017). A 2018 study that involved a large exome array and fine-mapping study across approximately 450,000 samples further limited the contribution of low-frequency coding variants to T2D susceptibility: the study identified 40 coding variant associations, with only

five with MAF < 5% and none with OR > 1.4 (Mahajan et al., 2018a). Most GWAS and GWAS meta-analyses have identified T2D risk loci in European-ancestry study populations, such as a recent GWAS meta-analysis by Mahajan et al across a total of approximately 900,000 study participants (74,124 T2D cases and 824,006 controls) (Mahajan et al., 2018b). This GWAS meta-analysis enabled robust detection of over 400 T2D risk signals, including 56 low-frequency (0.5% ≤ minor allele frequency (MAF) < 5%) and rare (MAF < 0.5%) index SNPs (Mahajan et al., 2018b). Most signals detected in this GWAS meta-analysis were common variants with modest effect sizes (Mahajan et al., 2018b). The odds ratios for low-frequency and rare variants detected in this study were mostly modest but ranged from 1.08-8.05, compared to 1.03-1.37 for common (MAF ≥ 5%) variants (Mahajan et al., 2018b).

Additionally, the inclusion of isolated study populations that have altered patterns of linkage disequilibrium (LD) and may have T2D-associated variants of relatively high frequency as compared with that of European-ancestry study populations that are over-represented in published T2D GWAS and meta-analyses (Andersen et al., 2016). This relatively high frequency of novel T2D-associated variants could result from genetic drift improves the power to detect association signals. Dominant-recessive GWAS models, which can also assist in the detection of rarer variants, are further described in Section 3.2.1. Table 1.3 provides an overview of a sampling of large-scale consortia that have conducted GWAS meta-analyses of T2D, including the consortium name, sequencing type, sample size, ancestry groups of study participants, and recent key publications resulting from these consortia (DeForest & Majithia, 2022).

Table 1.3. Summary of large-scale consortia that have conducted GWAS of T2D. Adapted from an article distributed under the terms of the CC Attribution 4.0 International Licence: http://creativecommons.org/licenses/by/4.0 (DeForest & Majithia, 2022).

Abbreviation	Name	Genotyping assay	Sample size	Ancestry	Citation
DIAGRAM	DIAbetes Genetics Replication And Meta-analysis	Genotype	26,676 T2D cases and 132,532 controls	European	(Mahajan et al., 2018b; Scott et al., 2017)
DIAMANTE	DIAbetes Meta-Analysis of Trans-Ethnic association studies	Genotype	180,834 T2D cases and 1,159,055 controls	African American, East Asian, European, Hispanic/Latino, South Asian	(Mahajan et al., 2018a)
SIGMA	Slim Initiative in Genomic Medicine for the Americas	Genotype, exome chip	8,227 T2D cases and 12,966 controls	Hispanic/Latino	(The SIGMA Type 2 Diabetes Consortium et al., 2014)
AGEN	Asian Genetic Epidemiology Network	Genotype	77,418 T2D cases and 356,122 controls	East Asian	(Spracklen et al., 2020)
T2D-GENES	Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples	Whole exome sequencing	20,791 T2D cases and 24,440 controls	Hispanic/Latino, European, African American, East Asian, South Asian	(Flannick et al., 2019)
MEDIA	MEta-analysis of type 2 Diabetes in African Americans	Genotype	8,284 T2D cases and 15,543 controls	African American	(Ng et al., 2014)
MVP	Million Veteran Program	Genotype	102,683 T2D cases and 170,726 controls	Predominantly European; also, African American, Hispanic/Latino, Asian, African, Caribbean	(Vujkovic et al., 2020)
UKB	UK Biobank	Genotype, whole exome, and whole genome	17,730 T2D cases and 210,752 controls	Predominantly European; also African, Hispanic/Latino, Asian, Caribbean	(O'Connor et al., 2022)
BBJ	BioBank Japan	Genotype, whole genome	36,614 T2D cases and 155,150 controls	Japanese	(Suzuki et al., 2019)
PAGE	Population Architecture Genomics and Epidemiology	Genotype	African American: 8,591 T2D cases and 16,887 controls; Asian: 3,124 cases and 4,313 controls; Native Hawaiian: 1,642 cases and 2,152 controls; Hispanic: 9,913 cases and 22,958 controls; European American: 3,156 cases and 14,837 controls	European, African American, Asian, Native Hawaiian, Hispanic/Latino	(Polfus et al., 2021)

Suzuki et al identified 88 T2D-associated loci ($p < 5.0 \times 10^{-8}$; 28 novel loci) with 115 independent signals ($p < 5.0 \times 10^{-8}$; 30 novel signals) in a meta-analysis of four GWAS (36,614 T2D cases and 155,150 controls) that included individuals from population-based cohorts in Japan (Suzuki et al., 2019). To gain biological insights, Suzuki et al also undertook a multi-ancestry comparison of associations of molecular pathways that were suggested in a previous GWAS meta-analysis of European-ancestry study populations (Scott et al., 2017). The analysis by Suzuki et al. highlighted that pathways of beta cell development and MODY were significantly associated with T2D in both European-ancestry and Japanese-ancestry study populations; however, the pathway of regulation of insulin secretion was only significantly associated with T2D in the Japanese-ancestry study population (Suzuki et al., 2019). These findings have not been replicated and lie in contrast to studies in European-ancestry study populations.

A larger GWAS meta-analysis of East Asian-ancestry study populations (77,418 T2D cases and 356,122 controls) identified 301 distinct association signals that met a locus-wide significance threshold ($p < 1.0 \times 10^{-5}$; 228 were genome-wide significant with $p < 5.0 \times 10^{-8}$) at 183 loci (51 novel) (Spracklen et al., 2020). Spracklen et al compared summary statistics from their meta-analysis with those within the European-ancestry T2D GWAS meta-analysis of Mahajan et al (Mahajan et al., 2018b): common variants associated with T2D in the East Asian-ancestry and European-ancestry study populations had strongly correlated effect sizes (Spracklen et al., 2020). Certain novel associations identified in this meta-analysis have higher allele frequencies, larger effect sizes in East Asian-ancestry study populations (Spracklen et al., 2020). This underscores the importance of including more diverse study populations and carefully accounting for population stratification to avoid biased test statistics (Uffelmann et al., 2021). More recently, Vujkovic et al conducted a multi-ancestry GWAS

meta-analysis of 228,499 T2D cases and 1,178,783 controls of European, African American, Hispanic, South Asian, and East Asian ancestry across the Million Veteran Program, DIAMANTE, Biobank Japan and other studies and reported 568 associations (283 novel), including 25 novel associations that were identified in ancestry-specific analyses (Vujkovic et al., 2020).

A recent preprint by Mahajan et al described a multi-ancestry GWAS meta-analysis of T2D in 180,834 T2D cases and 1,159,055 controls (48.9% non-European ancestry) in DIAMANTE, which identified 277 loci at genome-wide significance ($p < 5 \times 10^{-8}$), 11 of which had not been previously reported in recent T2D GWAS meta-analyses and 237 of which met a more stringent multi-ancestry analysis threshold ($p = 5 \times 10^{-9}$) (Mahajan et al., 2022). At the 58.6% of associations that were more precisely localized due to the more diverse study populations, the addition of non-European-ancestry GWAS to the multi-ancestry meta-regression was found to enhance fine-mapping resolution as compared with equally-sized European-ancestry GWAS (Mahajan et al., 2020). In the 2018 GWAS meta-analysis by Mahajan et al, conditional analysis revealed complexity at certain loci, including an observation of multiple association signals at the *TCF7L2* locus that were in weak LD with lead GWAS SNP rs7903146, which has been previously reported as the largest-effect common variant signal for T2D in European-ancestry populations (Mahajan et al., 2018b).

Previous evidence on how genetics researchers and clinical professionals conceptualise race, ethnicity, and ancestry has underscored that definitions of race and ancestry are context-specific and these terms are often used interchangeably (Popejoy et al., 2020). Thus far, evidence has indicated that common variants that are associated with T2D are largely shared across ancestry groups. However, studies have highlighted that study populations identified as the AFR (African ancestry) group experience limited transferability of some common and rare variants that are associated with T2D (Martin et al., 2019a). This has

resulted in calls to diversify GWAS and GWAS meta-analysis study populations, which have been increasingly addressed through efforts such as the DIAMANTE Consortium (Mahajan et al., 2022). Given unethical research practices of scientific researchers with respect to Indigenous study populations in the past (described further in Section 2.1), Indigenous bioethicists and statistical geneticists have developed ethical frameworks regarding the engagement of Indigenous communities in scientific research for scientists to consider (Claw et al., 2018).

1.4.1. Polygenic scores for T2D

Polygenic scores (PS) are used to capture in a single continuous estimate that summarises information from GWAS for a given trait or condition and an individual's genotypes for variants of interest. For example, a PS can be calculated to estimate an individual's genetic predisposition to a given condition. In the past 15 years, PS for complex, polygenic traits such as T2D began to be constructed using genome-wide significant variants that were identified in GWAS—most of which were common variants—and were found to associate strongly with T2D (Hivert et al., 2011; Langothe et al., 2008; Lyssenko et al., 2008; Meigs et al., 2008). These GWAS often use the widely accepted threshold of $p < 5.0 \times 10^{-8}$, which accounts for the testing burden of approximately 1 million independent tests across the genome in European-ancestry data from the International HapMap Consortium (Pe'er et al., 2008). These variants can be combined into a PS to quantify an individual's predisposition to a heritable trait (Wand et al., 2021), calculated as a weighted sum of the number of trait-increasing alleles that an individual carries. The summing across variants inherent to calculating a PS for a trait assumes an additive genetic architecture of the trait and, often, the independence of the included genetic variants (Lewis & Vassos, 2020).

Multiple methods can be used to construct PS and can attributed to three main categories: (1) polygenic scoring based on all genetic markers, (2) clumping/pruning and thresholding as well as (2) assessing the best prediction genome-wide (Lewis & Vassos, 2020). First, polygenic scoring based on all genetic markers employs summary statistics for a GWAS for a given trait as well as genotypes for a study population to calculate a PS across all markers in the summary statistics. Second, clumping/pruning and thresholding is an approach that involves pruning variants based on LD, followed by summing effect alleles over all SNPs that meet one or more p-value thresholds; this method is implemented by programs PRSice (Choi & O'Reilly, 2019) and PLINK (Chang et al., 2015). A recent review article on PS for T2D by Udler et al proposed terming the construction of PS by pruning (i.e., removing highly correlated variants) and thresholding (i.e., employing more liberal p-value thresholds) as 'restricted-to-significant PS' (rsPS) (Udler et al., 2019). Third, assessing the best prediction genome-wide entails explicitly modelling correlation structure between variants without identifying a minimum number of SNPs to be included in the subset used for the calculation of PS; this method is implemented by the program Bayesian LDpred (Vilhjálmsson et al., 2015). Udler et al termed this as 'global extended PS' (gePS) (Udler et al., 2019).

SBayesR is a methodology that implements Bayesian multiple regression, an extension of standard linear mixed model analyses, to combine likelihood estimation that connects multiple regression coefficients with GWAS summary statistics and accounts for LD (Lloyd-Jones et al., 2019). SBayesR was found to improve prediction accuracy for T2D and quantitative traits in large study populations of European ancestry, beyond the prediction accuracy of regression with summary statistics alone, LDPred, and pruning and threshold methods (Lloyd-Jones et al., 2019). PRS-CS uses a Bayesian regression framework and places continuous shrinkage priors—which can accommodate diverse genetic architectures and modelling of local LD patterns—on SNP effect sizes (Ge et al., 2019). Authors

demonstrated that PRS-CS substantially improved upon existing methods in the predicting common complex diseases, including T2D (Ge et al., 2019).

In the past 15 years, many studies have explored the utility of PS for T2D in predicting which individuals in a study population are at the highest risk of T2D. Improved sample sizes in T2D GWAS meta-analyses have enabled the construction of rsPS with increasing numbers of variants that improve modestly upon the predictive accuracy of previous iterations of T2D PS and upon that of clinical predictors (e.g., age, sex, BMI, fasting plasma glucose, HbA1c, etc.) alone for the incidence of T2D (Lyssenko et al., 2008; Mahajan, et al., 2018b; Meigs et al., 2008; Vassy et al., 2014). More recently, gePS for T2D have been constructed in two studies within the UK Biobank study using hundreds of thousands to millions of variants to predict T2D case-control status (Khera et al., 2018; Mahajan, et al., 2018b). Khera et al found that study participants in the top 3.5% of the gePS distribution had approximately threefold increased risk of T2D as compared with the remainder of the study population (Khera et al., 2018). Mahajan et al found that participants in the top 2.5% of the gePS distribution had 3.4-fold increased risk (prevalence 11.2%) compared with the median (prevalence 3.3%), and 9.4-fold increased risk compared with the bottom 2.5% of the distribution (prevalence 1.2%) (Mahajan et al., 2018b).

Further research on technical issues of transferability of T2D PS across study populations—as well as the broader utility of a PS relative to that of clinical variables—in predicting the incidence of T2D is needed. Currently, scientific literature suggests that common T2D-associated variants are largely shared across ancestry groups, but LD patterns are different. Due to the over-representation of participants from European ancestry groups, estimation of effects in non-EUR populations is suboptimal. An extension of PRS-CS, PRS-CSX, was recently developed to integrate GWAS summary statistics from multiple ancestry groups with the specific aim of improving polygenic prediction across population groups by using multi-

discovery methods to combine GWAS summary statistics from study populations of different ancestry groups (Ruan et al., 2022). The use of multi-discovery methods by PRS-CSX was found to significantly improve the prediction of T2D as done through single-discovery methods across multiple quantitative anthropometric and blood panel traits, including birth weight and HbA1c (Ruan et al., 2022). However, such improvement was modest when the target population was of European ancestry and smaller non-European-ancestry GWAS summary statistics were added to the discovery data set (Ruan et al., 2022). As PS are increasingly proposed to be used not only as research tools but also as clinical predictors, concerns surrounding reporting standards (Wand et al., 2021) and suboptimal transferability across study populations of PS (Martin et al., 2019) must be addressed to improve replicability and combat potential healthcare disparities.

1.4.2. Partitioned polygenic scores for T2D

This thesis will explore the various processes that contribute to the development of T2D. This section will provide an overview of previous research involving the clustering of genetic variants that are associated with various subphenotypes of T2D. Chapter 4 showcases my analyses of 'process-specific/partitioned' PS (pPS) for T2D that are composed of these variants. Discrete pPS can be systematically constructed using T2D-associated variants that share mediation of T2D risk through a given intermediary process (Udler et al., 2019). Early iterations of clustering approaches employed 'hard' unsupervised clustering to annotate each of dozens of T2D risk loci with respect to its physiological impact (Dimas et al., 2014; Scott et al., 2017; Voight et al., 2010). In the past decade, data from large GWAS of quantitative traits of interest and improved clustering approaches have enabled more sophisticated 'soft' clustering techniques that allow for each genetic variant to be annotated with more than one physiological impact.

Notably, Mahajan et al implemented a c-means 'soft' clustering approach using GWAS data from 10 T2D-related quantitative traits (Mahajan et al., 2018a), whilst Udler et al undertook a Bayesian nonnegative matrix factorization approach to identify variant clusters (Udler et al., 2018) using GWAS data from 47 diabetes-related traits within a partially overlapping set of 94 T2D-associated variants and identified broadly similar genetic clusters. Equipped with genetic cluster assignments that have previously been generated for a list of variants, individual-level T2D pPS can then be calculated within large, cross-sectional studies with genotypic and phenotypic data available by summing the number of risk alleles per physiological cluster. Then, individual T2D pPS can be tested for associations with T2D and related traits. Notably, a separate methodological paper on 'TACTICAL,' a framework for integrating tissue-specific gene expression and epigenomic maps to calculate 'tissues of action' scores for each association signal provided a complementary approach to partition risk (Torres et al., 2020). The tissue profiles emphasised that pancreatic islets were the main tissue in which signals operated, with respect to the development of T2D (Torres et al., 2020); this corresponds to the insulin action/beta cell clusters from the T2D pPS that Mahajan et al and Udler et al created . Chapter 4 will review the literature in greater depth and describe analyses of T2D pPS in the cohort study in the Indigenous community in the Southwestern US.

1.4.3. Transferability of findings across study populations

As described earlier, GWAS of T2D that have identified these genetic variants have predominantly included European-ancestry study populations (Mägi et al., 2017; Martin et al., 2019). While common SNPs driving genetic associations with complex human traits are largely shared across global populations (Rosenberg et al., 2010; Marigorta & Navarro, 2013)—with largely concordant directions of effect (Ntzani et al., 2012)—many studies have demonstrated that certain alleles have different effect sizes across populations (Wang et al.,

2012; Li & Keating, 2014). For example, *TCF7L2* rs7903146, identified in previous studies as a variant with high effect size and significance in association with T2D in European-ancestry study populations, associates more weakly with T2D in participants in the cohort study in which this thesis research was undertaken, an Indigenous population from the Southwestern US (Guo et al., 2007). Possible explanations for this discrepancy include differences in LD patterns, allele frequencies, and allelic effect heterogeneity. However, differences are largely due to differences in allele frequency. Differences in LD can appear to be differences in effect size; however, the effect size at the causal allele is similar. True examples of effect size heterogeneity are limited (Mahajan 2016).

The extent to which genetic differences among populations explain the higher risk for T2D experienced by Indigenous populations in the US is not wholly understood. However, studies that have examined the effects of these established T2D-susceptibility variants have found that, while there is some modest heterogeneity in effects across populations, the variants generally associate with diabetes in Indigenous populations consistently with the effects in Europeans (Carlson et al., 2013; Hanson et al., 2015). Some genetic variants with strong (i.e., high effect size) effects on diabetes risk are substantially more common in certain Indigenous study populations than in other continental ancestry groups (e.g., *ABCC8* R1420H and *HNF1A* G319S) (Baier et al., 2015; Hegele et al., 1999).

Certain variants in the gene *KCNQ1*—which are established as associated with T2D across multiple ancestry groups—exhibit parent-of-origin effects with respect to T2D in the PECRB cohort, an Indigenous population from the Southwestern US (Hanson et al., 2013). The strongest effect was in rs2299620 in *KCNQ1*, a partially imprinted locus that has demonstrated parent-of-origin effects in other study populations (Hanson et al., 2013). At this locus, the C allele was associated with T2D when maternally derived (OR 1.92; $p = 4.1 \times 10^{-12}$) but not when paternally derived (OR 0.93; $p = 0.47$), with $p = 9.9 \times 10^{-6}$ for difference in

maternal and paternal effects (Hanson et al., 2013). The ascertainment of T2D susceptibility variants remains incomplete; the extent to which the currently ascertained variants capture effects of causal variants across ancestry groups is unclear, due to variation in LD patterns and allelic heterogeneity (Chande et al., 2020; Chen et al., 2012; Hanson et al., 2015). Multiple efforts such as Diabetes Meta-Analysis of Trans-Ethnic association studies (DIAMANTE) Consortium have made substantial advancements in the inclusion of study populations beyond those of European ancestry alone, most recently with a 2022 multi-ancestry meta-analysis of GWAS of T2D across over 180,000 T2D cases and 1,150,000 controls (Mahajan et al., 2022).

1.5. Genetics of birth weight

Birth weight is an indicator of intrauterine nutrition and pregnancy outcomes and predictor of newborn and infant survival; it is also associated with risk of future cardiometabolic conditions in adulthood (Pettitt & Jovanovic, 2001). Parental diabetes has been shown to be associated with elevated risk of subsequent diabetes in offspring (Lindsay et al., 2000). Maternal diabetes, which may include diabetes during pregnancy, which confers *in utero* exposure of the offspring to a diabetic environment, has also been associated with subsequent risk of T2D and accelerated weight gain and T2D in offspring (Pettitt et al., 1988; Pettitt et al., 1983).

One study involving members of the Indigenous study population included in this thesis research (explained further in Section 1.6), as compared with offspring of people who developed diabetes after pregnancy or who remained non-diabetic, the offspring of people who had diabetes during pregnancy had greater risk of obesity and higher plasma glucose concentrations (Pettitt et al., 1993). A further study attributed *in utero* exposure to maternal diabetes with a substantial proportion of diabetes that occurred in offspring—and a greater

proportion of diabetes in offspring in more recent birth cohorts—above and beyond genetic effects (Dabelea et al., 2008). In this study population, maternal diabetes was associated with greater birth weight and greater risk of diabetes in offspring; by contrast, paternal diabetes was associated with lower birth weight and greater risk of diabetes in offspring (Lindsay et al., 2000). A positive association between maternal diabetes and offspring birth weight and negative association between paternal diabetes and offspring birth weight was also observed in an analysis of data from over 200,000 participants in the UK Biobank study (Tyrrell et al., 2013). In that study, birth weight in the offspring of parents who both had diabetes did not differ from that of offspring of parents who both did not have diabetes (Tyrrell et al., 2013).

In multiple study populations, including the Indigenous study population (Lindsay et al., 2000; McCance et al., 1994), the association between birth weight and the risk of T2D forms a U-shaped curve, in which lower and higher birth weights are associated with greater risk of T2D later in life due to a complex interplay of genetic and non-genetic influences (Harder et al., 2007; Whincup et al., 2008). This complex interplay will be discussed further in Section 1.7. More broadly, many studies have employed methods in clinical and genetic epidemiology to investigate the pairwise associations between parental T2D and offspring birth weight, between parental T2D and offspring T2D, and between own birth weight in offspring and their subsequent risk of T2D. One way to explore the relationship between birth weight and T2D is to further investigate the genetics of birth weight. For instance, an intergenerational study involving families—mothers, fathers, and offspring—and employing path analysis found that foetal (i.e., offspring) genetic factors explained 31% of the variation in birth weight (Lunde et al., 2007). T2D-associated SNPs can operate in different directions with respect to birth weight: some T2D risk alleles confer greater birth weight; others confer lower birth weight. The signals that have been identified in GWAS of birth weight implicate loci in biological pathways that may underlie the associations between birth weight and risk of diseases such

as T2D later in life (Horikoshi et al., 2016; Horikoshi et al., 2013; Beaumont et al., 2018; Freathy et al., 2010).

GWAS have identified over 400 independent signals that are associated with the complex trait T2D (Mahajan et al., 2018b), and a portion of these signals also associate significantly with birth weight (Zeng et al., 2017). Certain loci with foetal effects on both birth weight and type 2 diabetes—defined as loci that contain birth weight-associated SNPs (Warrington et al., 2019) that are in LD (R^2 > 0.3*)* with a T2D-associated SNP (Mahajan et al., 2018a)—have been shown to have underlying biological associations that confer T2D risk (Hughes et al., 2021). For instance, the birth weight and T2D risk locus rs2943641 in *IRS1* has been associated with higher insulin resistance (*IRS1*) (Mahajan et al., 2018b; Udler et al., 2018). Udler et al and Mahajan et al have also shown that loci in *ADCY5*, *CDKAL1*, *ANK1*, *HHEX/IDE*, *KCNQ1*, and *CCND2* were associated with reduced insulin secretion (Mahajan et al., 2018b; Udler et al., 2018); Thomsen et al showed that a locus in *HMGA2* was also associated with reduced insulin secretion (Thomsen et al., 2016).

To date, the most comprehensive genome-wide association analyses of own birth weight (foetal GWAS) and offspring birth weight (maternal GWAS) were conducted in study populations of European ancestry, including participants in the Early Growth Genetics (EGG) Consortium and UK Biobank study and identified 190 independent association signals (Warrington et al., 2019). The Warrington et al GWAS meta-analysis of own birth weight (N = 321,223 offspring; foetal GWAS) was conducted by combining summary statistics for the Early Growth Genetics Consortium meta-analysis and UK Biobank meta-analysis using a fixed-effects meta-analysis in the program Genome-Wide Association Meta-Analyses (GWAMA), excluding variants that failed quality control filters and conducting sensitivity analyses to confirm the quality of the results of the GWAS meta-analysis (Warrington et al., 2019).

The GWAS meta-analysis of offspring birth weight (N = 230,069 mothers; maternal GWAS) employed similar strategies and reference panels as applied to maternal genotypic data: in sum, two genomes—maternal and foetal—were investigated using GWAS meta-analyses and Mendelian randomisation (Warrington et al., 2019). Warrington et al constructed two GWAS meta-analyses due to previous evidence from a study of maternal and foetal mutations in the glucokinase gene (*GCK*) (Hattersley & Tooke, 1999) as well as an earlier meta-analysis of GWAS of birth weight in the Early Growth Genetics Consortium (Horikoshi et al., 2016). While the study by Horikoshi et al was limited by sample size constraints in the early GWAS meta-analysis, this study found that T2D risk alleles were associated with high birth weight, while others were associated with lower birth weight and prompted further investigation of the contrasting associations of T2D risk alleles with higher and lower birth weight, which could reflect different impacts of maternal and foetal genotype (Horikoshi et al., 2016).

The Warrington et al foetal GWAS identified 146 independent SNPs at genome-wide significance ($p < 6.6 \times 10^{-9}$); the maternal GWAS identified 72 independent SNPs according to that significance threshold (Warrington et al., 2019). SNPs at 30 loci (within 500 kb and LD $r^2 \geq 0.1$*)* were identified in both the foetal and maternal GWAS as associated with birth weight at a level of genome-wide significance (Warrington et al., 2019). Taken together, 209 lead SNPs (146 novel as of the 2019 publication of the study) were identified. Two-sample Mendelian randomisation (MR) (Pierce & Burgess, 2013) was applied to estimate causal effects of maternal exposures—height, glycaemic traits and blood pressure—on offspring birth weight. For example, an MR analysis that used 33 SNPs (maternal genotypes) that were associated with fasting plasma glucose estimated a 0.18 SD (95% CI 0.13-0.23) higher offspring birth weight per SD (0.4 mmol/L) higher maternal fasting glucose, independent of the direct foetal effects. Because some foetal and maternal effects act in opposite directions,

a structural equation modelling (SEM) method (Warrington et al., 2018) was used to partition the lead SNPs into five categories based on maternal and/or foetal genetic contributions to birth weight and estimate the contributions of these genetic effects at each locus (Figure 1.3) (Warrington et al., 2019).

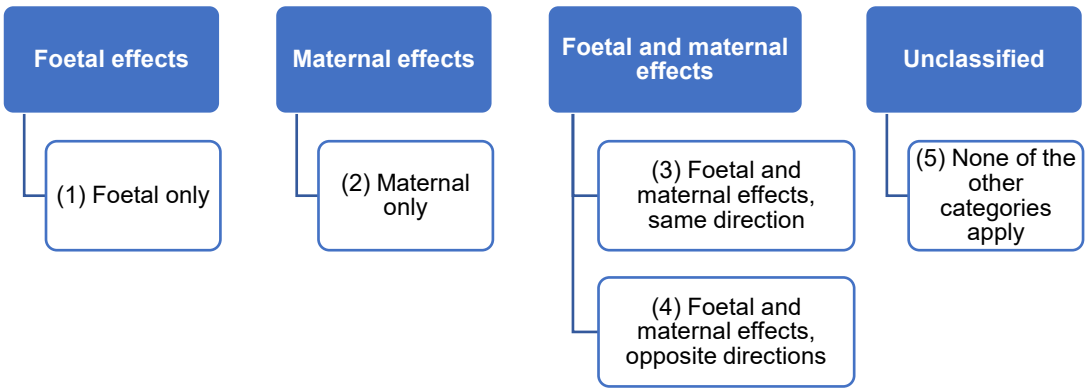


Figure 1.3. Five categories of maternal and/or foetal genetic contributions of the 209 lead SNPs to birth weight employed in the Warrington et al structural equation modelling method (Warrington et al., 2018, 2019).

A recent investigation within the Exeter Family Study of Child Health and Hyperglycemia and Adverse Pregnancy Outcomes (HAPO) tested the effects of maternal glycaemia and a foetal genetic score for birth weight (calculated using foetal genotypes across a set of birth-weight-associated SNPs) on birth weight and foetal insulin: it found that the score predominantly influences pathways independent of foetal insulin (Hughes et al., 2018). Mendelian randomization, which employs groups of genetic variants as instrumental variables (e.g., genetic scores for birth weight) to test causal effects, has been used in such attempts to demonstrate that birth weight is causally related to T2D, but results have been difficult to interpret (Huang et al., 2019a; T. Wang et al., 2016; Zanetti et al., 2018). More recently, structural equation modelling methods have been developed to parse out the maternal and foetal effects when testing for causal associations, as well as separate foetal GWAS of own birth weight, conditioned on maternal genotype; and maternal GWAS of offspring birth weight,

conditioned on foetal genotype (Warrington et al., 2018). The studies outlined in the present thesis expand upon this research: they include GWAS of own birth weight, construction of PS for birth weight, analyses of the association between PS for birth weight with subsequent T2D, and analyses of the contribution of birth weight data to the prediction of T2D incidence in a cohort study.

1.6. Longitudinal study of health in an Indigenous community from the Southwestern United States

In the past 50 years, due to the disparities in prevalence in diabetes and obesity experienced by Indigenous communities, there has been a proliferation of diabetes and obesity research involving the present study population and other Indigenous communities. Members of tribal leadership from the Indigenous community whose members have participated in the longitudinal study in which the research described in this thesis was undertaken have requested of National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases (NIH/NIDDK) Phoenix Epidemiology and Clinical Research Branch (PECRB) researchers that the community's name be omitted from scientific publications output by the research group. Thus, the original study name, as well as the tribe's name, are omitted from this thesis, and instead, the longitudinal study population will be referred to as the 'PECRB cohort' or 'PECRB study population.'

The analyses described in Chapters 3-5 are based on data collected in a longitudinal study of diabetes (1965-2007), which was conducted in an Indigenous study population from the Southwestern US. Notably, Knowler et al described the methods for conducting that study (Knowler et al., 1978); further details on genotypic and rigorously measured phenotypic data in the study population are described in Sections 2.1 and 2.2. Briefly, individuals who were at

least five years old were invited for health examinations every two years. Before participation, study volunteers had been fully informed of the nature and purpose of the study, and participants who were ≥ 18 years old had provided written informed consent. Participants who were < 18 years old had provided written assent and informed consent had been obtained from their parents or guardians. Protocols had been approved by the institutional review board of the NIH/NIDDK.

This is one of the few studies that have prospectively examined the role of exposure to maternal diabetes *in utero* on the early growth and risk of future T2D in offspring. In the past century, members of the tribe in which this study's participants are enrolled have experienced substantial changes from traditional lifestyles (Schulz et al., 2006)—dietary shifts from traditional crops to processed foods and reductions in physical activity (Esparza et al., 2000)—that contextualise their higher rates of diabetes and obesity today. This study is a unique and appropriate source of data for thesis research on genetic and non-genetic influences on birth weight and T2D for multiple reasons (Figure 1.4).

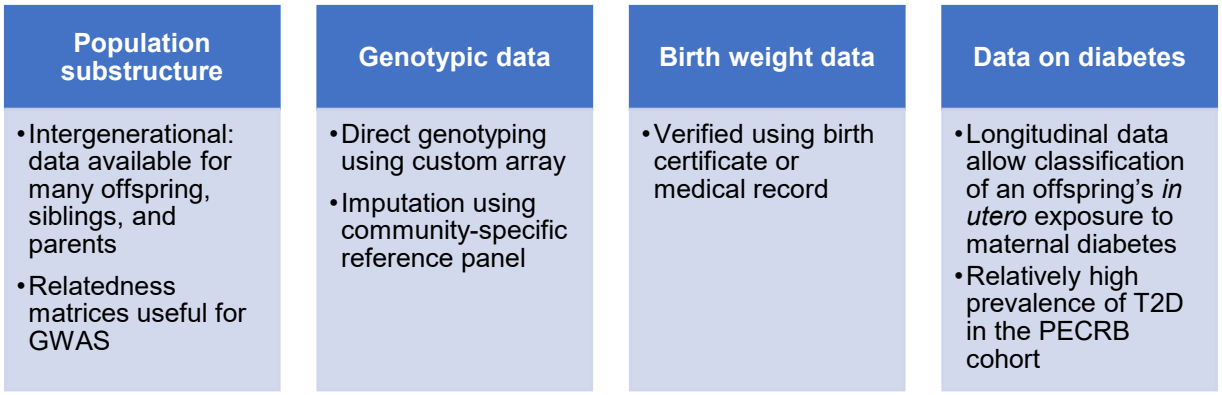


Figure 1.4. Key data in the longitudinal study that strengthen case for secondary studies of birth weight and T2D in the study population.

As summarised in Figure 1.4, the PECRB study also provides information on population substructure through detailed pedigrees, which allows researchers to investigate how

maternal and paternal genotypes and phenotypes associate with offspring genotypes and phenotypes. Study data have been analysed extensively in studies of genetic and non-genetic factors in association with complex traits (e.g., T2D, and obesity). The genotypic data and longitudinal phenotypic data described in this section have enabled the studies described in Chapters 3, 4, and 5 that employed methods in genetic and clinical epidemiology to study genetic and non-genetic influences on birth weight and T2D.

1.7. Birth weight and risk of T2D

The various quadrants in Figure 1.5, which summarize various genetic and non-genetic influences on low and high birth weight, which can be associated with greater risk of subsequent T2D in offspring, relate to various hypotheses that have emerged from decades of epidemiological research on this relationship.

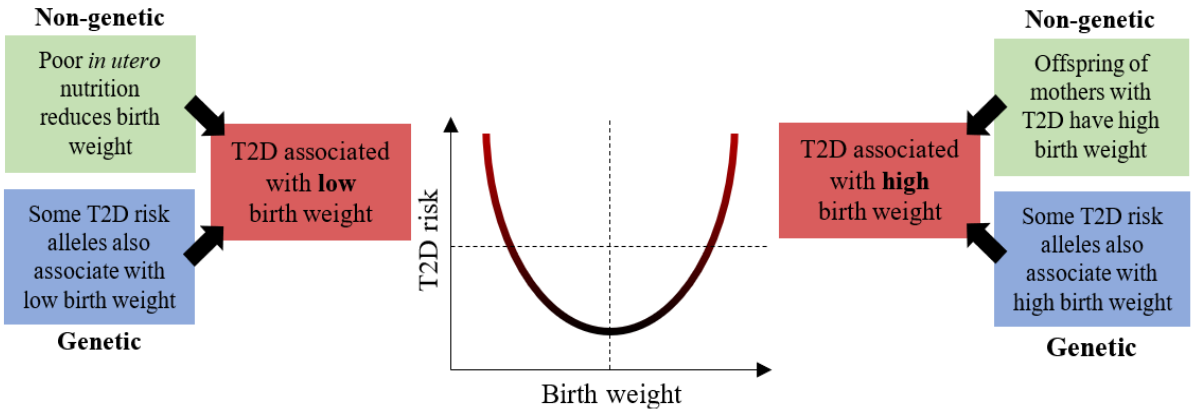


Figure 1.5. Graph depicting the U-shaped relationship between risk of T2D and birth weight, with examples of genetic and non-genetic influences on the upper tails of the curve at high T2D risk (Drong et al., 2012). This figure was adapted with Copyright Clearance Centre licence number 5261381133422.

The upper left quadrant (non-genetic) states that poor *in utero* nutrition reduces birth weight, which is associated with greater risk of subsequent risk of T2D in offspring, summarizes the Developmental Origins of Health and Disease (DOHaD) (i.e., foetal origins) hypothesis. The

upper right quadrant (non-genetic), which states that offspring of mothers with T2D have higher birth weight above and beyond the effects of genetic transmission alone, represents one of the tenets of the foetal overnutrition (i.e., developmental overnutrition) hypothesis. The lower left quadrant (genetic), which states that some T2D risk alleles also associate with lower birth weight, is most closely related to the foetal insulin hypothesis. By contrast, the lower right quadrant (genetic) represents evidence that some T2D risk alleles are associated with higher birth weight. Previous studies have demonstrated the heterogeneous directions of association between T2D- and birth weight-associated SNPs and birth weight (Horikoshi et al., 2016; Warrington et al., 2019). The complex genetic relationships between birth weight and T2D in the present Indigenous longitudinal study population will form the focus of Chapter 3. The DOHaD, foetal overnutrition and foetal insulin hypotheses and evidence in support of or that challenges them will be further explored in the following section.

1.7.1. The Developmental Origins of Health and Disease (DOHaD) Hypothesis

The DOHaD hypothesis, or the Barker hypothesis, is based on Hales and Barker's original proposition of 'foetal origins' or 'foetal programming,' that foetuses survive *in utero* malnutrition by adopting a 'thrifty phenotype' that predisposes them to develop metabolic syndrome later in life (Hales & Barker, 1992). The 'thrifty phenotype' hypothesis was proposed to confer a survival advantage in a setting of malnutrition by reducing glucose uptake and restricting body growth (Godfrey & Barker, 2000; Hales & Barker, 1992). Thus, maternal malnutrition gives rise to offspring with lower birth weight and higher postnatal energy intake and elevated risk of cardiometabolic conditions, including coronary heart disease (Stein et al., 1996) and T2D (Edwards, 2017). By contrast, adequate maternal nutrition gives rise to offspring with normal birth weight, normal postnatal energy intake, and the absence of metabolic syndrome later in life (Figure 1.6).

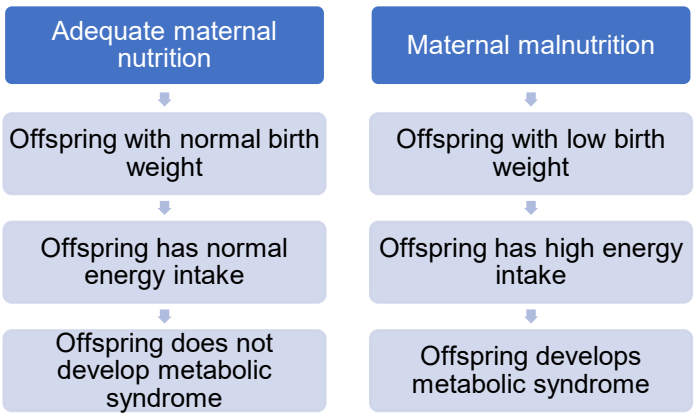


Figure 1.6. The DOHaD hypothesis schematic involving two scenarios: adequate maternal nutrition and maternal malnutrition lead to different outcomes for offspring.

The DOHaD hypothesis focuses on the role of the prenatal and postnatal nutritional environment on later metabolic disease risk, and physiological mechanisms underlying this relationship are not entirely clear. Numerous epidemiological studies support the DOHaD hypothesis. An investigation within the Dutch Famine Birth Cohort, which followed over 2400 individuals who were born around the time of the 1940s 'Dutch hunger winter,' demonstrated that exposure to famine (i.e., malnutrition) during any point in gestation was associated with glucose intolerance and higher rates of obesity (Roseboom et al., 2006).

Certain articles have challenged the DOHaD hypothesis, asserting that the inverse association between birth weight and adult disease might largely reflect the impact of different forms of bias. For instance, studies have suggested that bias could emerge from statistical artifacts of adjusting for subsequent body size (Huxley et al., 2002) and that publication bias could have influenced the fact that most related studies report inverse associations between birth weight and subsequent disease risk (Schluchter, 2003). Many previous studies of the DOHaD hypothesis have focused on the association between low birth weight and subsequent risk of cardiometabolic conditions in registries and cohort studies

(Hoet & Hanson, 1999). However, high birth weight is also associated with such increased risk (Dabelea, 2007).

1.7.2. The Foetal Overnutrition Hypothesis

The foetal overnutrition hypothesis, also termed the developmental overnutrition hypothesis, is based on the premise that *in utero* overnutrition creates conditions for later pathophysiological effects of an obesogenic environment and that subsequent postnatal overnutrition contributes to increased risk of T2D in the offspring (Dabelea & Crume, 2011). Maternal obesity or diabetes during pregnancy has also been described as an exposure with respect to the outcome of adiposity or diabetes in offspring (Lawlor et al., 2008). Different factors could contribute to developmental overnutrition of the foetus during pregnancy: maternal obesity, maternal diabetes, gestational weight gain in pregnancy (Perng et al., 2019). This may result in an increase the offspring's subsequent risk of obesity and T2D. It has been previously suggested that *in utero* exposure to maternal diabetes results in a 'vicious cycle,' in which the offspring has increased birth weight and risk of obesity, GDM, and T2D (Figure 1.7) (Pettitt et al., 1988).

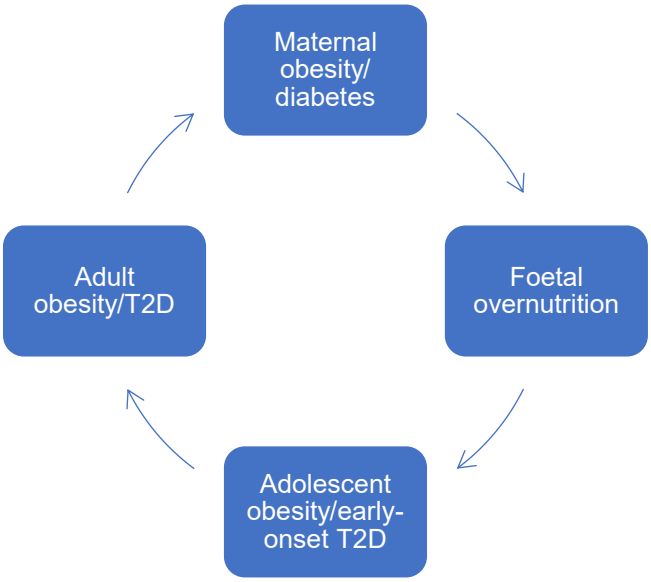


Figure 1.7. The 'vicious cycle' of diabetes and obesity proposed as a corollary of the foetal overnutrition hypothesis (Dabelea & Crume, 2011).

The 'vicious cycle' term denotes that maternal obesity and diabetes could lead to the intergenerational propagation of risk of diabetes. This could be facilitated by the exposure of the foetus to a hyperglycaemic environment *in utero*, which has been shown to associate with foetal macrosomia (Dabelea, Hanson, et al., 2000). Generally, offspring of a diabetic pregnancy have increased risk for developing type 2 diabetes and for developing obesity at an early age (Pettitt et al., 1988). Specifically, if the offspring of a diabetic pregnancy is female, the offspring has an elevated risk of developing gestational diabetes, further perpetuating the 'vicious cycle' (Innes et al., 2002). Further research is needed to elucidate the proposed mechanisms implied by the foetal overnutrition hypothesis and the associated 'vicious cycle' paradigm, including the complex genetic relationship between birth weight and T2D.

1.7.3. The Foetal Insulin Hypothesis

The foetal insulin hypothesis was first proposed as an alternative explanation to the original thrifty phenotype hypothesis (Hattersley & Tooke, 1999). It proposed that lower birth weight and adult-onset T2D had a shared genetic predisposition to reduced insulin secretion and action *in utero* and after birth, which lead to lower birth weight and greater risk of adult-onset T2D (Figure 1.7) (Hattersley et al., 1998). Insulin acts as a regulatory and trophic factor: hyperinsulinemia in the foetus (e.g., as a response to maternal hyperglycaemia) can result in foetal macrosomia and hypoinsulinemia can result in low birth weight (Hill & Milner, 1985).

Hattersley et al demonstrated in their initial study that proposed the foetal insulin hypothesis that the effects of maternal and foetal mutations in the glucokinase gene (*GCK*) were additive (Hattersley et al., 1998). Heterozygous mutations in *GCK* result in reduced sensing of glucose by the pancreatic beta cell, so individuals with *GCK-MODY* regulate glucose at a higher set-point (fasting plasma glucose 5.5-8 mmol/L; Stride 2002) and have stable, mild hyperglycaemia throughout life (Steele 2013). The authors demonstrated that maternal and paternal inheritance of the *GCK*-MODY, and maternal *GCK*-MODY in the absence of foetal *GCK*-MODY, have differing effects upon birth weight (Figure 1.8).

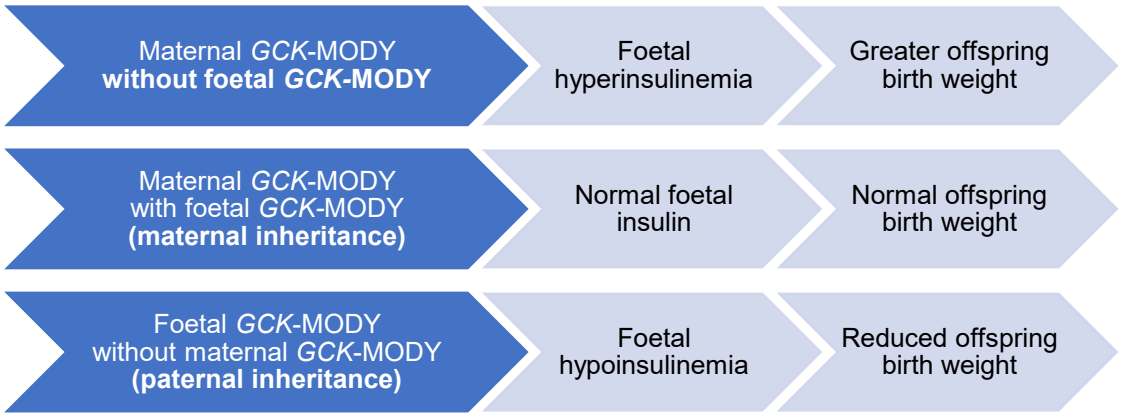


Figure 1.8. Summary of findings of study on *GCK*-MODY inheritance with respect to foetal insulin secretion and offspring birth weight (Hattersley et al., 1998).

As shown in Figure 1.8, maternal hyperglycaemia due to a *GCK* mutation (i.e., *GCK*-MODY) in the absence of *GCK*-MODY in the offspring resulted in a mean increase in offspring birth weight of 601g (p = 0.001), due to greater foetal insulin secretion in response to maternal hyperglycaemia (Hattersley et al., 1998). However, in the presence of both maternal *GCK*-MODY and foetal *GCK*-MODY (i.e., equal levels of glucose sensing in mother and foetus and therefore normal foetal insulin secretion), offspring birth weight was not found to be significantly different from average (Hattersley et al., 1998). In the case of foetal *GCK*-MODY and no maternal *GCK*-MODY (i.e., paternal inheritance of the *GCK* mutation), birth weight was approximately 533g lower than average (p = 0.002) (Hattersley et al., 1998). Maternal glucose that crosses the placenta—while insulin, a larger hormone, does not do so—is sensed at a higher threshold by the foetus, which results in foetal hypoinsulinaemia (Hughes et al., 2021).

Beyond the scope of rare monogenic alleles, other studies have investigated the influence of more common birth weight-associated alleles that have been identified in GWAS upon the risk of T2D in offspring. A 2021 review article synthesized evidence generated and questions that remain 20 years since the advent of the foetal insulin hypothesis (i.e., genetic predisposition to lower birth weight also confers higher risk of subsequent diabetes), stating that GWAS have shown that most variants in T2D risk loci have not been shown to associate with birth weight (Hughes et al., 2021). Methods such as structural equation modelling are necessary to estimate the maternal and foetal genetic effects on birth weight and diabetes (Hughes et al., 2021). However, as shown by Horikoshi et al, foetal and maternal effects that have been identified in maternal GWAS of offspring birth weight and foetal GWAS of own birth weight can be in opposite directions, cancelling each other out (Horikoshi et al., 2016). Thus, conducting separate foetal and maternal GWAS of own or offspring birth weight,

respectively, is important for understanding the complex genetic influences on birth weight, as well as the genetic relationship between birth weight and T2D.

1.8. Unanswered questions regarding the relationship between early growth and metabolic disease

This introductory chapter outlined evidence regarding genetic and non-genetic influences on birth weight and T2D, as well as evidence and hypotheses regarding the complex relationships between birth weight and T2D. Unanswered questions regarding different research methodologies used to study these relationships prompt the research described in this thesis.

- Most genome-wide association studies have been conducted in European-ancestry study populations; the genetic and non-genetic influences on the complex U-shaped relationship between birth weight and diabetes should be further investigated within diverse, non-European-ancestry study populations with maternal and offspring phenotypic and genotypic data available.

- Polygenic scores (PS) provide opportunities to estimate the contribution of genetic information to that of clinical variables in predicting the incidence of complex diseases: different PS for T2D should be generated and their predictive and clinical utility should be evaluated in a cohort study with longitudinal data across age groups.

- Clustering approaches have yielded proposed subgroups of T2D-associated genetic variants that associate with various processes in the aetiology of T2D: the associations of these partitioned/process-specific PS (pPS) should be validated in different study populations.

1.9. Aims of this thesis

The main aim of this thesis is to investigate genetic and non-genetic influences on birth weight and T2D using methods in genetic and clinical epidemiology. To accomplish this, specifically focussed on data from the PECRB study, this thesis describes research designed to:

- Identify genetic variants associated with birth weight and evaluate genetic relationships between birth weight and T2D.
- Evaluate the contribution of clinical variables in addition to polygenic scores for T2D to the prediction of T2D incidence in birth, youth, and adult cohorts.
- Assess the associations between partitioned/process-specific polygenic scores (pPS) for T2D with T2D and related traits.

2. Study participants and methods

The PECRB cohort described in this chapter consists of members of the Indigenous study population from the Southwestern US who had participated in a longitudinal study of diabetes (1965-2007). Study participants who were at least five years old had been invited for health examinations every two years. A total of 12,647 individuals had participated in the longitudinal study, with an average number of 4.2 examinations.

2.1. Genotypic data from the longitudinal study

Genotypic data for the present thesis were derived from a subset of the PECRB cohort (n = 7,701 participants) who had a DNA sample available (sample collection for DNA had begun in 1983). There was a relatively high degree of relatedness among individuals in this subset: 6,741 participants (87.5%) were a first-degree relative of at least one other individual in the sample. A previous population-based genome-wide linkage study among a subset of 1,024 of the 7,701 study participants who had a DNA sample available calculated identity-by-descent among individuals without known relationships (i.e., they are cryptically related). It showed that the average relatedness between random pairs of that subset of study participants was between that of fifth- and sixth-degree relatives, higher than that among a convenience sample of individuals of European ancestry who had also been included in another study by PECRB researchers (Hsueh et al., 2017).

Study participants had been previously genotyped using a custom genotyping Axiom array (Affymetrix; Santa Clara, CA) that was designed to capture common variation (MAF ≥ 5% at r^2 ≥ 0.85 in 300-kb windows), as well as low frequency coding variants (MAF 1-5%), across the genome in this population, using whole genome sequence data. The GWAS included 515,723 directly genotyped tag SNPs that passed all quality control metrics. These directly

genotyped SNPs captured approximately 91% of all common genetic variants (MAF ≥ 5%) and 56% of low-frequency variants (MAF 1-5%). Figure 2.1 is a histogram displaying percentages of SNPs by MAF category across these directly genotyped variants.

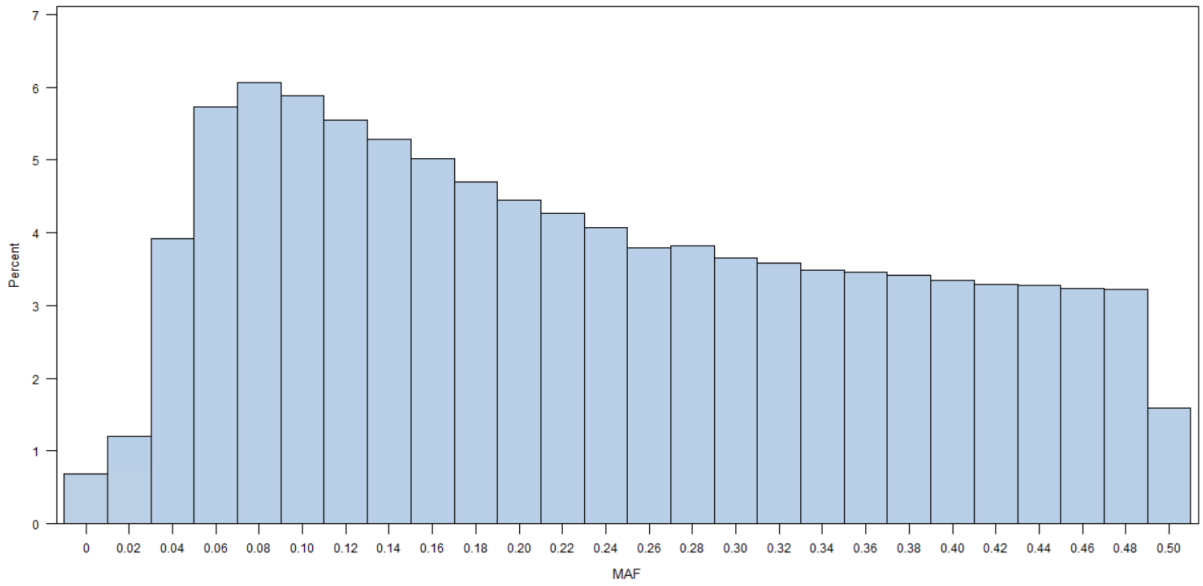


Figure 2.1. Histogram displaying the percentage of SNPs for each MAF bin across directly genotyped variants; bins are centred at increments of 0.02 between 0 and 0.50.

4,589,902 missing genotypes and untyped variants had been imputed, employing a scaffold composed of those 515,723 SNPs, using whole genome sequencing of 296 community members who were study participants. A previous study demonstrated that using members from this population as its own reference panel performed better than using an outside source, a reference panel from the HapMap Consortium (Malhotra et al., 2014). Similar, unpublished results had been found by NIH/NIDDK PECRB researchers when employing 1000 Genomes data, with no additional improvement found in combining the population-specific data with that of 1000 Genomes for imputation (correspondence with supervisor R. Hanson, 2022).

Sequencing reads for the reference panel had been aligned to the human genome build 37 (Hg37) using Burrows-Wheeler Alignment tool (BWA), a software package for mapping low-

divergent sequences against a large reference genome (e.g., the human genome) (H. Li & Durbin, 2009). Genotype calling had been performed by pre-calling genotypes using the HaplotypeCaller approach, then joint genotyping by the GATK GenotypeGVCFs approach, from GATK version 3.5 (Poplin et al., 2018; Van der Auwera & O'Connor, 2020). Genomic variants with more than 2 alleles were removed from the reference panel, and imputation was conducted using the IMPUTE version 2 (IMPUTE2) method (Howie et al., 2009). Genotypic data had been phased using the duoHMM method in SHAPEIT version 2—which incorporates pedigree information in the study population—to estimate haplotypes (Delaneau et al., 2013; O'Connell et al., 2014). Phased data had been divided into 5Mb segments and missing genotypes of the SNPs in the reference panel were typed using IMPUTE2, using default parameters (Howie et al., 2009). Imputed variants with MAF ≥ 0.01 and Qinfo ≥ 0.5 were included in the GWAS models described in Chapter 3. Figure 2.2 is a histogram displaying percentages of SNPs by Qinfo category across these imputed variants.

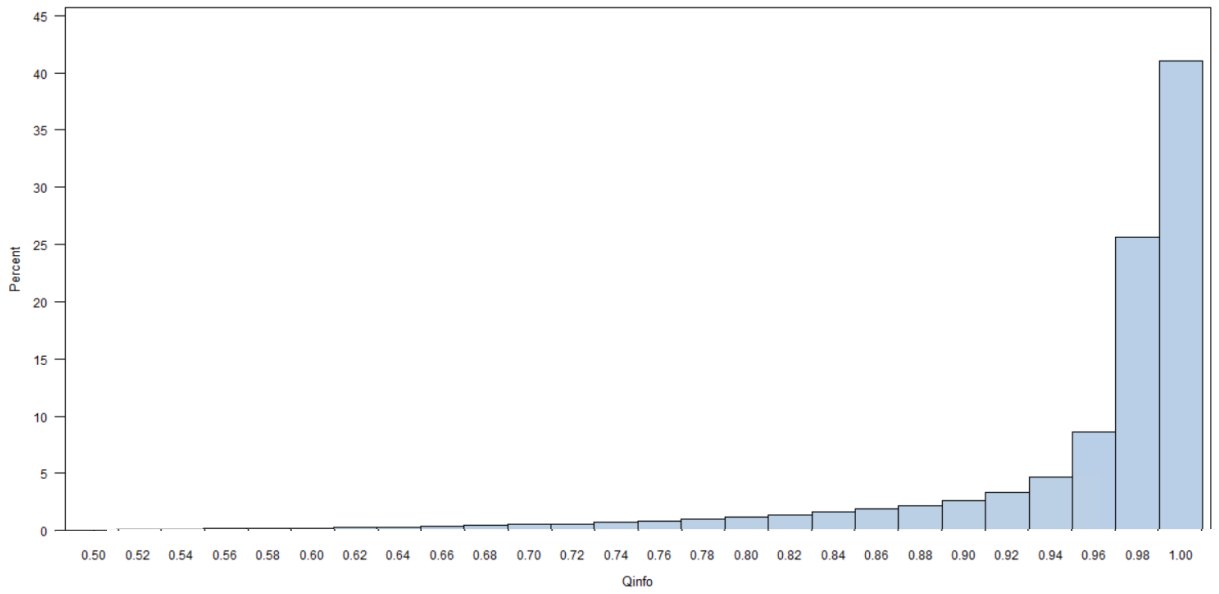


Figure 2.2. Histogram displaying the percentage of SNPs for each Qinfo bin across imputed variants; bins are centred at increments of 0.02 between 0.50 and 1.00 and the minimum Qinfo was 0.50.

The vast majority of imputed variants had Qinfo scores greater than 0.90. Figure 2.3 is a histogram displaying percentages of SNPs by MAF category across these imputed variants.

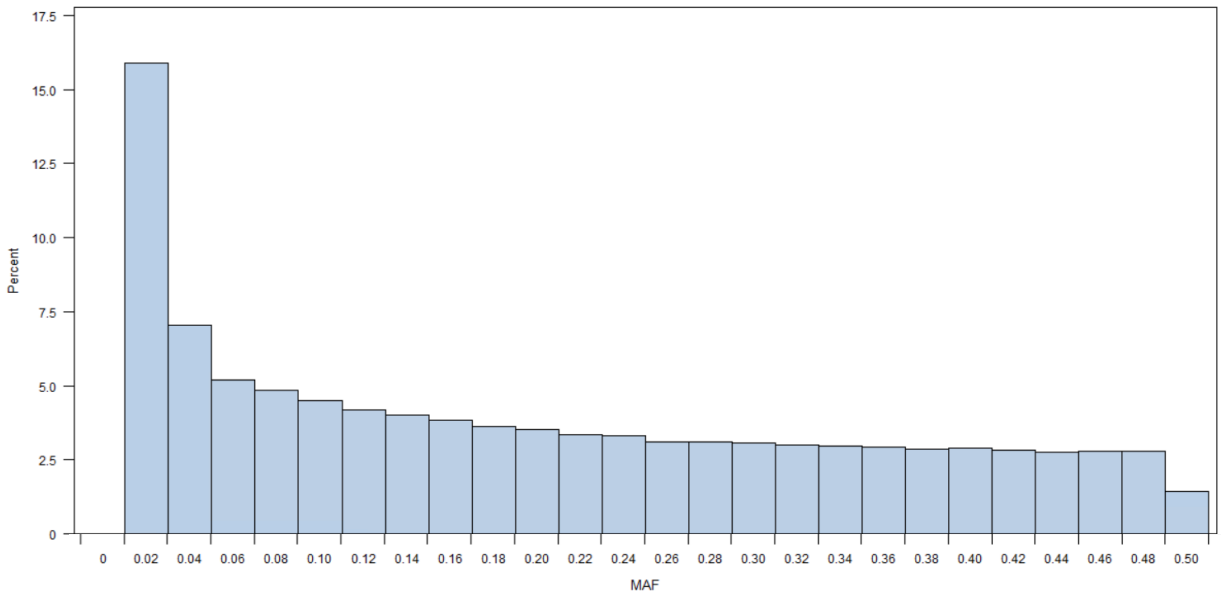


Figure 2.3. Histogram displaying the percentage of SNPs for each MAF bin across imputed variants; bins are centred at increments of 0.02 between 0 and 0.50 and the minimum MAF was 0.01.

GWAS that were undertaken in this thesis research involved association tests with adjustment for relevant covariates, including the first five genetic principal components (PCs). According to previous analyses by collaborators at the NIH/NIDDK, these first five genetic PCs capture most of the relevant structure in this study population, as reflected in non-Indigenous admixture and tribal affiliations. These genetic PCs had been generated by collaborators in 11,008 members of the Indigenous study population from 19,991 variants that were selected to be approximately 200 kb apart, to reduce the influence of LD. The first two genetic PCs had been found to distinguish participants with non-Indigenous admixture (largely European) and between the major linguistic groups of Indigenous peoples of the Americas (largely Uto-Aztecan and Athabascan in the Southwestern US) (Figure 2.4; from correspondence with supervisor R. Hanson, 2022).

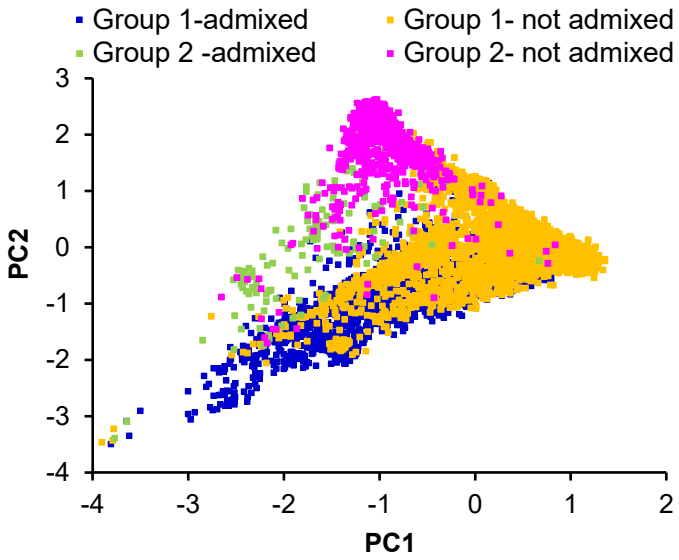


Figure 2.4. Genetic principal components 1 and 2 based on language group and stated heritage in longitudinal study participants with genotypic data (N = 7,701). Reproduced with permission from supervisor R. Hanson, 2022.

The third genetic PC had been found to separate study participants based upon tribal affiliation along a North-South gradient (Figure 2.5; correspondence with supervisor R. Hanson, 2022).

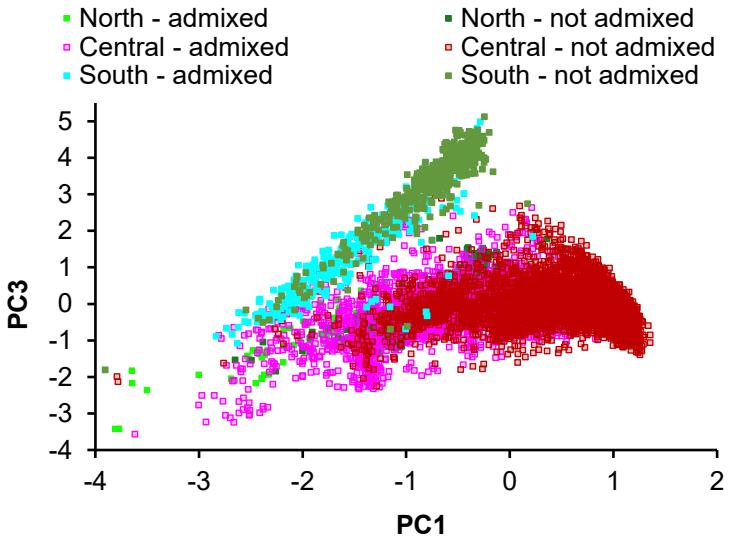


Figure 2.5. Genetic principal components 1 and 3 based on tribal geographic origin and stated heritage in longitudinal study participants with genotypic data (N = 7,701). Reproduced with permission from supervisor R. Hanson, 2022.

The fourth and fifth genetic PCs had been found to identify genetic differences between geographically nearby tribal groups, which comprise the current study population (Figure 2.6; correspondence with supervisor R. Hanson, 2022).

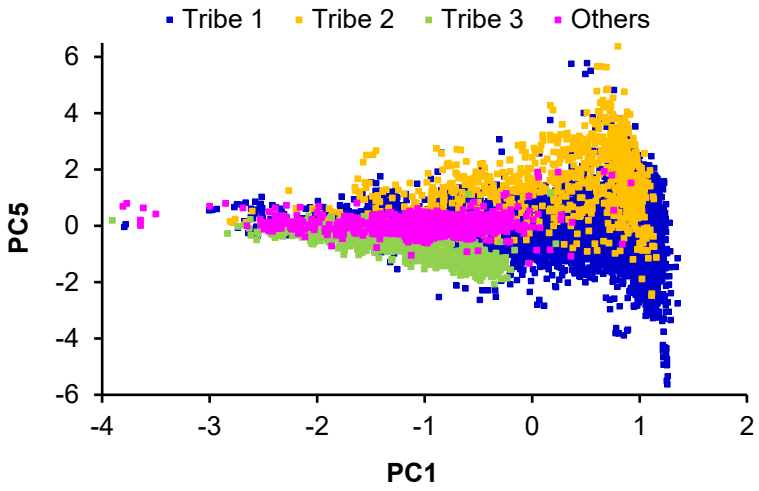


Figure 2.6. Genetic principal components 1 and 5 based on language group and stated heritage in longitudinal study participants with genotypic data (N = 7,701). Reproduced with permission from supervisor R. Hanson, 2022.

While adjusting for the top 10 genetic PCs is more conventional, NIH/NIDDK PECRB researchers had found that the sixth to tenth genetic PCs do not substantially improve the ability to distinguish between tribal groups that are relevant in analyses of the PECRB cohort. Furthermore, NIH/NIDDK PECRB researchers have found that adjustment for the first five PCs result in well-controlled genomic inflation factors for GWAS of most phenotypes if genetic relationships are accounted for in a linear mixed model. The table below displays a summary of basic information on birth year and the first five genetic principal components for individuals in the longitudinal study who had genotypic and phenotypic data available (N = 7,701) (Table 2.1); 4,294 were female and 3,407 were male. While genetic PCs are often derived to have a mean of 0, and sometimes an SD of 1, the summary statistics shown in Figure 2.1 do not have these properties because the 7,701 individuals in the aforementioned

'PECRB cohort' are a subset of all participants in PECRB studies who have genotypic data available.

Table 2.1. Basic information for individuals with genotypes in the longitudinal study (N = 7,701).

Variable (units)	Mean	SD
Birth year (year)	1967.09	18.01
PC1	0.36	0.85
PC2	-0.27	0.63
PC3	-0.16	0.63
PC4	0.03	1.16
PC5	0.04	1.10

PECRB researchers have generally controlled for the first five genetic PCs in most GWAS analyses; I have followed this practice on the analyses that are included in this thesis.

2.2. Phenotypic data from the longitudinal study

One of the core strengths of this longitudinal study for applications of genetic and clinical epidemiological studies of type 2 diabetes is the extensive phenotypic data for T2D and anthropometric and related glycaemic traits within study participants, many of whose family members were also study participants. The methods through which other members of the NIH/NIDDK Phoenix Epidemiology and Clinical Research Branch had collected these data from consenting participants in the longitudinal study are detailed below. Sample sizes for each trait are provided, including the number of individuals who had phenotypic data for that trait as well as genotypic data available. I also describe the covariates for which I adjusted in genetic analyses of these traits that are detailed in this thesis, such as GWAS and bivariate genetic correlation analyses.

2.2.1. Type 2 diabetes, fasting plasma glucose, and 2-hour plasma glucose

T2D status (which was available in N = 7,659 individuals; 4,277 female and 3,382 male) had been ascertained according to ADA criteria. At each exam, a 75-g OGTT was administered, with measurement of HbA1c and fasting and 2-hour plasma glucose (FPG, 2hPG). Diabetes was diagnosed using 1997 American Diabetes Association criteria (FPG ≥ 7.0mM, 2hPG ≥ 11.1mM, or clinical diagnosis) (Kahn, 1997). 2,571 individuals in the data set had developed documented diabetes by the time of their last examination. A summary of information in the T2D data set, a subset of the longitudinal study participants who had genotypic data available (N = 7,701), is provided below for the following continuous traits: age at exam, birth year, maximum BMI, and age of maximum BMI (Table 2.2).

Table 2.2. Basic information for continuous traits for individuals in the T2D data set (N = 7,659).

Variable (units)	Mean	SD
Age at exam (years)	34.05	16.55
Birth year (year)	1967.09	18.02
Maximum BMI (kg/m^2)	36.14	8.77
Age of maximum BMI (years)	32.79	13.17

Analyses of T2D, including bivariate genetic correlation analyses, were adjusted for sex, birth year, age, and the first five genetic principal components (PCs). Analyses of continuous measures of glycaemia (i.e., concentrations of fasting plasma glucose, FPG; and 2-hour post-load glucose, 2hPG) and insulinemia were conducted only among participants who had these measurements available at a nondiabetic examination. This enables the identification of factors that influence glucose metabolism without the confounding effects of diabetes or of its treatment.

FPG and 2hPG were analysed in participants without diabetes (N = 6,637; 3,648 female and 2,989 male). Table 2.3 below summarises the continuous traits of this subset of participants in the longitudinal study, including: age at exam, birth year, BMI, FPG, and 2hPG.

Table 2.3. Basic information for continuous traits for individuals without diabetes in the FPG and 2hPG data set (N = 6,637).

Variable (units)	Mean	SD
Age at exam (years)	31.12	13.21
BMI (kg/m^2)	33.83	8.40
FPG (mg/dL)	94.52	10.34
2hPG (mg/dL)	115.39	32.35

FPG was analysed with adjustment for sex, age, and the first five genetic PCs. 2hPG had been measured in a 75-g OGTT using methods described previously; it was analysed in participants without diabetes, with adjustment for sex, age and the first five genetic PCs (Narayan et al., 1996). Height and weight had been measured to calculate BMI. Maximum BMI observed during the longitudinal study (logmaxBMI) was also measured, defined as the highest recorded BMI from an exam at ≥ 15 years of age. It was analysed with a logarithmic transformation to reduce skewness, with adjustment for sex, age, birth year, and the first five genetic PCs.

2.2.2. Corrected insulin response

Corrected insulin response (CIR), a measure of insulin secretion, had been calculated from post-load plasma glucose and insulin concentrations (Sluiter et al., 1976) (Equation 2.1).

$$CIR = \frac{100 \cdot I}{G\,(G - 70)} \tag{2.1}$$

In the longitudinal study, 120-minute CIR (CIR120, N = 343) had been calculated from 2hPG and 2-hour insulin concentrations observed in the outpatient longitudinal studies; this measure was derived from the 2-hour insulin and glucose concentrations since these were available in the outpatient studies. CIR120 correlates moderately with more sophisticated measures of insulin secretion that are described below.

2.2.3. Birth weight and parental diabetes

Birth weight had been previously ascertained from Arizona state (birth certificates) records or from review of hospital records at each biennial visit during the longitudinal study, using methods described in greater detail by Lindsay et al (Lindsay et al., 2000). Before beginning the analyses of birth weight that are described in Chapters 3, 4, and 5, I updated the data set of participants who were singletons with birth weight and genotypic data available by adding participants who met these criteria but were not included in the original data sets utilised by Lindsay et al. I accomplished this and expanded the sample size of this data set through review of study participants' records, cleaning incorrectly input birth weight data (e.g., verifying that birth weight was recorded in grams rather than pounds) and verifying gestational age data (in weeks), where provided. Gestational age, which is often correlated with birth weight and infant mortality, is recognized as an important covariate in analyses of birth weight (Callaghan & Dietz, 2010). Notably, approximately 74.3% of the individuals in the birth weight data set (N = 3,700) had gestational age data available (N = 2,749); missing data were imputed using methods outlined in Section 2.2.4. A summary of the birth weight data set is provided below, for continuous traits: birth year, age of mother, gestational age, and birth weight (Table 2.4).

Table 2.4. Basic information for continuous traits for individuals in the birth weight data set (N = 3,700).

Variable (units)	Mean	SD
Birth year (year)	1976.09	11.98
Age of mother (years)	25.02	6.01
Gestational age (weeks)	39.64	1.22
Birth weight (grams)	3469.89	506.83

Birth weight, in grams, had been ascertained from birth and medical records and was normalised separately by sex via rank-based transformation (the Blom method in SAS 9.4) (Blom, 1958) for the purposes of analyses that are detailed in Chapters 3, 4, and 5.

Within the birth weight data set, maternal diabetes data had been available. Mothers and offspring in the longitudinal study were not specifically asked about gestational diabetes or gestational diabetes exposure. A combination of two categorical variables, INUTERO and NOT_IUE, was used as a proxy to assign study participants who had maternal diabetes data available to three categories of likeliness of exposure to diabetes *in utero*: exposed to maternal diabetes, likely unexposed, and unknown. The binary variable INUTERO indicated whether a mother was diagnosed with diabetes before the birth of the offspring. The binary variable NOT_IUE was defined as a documented nondiabetic examination in the mother at least one year after the child's birth (i.e., suggesting that exposure to diabetes in utero was unlikely).

These two variables were used to differentiate cases of more or less likely exposure to maternal diabetes *in utero*. Figure 2.7 below summarises the categories that are created by these two binary variables. For the INUTERO variable, a value of 1 indicates definite exposure, in which the mother was diagnosed before the child's birth; a value of 0 indicates that the offspring's mother was not diagnosed before the child's birth. For the NOT_IUE

variable, a value of 1 indicates that the mother was documented to not have diabetes more than one year after the child's birth. Otherwise, NOT_IUE was assigned a value of 0.

INUTERO \ NOT_IUE	1	0
1	• Mother **was** diagnosed with diabetes before birth of offspring • Mother **was** documented to not have diabetes >1 year after offspring's birth	• Mother **was** diagnosed with diabetes before birth of offspring • Mother **was not** documented to not have diabetes >1 year after offspring's birth
0	• Mother **was not** diagnosed with diabetes before birth of offspring • Mother **was** documented to not have diabetes >1 year after offspring's birth	• Mother **was not** diagnosed with diabetes before birth of offspring • Mother **was not** documented to not have diabetes >1 year after offspring's birth

Figure 2.7. Schematic summarising the categories of likelihood of an offspring's exposure to maternal diabetes *in utero*, based on diagnosis of diabetes before the birth of the offspring (INUTERO) and whether the mother was documented to not have diabetes more than one year after the offspring's birth (NOT_IUE).

Overall, diagnoses of diabetes that were observed in mothers were taken to be irreversible; thus, the three categories—exposed to maternal diabetes, likely unexposed, and unknown—were assumed to be mutually exclusive.

A summary of the binary traits in the birth weight data set—sex, INUTERO, and NOT_IUE—is provided below (Table 2.5).

Table 2.5. Basic information for binary traits for individuals in the birth weight data set (N = 3,700).

Variable	Number of individuals
Sex	
Female	2,037
Male	1,663
INUTERO	
0	3,538
1	162
NOT_IUE	
0	1,265
1	2,435

Study staff had not directly asked participants about parental diabetes. In addition to the variables used as a proxy for *in utero* exposure to maternal diabetes as described earlier in this section, parental diabetes had been defined from examinations of the parents of certain study participants, in the longitudinal study. This resulted in three categories (yes, no, or unknown) per parent. Given the ascertainment of the date of onset of diabetes in the longitudinal study, parental diabetes can be defined with respect to specific dates. For example, maternal diabetes that occurred before the offspring's birth could be used as a proxy for the offspring's exposure to maternal diabetes *in utero*.

2.2.4. Multiple imputation of gestational age

This section discusses the rationale behind and methodology of multiple imputation, which was used to address the missingness in the phenotype of gestational age within the PECRB birth weight cohort (25.7% of participants did not have data for gestational age). Missing data is a common occurrence in clinical epidemiological research; there are three main types of missingness: missing completely at random (MCAR), missing at random (MAR; missingness

depends on observed information), and missing not at random (MNAR; missingness depends upon unobserved information) (Pedersen et al., 2017). Multiple approaches to dealing with such missingness have been developed, including: complete-case analysis, the missing indicator method, single value imputation, sensitivity analyses with worst- and best-case scenarios, and multiple imputation (Pedersen et al., 2017). Multiple imputation is widely used in epidemiologic studies, as it addresses the issue of standard errors of extremely high or low values that arises from the aforementioned methods and aims to provide unbiased, valid estimates of association based on information from the available data (Klebanoff & Cole, 2008; Sterne et al., 2009). In comparison with single value imputation methods, which tend to inflate significance and underestimate standard errors, multiple imputation accounts appropriately for uncertainty in the imputed values.

Multiple imputation is normally done in three stages. First, independent variables that may help to impute variables with missing data are selected and used to generate multiple data sets with missing values imputed—potentially with individual values that vary among data sets—using the distribution of the missing data given in the observed data (Rubin, 1996). Such independent variables should include the variables in the analysis model, including the outcome variable (birth weight in the present case) (Pedersen et al., 2017). Five iterations of imputation in data sets have been suggested to be sufficient on theoretical grounds to fill in missing values (Carpenter & Kenward, 2007; Allison, 2000). Second, the analysis of interest is conducted in each imputed data set to calculate the association of interest (Pedersen et al., 2017). Third, measures of association from all imputed data sets are combined according to Rubin's rules (Rubin, 1996), with summary statistics reported to account for variations between and within imputed data sets (Pedersen et al., 2017).

For the analyses described in Chapter 3 that involve adjustment for imputed gestational age, multiple imputation was performed. Multiple imputation was done in SAS 9.4 using PROC MI to generate five multiple imputed data sets, using the following variables for imputation of gestational age: birth weight, sex, maternal age, likelihood of exposure to maternal diabetes (described further in Section 2.2.4), birth year, and the first five genetic PCs. PROC MI was developed for use with data in which missingness is assumed to be at random (Horton & Lipsitz, 2001). The relative efficiency of imputation was approximately 99% on average; this suggests that five cycles of imputation is sufficient, as additional cycles of imputation would add little information. The multiple R^2 between gestational age and the predictors was 0.1309, which is a multiple correlation (R) of 0.36. This correlation is moderate; if only the predicted value were used as if it were estimated without error, the standard errors would be underestimated.

2.2.5. Percentage body fat, insulin action, and insulin response measurements

A subset of the population that had genotypic data available in the longitudinal study also participated in detailed inpatient phenotyping studies that were conducted in the PECRB Clinical Research Centre (N = 557; 230 female and 327 male). Percentage body fat (PFAT) and insulin action as assessed by the M value by the hyperinsulinemic-euglycemic (H-E) 'clamp' (logM) are customarily analysed by NIH/NIDDK PECRB researchers only in individuals without diabetes. PFAT had been determined by underwater weighing as described by Lillioja et al (Lillioja et al., 1993) or by dual X-ray absorptiometry; a regression equation had been used to provide a similar scale for each measure (Tataranni & Ravussin, 1995). PFAT was analysed with adjustment for sex, age, and the first five genetic PCs. The M value was derived from the H-E clamp using methods described by Le et al (Le et al., 2009). The M-value measures the rate of glucose disappearance in response to an insulin

infusion, which thereby reflects insulin sensitivity. Data on these variables and related traits are summarised in Table 2.6.

Table 2.6. Basic information on PFAT, logM and related variables in longitudinal study participants who participated in inpatient phenotyping studies (N = 557).

Variable (units)	Mean	SD
Age at exam (years)	26.81	6.16
Birth year (year)	1963.43	8.01
BMI (kg/m^2)	33.46	8.49
PFAT (%)	32.33	8.49
logM ($\frac{mg}{kg\ EMBS \cdot min}$) †	0.56	0.13

† The units for logM are expressed as mg insulin per kilogram estimated metabolic body size (EMBS) per minute.

Detailed measures of insulin secretion had been measured or calculated only in a further subset of individuals with normal glucose tolerance, as glucotoxicity can influence these traits in those with impaired glucose tolerance or diabetes (Heinitz et al., 2018). This subset of study participants had been admitted to the NIH/NIDDK PECRB Clinical Research Centre for 8-16 days and fed a weight-maintaining diet. After three days, each subject had undergone an intravenous glucose tolerance test (IVGTT) to measure acute insulin response (AIR). Therein, a 25-gram intravenous injection of glucose (50% solution) had been administered; then, samples had been collected 3, 4, and 5 minutes after that timepoint to determine glucose and insulin concentrations. AIR is referred to throughout this thesis as logAIR as the variable was log-transformed to approximate a normal distribution. It had been calculated as the log of the mean plasma insulin concentration above basal from the third to fifth minute after the glucose bolus (Janssen et al., 1994; Thompson et al., 1995). logAIR (N = 404; 147 female and 257 male) was analysed with adjustment for age, sex, PFAT, logM, and the first five genetic PCs. Traits in this subset of participants are detailed in Table 2.7.

Table 2.7. Information on logAIR and related traits in longitudinal study participants who participated in inpatient phenotyping studies and had normal glucose tolerance.

Variable (units)	Mean	SD
Age at exam (years)	26.77	6.14
Birth year (year)	1964.25	7.73
BMI (kg/m^2)	32.73	7.16
PFAT (%)	31.04	8.47
logAIR ($\frac{\mu U}{mL}$) †	2.30	0.28

† The units for logAIR are expressed as the concentration of insulin, in micro-units (μU) insulin per mL.

30-minute corrected insulin response (CIR30, N = 401; 147 female and 254 male) had also been calculated in these same individuals from the inpatient studies from 30-minute plasma glucose and 30-minute insulin concentrations. Traits in this subset of participants are detailed in Table 2.8.

Table 2.8. Information on CIR30 and related traits in longitudinal study participants who participated in inpatient phenotyping studies and had normal glucose tolerance.

Variable (units)	Mean	SD
Age at exam (years)	26.77	6.14
Birth year (year)	1964.30	7.72
BMI ($\frac{kg}{m^2}$)	33.46	8.49
PFAT (%)	31.13	8.44
logAIR ($\frac{\mu U}{mL}$) †	2.30	0.28
CIR30 ($\frac{\mu U}{mg}$)	0.029	0.03

† The units for logAIR are expressed as the concentration of insulin, in micro-units (μU) insulin per mL.

Analyses with CIR120 were adjusted for sex, age, PFAT, logM, and the first 5 genetic PCs.

A schematic of individuals included in various data sets and subsets—defined based on the availability of genotypic and phenotypic data—is provided below (Figure 2.8).

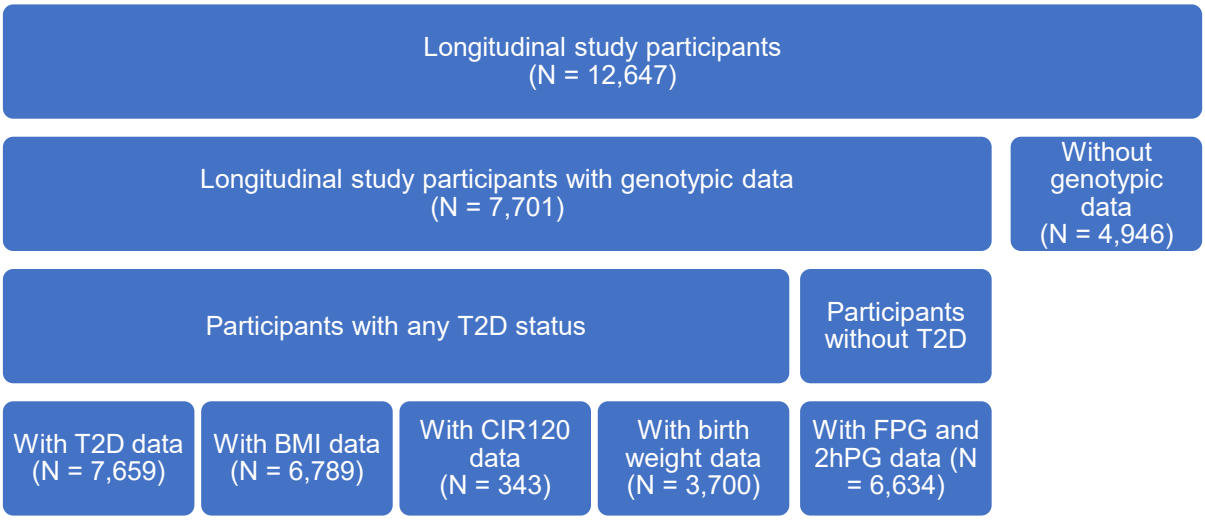


Figure 2.8. Schematic of subsets of individuals in the longitudinal study.

Additionally, a schematic of the study population that had participated in both the longitudinal study and had genotypic data available that had also participated in inpatient phenotyping studies that were conducted in the affiliated NIH/NIDDK PECRB Clinical Research Centre is provided below (Figure 2.9).

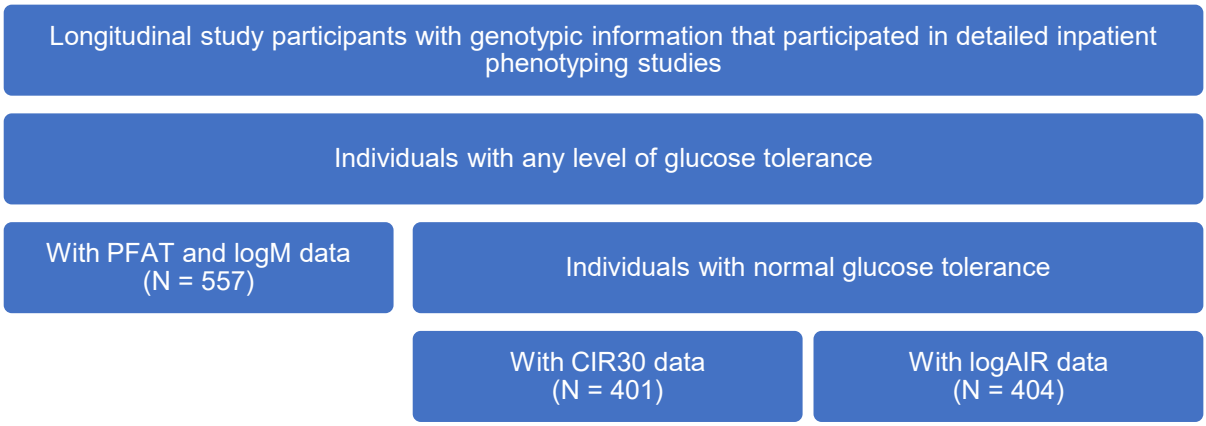


Figure 2.9. Schematic of subsets of individuals who participated in the longitudinal study and NIH/NIDDK PECRB Clinical Research Centre phenotyping studies.

2.3. Genetic approaches to studying birth weight and T2D

2.3.1. Heritability analyses

Heritability estimates allow discrimination between genetic and environmental contributors to phenotypic variance, which can be separated into multiple categories of sources of genetic variance (Figure 2.10) (Mayhew & Meyre, 2017).

Additive genetic influences	Non-additive genetic influences	Shared environmental factors	Unique environmental influences
• Sum of the effect of each allele at all loci that influence the phenotype	• Dominance: interaction between alleles at the same locus • Epistasis: interaction between alleles at different loci	• Influences common to members of a family • e.g., socioeconomic status	• e.g., differences in parental treatment, prenatal environment, life events

Figure 2.10. Summary of definitions of heritability (Mayhew & Meyre, 2017).

Narrow-sense heritability, which captures only additive genetic influences, is calculated as the fraction of the total variance in a trait (σ_B^2) that is explained by additive genetic variance (σ_A^2) (Equation 2.2):

$$h^2 = \frac{\sigma_A^2}{\sigma_B^2} \quad (2.2)$$

To calculate narrow-sense heritability (h^2), one can estimate σ_A^2 and σ_B^2 from a generalized linear model that assesses phenotypic resemblance based on a genetic relatedness matrix. One can test the significance of h^2 based on the comparison of log-likelihood values for two different generalized linear models: the sporadic model (which assumes no genetic effects)

and the polygenic model (which estimates genetic effects) (Mayosi et al., 2002). Heritability estimates reflect phenotypic resemblance among family members; in addition to genetic effects, they can also reflect clustering of environmental risk factors within families, depending upon study design. Thus, h^2 estimates the proportion of variance in the trait that is potentially attributable to additive genetic effects.

2.3.2. Bivariate genetic correlation analyses

LD score regression analyses were undertaken to assess the shared genetic variance between pairs of traits. This enables investigation as to whether a correlation between two traits, as opposed to a SNP-trait correlation, is attributable to the presence of genetic and/or environmental confounders (Lee et al., 2018). First, the conventional LDSC (Linkage Disequilibrium SCore regression) method, which calculates LD scores (i.e., the sum across all SNPs of each SNP's x^2 with a given focal SNP) statistics and regression analyses, was tested (Ni et al., 2018). However, this method returned impossible estimates for genetic correlations and estimates for heritability (observed scale; determined from the slope from the regression) and regression statistics that were lower than expected. This can likely be ascribed to the low sample sizes—by LDSC standards—that were available for each trait, as a total of 7,701 individuals in the longitudinal had genotypic data, and fewer had each phenotypic trait (maximum trait N = 7,659 for T2D) and the high degree of LD observed in the Indigenous study population. Thus, an alternative method for calculating genetic correlation among traits was undertaken within SOLAR-Eclipse; this method will be referred to as bivariate genetic correlation analyses.

SOLAR-Eclipse, an imaging genetics analysis software, incorporates phenotype, pedigree, and pairwise relatedness data to conduct each GWAS (Blangero et al., 2013), genome-wide heritability analysis (Ge et al., 2015), and bivariate genetic correlation analysis. SOLAR-

Eclipse conducts bivariate genetic correlations by calculating the significance of genetic (testrhog option) and environmental (testrhoe option) correlation and outputting heritability data per trait, after adjusting for specified covariates. This is accomplished by fitting a bivariate mixed model in which the phenotypic resemblance among pairs of individuals in the population is a function of their genetic relatedness, and the genetic covariance is a function of the shared variance between traits within pairs of individuals. In the present analyses, genetic relatedness was estimated empirically from the proportion of alleles shared identical by descent among all pairs of individuals in the population, as described below. The overall phenotypic correlation between the two traits, Pearson's *r*, is defined as ρ_P (Equation 2.3):

$$r = \rho_P = \sqrt{h_A^2}\sqrt{h_B^2} \cdot \rho_G + \sqrt{1 - h_A^2}\sqrt{1 - h_B^2} \cdot \rho_E \qquad (2.3),$$

where genetic correlation (ρ_G) is the proportion of variability between traits that is due to shared genetic effects, while environmental correlation (ρ_E) is the environmental contribution. When ρ_P and ρ_E are significant, traits in the model are suggested to be influenced by shared genetic and/or environmental factors, respectively (Seidlerová et al., 2008). Bivariate genetic correlation analyses were conducted within SOLAR-Eclipse using the 'polygenic' command; specific information on traits and covariates are provided in Section 3.2.

2.3.3. Genome-wide association studies

The advent of genome-wide association studies has enabled investigations of genetic associations across the entire genome and was made possible by the Human Genome Project and the International HapMap project (Belmont et al., 2003; Craig Venter et al., 2001; Lander et al., 2001). GWAS, in which large numbers (i.e., hundreds of thousands to millions)

of single-nucleotide polymorphisms (SNPs) can be assayed across many individuals to identify those significantly associated with a specific trait, were a key advance beyond candidate gene and family-based linkage studies in studying the genetics of complex traits (Manolio et al., 2009). In GWAS, linear or logistic regression models are often used to test for associations, including covariates to account for stratification and address confounding effects (Leiserson et al., 2013); for instance, they can employ logistic regression models to study binary traits such as T2D or linear regression models to test for genetic associations with continuous traits such as birth weight (Figure 2.11).

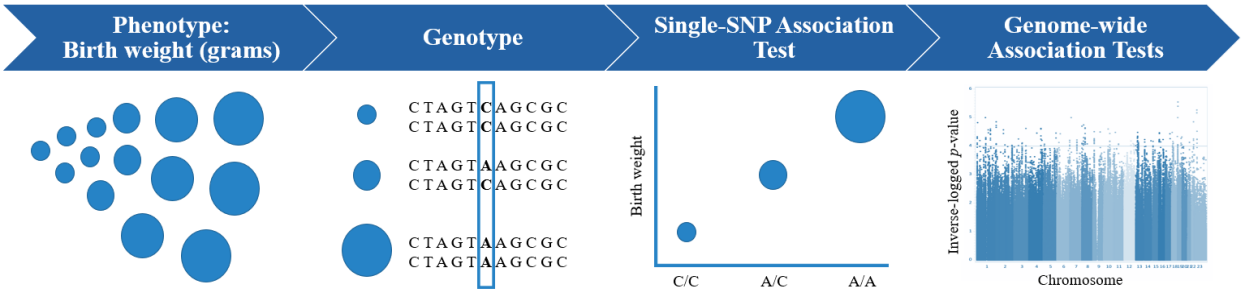


Figure 2.11. Summary schematic of GWAS, in which phenotypic data (e.g., birth weight) are regressed against genotype (and often covariates) across many genetic variants.

Power for genomic discovery can be increased by increasing sample size, which often entails conducting GWAS meta-analyses across multiple cohorts. GWAS rely upon the presence of blocks of high LD surrounding tagged SNPs to identify SNPs that are significantly associated with a given trait without necessarily typing every neighbouring genetic variant, reducing costs while limiting loss in statistical power (Daly et al., 2001). If the genotyping arrays are designed to tag most of the variants segregating in a population, then causal variants that influence a particular trait are likely to be captured. This can be done through direct genotyping or imputation and reflected in strong statistical associations between variants in the region surrounding the causal variant and the trait of interest.

Differences in LD patterns (e.g., observed differences in local LD block structure)—attributed to differences in recombination rates, demography, and genome and sampling variation—could lead to a decrease in power of GWAS based on tag SNPs and limit transferability across study populations (Greenwood et al., 2004; Gu et al., 2007; Liu et al., 2004; Osabe et al., 2007). Different continental ancestry groups are more likely to have divergent LD patterns based on different population histories; thus, transferability of GWAS results, as well as polygenic scores, across continental ancestry groups is particularly important to assess. Findings of a recent study of the extent to which differences in LD and MAF among ancestry groups can explain the loss of relative accuracy of polygenic scores suggested that causal variants that underlie the common genetic variation in European-ancestry GWAS are mostly shared across continents (Y. Wang et al., 2020). Notably, the study reported that differences in LD and MAF among ancestry groups can explain 70-80% of the loss of relative accuracy of European-based polygenic scores in African ancestry for T2D (Y. Wang et al., 2020).

GWAS and investigations into the 'missing heritability' (Manolio et al., 2009) of traits such as T2D, BMI and height emphasize that larger sample sizes, inclusion of study populations with different LD patterns and increased imputation density, coupled with functional data, will enable further genetic discovery for such complex traits (Mahajan et al., 2014; Scott et al., 2017; Yang et al., 2015). Previous studies have suggested that additional SNPs in the regions surrounding established SNPs—exemplifying allelic heterogeneity—are associated with the risk of T2D and other complex diseases (Morris et al., 2012; Yang et al., 2012). A study using summary statistics from the DIAGRAM Consortium meta-analysis of T2D GWAS and LD patterns inferred from a large reference sample identified novel SNPs surrounding previously identified T2D risk loci and demonstrated slightly improved prediction of T2D risk (Klimentidis et al., 2014).

GWAS analyses that are described in this thesis were undertaken in SOLAR-Eclipse (Sequential Oligogenic Linkage Analysis Routines) version 8.4.1 (University of Maryland School of Medicine; Baltimore, Maryland, US) (Blangero et al., 2013). SOLAR-Eclipse allows for the specification of a pedigree file, phenotypes file, and kinships matrix file (i.e., containing a genetic relatedness matrix). SOLAR-Eclipse can conduct mixed model association analyses that account for the specified relationships. In the current analyses, these relationships were estimated empirically based on the 482,616 directly genotyped autosomal GWAS markers with MAF > 0.05 for all pairs of individuals in the population who had been genotyped (29,648,850 pairs). Genomic segments shared identical by descent were identified using the *fastIBD* function of the genetic analysis software BEAGLE, as previously described by developers Browning & Browning as well as PECRB colleagues (Browning & Browning, 2011; Piaggi et al., 2017). Mixed model analyses are a useful tool for GWAS because they can assess associations while accounting for genetic relationships among individuals and avoid inflation of statistical significance that could appear due to the relatedness among study participants.

Accounting for relationships is particularly important in the present study population since many individuals are related. Conventional techniques for conducting mixed model analyses are computationally intensive and time-consuming with large sample sizes. Recent advances in statistical techniques have provided more rapid computations that are suitable for the repeated applications in large sample sizes, as required for GWAS. SOLAR-Eclipse employs an eigenvector decomposition method to increase the speed of the computations (Blangero et al., 2013). It provides efficient mixed model analyses for small- to medium-sized data sets and has been used historically to conduct genetic epidemiology studies within the current study population due to its relatively small study population size (7,701 individuals have genotypic data available; fewer individuals have phenotypic data available for T2D and

related traits) and relatedness among participants in the longitudinal study. The 'measured genotype association' (MGA) command was used within SOLAR-Eclipse to compute associations between genetic and phenotypic variables, with adjustment for relevant covariates.

2.3.4. Construction of polygenic scores for T2D

I compared associations of 10 different constructions of T2D PS—derived from results of GWAS conducted for populations from various world regions—with the risk of T2D and with related traits in cross-sectional (Chapters 3 and 4) and prospective analyses (Chapter 5). Each weighted T2D PS *i* (with *n* variants) was calculated using the sum of products of each effect size (*β,* i.e., log of the odds ratio) and dosage of effect alleles for each variant *j* (Equation 4.2).

$$T2D\ PS_i = \sum_{j=1}^{n_i} \beta_j \cdot dosage_j \qquad (4.2)$$

The current analyses, along with those detailed in Chapter 5, employed the use of 10 T2D PS that I constructed using the methods detailed in Section 2.4.4. In brief, genotypic data was derived from the Indigenous study population from the Southwestern US, 7,701 of whom had genotypes available in a GWAS of T2D that had been previously performed by collaborators at NIH/NIDDK PECRB. Ten different T2D PS were constructed, derived from the summary statistics of GWAS meta-analyses that were conducted among study populations from various world regions, or the current study population's GWAS of T2D, and the current study population's imputed genotypes. The 10 T2D PS, each named for the meta-analysis from which it was derived, are listed with basic accompanying details in Table 2.9.

Table 2.9. Summary of the 10 constructions of T2D employed in this thesis research.

T2D PS name	Number of SNPs	Variants derived from this GWAS or population	Weights derived from this GWAS or population
AGEN 2020	125	East Asian-ancestry populations (Spracklen et al., 2020)	East Asian-ancestry populations (Spracklen et al., 2020)
DIAGRAM 2018	293	European-ancestry populations (Mahajan et al., 2018b)	European-ancestry populations (Mahajan et al., 2018b)
DIAMANTE 2020 – Multi-ancestry African	276	Multi-ancestry meta-analysis – African-ancestry populations (Mahajan et al., 2020)	Multi-ancestry meta-analysis – African-ancestry populations (Mahajan et al., 2020)
DIAMANTE 2020 – Multi-ancestry East Asian	280	Multi-ancestry meta-analysis – East Asian-ancestry populations (Mahajan et al., 2020)	Multi-ancestry meta-analysis – East Asian-ancestry populations (Mahajan et al., 2020)
DIAMANTE 2020 – Multi-ancestry European	287	Multi-ancestry meta-analysis – European-ancestry populations (Mahajan et al., 2020)	Multi-ancestry meta-analysis – European-ancestry populations (Mahajan et al., 2020)
DIAMANTE 2020 – Multi-ancestry Hispanic/Latino	287	Multi-ancestry meta-analysis – Hispanic/Latino-ancestry populations (Mahajan et al., 2020)	Multi-ancestry meta-analysis – Hispanic/Latino-ancestry populations (Mahajan et al., 2020)
DIAMANTE 2020 – Multi-ancestry South Asian	282	Multi-ancestry meta-analysis – South Asian-ancestry populations (Mahajan et al., 2020)	Multi-ancestry meta-analysis – South Asian-ancestry populations (Mahajan et al., 2020)
DIAMANTE 2020 – Multi-ancestry Composite	287	Multi-ancestry meta-analysis – all ancestry groups (Mahajan et al., 2020)	Multi-ancestry meta-analysis – all ancestry groups (Mahajan et al., 2020)
Population-specific weight	287	Multi-ancestry meta-analysis (Mahajan et al., 2020)	Indigenous longitudinal study population
Population-specific variant	287	287 with lowest p-values of association with T2D across the genome, in the Indigenous study population	Indigenous longitudinal study population

First, the AGEN 2020 T2D PS included 125 SNPs and weights that were derived from a GWAS meta-analysis of East Asian-ancestry study populations (Spracklen et al., 2020). Summary statistics for the SNPs included in the AGEN 2020 T2D PS are in Table 7.2.1. Second, results from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium GWAS meta-analysis of European ancestry study populations were used to

construct the DIAGRAM 2018 T2D PS. In the original DIAGRAM 2018 meta-analysis, a total of 403 independent SNPs met the genome-wide significance threshold ($p < 5.0 \times 10^{-8}$); imputed genotypes for 293 of those SNPs were available in the present study population (Mahajan et al., 2018b). Summary statistics for the SNPs included in the DIAGRAM 2018 T2D PS are in Table 7.2.2.

Seven of the T2D PS were derived from the overlap of genotypes from the present Indigenous study population and genome-wide significant SNPs from the DIAMANTE 2020 multi-ancestry GWAS meta-analysis (Mahajan et al., 2020). The genome-wide significant ($p < 5.0 \times 10^{-8}$) SNPs from the GWAS meta-analysis constituted up to 287 variants. Weights for six DIAMANTE 2020 multi-ancestry GWAS T2D PS had weights derived from GWAS meta-analyses that were conducted among study participants of specific ancestry groups: composite, African, East Asian, European, Hispanic/Latino and South Asian (Mahajan et al., 2020). Weights for variants included in the one remaining T2D PS that was derived from the overlap between study population genotypes and the DIAMANTE 2020 multi-ancestry GWAS meta-analysis was the 'population-specific weight' T2D PS: the weights for those variants taken from the DIAMANTE analyses were derived from the GWAS of T2D that had previously been conducted in the Indigenous study population and underwent 10-fold cross-validation to address overfitting. Summary statistics for the SNPs included in these seven T2D PS are in Table 7.2.3.

To estimate population-specific weights for variants constituting the population-specific weight T2D PS, I used block-jackknifing among all participants in the GWAS (N = 7,701) to reduce overfitting associated with deriving the weights in the same dataset in which they are applied. The participants were randomly divided into 10 equally sized groups, and the weights for each group were calculated by analysing the association of each variant with T2D in the other 9 groups. Associations were analysed with prevalent T2D at last examination

(2571 cases among 7,659 participants with data available) with adjustment for age, sex, birth year and the first 5 genetic principal components (to account for population structure). Analyses were conducted with a linear mixed model, which was fit with SOLAR-Eclipse (Almasy & Blangero, 1998), that accounted for pairwise relationships among individuals. T2D was modelled as a continuous (0 to 1, inclusive) variable to reduce computation time, which exceeded the constraints of the National Institutes of Health High Performance Computing environment, Biowulf. The regression coefficient was converted to an odds ratio by the method of Haggstrom (Haggstrom, 1983). The population-specific weights, derived from the average logarithm of the odds ratio across all 10 sub-analyses, are shown in Table 7.2.3, along with weights for each ancestry group derived from the DIAMANTE 2020 meta-analyses.

To assess the utility of using further population-specific genetic information, I also derived a population-specific variant T2D PS by selecting T2D-associated variants from the 515,692 variants directly genotyped in the T2D GWAS in the Indigenous study population. For computational efficiency and to reduce overfitting with assessments over a large number of variants, I used 2-fold cross-validation in this situation, since lower values of k in k-fold cross-validation can help reduce overfitting when a large number of features is assessed (Scheinost et al., 2019). The GWAS participants were randomly divided into two equally sized groups, and a GWAS of T2D was conducted in each group (the ‘discovery’ group) to select the variants and weights used to construct the PS for the other group (the ‘target’ group). To reduce the influence of LD among variants in close physical proximity, the genome was divided into bins of ~200kb using the k-means procedure (PROC FASTCLUS in SAS), resulting in 25,763 bins across the genome.

Within each discovery group, the variant with the lowest p-value for association with T2D was taken within each bin. The number of variants that were used to construct this population-

specific variant score was 287: this number of SNPs was selected to obtain a PS with number of variants and information that were comparable to those from the DIAMANTE 2020 and DIAGRAM 2018 PS. To avoid loss of statistical power associated with further splitting of the sample, I did not attempt to optimize the number of variants; I also did not attempt further validation in additional samples. Thus, while the number of variants in this score is similar to those in other T2D PS, its generalizability is not certain. Summary statistics for the SNPs included in the population-specific variant T2D PS are located in Table 7.2.4.

3. Genetic analyses of birth weight

3.1. Introduction

Epidemiologic studies in many populations have shown that birth weight is associated with risk of type 2 diabetes (T2D) in adulthood. Multiple meta-analyses have reported that the relationship between birth weight and T2D is inversely related to T2D risk in some study populations; in others, the relationship is U-shaped (Harder et al., 2007; Whincup et al., 2008). Studies have used increasingly sophisticated methods to assess the genetic relationship between birth weight and T2D, with some variants conferring susceptibility to low birth weight and subsequent diabetes (i.e., aligning with the foetal insulin hypothesis) and some conferring susceptibility to high birth weight and subsequent diabetes (further described in Section 1.7). For instance, a study of the genetics of birth weight by Warrington et al included a foetal GWAS and maternal GWAS and employed structural equation modelling to assess the contributions of genetic effects of both foetus and mother (Warrington et al., 2019). In this study, the genetic correlation between birth weight and T2D susceptibility was negative for the foetal genotypes while it was near 0 for the maternal genotypes; for fasting plasma glucose, the genetic correlation was positive for the maternal genotype and negative for the foetal genotype (Warrington et al., 2019).

Previous studies involving the PECRB cohort have demonstrated that lower and higher birth weight groups are at higher risk for T2D than those with normal birth weight (Lindsay et al., 2000). Genome-wide association studies, genetic correlation analyses and PS allow for further investigation of the genetic epidemiology of birth weight and genetic relationships between this trait and T2D. In the cohort study in which this thesis research was undertaken, there is an R1420H loss-of-function variant in *SUR1* (*ABCC8*), which has a carrier rate of

3.3% in participants in the longitudinal study (of the 7,701 individuals with genotype data), and which is virtually absent in other study populations (Baier et al., 2015). This variant has been associated with higher mean birth weight (for the minor allele, β = 169.1 grams, p = 1.5×10^{-3}) and higher risk of T2D (OR = 2.02 [95% CI 1.45-2.82] per copy of the minor allele, p = 3.6×10^{-5}); however, neither of these results reach genome-wide significance (Baier et al., 2015). The R1420H loss-of-function variant in *SUR1* (*ABCC8*) has functional similarities to those known to cause foetal hyperinsulinemia and hyperinsulinemic hypoglycaemia of infancy (HHI) (Baier et al., 2015). HHI often results in increased birth weight--consistent with the foetal insulin hypothesis—and early-onset diabetes (Hattersley & Tooke, 1999). Missense variants in the genes *KCNJ11* and *ABCC8*, which encode subunits of K_{ATP} channels in pancreatic β cells, have previously been implicated in HHI and T2D (Gloyn et al., 2003; Kapoor et al., 2013).

The objective of the research in this chapter is to further explore the genetic relationship between birth weight and T2D and investigate genetic associations with birth weight using different models of GWAS, within the PECRB cohort. Specifically, seven models of GWAS of birth weight were conducted, including: additive and dominant-recessive models of GWAS, conditioning on maternal and foetal genotype, and different combinations of relevant phenotypic covariates. Bivariate genetic correlation analyses were undertaken with respect to birth weight and metabolic traits. In addition, a PS for birth weight and one for T2D were tested for relationships with birth weight and T2D.

3.2. Methods

3.2.1. Imputation of gestational age

As described in Section 2.2.3, 74.3% of the individuals in the birth weight data set (N = 3,700) had gestational age data available (N = 2,749). To address missingness for the 26.7% of participants who did not have gestational age data available, which was assumed to be at random (i.e., MAR), multiple imputation was performed. Then, five iterations of GWAS of birth weight were undertaken across five respective phenotypic data sets, among study participants who had birth weight and genotypic data available in the longitudinal study. Each GWAS was adjusted for birth year, the first five genetic PCs, and imputed gestational age.

3.2.2. Analysing the association between birth weight and T2D prevalence

The relationship between foetal (i.e., own) birth weight and the risk of T2D in the present study population has previously been reported to be U-shaped (Lindsay et al., 2000; McCance et al., 1994). Further details on this complex relationship, including genetic and non-genetic influences upon it, are explored further in Section 1.7. The analysis of the association between birth weight and T2D in this study population, with the addition of more individuals with birth weight data available since Lindsay and McCance's earlier studies (described further in Section 2.2.3), was performed in a slightly smaller subset of individuals that had GWAS data, as well as data for birth weight and T2D available (N = 3,690; 2,033 female and 1,657 male). For presentation purposes, the adjusted prevalence of T2D was calculated (adjusted to the mean value of all covariates).

SOLAR-Eclipse was used to test whether the relationship between birth weight and prevalence of T2D was linear or quadratic. The log likelihood, the natural log of the maximised model, was computed using the 'polygenic' command, for two models with different sets of covariates: (1) those in the linear model (age, sex, genetic PCs 1-5, birth year, and birth weight in grams (bwt)); and (2) those in the quadratic model (age, sex, genetic PCs 1-5, birth year, bwt, and bwt^2). The bwt^2 (i.e., birth weight squared) term was added to

the linear model to generate the quadratic model to test whether the relationship between birth weight and T2D risk was linear. The chi-square statistic was calculated as twice the difference of the log likelihood estimates and the one-tailed probability from the chi-square distribution was calculated from the x^2 distribution, with the x^2 statistic and one degree of freedom specified.

For plotting the relationship between birth weight and the estimated adjusted prevalence of T2D, birth weight was partitioned into six categories. These categories allowed for the U-shaped relationship between birthweight and the prevalence of T2D to be captured, and also aligned with previous studies' definitions of low (less than 2,500; 3,000; or 3,500 grams) and high birth weight (greater than or equal to 4,000 or 4,500 grams) (Egeland et al., 2000; Ørskou et al., 2003; Rode et al., 2007). Category 1 contained individuals with a birth weight less than 2,500 grams and with category 6 being individuals with a birth weight greater than 4,500 grams, as displayed in the table below (Table 3.1).

Table 3.1. Partitioning of birth weight into six categories.

Category of birth weight	Range of grams (g)	Number of individuals
1	< 2,500	82
2	2,500 ≤ bwt < 3,000	506
3	3,000 ≤ bwt < 3,500	1,411
4	3,500 ≤ bwt < 4,000	1,202
5	4,000 ≤ bwt < 4,500	373
6	> 4,500	116

The 'polygenic' command in SOLAR-Eclipse was used to analyse the association between birth weight and T2D, to generate estimates for T2D prevalence by category of birth weight, as adjusted for age, sex, birth year, and the first five genetic PCs. The standard errors for the adjusted prevalence of T2D were calculated using the delta method (Sokal & Rohlf, 1981).

Therein, the standard errors for each of the parameters that were used to calculate the adjusted prevalence of T2D (i.e., mean, sex coefficient, birth weight category coefficient), and the covariances between them, were employed. The standard error (*SE*) of the adjusted prevalence of T2D was then calculated as the square root of the delta value (Δ), which was the sum of the variances of each of the relevant parameters plus twice the covariance of each pair of parameters. For birth weight category 4 (with 4,000g ≤ bwt < 4,500g), which was the reference category, the adjusted prevalence of T2D depended only upon the mean prevalence of T2D (μ) and the sex coefficient (β_{sex}), and was calculated as (Equation 3.1):

$$Adjusted\ prevalence\ of\ T2D = \mu + (0.55\beta) \quad (3.1),$$

where 0.55 signifies the mean value of the sex covariate, as 55% of participants were female. The Δ for the reference category of birth weight was defined as the following (Equation 3.2):

$$\Delta_{ref\ cat\ bwt} = SE(\mu)^2 + 0.55^2 \times {SE_{\beta_{sex}}}^2 + 2 \times 0.55 \times R_{\mu\beta_{sex}} \times SE_{\mu} \times SE_{\beta_{sex}} \quad (3.2),$$

where $R_{\mu\beta_{sex}}$ is the estimated correlation between the mean estimate and β_{sex}. The *SE* of the adjusted prevalence of T2D was therefore (Equation 3.2):

$$SE_{ref\ cat\ bwt} = \sqrt{\Delta_{ref\ cat\ bwt}} \quad (3.2)$$

For the other, non-reference, categories of birth weight, the beta coefficient for each category *i* (β_i) was added to calculate the Δ (Equation 3.3):

$$\Delta_{cat_i\,bwt} = SE(\mu)^2 + 0.55^2 \times {SE_{\beta_{sex}}}^2 + {SE_{\beta_i}}^2 + 2 \times 0.55 \times R_{\mu\beta_{sex}} \times SE_{\mu} \times SE_{\beta_{sex}} + 2 \times 0.55 \times R_{\beta_{sex}\beta_i} \times SE_{\beta_{sex}} \times SE_{\beta_i} \quad (3.3)$$

The *SE* of the adjusted prevalence of T2D for each category *i* was therefore (Equation 3.4):

$$SE_{bwt\,cat_i} = \sqrt{\Delta_{bwt\,cat_i}} \quad (3.4)$$

The results of the aforementioned calculations are displayed in a graph of birth weight categories and the estimated adjusted prevalence of T2D (Section 3.3.1).

3.2.3. Heritability and bivariate genetic correlation analyses

Narrow-sense heritability and bivariate genetic correlation analyses were also undertaken; these methodologies are described further in Sections 2.3.1 and 2.3.2. Heritability analyses were conducted using the empirically determined genetic relationship matrix with the 'polygenic' command in SOLAR-Eclipse, for two main models: birth year and the first five genetic PCs (Model 1); and birth year, the first five genetic PCs, and *in utero* exposure to maternal diabetes (Model 2). Bivariate genetic correlation analyses, which are described in detail in Section 2.2.3, were undertaken to further explore the relationships between related covariates and T2D. These were completed using the 'polygenic' command in SOLAR-Eclipse, between birth weight and T2D, as well as related traits in the longitudinal study. Table 3.2 lists and describes the main outputs of the various bivariate genetic correlation analyses between birth weight and other traits (Bunt et al., 2007).

Table 3.2. Descriptions of main outputs of bivariate genetic correlation analyses.

Measure	Description
h^2	Narrow-sense heritability estimate for a trait
ρ_E	Environmental correlation: proportion of variability between traits due to shared environmental effects between traits
ρ_G	Genetic correlation: proportion of variability between traits due to shared genetic effects
$p(\rho_G$ **is nonzero)**	*p* value for the test of whether genetic correlation between traits is nonzero
ρ_P	Estimate of the phenotypic correlation between traits, derived from calculated values of ρ_E and ρ_G

Genetic correlation analyses were conducted between birth weight and traits collected in NIDDK Phoenix studies using methods previously described: diabetes, OGTT (listed as GTT fasting and 2-hour), fasting insulin, height, HOMA-β, HOMA-IR, log(M-value), log(maximum BMI), and log(acute insulin response) (Table 3.3).

Table 3.3. Overview of the bivariate genetic correlations and covariates.

Trait 1	Trait 1 is adjusted for these covariates	Trait 2	Trait 2 N	Trait 2 is adjusted for these covariates
Birth weight	Birth year, PC 1-5	T2D	7,659	Birth year, PC 1-5, age at exam, sex
Birth weight	Birth year, PC 1-5	GTT (fasting)	6,637	PC 1-5, age at exam, sex
Birth weight	Birth year, PC 1-5	GTT (2-hour)	6,637	PC 1-5, age at exam, sex
Birth weight	Birth year, PC 1-5	Fasting insulin	5,349	PC 1-5, age at exam, sex, BMI
Birth weight	Birth year, PC 1-5	Height (cm)	5,719	PC 1-5, sex
Birth weight	Birth year, PC 1-5	HOMA-β	5,342	PC 1-5, age at exam, sex, BMI
Birth weight	Birth year, PC 1-5	HOMA-IR	5,349	PC 1-5, age at exam, sex, BMI
Birth weight	Birth year, PC 1-5	log(M-value)	557	PC 1-5, age, sex, PFAT, $PFAT^2$
Birth weight	Birth year, PC 1-5	log(Maximum BMI)	6,789	Birth year, PC 1-5, sex, age at maximum BMI measurement
Birth weight	Birth year, PC 1-5	log(Acute insulin response)	404	PC 1-5, age, sex, PFAT, log(M-value)

Notably, the bivariate genetic correlation analysis for birth weight and T2D had to be modified to complete the analysis by opting to omit conversion to liability scale (i.e., to treat T2D as a continuous trait), as the numerical integrations required for analysis on the liability scale resulted in excessively long computation time (run times exceeding 100 days, which were interrupted by regular maintenance of the high performance computing server). Although the overall relationship between birth weight and T2D has been demonstrated to be nonlinear (i.e., U-shaped) in the PECRB cohort (described further in Section 3.2.2), row 1 in Table 3.3 represents the bivariate genetic correlation analysis that was conducted for T2D and for birth weight assumes that genetic and environmental correlations are linear (although they can be in opposite directions). Additionally, these analyses were presented without decomposing foetal and maternal genetic information.

3.2.4. GWAS of birth weight

GWAS of 3,700 participants' (2,037 female and 1,663 male) own birth weight were conducted using genotypic and phenotypic data from the PECRB study. Further information on this subset of individuals from the longitudinal study—including participant characteristics—are available in Section 2.2, which outlines the various subsets of the study population in which analyses throughout this thesis were undertaken. All were conducted using genotypic data that were directly genotyped using a custom Affymetrix Axiom array (512,093 single-nucleotide polymorphisms, SNPs) and imputed using a population-specific reference panel (4,589,902 SNPs), spanning over five million common SNPs altogether. Further details on ascertainment of genotypic data, which had been completed before the analyses that are outlined in this thesis by colleagues at NIH/NIDDK PECRB, are provided in Section 2.1. Birth weight had been ascertained from birth and medical records and was normalised separately by sex via rank-based transformation (the Blom method in SAS 9.4) (Blom, 1958).

Genetic associations with birth weight were analysed using a mixed model in SOLAR-Eclipse (Catonsville, MD), accounting for genetic relationships (inferred from marker data), and adjusting for covariates, in three main models across imputed variants (Table 3.4). Model 1 was an additive foetal GWAS of own birth weight, adjusted for the covariates birth year and the first five genetic PCs. Model 2 was additionally adjusted for *in utero* exposure to maternal diabetes (described further in Section 2.2.3). Model 3 was also an additive foetal GWAS of own birth weight, adjusted for the covariates birth year, the first five genetic PCs, and gestational age by multiple imputation. The covariates in GWAS Model 3 were similar to those in the additive foetal GWAS meta-analysis of own birth weight using data from other (European-ancestry) study populations, within the Early Growth Genetics Consortium: sex, gestational age and the first four genetic PCs (Warrington et al., 2019).

Table 3.4. GWAS of birth weight Models 1-3: conducted across all imputed variants.

	Model 1	Model 2	Model 3
Trait	Own birth weight	Own birth weight	Own birth weight
Genotype (maternal or foetal)	Foetal	Foetal	Foetal
Additive, dominant, or recessive	Additive	Additive	Additive
Adjusted for these covariates	Birth year, PCs 1-5‡	Birth year, PCs 1-5, *in utero* exposure to maternal diabetes	Birth year, PCs 1-5, gestational age (by multiple imputation)
N	3,700	3,700	3,700

‡ PC 1-5 signifies the first five genetic PCs that had been calculated previously by collaborators at NIH/NIDDK Phoenix. All models were adjusted for these first five genetic PCs to account for population stratification. Details are available in Section 2.1.3.

Because five separate phenotypic data sets with varying estimates for gestational age were generated in the multiple imputation process (discussed in Section 3.2.1), five iterations of GWAS were completed in succession, in which each analysis was adjusted for birth year, the first five genetic PCs, and the corresponding iteration of imputed gestational age. Next, the

average of standard errors for each variant across models was calculated according to effect allele frequency. Large test-based standard errors were replaced with this average value because the test-based standard errors used by SOLAR-Eclipse represent overestimates when $p \approx 1$. Finally, results were summarised across iterations by averaging the effect size and standard error (estimated by taking variability across iterations into account) across the five sets of summary statistics per SNP. To estimate significance, the x^2 statistics per SNP were averaged and the right-tailed probability of the x^2 distribution was calculated using the CHIDIST function in Microsoft Excel.

Three other models of GWAS of birth weight beyond those included in Table 3.4 were conducted only among directly genotyped variants, because the computational intensity of these analyses that would otherwise have encompassed over five million imputed variants would have required excessive processing time. These include a dominant-recessive foetal GWAS of birth weight, adjusted for birth year, the first five genetic PCs, and gestational age by multiple imputation (Model 4); an additive foetal GWAS of own birth weight, adjusted for birth year and the first five genetic PCs and conditioned on maternal genotype (Model 5); and an additive maternal GWAS of offspring birth weight, adjusted for birth year, the first five genetic PCs and conditioned on foetal genotype (Model 6; Table 3.5).

Table 3.5. GWAS of birth weight Models 4-7: conducted across directly genotyped variants only.

	Model 4	Model 5	Model 6	Model 7
Trait	Own birth weight	Own birth weight	Own birth weight	Offspring birth weight
Genotype (maternal, foetal)	Foetal	Foetal	Foetal	Maternal
Additive, dominant, or recessive	Dominant-recessive	Additive	Additive	Additive
Adjusted for these covariates	Birth year, PCs 1-5, gestational age (by multiple imputation)	Birth year, PCs 1-5, gestational age (without multiple imputation)	Birth year, PCs 1-5, maternal genotype	Birth year, PCs 1-5, foetal genotype
N	3,700	2,749	2,934	2,934

To assess the effects of multiple imputation of gestational age upon the GWAS of birth weight, a complete-case analysis (Model 4) was undertaken that was analogous to Model 3 (Table 3.4). Dominant and recessive models of GWAS of own birth weight, with adjustment for gestational age (by multiple imputation), were also conducted to investigate whether additional significant associations were identified with alternate genetic models (Model 4). Model 5 was a GWAS of birth weight that only included the 2,749 individuals in the PECRB cohort with non-missing gestational age data. These analyses were conducted because the standard additive model may not be sufficiently powerful to detect associations under alternate modes of inheritance, particularly when lower-frequency alleles are recessive (Freidlin et al., 2002; González et al., 2008). Model 5, the complete case analysis, was restricted to directly genotyped variants. To facilitate pairwise comparisons between Models 3 and 4 and Models 3 and 5 within results sections 3.3.5 and 3.3.6, respectively, QQ and Manhattan plots and tables summarising the top SNPs (per 200kb region, by p-value) were also shown for Model 3, limited to directly genotyped SNPs only.

The additive foetal GWAS of own birth weight was conditioned on maternal genotype, in the subset of individuals whose mothers had genotypic data available in the longitudinal study (Model 6). The additive maternal GWAS of offspring birth weight was conditioned on foetal genotype, in the subset of individuals whose mothers had genotypic data available in the longitudinal study (Model 7). Because substantial inflation was found in the results of Model 7, the results for both Models 6 and 7 were presented with genomic control. Genomic control—a quality control measure for GWAS and meta-analysis that addresses such inflation in test statistics—was conducted by dividing the test statistics across all variants by the genomic inflation factor (λ) (Zheng et al., 2006). λ was calculated as the ratio of the median of the observed x^2 statistics to the median of the expected x^2 statistics (Yang et al., 2011). The genomic inflation factor estimates the amount of inflation relative to test statistics that would be expected according to the global null hypothesis (Yang et al., 2011).

3.3. Results

3.3.1. Analysing the association between birth weight and T2D prevalence

The analyses of the association between birth weight and T2D in this study population were conducted to assess the extent to which the relationship in the present sample (participants in the GWAS) was similar to those that were reported previously in the PECRB cohort at large. These analyses included: (1) a test of whether the addition of a term representing the square of birth weight led to a significant difference in the estimate of the adjusted prevalence of T2D (i.e., whether or not the relationship was U-shaped) and (2) plotting the relationship between categories of birth weight and the adjusted prevalence of T2D. First, the log likelihood for the linear model was 1920.91 and that of the quadratic model was 1930.00. The chi-square statistic was calculated as twice the difference of the log likelihood estimates (x^2 =

18.19) and the *p*-value was taken from the x^2 distribution ($p = 2.0 \times 10^{-5}$). Because the addition of the square term is significant, this demonstrates that the relationship between birth weight and T2D is non-linear. The previously published U-shaped relationship between these traits was replicated in the subset of participants whose data was analysed in this chapter (Figure 3.1).

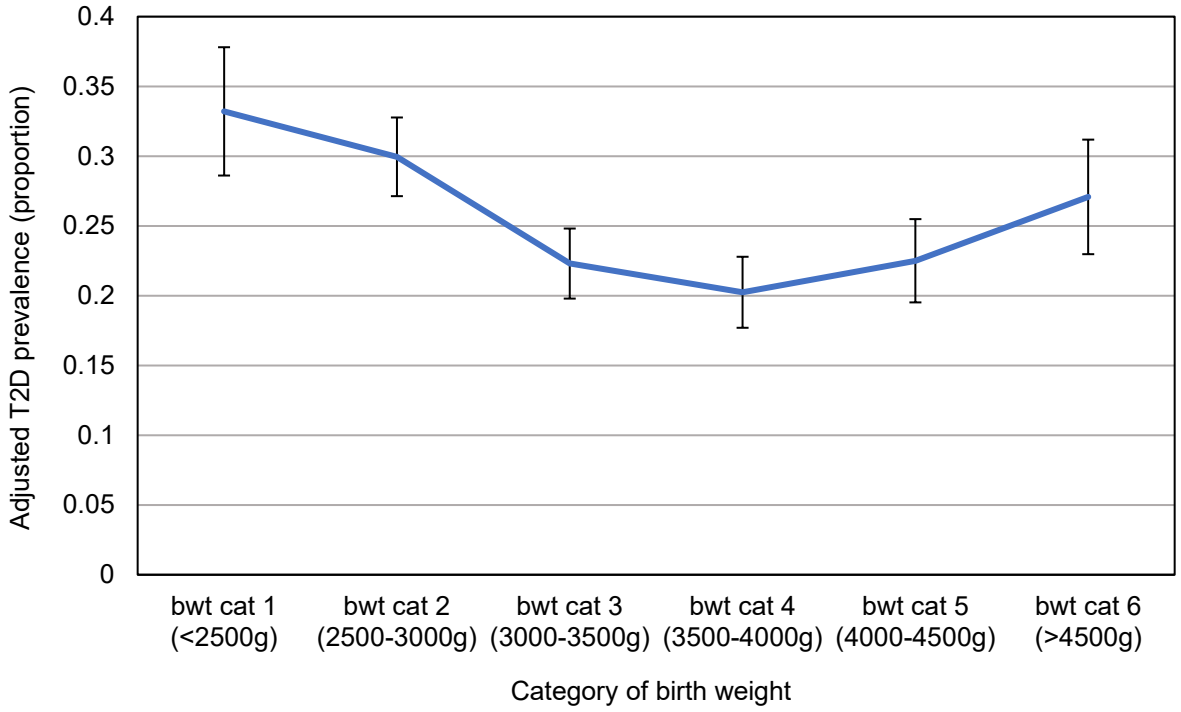


Figure 3.1. Adjusted prevalence of T2D by quintile of foetal birth weight.
The error bars displayed represent ± 1 SE (i.e., 68% CI).

Across the six categories of birth weight, a U-shaped curve was formed, with respect to the adjusted prevalence of T2D.

3.3.2. Heritability and bivariate genetic correlation and heritability analyses

The heritability estimates for Models 1, 2, and 3—the additive foetal GWAS of own birth weight adjusted for different covariates—were similar. First, for the model including the

covariates birth year and the first 5 genetic PCs, the h^2 estimate was 0.54 (SE = 0.03). Second, for the model including the covariates birth year, the first five genetic PCs, and *in utero* exposure to maternal diabetes, the h^2 was 0.53 (SE = 0.02). Third, for the model including the covariates birth year, the first five genetic PCs, and gestational age (by multiple imputation), the h^2 was 0.53 (SE = 0.03). Previous studies have reported that foetal genetic factors explain 10-40% of the normal birth weight (Lunde et al., 2007; Van Baal & Boomsma, 1998), substantially less than the h^2 estimates for Models 1, 2, and 3 in the PECRB cohort.

Bivariate genetic correlation analyses were undertaken to further investigate the relationships between birth weight and diabetes-related traits. For each analysis, Trait 1 (not displayed in the table) is birth weight (N = 3,700) and Trait 2 is named in the leftmost column. Results show that genetic correlation (ρ_G) statistics are relatively low across traits (Table 3.6).

Table 3.6. Bivariate genetic correlations between birth weight and related traits in the Indigenous study population. Trait 1 is birth weight (adjusted for birth year and PC 1-5).

Trait 2	Trait 2 N	Trait 2 adjusted for these covariates	Trait 2 h^2 (SE)	ρ_P	ρ_E (SE)	ρ_G (SE)	$p(\rho_G$ is nonzero)
T2D	7,659	sex, age, birth year, PC 1-5	0.27 (0.02)	-0.05	-0.13 (0.033)	0.079 (0.058)	0.18
log(true AIR)	404	sex, age, PFAT, log(M-value), PC 1-5	0.51 (0.15)	0.09	0.10 (0.16)	0.07 (0.15)	0.63
log(max BMI)	6,789	sex, age, birth year, PC 1-5	0.51 (0.02)	0.10	0.02 (0.04)	0.18 (0.05)	3.55×10^{-4}
log(M-value)	557	sex, age, PFAT, $PFAT^2$, PC 1-5	0.11 (0.11)	0.01	0.02 (0.09)	-0.02 (0.32)	1
HOMA-β	5,342	sex, age, BMI, PC 1-5	0.25 (0.03)	-0.06	-0.08 (0.04)	-0.05 (0.07)	0.49
HOMA-IR	5,349	sex, age, BMI, PC 1-5	0.25 (0.03)	-0.09	-0.11 (0.04)	-0.06 (0.07)	0.39
Fasting insulin	5,349	sex, age, BMI, PC 1-5	0.25 (0.03)	-0.08	-0.11 (0.04)	-0.06 (0.07)	0.43
GTT (fasting)	6,637	sex, age, PC 1-5	0.22 (0.03)	0.003	-0.05 (0.04)	0.10 (0.07)	0.16
GTT (2-hr)	6,637	sex, age, PC 1-5	0.17 (0.02)	-0.05	-0.05 (0.04)	-0.07 (0.08)	0.40
Height (cm)	5,719	sex, PC 1-5	0.72 (0.02)	0.24	0.09 (0.06)	0.34 (0.04)	1.92×10^{-14}

Height, a highly heritable trait with narrow-sense heritability h^2 = 0.72, had the strongest positive phenotypic correlation (ρ_P = 0.24), a strong positive genetic correlation with birth weight (ρ_G = 0.34); the *p*-value for the test of whether the genetic correlation ρ_G is nonzero was significant ($p = 1.92 \times 10^{-14}$). This was followed by the positive ρ_G of BMI with respect to birth weight (ρ_G = 0.18, SE = 0.05); the *p*-value for the test of whether ρ_G between birth weight and BMI is nonzero was significant (3.55×10^{-4}). Notably, T2D had a modest inverse phenotypic correlation (ρ_P = -0.05) and an inverse environmental correlation (ρ_E = -0.13, SE = 0.033); the genetic correlation was positive (ρ_G = 0.08) but the *p*-value for the test of whether the genetic correlation ρ_G is nonzero was not significant (*p* = 0.18).

3.3.3. Discussing significance in terms of p-value

Section 3.3.3 discusses the definition of significance thresholds that are employed in the following results sections. Chapter 3 in large part focusses upon the findings of various models of GWAS of foetal or offspring birth weight. As described in sections 3.3.4 and 3.3.5, none of the SNPs that were identified in any GWAS in this chapter reached the conventionally defined genome-wide level of significance ($p = 5.0 \times 10^{-8}$). This could partially be explained by the relatively low sample size of the PECRB cohort (maximum GWAS of birth weight N = 3,700), as compared with a maximum of over 200,000 for most SNPs in the Early Growth Genetics Consortium GWAS meta-analysis (Warrington et al., 2019). The lack of genome-wide significant loci identified in these GWAS could also be due to the observation that was previously made by Warrington et al that maternal and foetal genetic effects with respect to birth weight can work in opposite directions (Warrington et al., 2019). This could render a GWAS of own birth weight only, without accounting for maternal and foetal genotype, insufficiently powered to detect effects in which maternal and foetal directionality differs. Nonetheless, given the polygenic nature of complex traits, some of the top signals

that were observed in the current analyses may represent 'real' associations, despite the lack of genome-wide statistical significance.

The conventionally employed 'genome-wide significant' threshold is $p \leq 5.0 \times 10^{-8}$ (Dudbridge & Gusnanto, 2008): this originated from a study that estimated the effect of the N of independent tests (i.e., association tests in a GWAS across N variants) using Bonferroni correction (Bonferroni, 1935). Other studies specify more stringent levels of significance; for example, in reporting results from the EGG 2019 GWAS meta-analysis of birth weight, Warrington et al used $p < 6.6 \times 10^{-9}$ as the more stringent threshold of genome-wide significance—as calculated previously by Kemp et al in a GWAS of heel bone mineral density in a different, large study population (Kemp et al., 2017)—in addition to the more lenient and traditional $p < 5.0 \times 10^{-8}$.

Within Chapter 3, as none of the SNPs in any GWAS were found to have genome-wide significant associations with own or offspring birth weight, I took steps to define a level of 'suggestive significance' for GWAS of birth weight based on data from both the EGG 2019 GWAS (Warrington et al, 2019) that was adjusted for sex, PCs 1-4, and gestational age; and the PECRB GWAS of own birth weight that was adjusted for birth year, PCs 1-5, and gestational age (by multiple imputation; Model 3). Specifically, for various *p*-value thresholds for SNPs in the PECRB GWAS of birth weight Model 3, I identified the number of SNPs whose associations with birth weight reached that threshold. Then, among those SNPs that reached that level of significance in the PECRB GWAS of own birth weight, I determined the number of those SNPs that reached nominal significance (directionally consistent effect with 1-sided $p < 0.05$) in the EGG 2019 GWAS of own birth weight. I calculated the proportion of SNPs from the PECRB GWAS that were replicated (according to direction of effect and given significance thresholds) in the EGG 2019 GWAS under that definition, along with the 95% confidence interval of that proportion and the expected probability (two-sided *p*-value) that

these replications were significantly different from the expected proportion of 0.05. Results for this analysis are displayed in Table 3.7.

Table 3.7. Testing p-value thresholds for the PECRB and EGG birth weight GWAS.

PECRB *p*	N total PECRB	N replicated (PECRB & EGG)	Proportion replicated between PECRB and EGG (95% CI)	2-sided Pr(Z<|Z|)
0.00005	44	4	0.0909 (0.0253, 0.2167)	0.2131
0.0001	156	15	0.0962 (0.0548, 0.1536)	0.0082
0.0005	2,103	366	0.1740 (0.1581, 0.1909)	< 0.0001
0.001	4,382	551	0.1257 (0.1161, 0.1359)	< 0.0001
0.01	45,728	3,426	0.0749 (0.0725, 0.0774)	< 0.0001
0.05	220,522	18,214	0.0826 (0.0814, 0.0838)	< 0.0001
0.5	2,163,860	169,316	0.0782 (0.0779, 0.0786)	< 0.0001
0.99	4,258,599	324,021	0.0761 (0.0758, 0.0763)	< 0.0001

The EGG 2019 GWAS exhibits some inflation, which explains why even the PECRB $p < 0.99$ threshold itself exceeds 0.05. Because the proportion of replication (i.e., the sharing of variants between PECRB and EGG 2019 GWAS) reaches a local maximum (17.4%, 95% CI 15.8%-19.1%) at a PECRB *p*-value threshold of 0.0005 (i.e., 5.0×10^{-4}), $p < 5.0 \times 10^{-4}$ was taken to be a level of 'suggestive significance' level of significance for the reporting of GWAS results in the present chapter.

3.3.4. GWAS Models 1, 2, and 3: Additive foetal GWAS of own birth weight across imputed variants

The heterogeneity between the PECRB GWAS of birth weight Model 1 and the EGG 2019 GWAS was calculated, as measured by the mean heterogeneity index I^2 (Higgins & Thompson, 2002). The heterogeneity index describes the percentage of variation across studies that is due to heterogeneity as opposed to chance (Higgins & Thompson, 2002). The mean I^2 across all GWAS markers was 15.19%, which indicates that the percentage of

variability in effect estimates that is due to heterogeneity, rather than chance, is low (Higgins & Thompson, 2002). For the additive foetal GWAS of own birth weight with adjustment for birth year and PCs 1-5 (Model 1), 2,476 SNPs were found to have associations of suggestive significance ($p < 5.0 \times 10^{-4}$) with birth weight. The associations with the greatest significance in GWAS Model 1 per 200kb region are listed by ascending *p*-value in Table 3.8a. Imputation quality (Qinfo) scores are provided, which are of relevance to the imputed variants (i.e., GT = Imp) (Table 3.7). The strongest associations identified in GWAS Model 1 were with rs61515650 on chromosome 19 ($\beta = 0.16$ (SE = 0.03) SD birth weight per T allele; $p = 1.2 \times 10^{-6}$) and rs181228071 in *DMD* on chromosome X ($\beta = 0.23$ (SE = 0.05) SD birth weight per A allele; $p = 3.4 \times 10^{-6}$).

In GWAS Model 2, 2,577 variants were suggestively associated ($p < 5.0 \times 10^{-4}$) with birth weight. Relative to the results of GWAS Model 1, those of GWAS Model 2 (additionally adjusted for *in utero* exposure to maternal diabetes) shared multiple 'top' associations with Model 1, with modest attenuations in *p*-values for some of the SNPs associations of the greatest significance that were shared between the models (Tables 3.8a and b). For instance, a variant in *SUR1*/*ABCC8* (rs1272388614; chr11:17417205) was previously shown to be associated with T2D and birth weight in a subset of the PECRB cohort (Baier et al., 2015). In GWAS Models 1 and 2, the association of *SUR1/ABCC8* R1420H with birth weight was strong and reached suggestive significance ($p < 5.0 \times 10^{-4}$).

Model 3 of the birth weight GWAS consisted of an additive foetal GWAS of own birth weight with adjustment for birth year, PCs 1-5, and imputed gestational age (by multiple imputation). The results of Model 3 included 1,895 SNPs that reached the suggestive level of significance ($p < 5.0 \times 10^{-4}$). Notably, results for one SNP (*LOC105378114* rs73030292, chr6:166190683; $p = 4.83 \times 10^{-5}$) replicated a previously reported association at genome-wide significance ($p = 1.15 \times 10^{-10}$; Early Growth Genetics Consortium); however, associations of this, and all other

SNPs in the GWAS, did not reach genome-wide significance (p = 5.0 × 10^{-8}). A notable difference from GWAS Models 1 and 2 was observed in Model 3: missense variant *SUR1/ABCC8* R1420H (rs1272388614, chr11:17417205) had the strongest association with birth weight (β = 0.46 SD birth weight per H allele; p = 1.14 × 10^{-6}) (Table 3.8c).

The results displayed in Tables 3.8 appear similar to those for previous models of GWAS. One notable difference is that the *SUR1/ABCC8* R1420H variant in Model 3 is by far the highest on the y-axis in this model of GWAS that accounts for gestational age (by multiple imputation), representing the decrease from 1.5 × 10^{-5} in Model 1 to 1.2 × 10^{-6} in Model 3 (i.e., after additional adjustment for gestational age by multiple imputation).

Tables 3.8. SNPs with associations of greatest significance with own birth weight in Models 1, 2, and 3 (per 200kb region): additive foetal GWAS of own birth weight across imputed variants.
(a) GWAS Model 1: adjusted for birth year and PCs 1-5.
(b) GWAS Model 2: adjusted for birth year, PCs 1-5, and *in utero* exposure to diabetes.
(c) GWAS Model 3: adjusted for birth year, PCs 1-5, and gestational age (by multiple imputation).

(a)

SNP	Gene	Chr	Position†	GT‡	A1	A2	A1F	β (SE)	*p*-value	Qinfo
rs61515650	-	19	1729004	Imp	T	C	0.81	0.16 (0.03)	1.2×10^{-6}	0.98
rs181228071	*DMD*	X	32004006	Imp	A	G	0.95	0.23 (0.05)	3.4×10^{-6}	0.81
rs4799141	*PARD6G*	18	77945325	DG	T	C	0.18	-0.15 (0.03)	7.5×10^{-6}	1.00
rs17333221	-	7	15648013	DG	C	G	0.17	-0.15 (0.03)	1.0×10^{-5}	1.00
rs34655947	*TRIOBP*	22	38151100	Imp	TG	T	0.84	-0.15 (0.03)	1.0×10^{-5}	1.00
rs1272388614	*SUR1/ ABCC8*	11	17417205	DG	T	C	0.02	0.42 (0.10)	1.5×10^{-5}	1.00
rs2180184	*UBE4B*	1	10235946	DG	T	C	0.68	0.12 (0.03)	1.5×10^{-5}	1.00
rs6986387	-	8	21311558	Imp	T	C	0.91	-0.20 (0.05)	1.9×10^{-5}	0.97
rs145704214	*DNAJC15*	13	43637780	DG	A	G	0.11	0.17 (0.04)	2.0×10^{-5}	1.00
rs561846108	*DENND2C*	1	115148107	DG	T	C	0.94	-0.24 (0.06)	2.0×10^{-5}	0.94

(b)

SNP	Gene	Chr	Position†	GT‡	A1	A2	A1F	β (SE)	*p*-value	Qinfo
rs61515650	-	19	1729004	Imp	T	C	0.81	0.16 (0.03)	1.2×10^{-6}	0.98
rs17333221	-	7	15648013	DG	C	G	0.17	-0.15 (0.03)	1.0×10^{-6}	1.00
rs4799141	*PARD6G*	18	77945325	DG	T	C	0.18	-0.15 (0.03)	7.5×10^{-6}	1.00
rs561846108	*DENND2C*	1	115148107	DG	T	C	0.94	-0.25 (0.06)	1.0×10^{-5}	0.94
rs3036655	*LINC00298*	2	8100372	DG	-	CCA TCC	0.36	0.12 (0.03)	1.4×10^{-5}	1.00
rs34655947	*TRIOBP*	22	38151100	Imp	TG	T	0.84	-0.15 (0.03)	1.5×10^{-5}	1.00
rs2180184	*UBE4B*	1	10235946	DG	T	C	0.68	0.12 (0.03)	1.5×10^{-5}	1.00
rs6986387	-	8	21311558	Imp	T	C	0.91	-0.20 (0.05)	1.8×10^{-5}	0.97
rs11776266	*FBXO25*	8	425541	Imp	A	G	0.87	0.16 (0.04)	2.0×10^{-5}	0.98
rs143740821	*KAT6A*	8	41859534	DG	-	TA	0.24	0.12 (0.03)	2.3×10^{-5}	1.00

(c)

SNP	Gene	Chr	Position†	GT‡	A1	A2	A1F	β (SE)	*p*-value	Qinfo
rs1272388614	*SUR1/ ABCC8*	11	17417205	DG	T	C	0.02	0.46 (0.09)	1.1×10^{-6}	1.00
rs61515650	-	19	1729004	Imp	T	C	0.81	0.15 (0.03)	3.5×10^{-6}	0.98
rs35219516	-	13	40843260	Imp	C	T	0.90	-0.18 (0.04)	6.2×10^{-6}	0.99
rs2180184	*UBE4B*	1	10235946	DG	T	C	0.68	0.12 (0.03)	6.2×10^{-6}	1.00
rs5928238	*DMD*	X	33350078	Imp	T	A	0.71	-0.10 (0.02)	9.8×10^{-6}	0.99
rs17333221	-	7	15648013	DG	C	G	0.17	-0.14 (0.03)	1.5×10^{-5}	1.00
rs181228071	*DMD*	X	32004006	Imp	A	G	0.95	0.21 (0.05)	1.7×10^{-5}	0.81
rs7668488	-	4	157604720	Imp	C	T	0.51	0.10 (0.02)	2.2×10^{-5}	0.98
rs6986387	-	8	21311558	Imp	T	C	0.91	-0.19 (0.04)	4.5×10^{-5}	0.97
rs573146818	-	1	152449209	Imp	T	C	0.91	-0.19 (0.05)	2.3×10^{-5}	0.91

† Position is according to Genome Reference Consortium Human Build 37 (i.e., GRCh37 or hg19).
‡ GT signifies genotyping: directly genotyped (DG) or imputed (Imp) data available for each SNP.

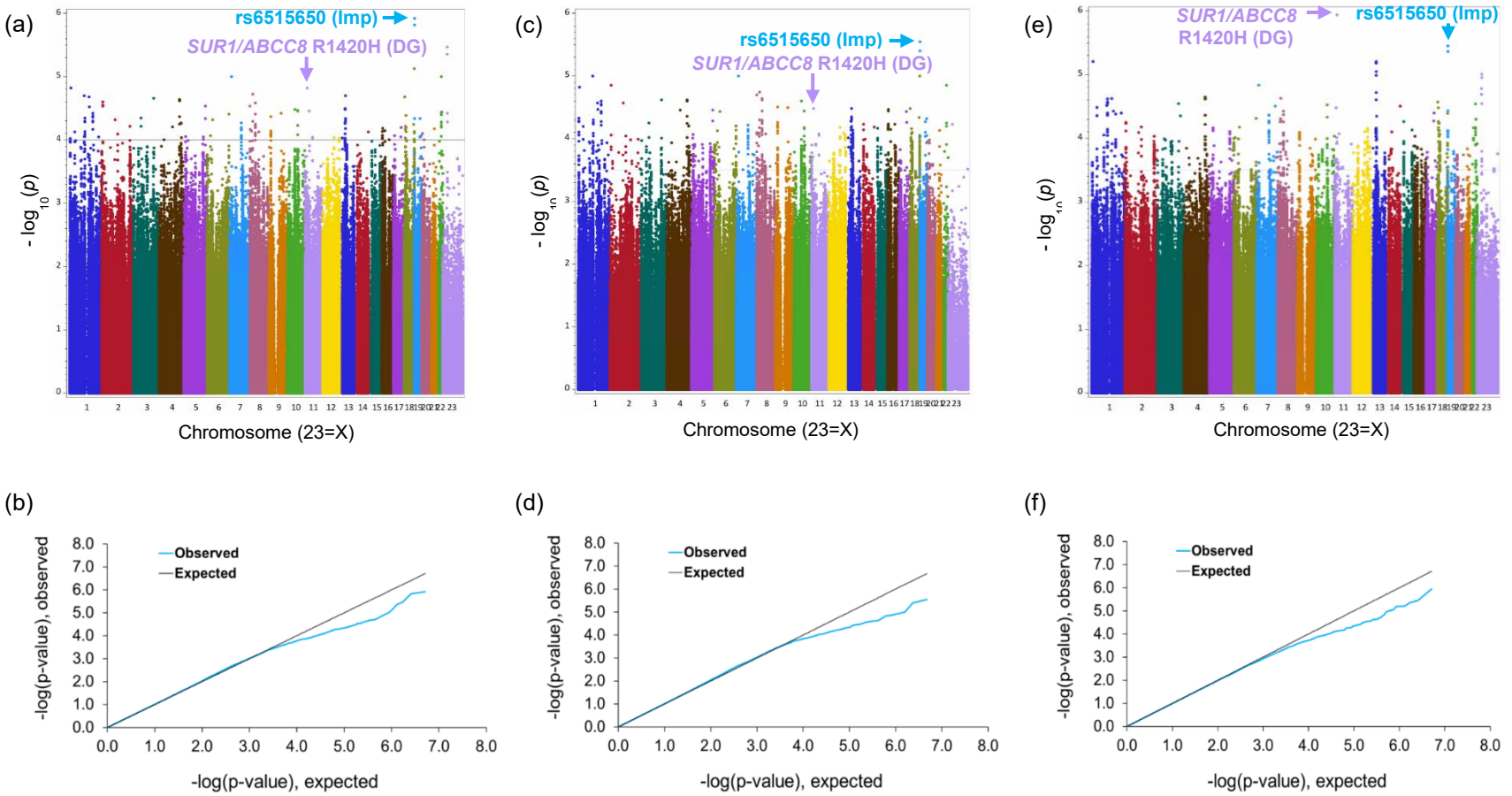


Figure 3.2. Manhattan and QQ plots for additive foetal GWAS of own birth weight: Model 1 (adjusted for birth year and PCs 1-5); Model 2 (adjusted for birth year, PCs 1-5, and *in utero* exposure to maternal diabetes); and Model 3 (adjusted for birth year, PCs 1-5, and gestational age by imputation). 'Imp' signifies a SNP that is imputed; 'DG' signifies directly genotyped.
Model 1 plots: (a) Manhattan plot, (b) QQ plot.
Model 2 plots: (c) Manhattan plot, (d) QQ plot.
Model 3 plots: (e) Manhattan plot, (f) QQ plot.

The Manhattan plots in Figure 3.2a and 3.2c display the results of GWAS Models 1 and 2, respectively. In Model 2, which was additionally adjusted for *in utero* exposure to maternal diabetes, p-values for some of the SNPs associated with own birth weight with greatest significance (Tables 3.6 and 3.7) were attenuated; however, the tables and Manhattan plots appear broadly similar. The quantile-quantile (QQ) plots for Models 1, 2, and 3 are displayed in Figure 3.2 (parts b, d, and f). The plots display a graph of the inverse logs of expected versus observed p-values in the GWAS of birth weight in each respective GWAS. Over the bulk of the distribution for all three models of GWAS, the plot of the observed inverse logged p-values matches that of the expectation. However, as the expected p-value decreases in magnitude, the observed p-value decreases more quickly, below the x=y line, lower than expected at the greater levels of significance. The QQ plot for Model 3 shows that the plot of the observed p-values dips slightly below $x=y$ (expectation) at lower inverse logged p-values than shown in the QQ plots for Models 1 and 2. Thus, the QQ plots for Models 1, 2, and 3 provide no evidence for significant associations with birth weight and largely conform to the null expectation.

3.3.5. GWAS Model 4: Dominant-recessive foetal GWAS of own birth weight, adjusted for imputed gestational age

GWAS Model 4 was conducted across directly genotyped variants only and was a dominant-recessive foetal GWAS of own birth weight that was adjusted for birth year, PCs 1-5, and gestational age by multiple imputation. The results of GWAS Model 4 are compared with those of GWAS Model 3, which was conducted across imputed variants and was an additive foetal GWAS of own birth weight that was adjusted for birth year, PCs 1-5, and gestational age with multiple imputation. The naïve x^2 across all variants in Model 4 was 1.41: this indicated substantial inflation of test statistics due to the multiple statistical tests that were

conducted for each variant (i.e., additive, dominant, and recessive models). To address this inflation, Sidak-corrected *p*-values were calculated, and Sidak-corrected *p*-values are presented in Table 3.9 (González et al., 2008). Within GWAS Model 4, 216 SNPs met the threshold of 'suggestive significance' discussed in Section 3.3.3 (Sidak-corrected $p = 5.0 \times 10^{-4}$). After Sidak correction, the mean x^2 of 1.07 across all SNPs in Model 4 did not indicate substantial inflation. Among directly genotyped SNPs in Model 3, 201 SNPs met the threshold of 'suggestive significance.'

Tables 3.9. SNPs with associations of greatest significance in GWAS Models 4 and 3 (per 200kb region).
(a) GWAS Model 4: dominant-recessive foetal GWAS of own birth weight adjusted for birth year, PCs 1-5, and gestational age (with multiple imputation), with Sidak correction.
(b) GWAS Model 3: additive foetal GWAS of own birth weight adjusted for birth year, PCs 1-5, and gestational age (with multiple imputation).

(a)

SNP	Gene	Chr	Position†	A1	A2	A1F	β (SE)	*p*-value
rs1272388614[d]	*SUR1/ABCC8*	11	17417205	T	C	0.02	0.47 (0.10)	2.2×10^{-6}
rs17527619[r]	-	10	72355806	A	G	0.02	0.40 (0.08)	2.9×10^{-6}
rs34287895[r]	*ADCY9*	16	4011008	-	G	0.02	0.37 (0.08)	7.0×10^{-6}
rs12754943[r]	-	1	226228627	T	C	0.12	-0.18 (0.04)	1.3×10^{-5}
rs2180184[a]	*UBE4B*	1	10235946	T	C	0.68	0.12 (0.03)	1.4×10^{-5}
rs826463[r]	-	10	72405504	A	G	0.04	0.28 (0.06)	1.7×10^{-5}
rs1335035[r]	*GRIK2*	6	102408680	T	C	0.06	-0.23 (0.05)	1.7×10^{-5}
rs2209815[r]	*PALM2AKAP2*	9	112561012	T	G	0.04	0.28 (0.06)	1.9×10^{-5}
rs4943748[a]	-	13	40838524	T	G	0.10	0.18 (0.04)	2.3×10^{-5}
rs10630839[r]	-	4	157611192	-	TG	0.13	0.17 (0.04)	2.3×10^{-5}

(b)

SNP	Gene	Chr	Position†	A1	A2	A1F	β (SE)	*p*-value
rs1272388614	*SUR1/ABCC8*	11	17417205	T	C	0.02	0.46 (0.09)	1.1×10^{-6}
rs2180184	*UBE4B*	1	10235946	T	C	0.68	0.12 (0.03)	6.2×10^{-6}
rs4943748	-	13	40838524	T	G	0.10	0.18 (0.04)	1.0×10^{-5}
rs17333221	-	7	15648013	C	G	0.17	-0.14 (0.03)	1.5×10^{-5}
rs58029094	*DMD*	X	33352563	A	G	0.31	0.09 (0.02)	2.8×10^{-5}
rs11704996	*TRIOBP*	22	38151170	A	G	0.16	0.14 (0.03)	2.9×10^{-5}
chr1:115148107	*DENND2C*	1	115148107	T	C	0.94	-0.23 (0.05)	3.0×10^{-5}
rs1005081	*ATP6V0A4*	7	138399376	T	C	0.73	-0.11 (0.03)	3.1×10^{-5}
rs230288	-	1	40260447	A	G	0.34	-0.11 (0.03)	3.2×10^{-5}
chr11:11373535	*CSNK2A3*	11	11373535	T	G	0.03	0.32 (0.08)	3.3×10^{-5}

† Position is according to Genome Reference Consortium Human Build 37 (i.e., GRCh37 or hg19).
[a] In Table 3.9.a, this symbol denotes that a given SNP's lowest *p*-value (Sidak-corrected) occurred in the additive GWAS. Thus, summary statistics displayed for this SNP originate from the additive model.
[d] In Table 3.9.a, this symbol denotes that a given SNP's lowest *p*-value (Sidak-corrected) occurred in the dominant GWAS. Thus, summary statistics displayed for this SNP originate from the dominant model.
[r] In Table 3.9.a, this symbol denotes that a given SNP's lowest *p*-value (Sidak-corrected) occurred in the recessive GWAS. Thus, summary statistics displayed for this SNP originate from the recessive model.

The top SNP in both GWAS Models 3 and 4 within *ABCC8*, at rs1272388614 (Model 4 β = 0.47 SD birth weight per copy of the T allele, SE = 0.10, dominant model $p = 2.2 \times 10^{-6}$; Model 3 β = 0.46 SD birth weight per copy of the T allele, SE = 0.09, $p = 1.1 \times 10^{-6}$) (Tables

3.9). Another SNP that was in both parts of Tables 3.9 was *UBE4B* rs2180184 on chromosome 1 (Model 4 β = 0.12 SD birth weight per copy of the T allele, SE = 0.03, additive model $p = 1.3 \times 10^{-5}$; Model 3 β = 0.12 SD birth weight per copy of the T allele, SE = 0.03, $p = 6.2 \times 10^{-6}$).

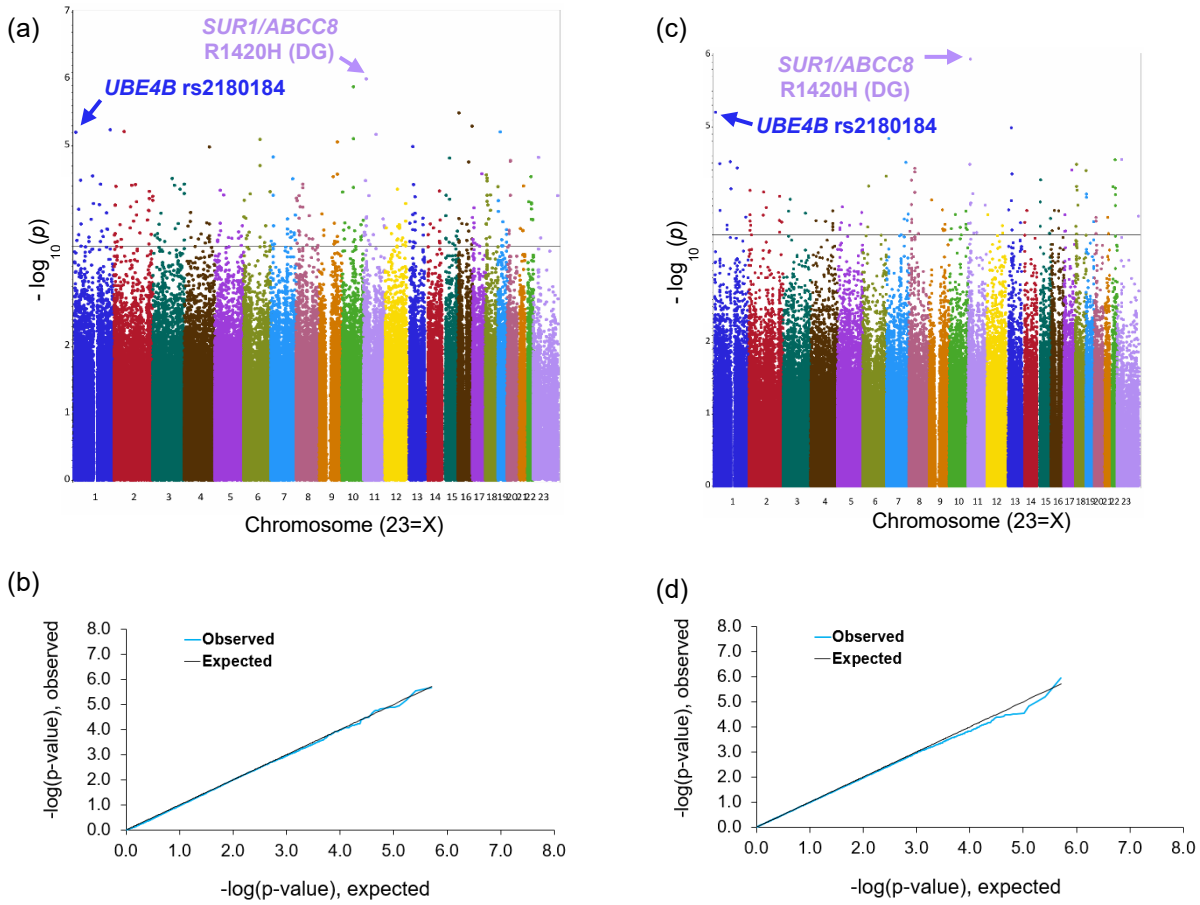


Figure 3.3. Manhattan and QQ plots for GWAS of birth weight Model 4 and Model 3. Model 4 was a dominant-recessive foetal GWAS of own birth weight adjusted for birth year, PCs 1-5, and gestational age (by multiple imputation). Model 3 was an additive foetal GWAS of own birth weight adjusted for birth year, PCs 1-5, and gestational age (by multiple imputation).
Model 4 plots: (a) Manhattan plot, (b) QQ plot.
Model 3 plots: (c) Manhattan plot (scaled to line up y-axis with that of Model 4), (d) QQ plot.

Notably, the QQ plot of the complete-case analysis Model 5 (Figure 3.3.b) adhered closely to the null. The QQ plot of Model 3 (Figure 3.3.d) showed no evidence of deviation from the null

over the bulk of the distribution; however, at higher expected inverse-logged *p*-values, the observed inverse-logged *p*-value curve dips below *x*=*y*.

3.3.6. GWAS Model 5: Additive GWAS of birth weight, adjusted for gestational age without multiple imputation

GWAS Model 5 was an additive foetal GWAS across directly genotyped variants of own birth weight that was adjusted for birth year, PCs 1-5, and gestational age without multiple imputation (Table 3.10.a). Thus, only the 2,749 individuals with gestational age data of the 3,700 in the birth weight data set were included in Model 5. Results of GWAS Model 5 are compared with those of directly genotyped variants only from GWAS Model 3, an additive foetal GWAS of own birth weight, adjusted for birth year, PCs 1-5, and gestational age with multiple imputation. As shown in Figure 3.4, the correlation between the beta coefficients of Models 5 and 3 was strong and positive: the Pearson correlation coefficient, *r*, was 0.88 ($p < 10^{-4}$).

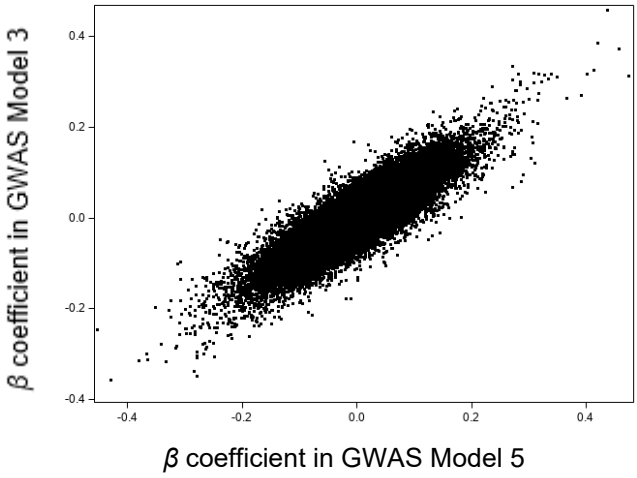


Figure 3.4. Beta-versus-beta plot for GWAS Model 5 (without multiple imputation of the gestational age covariate) and Model 3 (with multiple imputation of gestational age).

Model 5 had 272 SNPs reach a suggestive level of significance (p = 5.0 × 10^{-4}); among directly genotyped SNPs only within Model 3, 201 SNPs met this threshold. There was no evidence of test statistic inflation in Model 5 (mean x^2 = 1.00).

Tables 3.10. SNPs with associations of greatest significance in GWAS Models 5 and 3 (per 200kb region).
(a) GWAS Model 5: additive foetal GWAS of own birth weight adjusted for birth year, PCs 1-5, and gestational age (without multiple imputation).
(b) GWAS Model 3: additive foetal GWAS of own birth weight adjusted for birth year, PCs 1-5, and gestational age (with multiple imputation).

(a)

SNP	Gene	Chr	Position†	A1	A2	A1F	β (SE)	*p*-value
rs230288	-	1	40260447	A	G	0.33	-0.15 (0.03)	1.2×10^{-6}
rs72820019	-	5	172795441	A	G	0.76	0.15 (0.03)	4.2×10^{-6}
rs4799141	*PARDG6*	18	77945325	T	C	0.19	-0.16 (0.04)	6.2×10^{-6}
rs4807302	*GNG7*	19	2591597	A	G	0.39	-0.13 (0.03)	1.1×10^{-5}
rs11339383	-	13	54359682	-	T	0.14	-0.18 (0.04)	1.1×10^{-5}
rs2180184	*UBE4B*	1	10235946	T	C	0.68	0.13 (0.03)	1.1×10^{-5}
rs12953180	-	17	12962911	A	G	0.16	0.17 (0.04)	1.2×10^{-5}
rs73030292	*PDE10A, LOC105378114*	6	166190683	C	G	0.26	-0.14 (0.03)	1.7×10^{-5}
rs4943748	-	13	40838524	T	G	0.11	0.20 (0.05)	1.7×10^{-5}
rs4668743	-	2	12039694	T	C	0.07	0.24 (0.06)	1.9×10^{-5}

(b)

SNP	Gene	Chr	Position†	A1	A2	A1F	β (SE)	*p*-value
rs1272388614	*SUR1/ABCC8*	11	17417205	T	C	0.02	0.46 (0.09)	1.1×10^{-6}
rs2180184	*UBE4B*	1	10235946	T	C	0.68	0.12 (0.03)	6.2×10^{-6}
rs4943748	-	13	40838524	T	G	0.10	0.18 (0.04)	1.0×10^{-5}
rs17333221	-	7	15648013	C	G	0.17	-0.14 (0.03)	1.5×10^{-5}
rs58029094	*DMD*	X	33352563	A	G	0.31	0.09 (0.02)	2.8×10^{-5}
rs11704996	*TRIOBP*	22	38151170	A	G	0.16	0.14 (0.03)	2.9×10^{-5}
chr1:115148107	*DENND2C*	1	115148107	T	C	0.94	-0.23 (0.05)	3.0×10^{-5}
rs1005081	*ATP6V0A4*	7	138399376	T	C	0.73	-0.11 (0.03)	3.1×10^{-5}
rs230288	-	1	40260447	A	G	0.34	-0.11 (0.03)	3.2×10^{-5}
chr11:11373535	*CSNK2A3*	11	11373535	T	G	0.03	0.32 (0.08)	3.3×10^{-5}

† Position is according to Genome Reference Consortium Human Build 37 (i.e., GRCh37 or hg19).

UBE4B rs2180184 on chromosome 1 (Model 5 β = 0.13 per copy of the T allele, SE = 0.03, p = 1.1 × 10^{-5}; Model 3 β = 0.12 per copy of the T allele, SE = 0.03, p = 6.2 × 10^{-6}) had relatively high significance in both analyses.

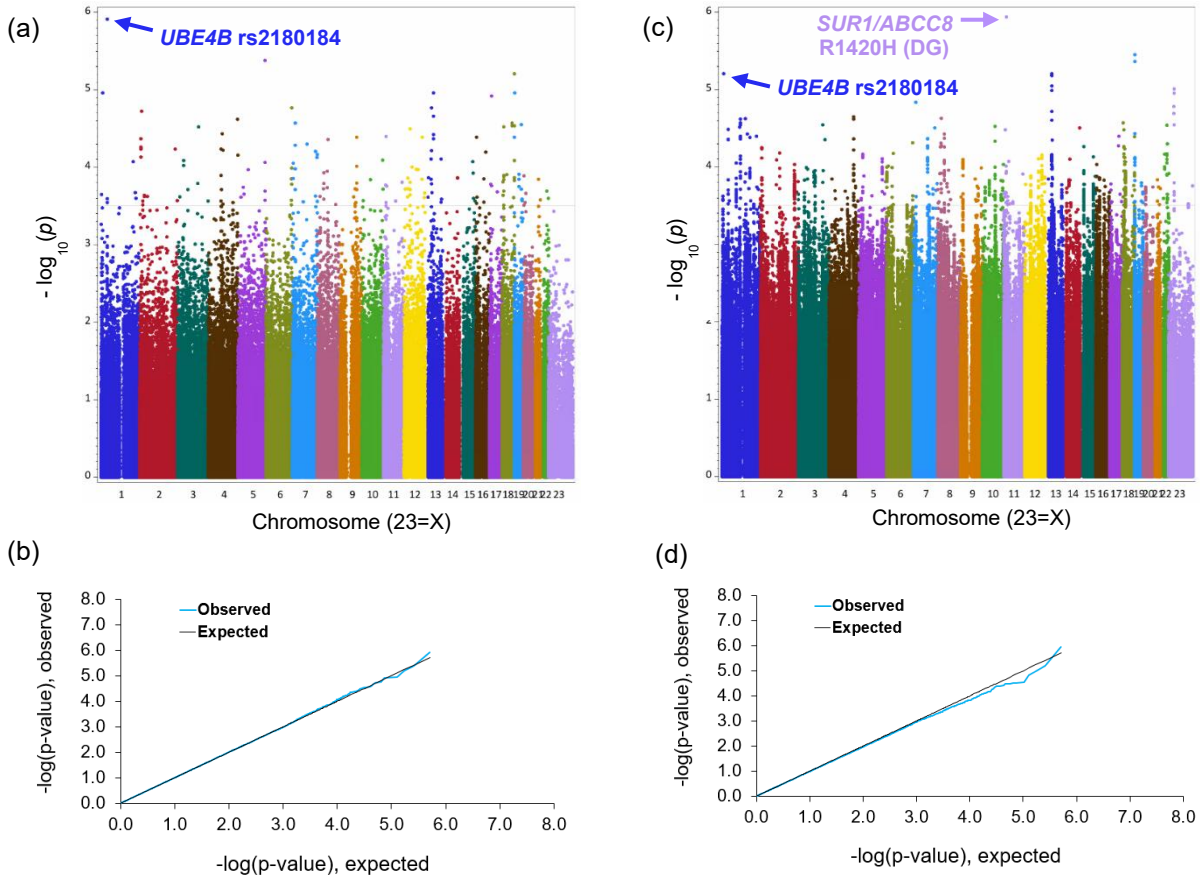


Figure 3.5. Manhattan and QQ plots for GWAS of birth weight Model 5 and Model 3. Model 5 was an additive foetal GWAS of own birth weight adjusted for own birth year, PCs 1-5, and gestational age (without multiple imputation). Model 3 was an additive foetal GWAS of own birth weight adjusted for birth year, PCs 1-5, and gestational age (by multiple imputation).
Model 5 plots: (a) Manhattan plot, (b) QQ plot.
Model 3 plots: (c) Manhattan plot, (d) QQ plot.

Notably, the QQ plot for the complete-case analysis Model 5 (Figure 3.5.b) appeared similar to that of the directly genotyped variants in Model 3 (Figure 3.5.d). In both models' QQ plots, there was no evidence for deviation from the null across the bulk of the distribution, and a slight increase of the observed inverse-logged *p*-values above the expected inverse-logged *p*-values. However, the QQ plot for Model 3 displayed a decline in the observed inverse-logged *p*-values below the expected inverse-logged *p*-values at slightly lower values on the x-axis.

3.3.7. GWAS Models 6 and 7: Additive foetal GWAS of own birth weight conditioned on maternal genotype and additive maternal GWAS of offspring birth weight conditioned on foetal genotype

GWAS Models 6 and 7 were conducted across a lower sample size of 2,934 individuals who had both their own and their maternal genotypic data within the original birth weight data set (N = 3,700; described further in Section 3.2.4), across directly genotyped variants. GWAS Model 6 was an additive foetal GWAS of own birth weight, adjusted for own birth year and own PCs 1-5, and conditioned on maternal genotype. GWAS Model 7 was an additive maternal GWAS of offspring birth weight, adjusted for offspring birth year and offspring PCs 1-5, and conditioned on foetal genotype. Because substantial inflation, as measured by mean x^2 across all included variants within a given GWAS, was found in the results of Model 7, the results for both Models 6 (mean x^2 = 1.02) and 7 (mean x^2 = 1.48) are presented with genomic control to facilitate comparison of results between models.

In GWAS Model 6, the additive foetal GWAS of own birth weight conditioned on maternal genotype, 268 SNPs reached the threshold of suggestive significance ($p = 5.0 \times 10^{-4}$). As shown in Table 3.11, *UBE4B* rs2180184 appeared as the top association with regard to *p*-value (β = 0.15 SD own birth weight per copy of the T allele, SE = 0.03; $p = 4.2 \times 10^{-6}$), and also appeared in multiple other models of birth weight GWAS. In GWAS Model 7, the additive maternal GWAS of offspring birth weight conditioned on foetal genotype, 205 SNPs reached suggestive significance ($p = 5.0 \times 10^{-4}$). The SNP that had the greatest significance of association with offspring birth weight was rs183665761 on chromosome 6 (β = -0.29 SD birth weight per copy of the T allele, SE = 0.05; $p = 2.9 \times 10^{-6}$).

Tables 3.11. SNPs with associations of greatest significance in GWAS Models 6 and 7 (per 200kb region).
(a) GWAS Model 6: additive foetal GWAS of own birth weight adjusted for own birth year and own PCs 1-5 and conditioned on maternal genotype.
(b) GWAS Model 7: additive maternal GWAS of offspring birth weight adjusted for offspring birth year, offspring PCs 1-5, and conditioned on foetal genotype.

(a)

SNP	Gene	Chr	Position†	A1	A2	A1F	β (SE)	*p*-value
rs2180184	*UBE4B*	1	10235946	T	C	0.67	0.15 (0.03)	4.2×10^{-6}
rs149769	-	20	57863079	C	G	0.26	0.15 (0.03)	5.3×10^{-6}
rs62541849	*GLIS3*	9	4297603	T	C	0.57	0.14 (0.03)	9.4×10^{-6}
rs819350	-	5	89102929	A	C	0.75	0.16 (0.03)	1.1×10^{-5}
rs145643465	*MYT1L*	2	2118282	T	C	0.90	-0.23 (0.05)	1.4×10^{-5}
rs6847964	*GALNTL6*	4	173071965	A	T	0.31	0.14 (0.03)	1.7×10^{-5}
rs140327985	-	4	15950790	T	G	0.18	-0.17 (0.04)	1.9×10^{-5}
rs949157190	*LOC101929996*	4	184470265	C	G	0.94	-0.32 (0.08)	2.5×10^{-5}
rs1007120892	*LOC105374902*	6	6898464	A	G	0.07	0.27 (0.06)	2.8×10^{-5}
rs7181142	*LOC105371024*	15	101311503	A	C	0.52	-0.13 (0.03)	2.9×10^{-5}

(b)

SNP	Gene	Chr	Position†	A1	A2	A1F	β (SE)	*p*-value
rs183665761	-	6	153892587	T	C	0.89	-0.29 (0.05)	2.9×10^{-6}
rs403706	*PKNOX1*	21	44401317	T	C	0.78	0.22 (0.04)	4.4×10^{-6}
rs10142227	-	14	42521704	A	G	0.82	-0.21 (0.04)	8.0×10^{-6}
chr4:184637033	-	4	184637033	T	G	0.81	-0.22 (0.04)	1.1×10^{-5}
rs34662285	*CFAP52*	17	9487085	A	G	0.50	0.17 (0.03)	1.1×10^{-5}
rs17140367	*HGD*	3	120391944	T	C	0.56	0.17 (0.03)	1.4×10^{-5}
rs9787528	-	10	95590639	T	G	0.24	-0.19 (0.04)	2.5×10^{-5}
rs71705180	*LOC105 369715*	12	30145352	-	TTCCAG GCTGT	0.14	-0.22 (0.04)	2.6×10^{-5}
rs59849180	-	4	25471830	A	G	0.15	0.24 (0.05)	3.0×10^{-5}
rs351360	*WNT2B*	1	113040421	T	C	0.48	0.16 (0.03)	3.1×10^{-5}

† Position is according to Genome Reference Consortium Human Build 37 (i.e., GRCh37 or hg19).

Manhattan plots for GWAS Models 6 and 7—which are shown after correction for inflation—are displayed in Figure 3.5 (parts a and c, respectively), below.

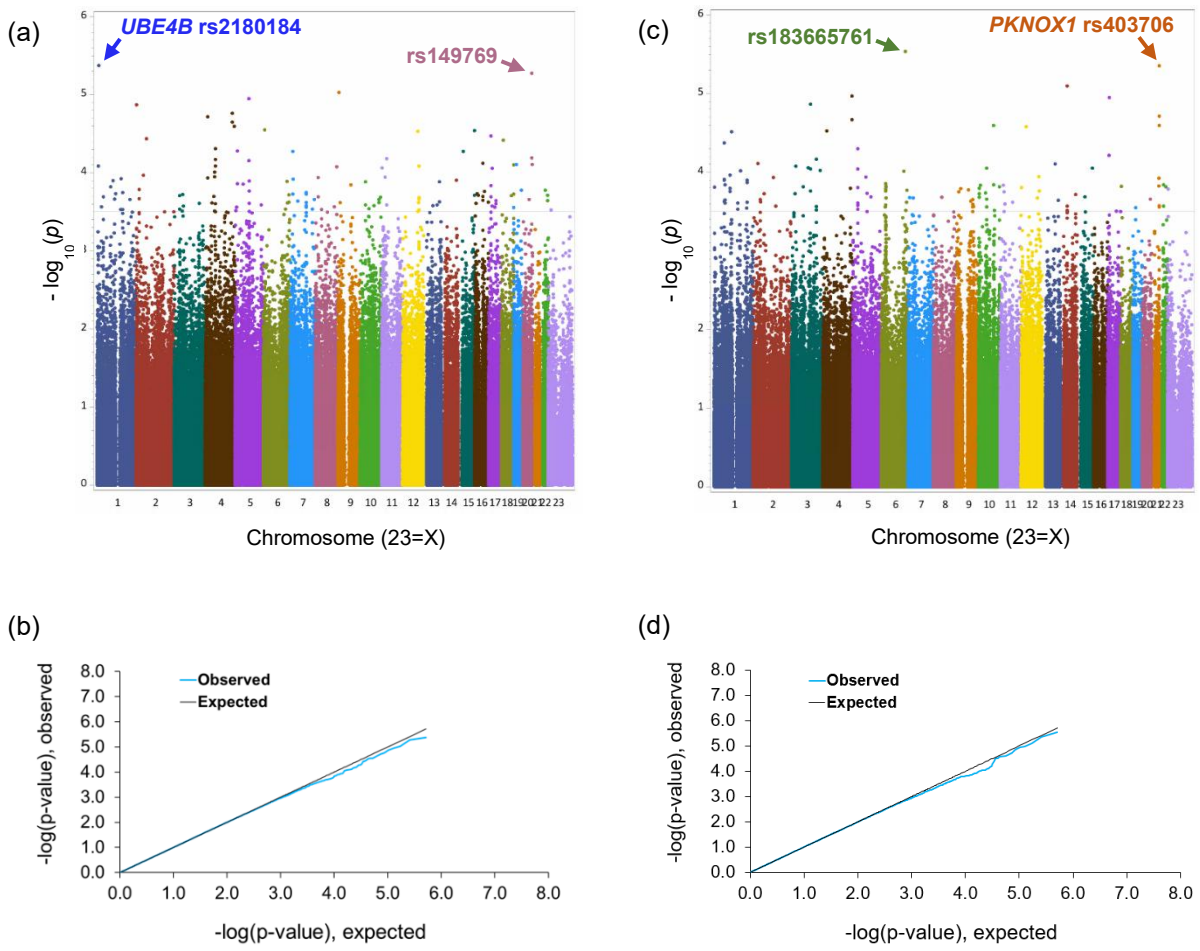


Figure 3.6. Manhattan and QQ plots for GWAS of birth weight Model 6 and Model 7. Model 6 was an additive foetal GWAS of own birth weight adjusted for own birth year and own PCs 1-5 and conditioned on maternal genotype. Model 7 was an additive maternal GWAS of offspring birth weight adjusted for offspring birth year, offspring PCs 1-5, and conditioned on foetal genotype. The images shown in the figure are after correction for genomic inflation.
Model 6 plots: (a) Manhattan plot, (b) QQ plot.
Model 7 plots: (c) Manhattan plot, (d) QQ plot.

The quantile-quantile (QQ) plots for Models 6 and 7 are displayed in Figure 3.6 (parts b and d) and appear largely similar. Over the bulk of the distribution for both models of GWAS, the plot of the observed inverse logged p-values (after the performance of genomic control) matches that of the expectation. Thus, the QQ plots for Models 6 and 7 largely conform to the null expectation, which indicates that there was no evidence of genetic effects on birth weight in these analyses.

3.4. Summary

3.4.1. GWAS of birth weight

Seven different models of GWAS of birth weight were conducted, using various combinations of covariates, imputed or directly genotyped variants, and maternal and foetal genotypes. None of the GWAS models yielded genome-wide significant ($p < 5.0 \times 10^{-8}$) associations. A substantial number of SNPs met the threshold of suggestive significance ($p < 5.0 \times 10^{-4}$) in each of the seven models. However, according to their QQ plots, only Model 4 and 5 displayed deviations from the null hypothesis. While none of the models of GWAS of birth weight identified SNPs that reached genome-wide significance, certain signals that have high levels of significance across multiple models could still merit further exploration.

Results were similar across Models 1, 2, and 3; notably, in Model 3, which was adjusted for gestational age (by multiple imputation), the *SUR1/ABCC8* R1420H variant (rs1272388614, chr11:17417205) had the greatest significance of association with birth weight among imputed variants. This variant had previously been identified by Baier et al as strongly associated with birth weight and diabetes in terms of effect size, but not reaching genome-wide significance in the PECRB cohort (Baier et al., 2015). Other variants that appeared across GWAS Models 1, 2, and 3, as well as multiple other models that were conducted across directly genotyped variants, were a variant in *UBE4B* (rs2180184, chr1:10235946), as well as one on chromosome 19 (rs61515650, chr19:1729004). Notably, the narrow-sense heritability (h^2) estimates across Models 1, 2, and 3, were nearly identical: there was no difference among the different adjustment strategies. While the top birth weight-associated SNPs appeared similar across the three models, there was not robust evidence that these were true associations with birth weight.

GWAS Model 4, the dominant-recessive foetal GWAS of own birth weight adjusted for birth year, PCs 1-5, and gestational age (with multiple imputation) was conducted in the same sample as Model 3 (N = 3,700). Because of the selection of one set of summary statistics for each SNP with the lowest p-value among additive (i.e., Model 3), dominant, and recessive models of this GWAS, the QQ plot for Model 4 displayed the observed *p*-value distribution adhering to that of the null for a majority of the distribution, and a slight increase in observed inverse-logged *p*-values above the expected. The mean x^2 was 1.07, indicating that this could partially be due to inflation conferred by the method of selecting the p-values of greatest significance across three separate GWAS for inclusion in the summary statistics of GWAS Model 4. GWAS Model 5, the additive foetal GWAS of own birth weight adjusted for birth year, PCs 1-5, and gestational age (without multiple imputation) was conducted in a smaller sample because approximately 70% of individuals had gestational age data available. It functioned as a complete-case analysis relative to GWAS Model 3, which included gestational age (by multiple imputation) as a covariate. In Model 4, the aforementioned variant in *SUR1/ABCC8* was also the SNP with the greatest significance of association with own birth weight, and the variant in *UBE4B* also appeared among the associations of greatest significance.

I conducted GWAS Models 6 and 7, a foetal GWAS of own birth weight and a maternal GWAS of offspring birth weight, respectively, because previous studies have shown that maternal and foetal genetic effects with respect to birth weight can act in opposite directions (Tyrrell et al., 2013; Warrington et al., 2019). GWAS Model 6 was an additive foetal GWAS of own birth weight adjusted for own birth year and own PCs 1-5 and conditioned on maternal genotype. GWAS Model 7 was an additive maternal GWAS of offspring birth weight adjusted for offspring birth year and offspring PCs 1-5 and conditioned on foetal genotype. Substantial genomic inflation was exhibited in GWAS Model 7. However, Warrington et al and Horikoshi

et al have previously identified signals in other study populations: this is likely due to the much smaller sample sizes of the PECRB cohort (N = 2,934 in both Models 6 and 7) as compared to those in the Early Growth Genetics Consortium GWAS meta-analyses. Future analyses could explore the inclusion of maternal genetic principal components as covariates and adding an additional variance component for a shared maternal effect to address this inflation. However, despite the potential increase in power conferred by partitioning foetal and maternal effects on birth weight, the present analyses did not find substantial evidence for associations, as statistical power to detect associations was likely attenuated due to the additional reduction in sample size.

3.4.2. Bivariate genetic correlation analyses

Across T2D and other glycaemic and anthropometric traits included in the bivariate genetic correlation analyses, environmental correlation estimates were relatively strong, as expected. On the whole, genetic correlations were rather modest, with some exceptions. Maximum BMI had the highest positive genetic correlation with respect to birth weight; this appears to be a significant correlation. Height had a relatively strong positive phenotypic correlation and genetic correlation with birth weight; this also appeared to be a significant correlation. Notably, T2D had an inverse environmental correlation with birth weight. The findings regarding T2D are difficult to parse and prompt further analyses of the genetic relationship between birth weight and T2D using other methods. Because the relationship between T2D and birth weight in the PECRB cohort was shown to be nonlinear (Section 3.3.1), a limitation of the bivariate genetic correlation analysis between birth weight and T2D is that it assumed a linear relationship. This limitation could be addressed by conducting a ‘trivariate’ genetic correlation analysis for birth weight and T2D that also included a quadratic term for birth weight, to account for the U-shaped relationship between birth weight and T2D. Additionally,

separate maternal and foetal genetic correlation analyses of birth weight and T2D could be conducted to address the fact that maternal and foetal genetic effects on birth weight may operate in opposite directions.

4. Associations of polygenic scores (PS) and partitioned polygenic scores (pPS) with prevalent T2D

4.1. Introduction

4.1.1. T2D PS

In the past two decades, polygenic scores have taken different names and forms and emerged as a modality for investigating genetic associations with the prevalence and incidence of complex, polygenic traits such as T2D. An early study by Weedon et al employed logistic regression to calculate odds ratios for T2D for combinations of risk alleles at 3 SNPs, notably including rs7903146 in *TCF7L2* (which has a relatively large influence on T2D risk), in a case-control study of over 6,000 individuals (2,409 cases and 3,668 controls) (Weedon et al., 2006). Weedon et al reported that participants with six risk alleles across the three SNPs had an OR of 2.84 (95% CI 1.21-1.35) times that of the reference group, which had zero or one risk alleles (Weedon et al., 2006). Another study by Langothe et al also employed logistic regression to assess the OR for combinations of 18 established T2D-associated SNPs—also including the top T2D-associated SNP in *TCF7L2* as well as rs8050135 in *FTO*—in a case-control study of nearly 5,000 participants (2,309 cases and 2,598 controls) (Langothe et al., 2008). Similarly to Wheedon et al., Langothe et al reported that individuals who carried more risk alleles had a higher risk of T2D, and the 1.2% of individuals with more than 24 risk alleles across the 18 SNPs had an odds ratio for T2D of 4.2 (95% CI 2.11-8.56) as compared with the 1.8% of participants who had 10-12 risk alleles (Lango et al., 2008). However, Langothe et al also stated that the area under the receiver operating characteristic curve (AUC) (i.e., predictive accuracy with respect to T2D) for these

18 variants was 0.60; the AUC for age, BMI and sex was 0.78, and adding the 18 SNPs as predictors only modestly increased the AUC to 0.80 (Langothe et al., 2008).

More recently, more sophisticated modes of construction of T2D PS have included testing and validation data sets, accounted for LD among T2D-associated SNPs (e.g., LDPred), and integrated odds ratios as weights from the GWAS summary statistics from which they are derived. Coupled with GWAS that have increased power to detect lower-frequency variants, these modes of constructing T2D PS have enabled the generation of PS that contain greater numbers of SNPs and that demonstrate greater accuracy. For example, Khera et al employed summary statistics from a previously conducted GWAS meta-analysis of T2D in study populations of European ancestry (Scott et al., 2017) and employed the LDPred algorithm to generate and test a T2D PS that contained 6,917,436 SNPs within a discovery data set in UK BioBank (5,853 cases and 288,978 controls) (Khera et al., 2018). The AUC of this T2D PS was 0.72 (95% CI 0.72-0.73) in the validation data set and 0.73 (95% CI 0.72-0.73) in the testing data set, substantially higher than that reported by Langothe et al (Khera et al., 2018). 3.5% of individuals were identified using the T2D PS as having at least threefold risk (OR ≥ 3.0) of T2D (Khera et al., 2018).

While T2D PS have increasingly integrated study populations that represent more diverse ancestry groups in the past two decades, to date, they have largely been developed and validated in study populations of European ancestry (Ge et al., 2021). A recent study reported that a PS derived from a multi-ancestry meta-analysis for coronary artery disease outperformed PS that were derived from Japanese and European GWAS (Koyama et al., 2020). Chikowore et al recently used PRSice to construct ethnic-specific T2D PS from a multi-ancestry T2D GWAS meta-analysis (228,499 cases and 1,178,783 controls) (Vujkovic et al., 2020) for the target data set as the South African Zulu study (1,602 cases and 981 controls) and for validation and testing in the African American Diabetes Mellitus study (2,148

cases and 2,161 controls) (Chikowore et al., 2022). The African American-derived T2D PS and multi-ethnic T2D PS performed similarly, as measured by AUC (Chikowore et al., 2022). As indicated by sensitivity and specificity, the African American-derived T2D PS was more transferable within and across countries represented by African American Diabetes Mellitus study participants, as compared with the European- and multi-ethnic-derived T2D PS (Chikowore et al., 2022).

This phenomenon prompts further investigation on the transferability of PS for T2D across study populations of different ancestry groups; cross-sectional associations between T2D PS derived from different ancestry groups and T2D are explored in Section 4.3.1. One solution to address the transferability problem of PS is to diversify genome-wide association studies to enable well-powered GWAS across multiple ancestry groups. Specifically, transferability of PS for T2D has not been widely assessed in populations Indigenous to the US. Thus far, evidence from a worldwide survey of haplotype variation and LD in the human genome has suggested that tagging of SNPs may be relatively extensive in these populations due to extended LD despite limited haplotype sharing (Conrad et al., 2006). Future studies could also contribute information on the predictive utility of T2D PS in predicting the incidence of T2D. They could do so by by employing net reclassification improvement (i.e., a quantification of correct reclassification introduced by including an additional variable in a model) and decision curve analyses (i.e., quantifying the marginal benefit of including an additional variable in a predictive model). Such analyses involving the PECRB cohort and results will be described in Chapter 5.

4.1.2. T2D pPS

While ‘global’ T2D PS using all T2D susceptibility variants are suitable for risk prediction, ‘partitioned’ or ‘process-specific’ polygenic scores (pPS) that capture the aetiological diversity

of T2D may have more traction for clinical decision making with respect to the heterogeneity of T2D. T2D pPS can be calculated for individuals using their genotypes and genetic cluster assignments for a set of T2D-associated variants using GWAS summary statistics for T2D and related traits. McCarthy's proposed 'palette model' of diabetes predisposition provides a schematic to conceptualise this heterogeneity: it posits that multiple pathophysiological processes (e.g., obesity, fat distribution, islet function, insulin sensitivity, etc.)—each influenced by genetic and non-genetic factors—can in parallel lead to the development of T2D and be depicted as multicoloured points within a multidimensional space (McCarthy, 2017). This 'palette model' assumes disease heterogeneity and is consistent with current understandings of the pathogenesis and genetic architecture of T2D as a complex disease that is associated with many genetic variants influencing different processes, most of which have relatively small effect sizes (Fuchsberger et al., 2016; Udler et al., 2018). Thus, for individuals in whom particular pathways seem to be more important, 'personalised' treatments or preventive strategies that are targeted to these pathways may be particularly effective.

Previous studies have shown that distinct physiological processes are known to contribute to the development of T2D and many centre around impaired insulin action and insulin secretion. T2D-associated variants have been shown to increase or decrease risk of T2D, working through multiple processes (process and genetic variant that has been linked with each listed); for example, muscle dysfunction (*TBC1D4*), adipose dysfunction (*KLF14*), liver dysfunction (*GCKR*), obesity (*FTO*), and fat distribution (*TBX15*) (Franks & McCarthy, 2016). Similarly, multiple processes can impair insulin secretion: defects in islet development (*PAX4*), which can contribute to islet dysfunction (*KCNJ11*), and failure of the incretin system (*GIPR*) (Franks & McCarthy, 2016). The proliferation of GWAS and GWAS meta-analyses through such efforts as the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM)

Consortium and Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) in the past decade—in combination with 'soft' clustering approaches (Bezdek et al., 1984)—have enabled recent efforts to generate clusters of T2D-associated loci with respect to related metabolic traits (Mahajan et al., 2018a; Udler et al., 2018, 2019), irrespective of known or presumed function of each locus.

4.1.3. Previous genetic clustering analyses among T2D-associated loci

Mahajan et al analysed the patterns of association between T2D and other metabolic traits by performing a hierarchical clustering analysis among 94 coding and non-coding variants that were significantly associated with T2D in their expanded GWAS meta-analysis as well as publicly available association summary statistics and collaborations (Mahajan et al., 2018a). This hierarchical clustering process implemented a fuzzy c-means algorithm, the complete agglomeration method, which employed the distance matrix among the *z* scores of the loci and traits (Mahajan et al., 2018a). The clustering algorithm enabled Mahajan et al to assign 71 of the 94 T2D-associated loci to one of three physiological clusters, which emerged from the data and were not predefined: insulin action, insulin secretion, and BMI-dyslipidaemia (Mahajan et al., 2018a). The remaining 23 loci were relegated to an 'undetermined' cluster, which lacked any striking phenotypic associations (Mahajan et al., 2018a). In aggregate, this generated multi-trait association patterns across glycaemic traits, anthropometric traits, and lipid levels to yield five different clusters (which included subsets of the original three clusters), plus one cluster that contained a combination of features from two clusters (Mahajan et al., 2018a).

In parallel, Udler et al used a complementary soft-clustering approach with 94 T2D association signals—which partially overlaps with those used by Mahajan et al—and publicly-available GWAS data for 47 diabetes-related traits to generate five clusters and assignments

for each of the 94 variants (Udler et al., 2018). Bayesian nonnegative matrix factorisation (bNMF) enables the calculation of weights and the assignment of individual signals to one or more clusters (Tan & Févotte, 2013); Udler et al modified the original algorithm to account for positive and negative associations between signals and traits (Udler et al., 2018). Using GWAS summary statistics for 47 diabetes-related traits, Udler et al investigated cluster-trait associations and confirmed clustering results. Among the two clusters under the 'β-cell dysfunction' category, both were significantly associated with decreased HOMA-β ($p < 10^{-10}$) and decreased fasting insulin ($p < 5 \times 10^{-4}$); the 'β-cell' cluster was significantly associated with increased proinsulin ($p < 10^{-9}$), while the 'proinsulin' cluster was significantly associated with decreased proinsulin ($p < 10^{-17}$). The 'lipodystrophy' cluster shared multiple loci with the 'lipodystrophy-like' insulin resistance cluster that was previously proposed by Yaghootar et al (Yaghootkar et al., 2016); Udler et al.'s lipodystrophy cluster was significantly associated with decreased ISI adjusted for BMI, adiponectin, HDL cholesterol, and increased TG ($p < 10^{-20}$). The obesity cluster may represent obesity-mediated insulin resistance: the cluster was significantly associated with increased fasting insulin unadjusted for BMI ($p < 10^{-7}$) but not fasting insulin adjusted for BMI ($p = 0.57$). A recent study constructed individual-level T2D pPS for the five clusters from Udler et al in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium and the UK Biobank (N = 454,193) (DiCorpo et al., 2022). The T2D pPS had differential associations with various cardiometabolic outcomes: for instance, increased pPS for adiposity and lipodystrophy were significantly associated with increased blood pressure and hypertension, but had opposite directions of association with measures of adiposity (e.g., BMI) (DiCorpo et al., 2022).

Udler et al.'s use of bNMF and thresholding of the resulting clusters provided a novel approach to generating T2D genetic clusters based on phenotypic and genotypic information and will be compared with Mahajan et al.'s related study. The physiological impacts and

phenotypic features associated with each corresponding cluster developed by Udler et al and Mahajan et al appear broadly consistent; with differences in assignments of individual loci across clusters between the studies potentially attributable to the differing traits whose GWAS were used in the clustering approaches. As shown in Table 4.1, each of Udler et al.'s five clusters appears to share phenotypic features with Mahajan et al.'s six clusters; specifically, the clusters associated with reduced insulin sensitivity in both clusters appeared to have the greatest amount of overlap (Udler et al., 2019).

Table 4.1. Synthesis of T2D genetic clusters for T2D that capture heterogeneity in physiological impacts and phenotypic features (Mahajan et al., 2018a; Udler et al., 2018, 2019). Reproduced from Udler et al., 2019, an open access article distributed under the terms of the Creative Commons CC BY license.

Physiological impact		Phenotypic features	Cluster name		Examples of T2D loci
			Udler et al., 2018	Mahajan et al., 2018a	
Adverse impact on β-cell function	High proinsulin	Low fasting insulin (+ high proinsulin)	β-cell	Insulin secretion 1	*ABO, ADCY5, HNF1A, HNF1B, MTNR1B, SLC30A8, TCF7L2*
	Low proinsulin	Low fasting insulin (+ low proinsulin)	Proinsulin	Insulin secretion 2	*IGF2BP2, CENTD2/ARAP1, CCND2*
Reduced insulin sensitivity	Mediation with fat distribution	High fasting insulin, low BMI, low waist circumference (WC), high triglycerides (TG)	Lipodystrophy	Insulin action	*MACF1, GRB14, IRS1, PPARG, ANKRD55, KLF14, LPL, CMIP*
	Mediation via obesity	High fasting insulin, high BMI, high WC	Obesity	Adiposity	*NRXN3, FTO, MC4R*
	Mediation via lipid metabolism	Low TG	Liver/lipid	Dyslipidaemia	*GCKR, TM6SF2*
Undetermined	N/A	No striking phenotype association	N/A	Mixed features	*BCL11A, TLE1, PLEKHA1, HMGA2, MTMR3*

By contrast, the clusters associated with adverse impacts on β-cell function (β-cell and proinsulin in Udler et al.; insulin secretion 1 and 2 in Mahajan et al.) exhibited more inconsistencies between the studies. Notably, the method used by Mahajan et al generated these two insulin secretion clusters without explicitly including summary statistics for a GWAS of proinsulin; this suggests that there is a broader distinction between these clusters that is not captured by the inclusion of proinsulin in the clustering approach, alone. Some variants with weaker associations with proinsulin were not consistently assigned to the corresponding clusters in the other study; however, T2D loci in *SLC30A8, ADCY5, HNF1A, TCF7L2,* and *MTNR1B* were consistently mapped to a cluster that had increased proinsulin and *ARAP1, IGF2BP2, DGKB,* and *CCND2* were consistently mapped to one with decreased proinsulin (Udler et al., 2019). While T2D has over 400 genome-wide significant association signals to date (Mahajan et al., 2018b), these two analyses by Udler et al and Mahajan et al were restricted to a subset of the most robust loci, which were primarily identified in GWAS meta-analyses in European-ancestry study populations (Udler et al., 2019). Further work is needed to conduct GWAS and large-scale GWAS meta-analyses of T2D and related traits in diverse study populations and to extend analyses of T2D pPS across more variants.

4.1.4. Genetic clusters used in the current study

Before the present study, Mahajan et al had conducted T2D genetic cluster analyses among a set of 307 T2D signals—a robust subset of the 403 genome-wide significant signals that were identified in Mahajan et al.'s large-scale 2018 GWAS meta-analysis in European-ancestry study population (Mahajan et al., 2018b) (Mahajan, personal communication). While the analyses among the 307 T2D signals are not yet published, I accessed preliminary results with permission of Dr Anubha Mahajan, study lead and my DPhil co-supervisor. As in their 2018 study of genetic clusters associated with T2D and various traits that included 94

T2D signals (Mahajan et al., 2018a), Mahajan et al have since implemented hierarchical clustering among the expanded set of 307 T2D signals, which identified six genetic clusters (GCs) that were defined based on multi-trait association patterns across 10 core traits. The six GCs captured loci with primary effects on: adiposity (GC1-BMI), lipids (GC2-lipids), insulin action (GC3-IA), and beta-cell function (GC4-BCF, GC5-BCF), plus a GC with mixed features (GC6-MIX). While both GC4-BCF and GC5-BCF represent beta-cell defects, high proinsulin is thought to reflect a defect in insulin processing, while low proinsulin is thought to reflect a defect in insulin synthesis or low beta-cell mass.

Mahajan et al investigated the pairwise associations between these individual T2D pPS loadings (i.e., clusters) with the risk of diabetes complications using publicly available GWAS summary statistics, and those available through collaborations, for coronary artery disease (CAD), atrial fibrillation (AF), hypertension (HTN), chronic kidney disease (CKD), estimated glomerular filtration rate (eGFR), albuminuria, stroke, and liver fat percentage. These results demonstrated that the clusters (i.e., pPS) can capture individual differences in subphenotypes that are related to the development of T2D and can be related to clinically relevant outcomes (Mahajan, personal communication). Further research on the transferability of T2D pPS and associations with T2D-related traits in other study populations is needed, which prompted the analyses that are outlined in this chapter.

4.2. Methods

4.2.1. Construction of T2D PS

The current analyses, along with those detailed in Chapter 5, employed the use of 10 T2D PS that I constructed using the methods detailed in Section 2.4.4. In brief, genotypic data was derived from the Indigenous study population from the Southwestern US, 7,701 of whom had

genotypes available in a GWAS of T2D that had been previously performed by collaborators at NIH/NIDDK PECRB. 10 different T2D PS were constructed, independent signals derived from the summary statistics of GWAS meta-analyses that were conducted among study populations from various world regions, and the current study population's imputed genotypes. Because the PS are standardised, the results are scaled to the effect per SD of the score.

(1) The AGEN 2020 T2D PS (N = 125 SNPs) was derived from the genome-wide significant SNPs identified in a GWAS meta-analysis of East Asian-ancestry study populations;

(2) the DIAGRAM 2018 T2D PS (N = 245 SNPs), derived from a GWAS meta-analysis of European-ancestry study populations;

The six DIAMANTE 2020-derived T2D PS, with variants derived from a GWAS meta-analysis of multiple ancestry groups and with weights derived from meta-analyses among study participants from ancestry groups (multi-ancestry, African, East Asian, European, Hispanic/Latino, and South Asian) for which they are named, and the population-specific weight PS with weights sourced from a previous T2D GWAS in the Indigenous study population.

(3) DIAMANTE 2020 multi-ancestry – composite (N = 287 SNPs);

(4) DIAMANTE 2020 multi-ancestry – African (N = 276 SNPs);

(5) DIAMANTE 2020 multi-ancestry – East Asian (N = 280 SNPs);

(6) DIAMANTE 2020 multi-ancestry – European (N = 287 SNPs);

(7) DIAMANTE 2020 multi-ancestry – Hispanic/Latino (N = 287 SNPs);

(8) DIAMANTE 2020 multi-ancestry – South Asian (N = 282 SNPs);

Two PS were constructed for the Indigenous study population:

(9) the population-specific variant PS (N = 287), which included the same number of SNPs as T2D PS 3-8, constructed using a method that limited the influence of LD among variants in close physical proximity, dividing the genome into bins of ~200kb using the k-means procedure (PROC FASTCLUS in SAS), resulting in 25,763 bins across the genome. Within each discovery group, the variant with the lowest p-value for association with T2D was taken within each bin. I chose 287 variants for constructing this population-specific variant score to obtain a PS with number of variants and information comparable to those from the DIAMANTE 2020 and DIAGRAM 2018 PS.

(10) the population-specific weight PS (N = 287), with variants with the lowest p-values of association with T2D, and corresponding weights, derived from a previously conducted T2D GWAS in the Indigenous study population.

One main distinction was made for the purposes of analyses for this chapter. There are two iterations of the DIAGRAM 2018 T2D PS (i.e., the first T2D PS in the list above): the additional one described in this chapter (N = 245 SNPs) and the one described in Chapters 2, 3, and 5 (N = 293). T2D GWAS summary statistics including the effect size (odds ratio for T2D in Europeans) were sourced from publicly available GWAS meta-analysis summary statistics that were generated by Mahajan et al (Mahajan et al., 2018b); this process is described in greater detail in Section 2.4.4.

However, only a subset of 307 of the original 403 genome-wide significant ($p < 5 \times 10^{-8}$) SNPs were partitioned into various genetic clusters to calculate the various T2D pPS by Mahajan et al. The overlap of these 307 SNPs and imputed genotypes from the Indigenous

study participants in the longitudinal study (N = 245 SNPs) is used in Chapter 4, rather than the 293 SNPs that were included in the DIAGRAM 2018 T2D PS as specified in Section 2.4.4 and Chapters 3 and 5. The T2D PS was constructed for each individual using the sum of products of each logged odds ratio (OR) and number of effect alleles (*dosage*) for each variant, across the 245 variants that were included in the DIAGRAM 2018 T2D PS that is used in this chapter (N = 245 SNPs) (Equation 2.3). The 293-SNP DIAGRAM 2018 T2D PS was used in analyses with T2D in comparison with the other nine T2D PS; the 245-SNP DIAGRAM 2018 T2D PS was used in analyses of associations with T2D and related traits alongside the T2D pPS.

4.2.2. Associations of T2D and birth weight PS with T2D and birth weight

Weighted T2D polygenic scores (PS) were constructed per study participant using imputed genotypes that were available for participants in the longitudinal study and genotypic data for genome-wide significant SNPs from large GWAS meta-analyses for birth weight and T2D. Details on this genotypic data are provided in Section 2.1 and on the construction of the T2D PS are in Section 2.3.3.

The birth weight PS was constructed using GWAS summary statistics from a meta-analysis of GWAS in European-ancestry study populations in the Early Growth Genetics (EGG) Consortium: this was a foetal GWAS of own birth weight with adjustment for sex, the first four principal components, and gestational age (Warrington et al., 2019). A total of 193 independent SNPs met the genome-wide significance threshold ($p < 5 \times 10^{-8}$); imputed genotypes for 126 of those SNPs were available in the present study population. The birth weight PS was calculated for each participant as the weighted sum of the product of each SNP's beta coefficient derived from EGG data and the number of effect alleles that participant possesses, across those 126 SNPs. The T2D polygenic score was constructed

using GWAS summary statistics from a meta-analysis of GWAS in European-ancestry study populations in the DIAGRAM Consortium (Mahajan et al., 2018b). The T2D PS was calculated for each participant as the weighted sum of the product of each SNP's beta coefficient derived from DIAGRAM data and the number of effect alleles that participant possesses, across those 293 SNPs. Further details on the construction of the DIAGRAM 2018 T2D PS are in Section 2.3.

Associations of EGG 2019 birth weight PS and the DIAGRAM 2018 T2D PS with T2D and birth weight were calculated, accounting for participants' pairwise relationships, using linear mixed models. Phenotypic data (birth weight, N = 3,700; T2D, N = 7,659) were derived from study participants in the longitudinal study who had genotypic data available for the calculation of the PS (derivation of this phenotypic data and participant characteristics for these subsets of longitudinal study participants are described further in Section 2.2). Birth weight data were normalized separately by sex and analysed for associations with the EGG 2019 birth weight PS and the DIAGRAM 2018 T2D PS, adjusted for birth year, gestational age and the first 5 genetic PCs. T2D data were analysed for associations with the EGG 2019 birth weight PS and the T2D PS, adjusted for age, sex, birth year and the first 5 genetic PCs. A chi-square test for whether the EGG 2019 birth weight PS had a nonlinear (i.e., U-shaped) relationship with T2D was also tested by adding a quadratic term (i.e., the square of the EGG 2019 birth weight PS) to the aforementioned model and comparing the log likelihood estimates for the aforementioned and present model.

4.2.3. Construction of T2D pPS based on Mahajan et al.'s clusters

The current analyses used the genetic cluster assignments that Mahajan et al had previously generated, using genotypes for the 245 of the original 307 variants that were imputed and available from in the Indigenous study population from the Southwestern U.S (N = 7,996,223

available imputed genotypes; details in Chapter 2), to calculate T2D pPS corresponding to each of the clusters, for all individuals with genotypic data (N = 7,701). The 62 variants that were missing were mostly monomorphic, and otherwise, unavailable within the Indigenous study population's genotypic data due to differing LD patterns. To calculate the six T2D pPS, I employed the genetic cluster assignments the original set of 307 variants that were provided directly to me by Dr Mahajan through collaboration (Mahajan A.; personal correspondence, 2020). Weighted pPS were calculated for each genetic cluster *i* (with *n* variants) using the sum of products of each effect size (*β*) and number of effect alleles (*dosage*) for each variant *j* (Equation 4.2).

$$pPS_i = \sum_{j=1}^{n_i} \beta_j \cdot dosage_j \tag{4.2}$$

A summary of T2D pPS clusters and the T2D PS, including phenotypic features, the number of SNPs from Mahajan et al.'s study (totalling 307 SNPs among the genetic clusters), the subset of SNPs available for the Indigenous study population's scores (totalling 245 SNPs) and examples of loci included in each score are summarised in Table 4.2. pPS effect sizes are scaled to the effect per SD of the pPS.

Table 4.2. Data for the T2D pPS and T2D PS that are available in the Indigenous study population.

GC #	Cluster name	Phenotypic features	N SNPs from Mahajan et al	N SNPs available in PECRB cohort
1	**Adiposity**	High BMI	23	21
2	**Lipids**	Low TG, low TC, high HDL	3	3
3	**Lipodystrophy-like, Insulin action**	High WHR, high TG, low HDL; no strong BMI association	34	28
4	**Insulin secretion 1**	Low fasting insulin; strongest association w/ T2D, OGTT/2hPG, FPG	11	11
5	**Insulin secretion 2**	Low fasting insulin	94	74
6	**Mixed features**	No prominent feature	142	108
7	**All**	**Assoc w/ T2D**	**307**	**245**

Associations of the T2D PS and each of the T2D pPS with T2D and related metabolic traits were calculated using phenotypic data from the Indigenous study population.

4.2.4. Associations between the T2D pPS, T2D PS, T2D and related metabolic traits

Pairwise associations between the T2D PS and T2D pPS and T2D and related metabolic traits were computed. This was done using the measured genotype association (MGA) function within the program SOLAR-Eclipse (version 8.4.1). All the phenotypic data collection for traits whose data was involved in this study was conducted previously by other researchers at the NIH/NIDDK Phoenix Epidemiology and Clinical Research Branch and Clinical Research Centre; details are available in Section 2.3.

4.3. Results

4.3.1. Associations between 10 constructions of T2D PS and T2D

The 10 T2D PS and their associations with T2D prevalence (N = 7,659 individuals) were assessed. Odds ratios (OR) were calculated from beta coefficients and variance explained as output by SOLAR-Eclipse, according to the Haggstrom method (Haggstrom, 1983). Results for these analyses are displayed in Table 4.3.

Table 4.3. Associations between T2D PS and T2D

	OR†	SE(OR)	*p*
AGEN 2020 T2D PS (N = 125 SNPs)	2.03	0.07	1.76×10^{-24}
DIAGRAM 2018 T2D PS (N = 293 SNPs)	1.95	0.06	8.10×10^{-31}
DIAMANTE 2020 – Multi-ancestry – African T2D PS (N = 276 SNPs)	1.57	0.05	4.14×10^{-18}
DIAMANTE 2020 – Multi-ancestry – East Asian T2D PS (N = 280 SNPs)	1.74	0.05	4.36×10^{-27}
DIAMANTE 2020 – Multi-ancestry – European T2D PS (N = 287 SNPs)	2.08	0.06	9.93×10^{-34}
DIAMANTE 2020 – Multi-ancestry – Hispanic/Latino T2D PS (N = 287 SNPs)	1.78	0.05	4.61×10^{-28}
DIAMANTE 2020 – Multi-ancestry – South Asian T2D PS (N = 282 SNPs)	2.03	0.06	8.01×10^{-26}
DIAMANTE 2020 – Multi-ancestry – Composite T2D PS (N = 287 SNPs)	2.14	0.06	7.65×10^{-33}
Population-specific weight T2D PS (N = 287 SNPs)	1.30	0.03	7.52×10^{-15}
Population-specific variant T2D PS (N = 287 SNPs)	1.14	0.04	4.55×10^{-4}

† Expressed in SD of the respective T2D PS.

The DIAMANTE 2020 – multi-ancestry – composite T2D PS and multi-ancestry – European T2D PS and the DIAGRAM 2018 T2D PS had the greatest degree of significance, in terms of p-value. The DIAMANTE 2020 – multi-ancestry – composite T2D PS and DIAMANTE 2020 multi-ancestry – European T2D PS had the greatest odds ratios. The 10 T2D PS had a broad range of OR with respect to T2D in the Indigenous study population: between 1.14 for the population-specific weight T2D PS and 2.14 for the DIAMANTE 2020 – multi-ancestry – composite T2D PS.

4.2.5. Associations of T2D and birth weight PS with T2D and birth weight

The heterogeneity between summary statistics for the PECRB GWAS and the EGG 2019 GWAS for the 126 SNPs included in the EGG 2019 birth weight PS was calculated, as measured by the mean heterogeneity index I^2 (Higgins & Thompson, 2002). The heterogeneity index describes the percentage of variation across studies that is due to heterogeneity as opposed to chance (Higgins & Thompson, 2002). I^2 was 15.54%, which indicate that the percentage of variability in effect estimates that is due to heterogeneity, rather than chance, is low (Higgins & Thompson, 2002). The EGG 2019 birth weight PS (126 SNPs) had a significant, positive association with birth weight (β = 0.134 SD birth weight per SD birth weight PS (95% CI 0.101, 0.168); $p = 4.1\times10^{-15}$) in the PECRB cohort (N = 3,700 with birth weight and genotypic data). The T2D PS was likewise significantly, positively associated with T2D (OR = 1.48 per SD T2D PS (1.38, 1.58); $p = 8.1\times10^{-31}$). The birth weight PS was significantly, inversely associated with T2D (OR = 0.91 per SD birth weight PS (0.85, 0.97); p = 0.0043). By adding a square term for the birth weight PS to the model, I found no evidence for a non-linear (i.e., U-shaped) relationship between birth weight PS and T2D (p = 0.80). The T2D PS was not significantly associated with birth weight (β = -0.005 per SD T2D PS (-0.040, 0.030); p = 0.80).

4.2.6. Associations between T2D pPS and T2D and related traits

Pairwise associations between the T2D pPS and T2D and related metabolic traits were computed. This was done using the measured genotype association (MGA) function within the program SOLAR-Eclipse (version 8.4.1). The adiposity pPS (N = 21 SNPs), which was characterised by elevated BMI in Mahajan et al.'s clustering process, was strongly and significantly positively associated with PFAT (β = 1.0 per SD pPS, $p = 5.5 \times 10^{-4}$) and

significantly associated with logmaxBMI (β = 0.016, p = 2.6 × 10^{-7}) as well as T2D (β = 0.014, p = 0.0054) in the Indigenous study population (Table 4.4).

Table 4.4. Associations of the adiposity T2D PS (21 SNPs) with T2D and related traits.

	β	SE(β)	p
T2D (N = 7,659)	0.014	0.0050	0.0054
logmaxBMI (N = 6,789)	0.016	0.0031	2.6 × 10-7
FPG (N = 6,637)	0.22	0.15	0.13
2hPG (N = 6,637)	0.23	0.46	0.62
logAIR (N = 404)	0.11	0.014	0.44
CIR30 (N = 401)	0.0029	0.0012	0.019
CIR120 (N = 343)	-6.2 × 10-5	0.0029	0.98
logM (N = 557)	0.0043	0.004	0.34
PFAT (N = 557)	1.0	0.29	5.5 × 10-4

† Expressed in SD of the respective T2D PS.

The lipids pPS (N = 3 SNPs), which was characterised by lower triglycerides in Mahajan et al.'s clustering process, did not significantly associate with any of the metabolic traits analysed (Table 4.5). This could be due to the relatively low number of SNPs that comprise this pPS.

Table 4.5. Associations of the lipids T2D pPS (3 SNPs) with T2D and related traits.

	β	SE(β)	p
T2D (N = 7,659)	-0.0062	0.0049	0.21
logmaxBMI (N = 6,789)	-0.0030	0.0031	0.33
FPG (N = 6,637)	0.025	0.15	0.86
2hPG (N = 6,637)	0.14	0.45	0.76
logAIR (N = 404)	-0.0024	0.015	0.87
CIR30 (N = 401)	3.8 × 10-4	0.0014	0.78
CIR120 (N = 343)	0.0047	0.0031	0.13
logM (N = 557)	0.0013	0.005	0.78
PFAT (N = 557)	0.11	0.31	0.73

† Expressed in SD of the respective T2D PS.

The lipodystrophy-like/insulin action pPS (N = 28 SNPs), characterised by elevated WHR, elevated TG, and decreased HDL cholesterol by Mahajan et al., was strongly and significantly associated with 2hPG (β = 2.1 per SD pPS, $P = 4.6 \times 10^{-6}$) and significantly associated with FPG ($P = 0.045$), logM ($P = 5.3 \times 10^{-5}$), and T2D ($P = 2.0 \times 10^{-7}$) (Table 4.6).

Table 4.6. Associations of the lipodystrophy-like/insulin action pPS (28 SNPs) with T2D and related traits.

	β	SE(β)	p
T2D (N = 7,659)	0.025	0.0049	2.0 × 10-7
logmaxBMI (N = 6,789)	-0.0041	0.0031	0.19
FPG (N = 6,637)	0.29	0.15	0.045
2hPG (N = 6,637)	2.1	0.45	4.6 × 10-6
logAIR (N = 404)	0.0096	0.014	0.48
CIR30 (N = 401)	7.8 × 10-4	0.0013	0.55
CIR120 (N = 343)	0.0038	0.0028	0.18
logM (N = 557)	-0.018	0.004	5.3 × 10-5
PFAT (N = 557)	-0.28	0.29	0.34

† Expressed in SD of the respective T2D PS.

Associations with traits in the Indigenous study population for the two insulin secretion pPS are listed in Table 4.7. In the original clustering procedure, Mahajan et al had found that the insulin secretion 1 cluster was associated with higher proinsulin than the insulin secretion 2 cluster.

Table 4.7. Associations of the insulin secretion 1 pPS (11 SNPs) and insulin secretion 2 pPS (74 SNPs) with T2D and related traits.

	Insulin secretion 1 pPS			Insulin secretion 2 pPS		
	β	SE(β)	p	β	SE(β)	p
T2D (N = 7,659)	0.028	0.0051	2.7×10^{-8}	0.038	0.0051	6.2×10^{-14}
logmaxBMI (N = 6,789)	-0.012	0.0032	2.5×10^{-4}	-0.015	0.0032	4.9×10^{-6}
FPG (N = 6,637)	0.64	0.15	1.8×10^{-5}	0.45	0.15	0.0031
2hPG (N = 6,637)	1.4	0.46	0.0027	2.1	0.47	7.4×10^{-6}
logAIR (N = 404)	-0.056	0.014	5.3×10^{-5}	-0.062	0.014	1.5×10^{-5}
CIR30 (N = 401)	-0.0032	0.0013	0.014	-0.0058	0.0013	6.9×10^{-6}
CIR120 (N = 343)	-0.0048	0.0030	0.11	-0.0039	0.0032	0.22
logM (N = 557)	-3.2×10^{-4}	0.003	0.90	0.0046	0.005	0.34
PFAT (N = 557)	-0.068	0.30	0.82	-0.53	0.32	0.98

† Expressed in SD of the respective T2D PS.

Both the insulin secretion 1 and insulin secretion 2 pPS were significantly ($P < 0.05$) associated with T2D, logmaxBMI, FPG, 2hPG, logAIR, and CIR30, in the Indigenous study population. Between the two pPS, they had consistent directions of association: both associated positively with T2D, FPG, and 2hPG; and negatively with logmaxBMI, logAIR, and CIR30.

The mixed features pPS (108 SNPs) was significantly associated with the same traits: T2D, logmaxBMI, FPG, 2hPG, and logAIR; the association of greatest significance was with T2D ($P = 2.4 \times 10^{-6}$) (Table 4.8).

Table 4.8. Associations of the mixed features PS (108 SNPs) with T2D and related traits.

	β	SE(*β*)	*p*
T2D (N = 7,659)	0.024	0.0050	2.4 × 10-6
logmaxBMI (N = 6,789)	0.0073	0.0031	0.020
FPG (N = 6,637)	0.44	0.15	0.0028
2hPG (N = 6,637)	1.1	0.45	0.012
logAIR (N = 404)	-0.029	0.014	0.037
CIR30 (N = 401)	-0.0017	0.0013	0.19
CIR120 (N = 343)	9.2 × 10-4	0.0030	0.76
logM (N = 557)	0.0029	0.005	0.54
PFAT (N = 557)	0.26	0.31	0.40

† Expressed in SD of the respective T2D PS.

It was also significantly ($P < 0.05$) associated with logmaxBMI ($\beta = 0.0073$, $P = 0.020$), FPG ($\beta = 0.44$, $P = 0.0028$), 2hPG ($\beta = 1.1$, $P = 0.012$), and logAIR ($\beta = -0.029$, $P = 0.037$). Apart from the association with logAIR, all associations were positive in direction.

Lastly, the T2D PS (N = 245 SNPs), which was constructed using summary statistics from the DIAGRAM Consortium 2018 GWAS meta-analysis of study participants of European ancestry, was relatively strongly and highly significantly associated with the most traits. The strongest associations were with T2D ($\beta = 0.057$, $P = 1.0 \times 10^{-29}$), as expected. It was also significantly ($P < 0.05$) associated with logmaxBMI ($\beta = 0.0073$, $P = 0.020$), FPG ($\beta = 0.93$, $P = 8.0 \times 10^{-10}$), 2hPG ($\beta = 3.1$, $P = 2.5 \times 10^{-11}$), logAIR ($\beta = -0.074$, $P = 2.3 \times 10^{-7}$), CIR30 ($\beta = -0.0048$, $P = 2.6 \times 10^{-4}$) (Table 4.9).

Table 4.9. Associations of the DIAGRAM 2018 T2D PS (245 SNPs) with T2D and related traits.

	β	SE(*β*)	*p*
T2D (N = 7,659)	0.057	0.0051	1.0 × 10-29
logmaxBMI (N = 6,789)	-0.0075	0.0032	0.020
FPG (N = 6,637)	0.93	0.15	8.0 × 10-10
2hPG (N = 6,637)	3.1	0.46	2.5 × 10-11
logAIR (N = 404)	-0.074	0.014	2.3 × 10-7
CIR30 (N = 401)	-0.0048	0.0013	2.6 × 10-4
CIR120 (N = 343)	-0.0024	0.0031	0.45
logM (N = 557)	-4.5 × 10-4	0.005	0.93
PFAT (N = 557)	-0.42	0.32	0.99

† Expressed in SD of the respective T2D PS.

Associations with T2D, FPG and 2hPG were positive in direction; those with FPG and 2hPG were relatively great in magnitude. Those with logmaxBMI, logAIR, and CIR30 were negative.

4.4. Summary

Results of the T2D PS-T2D association study demonstrated that the 10 constructions of T2D PS had a broad range of OR and p-values. In terms of p-value, the DIAMANTE 2020 – multi-ancestry – composite T2D PS and DIAMANTE 2020 – multi-ancestry – European T2D PS and the DIAGRAM 2018 T2D PS had the greatest degree of significance of association. Data from the large-scale DIAGRAM 2018 GWAS meta-analysis had been integrated into the DIAMANTE 2020 multi-ancestry meta-analysis (Mahajan et al., 2018b), which could contribute to these similarities in the corresponding T2D PS-T2D associations. Overall, the T2D PS that were derived from the overlap of PECRB imputed genotypes and summary statistics from larger GWAS meta-analyses performed best. This may reflect the relative ease of tagging causal variants identified in these larger GWAS within the PECRB cohort, which

has a relatively high degree of LD. The DIAGRAM 2018 T2D PS was highlighted in further analyses with T2D-related traits along with T2D pPS, as described further below.

The results of tests of association of the T2D PS and birth weight PS with T2D and birth weight demonstrate that there is some relationship between low birth weight and T2D susceptibility that was not seen in the bivariate analysis as conducted. Specifically, the birth weight PS had a significant, positive association with birth weight, and the T2D PS had a significant, positive association with T2D. However, the birth weight PS was significantly, inversely associated with T2D and the T2D PS was not significantly associated with birth weight. Results indicate that genetic variants that associate with birth weight and T2D from larger European GWAS largely also influence birth weight and T2D, respectively, in the PECRB cohort. Notably, there was no evidence for a non-linear relationship between the birth weight PS and T2D (i.e., a U-shaped curve, as in the relationship between birth weight itself and T2D in this same study population).

These findings support the notion that some variants that confer susceptibility to low birth weight also confer susceptibility to T2D, which was reported in a previous Mendelian randomisation study in a European-ancestry study population (Huang et al., 2019b). While these findings prompt further investigation and biological mechanisms that may underlie them are unclear, they may align with the foetal insulin hypothesis, which posits that genetic predisposition to insulin resistance results in lower birth weight and subsequent diabetes (Hattersley & Tooke, 1999). The difference in results of the bivariate genetic correlation analyses and these association tests may be because certain variants may mediate the relationship between birth weight and risk of T2D. However, a key limitation of this analysis is that an overall birth weight polygenic score, as opposed to separate maternal and foetal birth weight PS, was used. As discussed by Warrington et al and Tyrrell et al, maternal and foetal effects with respect to birth weight may operate in opposite directions (Tyrrell et al., 2013;

Warrington et al., 2019); thus, future analyses could create separate maternal and foetal birth weight PS to further explore the genetic relationship between birth weight and subsequent T2D in offspring.

Further research is needed to assess the maternal and foetal genetic influences on birth weight and subsequent diabetes, using foetal and maternal GWAS of birth weight and PS. Previous studies were undertaken in European-ancestry populations: in the absence of Mendelian randomization techniques due to the present study population's relatively small sample size, the analyses described in this section leveraged summary statistics from large-scale GWAS meta-analyses of birth weight and T2D in European-ancestry populations to construct polygenic scores for the traits as instruments to explore the genetic relationship between birth weight and T2D.

Results of the T2D pPS- and PS-metabolic trait association study broadly agree with the putative pathophysiological mechanisms that underlie each of the six clusters by Mahajan et al. For instance, the adiposity pPS—which contains loci in the well-known obesity-associated genes *FTO* and *MC4R*—was strongly and significantly associated with PFAT, and significantly associated with logmaxBMI and T2D, in the Indigenous study population. However, the lipids cluster, which contained 3 SNPs in both the extended Mahajan et al cluster and the cluster once merged with the Indigenous study population's genotypes, was not significantly associated with any of the traits (Table 4.5). This prompts future large-scale GWAS and GWAS meta-analyses on T2D and related metabolic traits that could enable further extension of these pPS to explore pPS-trait associations. Additionally, analytical approaches to account for disparities in the sizes of various pPS in such association studies could also be implemented.

Broadly, both Udler et al and Mahajan et al included predominantly European-ancestry study populations in their pPS cluster generation and analytical processes. Associations between T2D pPS and T2D-related traits should also be investigated in other, more diverse, populations to assess transferability. The present association study demonstrates an instance of transferability of the phenotypic classifications of the T2D pPS and the T2D PS to a non-European-ancestry study population and suggest that genetic influences on adiposity, insulin action, and insulin secretion contribute to development of T2D in this population. However, a limitation of this study was that only 245 of the original 307 variants from Mahajan et al.'s clusters were available in the Indigenous study population, potentially due to differing patterns of LD across study populations and due to some variants in the European-ancestry T2D GWAS meta-analysis appearing as rare to monomorphic within the Indigenous population.

5. Predicting the incidence of T2D using clinical factors and a polygenic score (PS) for T2D in a cohort study

5.1. Introduction

T2D-associated genetic variants, derived from GWAS, have largely been reproducible across populations. Chapter 2 provided details on how PS for T2D have been constructed in other studies and in the research outlined in Chapters 4 and 5. Chapter 4 focussed on the use of T2D PS and pPS for the prediction of T2D prevalence (i.e., the integral of incidence, with the exclusion of deaths and other loss to follow up, among a cohort), cross-sectionally. There is limited information on how T2D PS based on T2D-associated variants from GWAS meta-analyses add to clinical factors for predicting T2D incidence (Vassy et al., 2014). Such prediction could help identify individuals at increased risk of T2D for targeted prevention efforts.

Previous studies that assessed contributions of a T2D PS to clinical predictors in the prediction of T2D incidence have involved relatively short follow-up time in adult populations and have mostly been conducted in European-ancestry populations (He et al., 2021; Lyssenko et al., 2008; Mars et al., 2020; Meigs et al., 2008; Park et al., 2015; Vaxillaire et al., 2014). These studies, using PS constructed from 15 variants to over six million common variants, have generally found that PS were significantly associated with T2D incidence but contributed little beyond clinical factors to overall prediction of T2D (He et al., 2021; Lyssenko et al., 2008; Mars et al., 2020; Meigs et al., 2008; Park et al., 2015; Vaxillaire et al., 2014).

Previous studies were largely conducted in adults, but utility of PS for prediction of subsequent T2D may be greater earlier in life (in youth or even at birth). Clinical variables (e.g., obesity and dysglycaemia) may not manifest earlier on in life; thus, earlier, genetic information may exhibit greater utility. Indeed, previous work in the present Indigenous study

population has shown that family recurrence risk of diabetes is higher in participants of younger ages, which suggests a relatively strong role of genetics earlier on in life (Knowler et al., 1990). In the present study, I assessed the utility of a T2D PS for prediction of T2D incidence in an Indigenous population from the Southwestern US with a high prevalence of T2D and in which long-term follow-up data are available. I aimed to analyse how genetic and clinical factors could inform strategies for screening and prevention in three cohorts of individuals in different age groups (birth, youth, and adulthood) at baseline.

5.2. Methods

The analyses in this chapter are based on data that had been collected in a longitudinal study of diabetes that was conducted in an Indigenous study population from the Southwestern US. Methods and study participant data are further described in Chapter 2. Genotypic and phenotypic data and the study design of the presently described analyses of T2D PS, clinical factors, and T2D incidence are outlined in the following subsections.

5.2.1. Study design and participants

Of the 7,701 individuals with genotypes available, I constructed three cohorts based on age at baseline examination for those who had data for at least two exams with availability of clinical variables. Those with diabetes at baseline were excluded from the cohorts. The cohorts overlapped: a total of 4,770 individuals were included among the three cohorts, which were not mutually exclusive. There were 2,333 participants who were followed from first examination in adulthood (age ≥ 20 years); within the adult cohort, 640 cases of T2D occurred over 16,686 person-years of follow-up. There were 2229 participants followed from first examination in youth (age 5-19 years); within the youth cohort, 228 cases of T2D occurred over 17,803 person-years of follow-up. There were 2,894 participants with birth

weight data available who were considered to be followed from birth; within this 'birth cohort,' 438 cases of T2D occurred over 61,591 person-years of follow-up. Individuals were included in multiple cohorts if suitable data were available (phenotypic data summarized in Table 5.1).

5.2.2. Phenotypic data

Individuals who were at least five years old had been invited for health examinations every two years. At each exam, a 75-gram oral glucose tolerance test had been administered with measurement of HbA1c and fasting and 2-hour plasma glucose (FPG and 2hPG). Diabetes was diagnosed using 1997 American Diabetes Association criteria (FPG ≥ 7.0mM, 2hPG ≥ 11.1mM, or clinical diagnosis) (Kahn, 1997). Birth weight had been collected from clinical information and Arizona State birth certificates. NIDDK PECRB research staff had not directly asked participants about parental diabetes. To approximate the information that would be available in clinical encounters, parental diabetes had been defined using three categories—yes, no, or unknown—per parent, using data from the longitudinal study obtained prior to participants' entry into the analysis. Further details on phenotypic data are in Tables 5.1.

5.2.3. Genotypic data

7,701 study participants had genotypes available from previous GWAS, generated using a custom Axiom array designed to capture common variation in members of this community (MAF ≥ 0.05, or MAF ≥ 0.01 for coding variants), using methods described in Section 2.1. The custom array identified tag single-nucleotide polymorphisms (SNPs) for approximately 5 million variants; 515,692 SNPs were directly genotyped. This genotyping captured approximately 92% of variation with MAF ≥ 0.05 and 50% of variation with MAF 1-5%. Missing and ungenotyped variants were imputed with whole genome sequence data for 266 community members as a reference panel using Impute 2 (Howie et al., 2009). Variants were

excluded if imputation quality score < 0.5 or MAF < 0.01. A total of 4,589,902 variants were imputed.

Tables 5.1. Summary of continuous and binary traits of study participants in adult, youth, and birth cohorts.

Table 5.1a. Summary of continuous traits of participants in adult, youth, and birth cohorts.

	Birth cohort (N = 2,894)[b]		Youth cohort (N = 2,229)[c]		Adult cohort (N = 2,333)[c]	
Variable	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Duration of follow-up time (years)	21.28	7.81	7.99	4.34	7.15	4.45
Age (years)			12.05	3.73	31.04	10.43
BMI (kg/m^2)			24.83	7.02	34.33	7.68
FPG (mM)			4.93	0.40	5.21	0.54
2hPG (mM)			5.64	1.28	6.29	1.68
HbA1c (%)			5.06	0.38	5.25	0.46
Birth weight (g)	3456.86	532.09				
Unweighted DIAGRAM 2018 T2D PS[a]	306.88	9.77	306.41	9.68	306.97	9.34

[a] The unweighted DIAGRAM 2018 PS is the sum of the number of T2D risk alleles across the 293 SNPs in the DIAGRAM 2018 PS (Table S1).
[b] The clinical variables that were included in analyses for the birth cohort included only those that would have been available for individuals at birth; thus, data for age, BMI, FPG, 2hPG, and HbA1c are not included in those analyses, and their summary statistics not listed here.
c The clinical variables that were included in analyses for the youth and adult cohorts did not include birth weight; thus, data for birth weight are not included in those analyses, and their summary statistics not listed here.

Table 5.1b. Summary of binary traits of participants in adult, youth, and birth cohorts.

	Birth cohort (N = 2,894)		Youth cohort (N = 2,229)		Adult cohort (N = 2,333)[c]	
Variable	**Frequency**	**Percent**	**Frequency**	**Percent**	**Frequency**	**Percent**
sex female	1569	54.22	1205	54.06	1395	59.79
maternal diabetes prior to participant's entry into study						
with diabetes	96	3.32	490	21.98	771	33.05
without diabetes	1837	63.48	1417	63.57	655	28.08
diabetes status unknown	961	33.21	322	14.45	907	38.88
paternal diabetes prior to participant's entry into study						
with diabetes	51	1.76	220	9.87	327	14.02
without diabetes	1015	35.07	764	34.28	350	15.00
diabetes status unknown	1828	63.17	1245	55.85	1656	70.98
low birth weight (< 3000g)	493	17.04				
high birth weight (≥ 4000g)	391	13.51				

5.2.4. Construction of 10 polygenic scores for T2D

10 T2D PS were constructed using summary statistics from GWAS that had been previously conduced for populations that included study populations of different ancestry groups. None of the T2D PS was trained on data from the Indigenous study population. These included the following T2D PS, each named for the meta-analysis from which it was derived (Table 5.2).

Table 5.2. Summary of the 10 constructions of T2D employed in this thesis research.

T2D PS name	Number of SNPs	Variants derived from this GWAS or population	Weights derived from this GWAS or population
AGEN 2020	125	East Asian-ancestry populations (Spracklen et al., 2020)	East Asian-ancestry populations (Spracklen et al., 2020)
DIAGRAM 2018	293	European-ancestry populations (Mahajan et al., 2018b)	European-ancestry populations (Mahajan et al., 2018b)
DIAMANTE 2020 – Multi-ancestry African	276	Multi-ancestry meta-analysis – African-ancestry populations (Mahajan et al., 2020)	Multi-ancestry meta-analysis – African-ancestry populations (Mahajan et al., 2020)
DIAMANTE 2020 – Multi-ancestry East Asian	280	Multi-ancestry meta-analysis – East Asian-ancestry populations (Mahajan et al., 2020)	Multi-ancestry meta-analysis – East Asian-ancestry populations (Mahajan et al., 2020)
DIAMANTE 2020 – Multi-ancestry European	287	Multi-ancestry meta-analysis – European-ancestry populations (Mahajan et al., 2020)	Multi-ancestry meta-analysis – European-ancestry populations (Mahajan et al., 2020)
DIAMANTE 2020 – Multi-ancestry Hispanic/Latino	287	Multi-ancestry meta-analysis – Hispanic/Latino-ancestry populations (Mahajan et al., 2020)	Multi-ancestry meta-analysis – Hispanic/Latino-ancestry populations (Mahajan et al., 2020)
DIAMANTE 2020 – Multi-ancestry South Asian	282	Multi-ancestry meta-analysis – South Asian-ancestry populations (Mahajan et al., 2020)	Multi-ancestry meta-analysis – South Asian-ancestry populations (Mahajan et al., 2020)
DIAMANTE 2020 – Multi-ancestry Composite	287	Multi-ancestry meta-analysis – all ancestry groups (Mahajan et al., 2020)	Multi-ancestry meta-analysis – all ancestry groups (Mahajan et al., 2020)
Population-specific weight	287	Multi-ancestry meta-analysis (Mahajan et al., 2020)	Indigenous longitudinal study population
Population-specific variant	287	287 with lowest p-values of association with T2D across the genome, in the Indigenous study population	Indigenous longitudinal study population

Further details on the construction of these T2D PS are explained in Section 2.4.4.

5.2.5. Statistical analyses

Analyses were completed in SAS 9.4 (SAS Institute; Cary, NC). For each cohort, individuals were followed from inception (first examination with clinical data available for youth and adult cohorts; birth for birth cohort) until they developed T2D or until their last examination—whichever came first. I evaluated the relative contributions of various combinations of clinical factors and/or the PS in the following analyses: cumulative incidence, survival, area under the receiver operating characteristic curve (AUC), net reclassification improvement (NRI), and decision curve. The variables available for the adult cohort included: age, sex, parental diabetes, BMI, HbA1c, FPG, 2hPG, and the T2D PS (Vijayakumar et al., 2017). Those for the youth cohort included: age, sex, parental diabetes, modified BMI *z* score (Chambers et al., 2017), HbA1c, FPG, 2hPG, and PS. Those for the birth cohort included: sex, parental diabetes, birth weight, and PS. Since HbA1c was only measured for examinations after 1989, I conducted additional analyses that did not require HbA1c to allow for longer follow-up and greater sample size: these analyses returned similar findings to those that included HbA1c, which are displayed in Table 5.7. For the analyses in the birth cohort that are detailed in Chapter 5 of this thesis, I analysed birth weight using two binary variables, one denoting birth weight < 3000g and another denoting birth weight ≥ 4000g. I had also analysed birth weight as a continuous variable, with linear and quadratic terms to account for the U-shaped relationship between birth weight and T2D and obtained similar results as without these terms.

I also conducted analyses including stated admixture as a covariate; its inclusion returned virtually the same results as without. Cumulative incidence, survival, decision curve, and NRI analyses required calculation of the predicted occurrence of T2D at a specified follow-up time for all individuals to ensure comparability: a follow-up of ten years was used for the adult and youth cohorts, and 30 years for the birth cohort. Previous studies in this population

demonstrated that genetic variants at *KCNQ1* rs2237895 (effect allele frequency = 0.49, OR = 1.31; exhibits parent-of-origin effects) (Hanson et al., 2013) and *ABCC8* rs1272388614 (effect allele frequency = 0.017, OR = 2.02) (Baier et al., 2015) are significantly and strongly associated with T2D. I conducted further analyses to assess the contributions of these genotypes in addition to the DIAGRAM 2018 T2D PS for the prediction of T2D incidence.

5.2.6. Cumulative incidence and survival analyses

I used Cox proportional hazards regression to evaluate associations of clinical variables and PS with T2D incidence. Cumulative incidence of T2D was calculated as proportion of individuals that developed T2D over the specified follow-up time, using Breslow's method (PROC PHREG in SAS). To assess separate contributions of PS and clinical risk, I calculated the predicted cumulative incidence of T2D in the longitudinal study for the adult, youth, and birth cohorts according to different levels of PS and of clinical risk. A linear predictor was constructed for the clinical variables (age, sex, BMI, HbA1c, and FPG) and the value of the DIAGRAM 2018 T2D PS in the proportional hazards model was scaled to have a range of -2 to 2 and centre at 0.

5.2.7. AUC analyses

I compared the AUC of models that included clinical variables alone with the AUC of those that included clinical variables and the various PS. The AUC expresses the probability within a pair of individuals—one who developed T2D and one who did not—that the individual who developed T2D had a higher predicted probability of doing so (Pencina & D'Agostino, 2004). It provides a single aggregate estimate of the predictive accuracy of one or more predictors in a model, with respect to a given outcome.

5.2.8. NRI analyses

Continuous-variable net reclassification improvement (NRI) quantifies the amount of correct reclassification introduced by using a model with an additional variable (Pencina et al., 2008). I analysed NRI by calculating the net proportion of events reclassified correctly (i.e., assigned a higher probability value) plus the net proportion of non-events reclassified correctly (i.e., assigned a lower probability value) (Chambless et al., 2011). Rather than providing a single aggregate estimate of predictive accuracy as stated in the description of the AUC, NRI provides insight into the predictive performance of a single variable, with respect to one or more others and a given outcome.

5.2.9. Decision curve analyses

I employed decision-analytic methods to assess consequences of clinical decisions and expected outcomes of alternative clinical management (i.e., including various combinations of clinical variables with and without the PS in prediction models). These analyses assume that the threshold probability (p_t) of developing T2D at which one would opt for an intervention is informative of how one weighs the relative benefits and harms of true-positive and false-positive predictions, and the net benefit of using a predictive model to select individuals above a given p_t is calculated accordingly (Vickers & Elkin, 2006). I used extensions to decision curve methods for survival analysis to plot net benefit across a range of p_t values to evaluate for which p_t ranges and what corresponding proportion of the population the PS had marginal net benefit (Vickers et al., 2008).

5.2.10. Age group comparison analyses

To test the significance of differences in predictive properties of the PS among the three cohorts (adult, youth, and birth cohorts), a bootstrapping procedure was employed (Efron, 1981). In each iteration, the 4,770 individuals who were included in at least one of the three cohorts—adult (N = 2,333), youth (N = 2,229), and/or birth (N = 2,894) cohorts—were resampled with replacement. The survival, area under the ROC curve, and net reclassification improvement analyses were then repeated for each cohort, and the pairwise differences in the logarithm of the hazard ratio, AUC and NRI associated with the PS were calculated. The standard errors of these differences among cohorts were taken from their standard deviations across 2000 replicate samples, and the standard errors were used to calculate the statistical significance of the differences (this assumes the differences are normally distributed). This provides a robust approach to assessing the differences among cohorts, which accounts for the partial overlap among them.

5.3. Results

5.3.1. Ten constructions of T2D PS

All 10 T2D PS, constructed using the overlap of published T2D GWAS summary statistics and genotypes available in this study population, had statistically significant associations with T2D incidence in the study population. HRs for the PS in models adjusted for clinical variables (age, sex, BMI, FPG, HbA1c, and parental diabetes for the adult cohort; age, sex, modified BMI *z* score, FPG, HbA1c, and parental diabetes for the youth cohort; and sex, birth weight, and parental diabetes for the birth cohort) ranged from 1.13 to 1.27 per SD of the respective T2D PS for the adult cohort, from 1.19 to 1.49 for the youth cohort, and from 1.27 to 1.48 for the birth cohort (Table 5.3). The PS that consistently had the strongest associations with T2D incidence (highest HRs) was that constructed using the DIAGRAM 2018 GWAS: 1.27 per SD DIAGRAM 2018 T2D PS (95% CI 1.17, 1.38) in the adult cohort,

1.49 (95% CI 1.29, 1.72) in the youth cohort, and 1.48 (95% CI 1.35, 1.63) in the birth cohort. The DIAMANTE 2020 multi-ancestry PS (Table 5.5) and the population-specific variant PS (Table 5.6) also had strong associations with T2D incidence, though not quite as strong.

Table 5.3. Cox regressions comparing associations of PS constructions with T2D incidence: birth, youth, and adult cohorts.

	Birth cohort (N = 2894)[b]		Youth cohort (N = 2229)[c]		Adult cohort (N = 2333)[d]	
PS Construction	**HR (per SD PS)**	**95% HR CI**	**HR (per SD PS)**	**95% HR CI**	**HR (per SD PS)**	**95% HR CI**
AGEN 2020 (N = 125 SNPs)	1.27	1.15, 1.39	1.28	1.12, 1.46	1.14	1.06, 1.24
DIAGRAM 2018 (N = 293 SNPs)	1.48	1.35, 1.63	1.49	1.29, 1.72	1.27	1.17, 1.38
DIAMANTE 2020 multi-ancestry composite (N = 287 SNPs)	1.45	1.32, 1.60	1.34	1.17, 1.54	1.20	1.10, 1.30
DIAMANTE 2020 multi-ancestry African (N = 276 SNPs)	1.29	1.16, 1.41	1.21	1.06, 1.38	1.13	1.04, 1.22
DIAMANTE multi-ancestry East Asian (N = 280 SNPs)	1.33	1.22, 1.46	1.19	1.04, 1.36	1.13	1.04, 1.23
DIAMANTE 2020 multi-ancestry European (N = 287 SNPs)	1.47	1.33, 1.62	1.37	1.19, 1.58	1.21	1.12, 1.32
DIAMANTE 2020 multi-ancestry Hispanic/Latino (N = 287 SNPs)	1.40	1.28, 1.55	1.32	1.15, 1.51	1.18	1.09, 1.29
DIAMANTE 2020 multi-ancestry South Asian (N = 282 SNPs)	1.32	1.20, 1.45	1.35	1.18, 1.56	1.19	1.09, 1.29
DIAMANTE 2020 population-specific weight (N = 287 SNPs)	1.32	1.20, 1.45	1.24	1.08, 1.41	1.14	1.05, 1.24
Population-specific variant (N = 287 SNPs)	1.41	1.28, 1.54	1.44	1.26, 1.64	1.17	1.09, 1.27

[b] Model in birth cohort was adjusted for clinical variables: sex, parental diabetes, and birth weight.
[c] Model in youth cohort was adjusted for clinical variables: age, sex, parental diabetes, modified BMI *z* score and HbA1c.
[d] Model in adult cohort was adjusted for clinical variables: age, sex, parental diabetes, BMI and HbA1c.

Results for Cox regressions, AUC, and NRI analyses for the multi-ancestry composite DIAMANTE 2020 T2D PS (Table 5.4), population-specific variant T2D PS (Table 5.5), and the DIAGRAM 2018 T2D PS (Table 5.6) are displayed below for the adult cohort. For the rest of Section 5.3, because the DIAGRAM 2018 T2D PS performed the best in survival analyses, results for the DIAGRAM 2018 PS, which performed best according to survival analyses with respect to the incidence of T2D in the longitudinal study, are presented: results for survival, AUC, and NRI analyses for youth and birth cohorts for this T2D PS are displayed in Table 5.6.

Table 5.4. Survival (i.e., Cox regression), AUC and NRI analyses of the multi-ancestry composite DIAMANTE 2020 T2D PS: adult cohort (N = 2,333).

	Clinical Factors[f] (AUC = 0.728)			Clinical Factors + PS (AUC = 0.732)			Clinical Factors + 2hPG (AUC = 0.760)			Clinical Factors + 2hPG + PS (AUC = 0.764)		
	HR (95% CI)	p-value	NRI[g]	HR (95% CI)	p-value	NRI	HR (95% CI)	p-value	NRI	HR (95% CI)	p-value	NRI
Age (years)	1.00 (0.992, 1.01)	0.869	-1.27×10^{-3}	1.00 (0.995, 1.01)	0.385	0.0388	1.00 (0.992, 1.01)	0.992	-0.0494	1.00 (0.994, 1.01)	0.531	0.671
Sex (F/M)	1.35 (1.14, 1.59)	5.48×10^{-4}	0.115	1.31 (1.10, 1.55)	1.83×10^{-3}	0.112	1.13 (0.953, 1.34)	0.157	0.104	1.10 (0.927, 1.31)	0.271	0.105
Mother Diabetic/NonDb	1.52 (1.21, 1.90)			1.51 (1.21, 1.89)			1.43 (1.14, 1.80)			1.42 (1.14, 1.78)		
Mother Unknown/NonDb	1.45 (1.15, 1.82)	2.75×10^{-4}	0.167	1.41 (1.12, 1.77)	3.12×10^{-4}	0.138	1.39 (1.10, 1.74)	2.60×10^{-3}	0.185	1.36 (1.08, 1.71)	2.88×10^{-3}	0.154
Father Diabetic/NonDb	1.45 (1.06, 1.96)			1.45 (1.07, 1.97)			1.38 (1.02, 1.88)			1.40 (1.03, 1.90)		
Father Unknown/NonDb	1.20 (0.921, 1.56)			1.17 (0.901, 1.52)			1.20 (0.925, 1.56)			1.18 (0.905, 1.53)		
BMI (kg/m²)	1.02 (1.01, 1.03)	1.22×10^{-4}	0.157	1.02 (1.01, 1.03)	1.33×10^{-5}	0.202	1.02 (1.01, 1.03)	2.11×10^{-5}	0.259	1.03 (1.01, 1.04)	2.94×10^{-6}	0.265
Fasting Glucose (mM)	1.99 (1.70, 2.33)	2.78×10^{-17}	0.125	1.96 (1.90, 2.78)	3.27×10^{-16}	0.280	1.44 (1.22, 1.71)	2.51×10^{-5}	0.280	1.44 (1.21, 1.71)	2.90×10^{-5}	0.125
HbA1c (%)	2.36 (1.95, 2.85)	1.18×10^{-18}	0.306	1.20 (1.10, 1.30)	1.13×10^{-17}	0.269	1.98 (1.64, 2.39)	9.64×10^{-13}	0.196	1.95 (1.61, 2.35)	3.72×10^{-12}	0.190
2-hour Glucose (mM)							1.32 (1.25, 1.39)	3.74×10^{-24}	0.381	1.31 (1.24, 1.39)	2.92×10^{-23}	0.348
Polygenic Score (SD)				1.20 (1.10, 1.30)	2.43×10^{-5}	0.251				1.17 (1.08, 1.28)	1.85×10^{-4}	0.245

[f] 'Clinical factors' for adult cohort refers to all the following factors: age, sex, parental diabetes, BMI, HbA1c and FPG.
[g] NRI values were calculated for each factor by assessing its contribution to other factors (e.g., clinical factors and/or 2hPG), as specified by column headers.

Table 5.5. Cox regressions, AUC and NRI analyses of the population-specific variant T2D PS: adult cohort (N = 2,333)

	Clinical Factors (AUC = 0.728)			Clinical Factors + PS (AUC = 0.730)			Clinical Factors + 2hPG (AUC = 0.760)			Clinical Factors + 2hPG + PS (AUC = 0 762)		
	HR (95% CI)	p-value	NRI	HR (95% CI)	p-value	NRI	HR (95% CI)	p-value	NRI	HR (95% CI)	p-value	NRI
Age (years)	1.00 (0.992, 1.01)	0.869	-1.26×10^{-3}	1.00 (0.994, 1.01)	0.543	0.0189	1.00 (0.992, 1.01)	0.992	-0.0494	1.00 (0.993, 1.01)	0.702	0.0665
Sex (F/M)	1.35 (1.14, 1.59)	5.48×10^{-4}	0.115	1.34 (1.13, 1.58)	7.51×10^{-4}	0.115	1.13 (0.953, 1.34)	0.157	0.104	1.14 (0.956, 1.35)	0.149	0.104
Mother Diabetic/NonDb	1.52 (1.21, 1.90)			1.47 (1.18, 1.84)			1.43 (1.14, 1.80)			1.40 (1.12, 1.76)		
Mother Unknown/NonDb	1.45 (1.15, 1.82)	2.75×10^{-4}	0.167	1.42 (1.13, 1.79)	3.48×10^{-3}	0.194	1.39 (1.10, 1.74)	2.60×10^{-3}	0.185	1.37 (1.09, 1.72)	0.0142	0.171
Father Diabetic/NonDb	1.45 (1.06, 1.96)			1.29 (0.947, 1.77)			1.38 (1.02, 1.88)			1.27 (0.930, 1.73)		
Father Unknown/NonDb	1.20 (0.921, 1.56)			1.12 (0.859, 1.46)			1.20 (0.925, 1.56)			1.14 (0.872, 1.48)		
BMI (kg/m²)	1.02 (1.01, 1.03)	1.22×10^{-4}	0.157	1.02 (1.01, 1.03)	3.09×10^{-5}	0.182	1.02 (1.01, 1.03)	2.11×10^{-5}	0.259	1.02 (1.01, 1.03)	5.05×10^{-6}	0.266
Fasting Glucose (mM)	1.99 (1.70, 2.33)	2.78×10^{-17}	0.286	1.97 (1.68, 2.31)	6.84×10^{-17}	0.300	1.44 (1.22, 1.71)	2.51×10^{-5}	0.110	1.45 (1.22, 1.71)	2.09×10^{-5}	0.115
HbA1c (%)	2.36 (1.95, 2.85)	1.18×10^{-18}	0.306	2.30 (1.90, 2.78)	8.57×10^{-18}	0.270	1.98 (1.64, 2.39)	3.74×10^{-24}	0.196	1.94 (1.60, 2.33)	5.31×10^{-12}	0.237
2-hour Glucose (mM)							1.32 (1.25, 1.39)	9.64×10^{-13}	0.381	1.31 (1.24, 1.39)	3.34×10^{-23}	0.374
Polygenic Score (SD)				1.17 (1.09, 1.27)	7.13×10^{-5}	0.114				1.15 (1.06, 1.25)	5.93×10^{-4}	0.116

[f] 'Clinical factors' for adult cohort refers to all the following factors: age, sex, parental diabetes, BMI, HbA1c and FPG.
[g] NRI values were calculated for each factor by assessing its contribution to other factors, as specified by column headers.

Table 5.6. Results of survival, area under the ROC curve (AUC), and net reclassification improvement (NRI) analyses for the DIAGRAM 2018 T2D PS in the adult (N = 2,333); youth (N = 2,229); and birth cohorts (N = 2,894).

	Models in Adult Cohort (age ≥ 20 years)											
	Clinical Factors[a] (AUC = 0.728)			Clinical Factors + PS (AUC = 0.735)			Clinical Factors + 2hPG (AUC = 0.760)			Clinical Factors + 2hPG + PS (AUC = 0.765)		
	HR (95% CI)	p-value	NRI	HR (95% CI)	p-value	NRI	HR (95% CI)	p-value	NRI	HR (95% CI)	p-value	NRI
Age (years)	1.00 (0.992, 1.01)	0.869	0.0118	1.00 (0.995, 1.01)	0.393	0.0672	1.00 (0.992, 1.01)	0.992	-0.0448	1.00 (0.994, 1.01)	0.540	0.112
Sex (F/M)	0.345 (1.14, 1.59)	5.48×10^{-4}	0.116	1.34 (1.13, 1.58)	7.70×10^{-4}	0.116	1.13 (0.953, 1.34)	0.157	0.110	1.13 (0.954, 1.35)	0.154	0.107
Mother Diabetic/NonDb‡	1.52 (1.21, 1.90)	2.75×10^{-4}	0.167	1.46 (1.17, 1.83)	2.08×10^{-3}	0.140	1.43 (1.14, 1.80)	2.60×10^{-3}	0.185	1.38 (1.10, 1.73)	0.0131	0.146
Mother Unknown/NonDb	1.45 (1.15, 1.82)			1.38 (1.09, 1.74)			1.39 (1.10, 1.74)			1.33 (1.06, 1.67)		
Father Diabetic/NonDb	1.45 (1.06, 1.96)			1.38 (1.01, 1.87)			1.38 (1.02, 1.88)			1.33 (0.975, 1.81)		
Father Unknown/NonDb	1.20 (0.921, 1.56)			1.17 (0.900, 1.52)			1.20 (0.93, 1.56)			1.17 (0.900, 1.52)		
BMI (kg/m²)	1.02 (1.01, 1.03)	1.22×10^{-4}	0.158	1.03 (1.01, 1.03)	1.03×10^{-5}	0.295	1.02 (1.01, 1.03)	2.11×10^{-5}	0.242	1.03 (1.02, 1.04)	1.88×10^{-6}	0.269
Fasting Glucose (mM)	1.99 (1.70, 2.33)	2.78×10^{-17}	0.303	1.95 (1.66, 2.29)	3.13×10^{-16}	0.275	1.44 (1.22, 1.71)	2.51×10^{-4}	0.176	1.44 (1.21, 1.71)	2.93×10^{-5}	0.191
HbA1c (%)	2.36 (1.95, 2.85)	1.18×10^{-18}	0.306	2.29 (1.89, 2.77)	1.14×10^{-17}	0.292	1.98 (1.64, 2.39)	9.64×10^{-13}	0.196	1.93 (1.60, 2.33)	5.67×10^{-12}	0.209
2-hour Glucose (mM)							1.32 (1.25, 1.40)	3.74×10^{-24}	0.420	1.31 (1.24, 1.38)	6.71×10^{-23}	0.426
Polygenic Score (SD)				1.27 (1.17, 1.38)	1.61×10^{-8}	0.270				1.24 (1.15, 1.35)	2.90×10^{-7}	0.251
	Models in Youth Cohort (age 5-19 years)											
	Clinical Factors (AUC = 0.805)			Clinical Factors + PS (AUC = 0.812)			Clinical Factors + 2hPG (AUC = 0.820)			Clinical Factors + 2hPG + PS (AUC = 0.825)		
	HR (95% CI)	p-value	NRI	HR (95% CI)	p-value	NRI	HR (95% CI)	p-value	NRI	HR (95% CI)	p-value	NRI
Age (decades)	1.08 (1.04, 1.12)	6.50×10^{-5}	0.313	1.08 (1.04, 1.12)	2.02×10^{-5}	0.326	1.08 (1.04, 1.12)	4.96×10^{-5}	0.335	1.09 (1.05, 1.13)	1.37×10^{-5}	0.318
Sex (F/M)	1.40 (1.06, 1.84)	0.0170	0.0459	1.32 (1.01, 1.74)	0.0458	0.046	1.09 (0.816, 1.45)	0.562	0.0262	1.04 (0.78, 1.39)	0.794	0.0241
Mother Diabetic/NonDb	2.46 (1.83, 3.29)	1.21×10^{-11}	0.524	2.27 (1.70, 3.05)	1.28×10^{-9}	0.524	2.26 (1.68, 3.04)	2.38×10^{-9}	0.524	2.10 (1.56, 2.83)	1.29×10^{-7}	0.516
Mother Unknown/NonDb	1.84 (1.27, 2.68)			1.87 (1.29, 2.72)			1.82 (1.26, 2.64)			1.83 (1.26, 2.65)		
Father Diabetic/NonDb	1.88 (1.27, 2.78)			1.72 (1.16, 2.55)			1.74 (1.18, 2.57)			1.61 (1.09, 2.39)		
Father Unknown/NonDb	1.01 (0.740, 1.365)			0.972 (0.714, 1.32)			1.01 (0.743, 1.37)			0.977 (0.719, 1.329)		
Modified BMI Z-score	1.50 (1.37, 1.66)	1.43×10^{-16}	0.577	1.53 (1.39, 1.69)	4.48×10^{-17}	0.640	1.44 (1.30, 1.59)	6.63×10^{-13}	0.560	1.47 (1.33, 1.63)	1.39×10^{-13}	0.582
Fasting Glucose (mM)	2.32 (1.66, 3.23)	7.87×10^{-7}	0.170	2.06 (1.48, 2.88)	2.10×10^{-5}	0.166	1.46 (1.01, 2.10)	0.0435	-0.0391	1.31 (0.904, 1.88)	0.155	-0.0911
HbA1c (%)	1.96 (1.36, 2.81)	2.77×10^{-4}	0.177	1.87 (1.31, 2.68)	6.29×10^{-4}	0.150	1.63 (1.15, 2.31)	6.19×10^{-3}	0.103	1.58 (1.12, 2.24)	0.00910	0.104
2-hour Glucose (mM)							1.35 (1.21, 1.49)	1.71×10^{-8}	0.333	1.34 (1.21, 1.48)	4.20×10^{-8}	0.332
Polygenic Score (SD)				1.49 (1.29, 1.72)	4.31×10^{-8}	0.249				1.48 (1.28, 1.71)	1.06×10^{-7}	0.258
	Models in Birth Cohort (birth)											
	Sex, parental diabetes (AUC = 0.597)			Sex, parental diabetes + PS (AUC = 0.683)			Sex, parental diabetes, birth weight (AUC = 0.614)			Sex, parental diabetes, birth weight + PS (AUC = 0.685)		
	HR (95% CI)	p-value	NRI	HR (95% CI)	p-value	NRI	HR (95% CI)	p-value	NRI	HR (95% CI)	p-value	NRI
Sex (F/M)	1.18 (0.974, 1.43)	0.0915	0.090	1.18 (0.972, 1.43)	0.0947	0.090	1.15 (0.951, 1.40)	0.146	0.090	1.14 (0.939, 1.38)	0.184	0.090
Mother Diabetic/NonDb	7.28 (5.00, 10.6)	6.79×10^{-30}	0.212	7.86 (5.39, 11.5)	1.12×10^{-30}	0.231	7.40 (5.08, 10.77)	4.00×10^{-30}	0.302	7.95 (5.44, 11.6)	5.20×10^{-31}	0.271
Mother Unknown/NonDb	0.871 (0.711, 1.07)			0.887 (0.724, 1.09)			0.875 (0.715, 1.07)			0.886 (0.723, 1.09)		
Father Diabetic/NonDb	3.93 (2.36, 6.54)			3.98 (2.39, 6.64)			3.98 (2.38, 6.65)			4.05 (2.42, 6.78)		
Father Unknown/NonDb	1.01 (0.818, 1.26)			1.07 (0.863, 1.33)			0.977 (0.787, 1.21)			1.03 (0.829, 1.28)		
Low Birth weight (< 3000g)							1.56 (1.24, 1.96)	6.11×10^{-4}	0.164	1.16 (0.877, 1.53)	5.99×10^{-4}	0.164
High Birth weight (≥ 4000g)							1.17 (0.887, 1.55)			1.57 (1.25, 1.97)		
Polygenic Score (SD)				1.48 (1.35, 1.63)	2.83×10^{-16}	0.309				1.48 (1.35, 1.63)	2.77×10^{-16}	0.309

[a] 'Clinical factors' refers to: age, sex, parental diabetes, BMI, and FPG for the adult cohort; and age, sex, parental diabetes, modified BMI *z* score, and FPG for the youth cohort.

‡ 'NonDb' stands for not having documented diabetes, regarding parental diabetes for offspring who are study participants.

5.3.2. Association of the DIAGRAM 2018 T2D PS and clinical predictors with the incidence of T2D

The best-performing PS—the DIAGRAM 2018 T2D PS—was significantly associated with T2D incidence in adult, youth, and birth cohorts. In the adult cohort, ten-year cumulative incidence of T2D in the lowest decile of PS was 20.5%; in the highest, 42.5%. The HR for the DIAGRAM 2018 PS in a survival analysis without adjustment for any covariates was 1.31 per SD T2D PS ($p = 6.9 \times 10^{-11}$) (Figure 5.1).

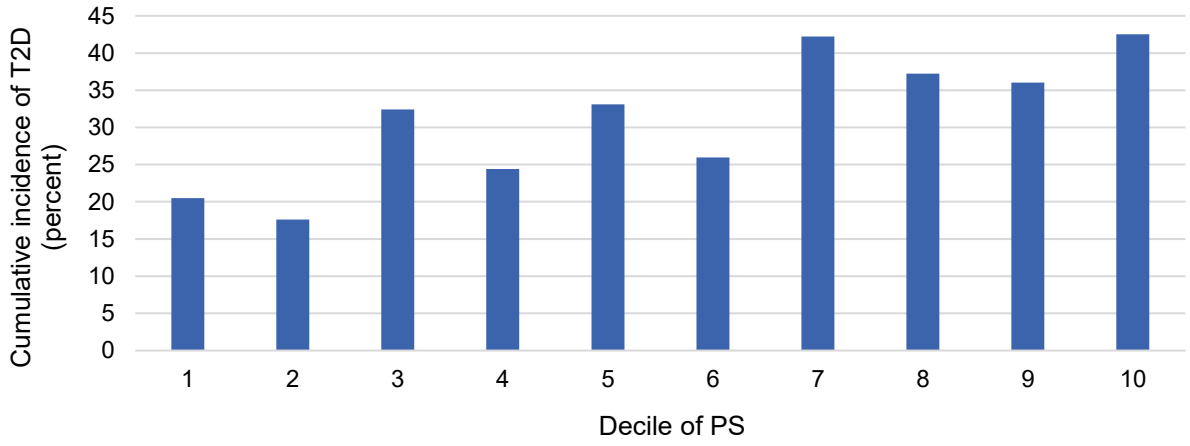


Figure 5.1. Cumulative incidence of T2D over 10 years of follow-up, stratified by decile of the DIAGRAM 2018 T2D PS, in the adult cohort (N = 2,333).

At 10 years follow-up, 504 individuals had developed T2D and 635 remained at risk.

In the adult cohort, the clinical predictors—age, sex, BMI, HbA1c, and FPG—were also strongly associated with the incidence of T2D. The 10-year cumulative incidence of T2D was 7.8% in the lowest decile of the linear predictor of clinical variables; in the highest, 68.7% (Figure 5.2).

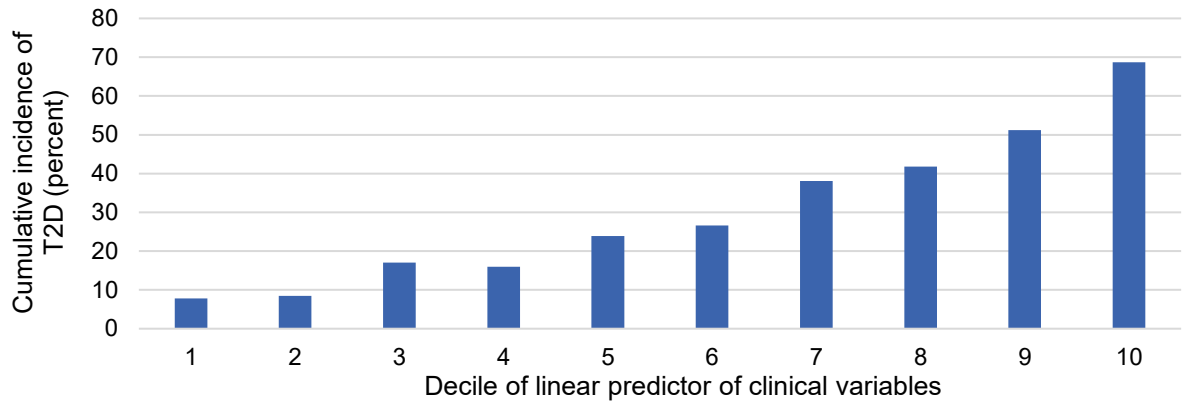


Figure 5.2. Cumulative incidence of T2D over 10 years of follow-up, stratified by decile of linear predictor of clinical variables (age, sex, parental diabetes, BMI, FPG, HbA1c), in the adult cohort (N = 2,333).

At 10 years follow-up, 504 individuals had developed T2D and 635 remained at risk.

In the youth cohort, ten-year cumulative incidence of T2D in the lowest decile of the DIAGRAM 2018 T2D PS was 2.4%; in the highest, 21.5% (unadjusted HR = 1.59 per SD T2D PS, $p = 6.8 \times 10^{-12}$) (Figure 5.3).

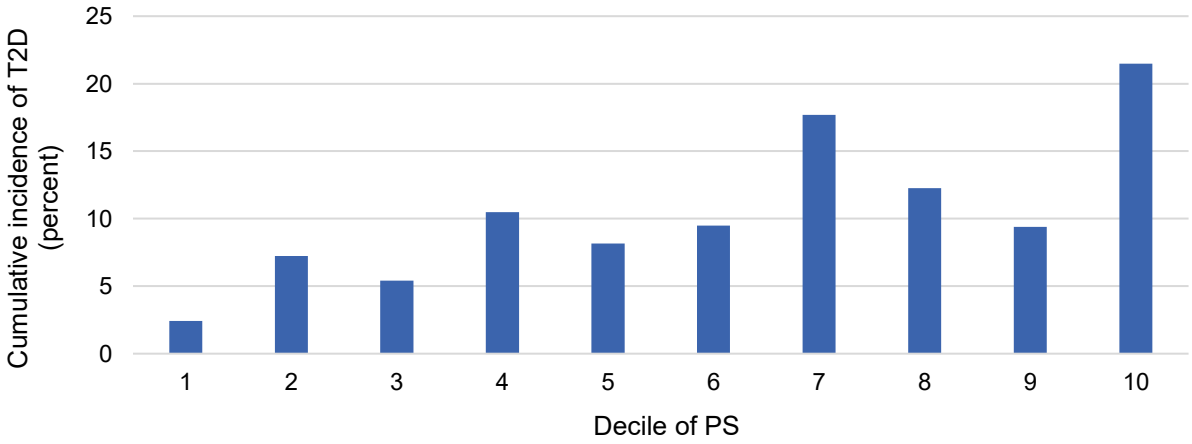


Figure 5.3. Cumulative incidence of T2D over 10 years of follow-up, stratified by decile of the DIAGRAM 2018 PS, in the youth cohort (N = 2,229).

At 10 years follow-up, 152 had developed T2D and 745 remained at risk.

In the youth cohort, the 10-year cumulative incidence of T2D was 1.9% in the lowest decile of the linear predictor of clinical variables; in the highest, 39.5% (Figure 5.4).

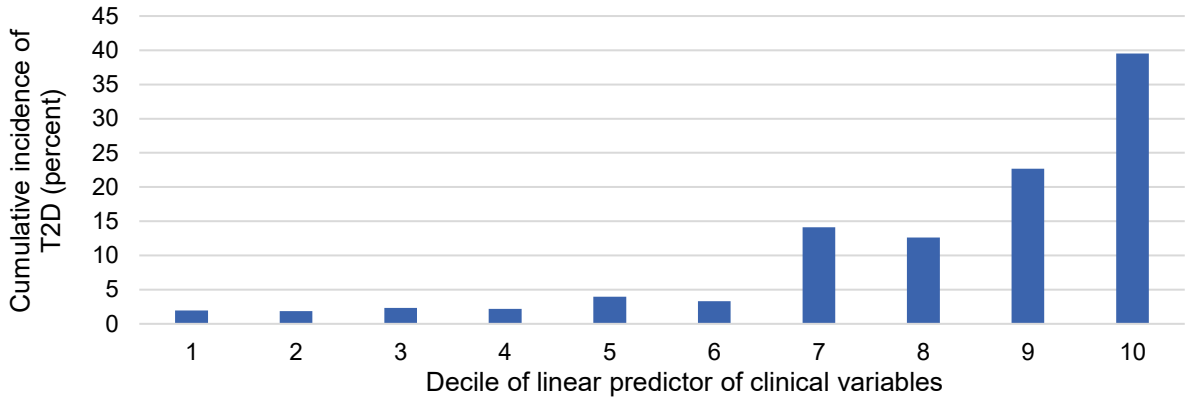


Figure 5.4. Cumulative incidence of T2D over 10 years of follow-up, stratified by decile of linear predictor of clinical variables (age, sex, parental diabetes, modified BMI *z*, FPG, HbA1c), in the youth cohort (N = 2,229).

At 10 years follow-up, 152 had developed T2D and 745 remained at risk.

In the birth cohort, the 30-year cumulative incidence of T2D in the lowest decile of the DIAGRAM 2018 T2D PS was 15.1%; in the highest, 37.3% (unadjusted HR = 1.47 per SD T2D PS, $p = 1.7 \times 10^{-15}$) (Figure 5.5).

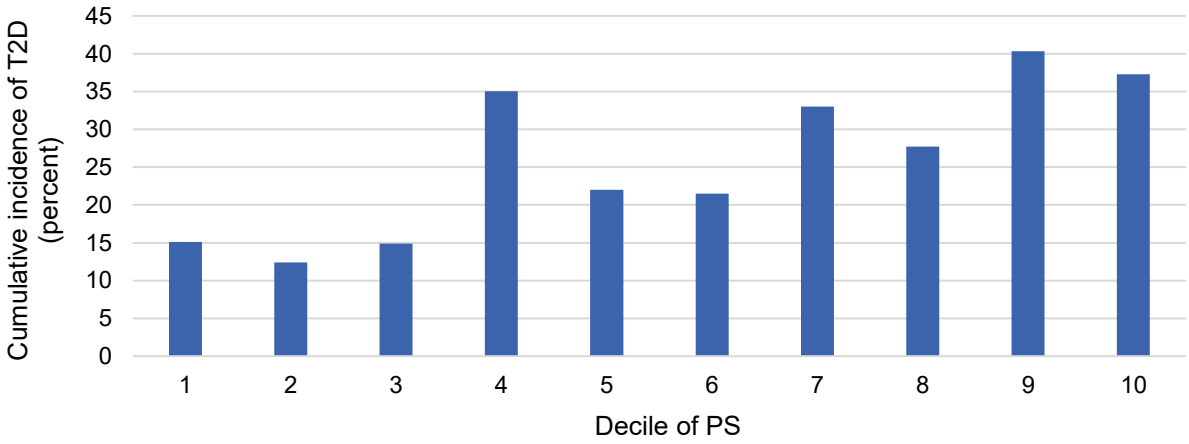


Figure 5.5. Cumulative incidence of T2D over 30 years of follow-up, stratified by decile of the DIAGRAM 2018 PS, in the birth cohort (N = 2,894).

At 30 years follow-up, 340 had developed T2D and 474 remained at risk.

In the birth cohort, the 30-year cumulative incidence of T2D was 21.0% in the lowest decile of the linear predictor of clinical variables; in the highest, 54.0% (Figure 5.6).

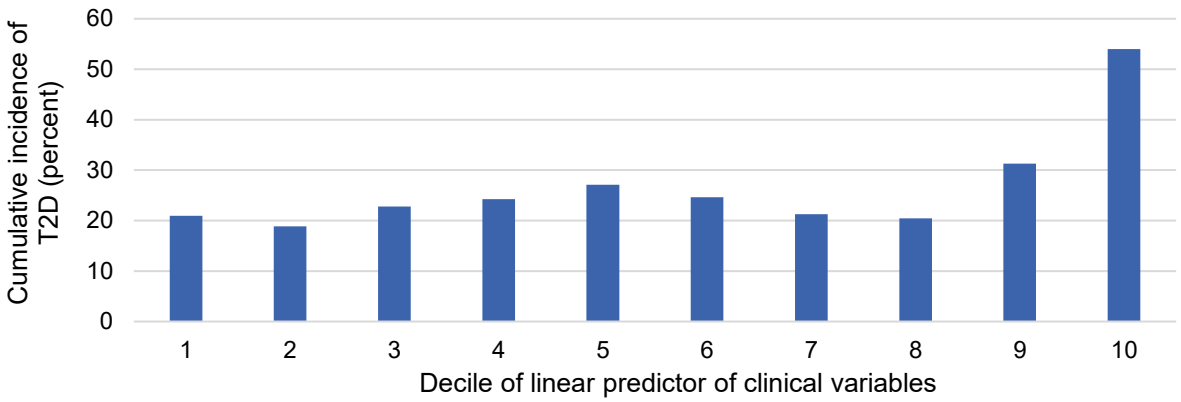


Figure 5.6. Cumulative incidence of T2D over 30 years of follow-up, stratified by decile of linear predictor of clinical variables (sex, parental diabetes, birth weight), in the birth cohort (N = 2,894).

At 30 years follow-up, 340 had developed T2D and 474 remained at risk.

5.3.3. Survival analyses with adjustment for clinical factors

Survival analyses were conducted to assess associations between individual clinical variables and the PS with T2D incidence. In the adult cohort, in the model with clinical variables, the HR of the PS was 1.27 per SD (p = 1.6 × 10^{-8}; 95% CI 1.17, 1.38) (Table 5.6). In the youth cohort, in the model with clinical variables, the HR of the PS was 1.49 (p = 4.3 × 10^{-8}; 95% CI 1.29, 1.72). In the birth cohort, in the model with clinical variables, the HR of the PS was 1.48 (p = 2.8 × 10^{-16}); 95% CI 1.35, 1.63). Adding 2hPG to adult and youth cohorts' models did not substantially alter the HRs of the PS.

5.3.4. AUC analyses

AUC analyses were conducted to evaluate predictive accuracy of models containing combinations of clinical variables and the PS. Their results are described below in-text, as well as in Table 5.6. In the adult cohort, AUC for the model with age and sex was 0.590; with the PS, 0.619 (difference in AUC, i.e., ΔAUC, 0.029; 95% CI 0.010, 0.048; p = 0.003). In the youth cohort, corresponding AUCs were 0.625 and 0.682 (ΔAUC 0.057; 95% CI 0.026, 0.089; $p = 3.96 \times 10^{-4}$). In the birth cohort, AUC for the model with sex (age was not included as a covariate as this birth cohort included only predictors that would be available at one's birth) was 0.537; with the PS, 0.638 (ΔAUC = 0.101; 95% CI 0.068, 0.135; $p < 10^{-5}$).

Though the PS was strongly associated with incident T2D, the improvement in AUC compared with clinical variables alone was modest (Table 5.6). In the adult cohort, the AUC for the model with clinical variables was 0.728; with the PS, 0.735 (ΔAUC = 0.007; 95% CI 0.001, 0.014; $p = 0.023$). In the youth cohort, AUC for the model with clinical variables was 0.805; with the PS, 0.812 (ΔAUC = 0.007; 95% CI -0.003, 0.017; $p = 0.173$). For the birth cohort, the increment in AUC with addition of the PS was greater: the AUC for the model including sex and parental diabetes was 0.597; with the PS, 0.683 (ΔAUC = 0.086; 95% CI 0.057, 0.115; $p < 10^{-5}$).

The improvement in AUC compared with clinical variables and 2hPG, was also modest (Table 5.6). In the adult cohort, the AUC for the model with clinical variables and 2hPG, was 0.760; with the PS, 0.764. (ΔAUC = 0.004; 95% CI -0.001, 0.009; $p = 0.090$). In the youth cohort, AUC for the model with clinical variables and 2hPG, was 0.820; with the PS, 0.825 (ΔAUC = 0.005; 95% CI -0.003, 0.014; $p = 0.239$). For the birth cohort, the increment in AUC with addition of the PS was greater. The AUC for the model including sex, parental diabetes, and birth weight was 0.614; with the PS, 0.685 (ΔAUC = 0.072; 95% CI 0.045, 0.099; $p < 10^{-5}$).

5.3.5. Scaled value of the T2D PS, percentile of the clinical linear predictor, and predicted cumulative incidence of T2D

While AUC provides a measure of overall predictive accuracy, it does not fully capture the extent to which addition of a variable can affect individual risk estimates. To examine this, I calculated predicted cumulative incidence of T2D according to PS for various levels of clinical risk. Across all cohorts, greater T2D PS and greater percentiles of clinical linear predictor were both directly associated with predicted cumulative incidence of T2D (Figures 5.7, 5.8, 5.9).

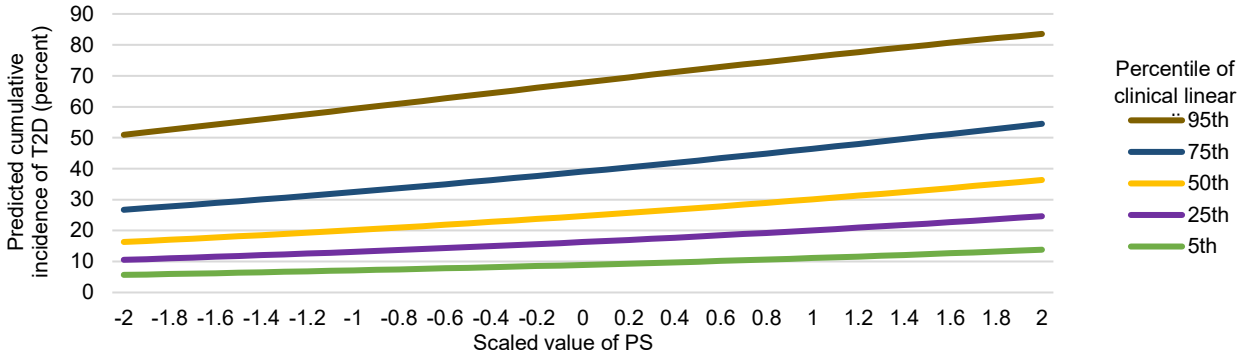


Figure 5.7. Predicted cumulative incidence of T2D over 10 years of follow-up in adult cohort (N = 2,333) for the scaled DIAGRAM 2018 T2D PS and specified percentiles of the clinical linear predictor: clinical variables include age, sex, parental diabetes, BMI, FPG, and HbA1c.

At 10 years follow-up, 504 individuals had developed T2D; 635 remained at risk.

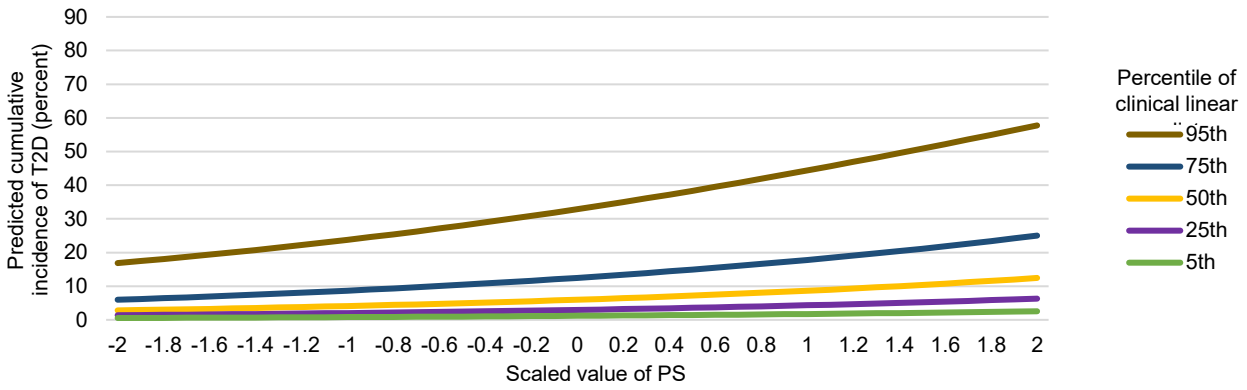


Figure 5.8. Predicted cumulative incidence of T2D over 10 years of follow-up in youth cohort (N = 2,229) for the scaled DIAGRAM 2018 T2D PS and specified percentiles of the clinical linear predictor: clinical variables include age, sex, parental diabetes, modified BMI *z*, FPG, and HbA1c.

At 10 years follow-up, 152 had developed T2D; 745 remained at risk.

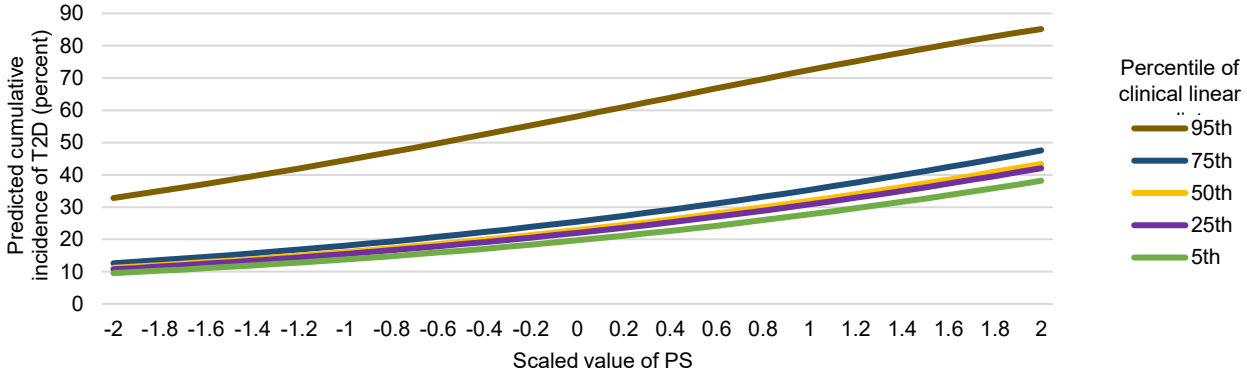


Figure 5.9. Predicted cumulative incidence of T2D over 30 years of follow-up in birth cohort (N = 2,894) for the scaled DIAGRAM 2018 T2D PS and specified percentiles of the clinical linear predictor: clinical variables include sex, parental diabetes, and birth weight.

At 30 years follow-up, 340 developed T2D; 474 remained at risk.

To further quantify the contribution of each variable to the model's risk classification, I calculated the NRI of each variable (Table 5.6). NRI quantifies the extent to which T2D cases and non-cases are consequently reclassified upon inclusion of an additional variable. The NRI for adding the PS to clinical variables was 0.264 in the adult cohort; in the youth cohort,

0.249; in the birth cohort, 0.309. In comparison, NRI for HbA1c was 0.292 in the adult cohort and 0.150 in the youth cohort.

5.3.6. Age group comparison analyses

Bootstrapping was performed to estimate pairwise differences among cohorts in results of survival, AUC, and NRI analyses by rerunning models with the DIAGRAM 2018 T2D PS. To avoid differences that may arise from the use or availability of different clinical covariates across cohorts, I first ran these analyses without including any clinical predictors. Results of survival, AUC and NRI analyses for the adult, youth and birth cohorts are shown in Table 5.7.

Table 5.7. Bootstrapping analyses: results of survival, AUC, and NRI analyses for the DIAGRAM 2018 T2D PS alone, across age groups (N = 4,770).

Cohort	log(HR) (per SD PS)	AUC	NRI for PS
Adult	0.239	0.007	0.270
Youth	0.399	0.007	0.268
Birth	0.395	0.072	0.362

The AUC for the baseline model does not contain the T2D PS; the ΔAUC describes the difference in AUC between the model with versus without the T2D PS. The continuous NRI was computed for the T2D PS.

Differences in results of survival, AUC, and NRI analyses for individuals included in the adult, youth, and birth cohorts are summarized below (Tables 5.8, 5.9, 5.10, 5.11). Values were calculated for adult, youth, and birth cohorts for one model with only the PS (i.e., without adjustment for clinical variables) and one model with the PS and adjustment for clinical covariates. Pairwise differences (diff) for each of these results were calculated in the

following format, with negative values possible because '1' signifies 'cohort 1' and '2' signifies 'cohort 2' (Equations 5.1, 5.2, 5.3, 5.4).

$$diff_{logHR12} = \log HR_1 - \log HR_2 \quad (5.1)$$

$$diff_{AUC12} = AUC_1 - AUC_2 \quad (5.2)$$

$$diff_{\Delta AUC12} = \Delta AUC_1 - \Delta AUC_2 \quad (5.3)$$

$$diff_{NRI12} = NRI_1 - NRI_2 \quad (5.4)$$

For the purposes of discussing the results of these bootstrapping analyses, $p < 0.05$ was taken to be a level of nominal significance. As shown in Table 5.9, the difference in hazard ratios for the DIAGRAM 2018 T2D PS between youth and adult cohorts was relatively great ($diff_{logHR12} = 0.192$, SE = 0.072) and significant ($p = 0.006$). The difference in hazard ratios for the DIAGRAM 2018 T2D PS between birth and adult cohorts was also significant ($diff_{logHR12} = 0.115$, SE = 0.055, $p = 0.037$).

Table 5.8. Bootstrapping analyses for DIAGRAM 2018 T2D PS for models without adjustment for clinical variables: logHR and differences (N = 4,770).

Cohort 1	Cohort 2	logHR1	logHR2	$diff_{logHR12}$ (SE)	p-value for $diff_{logHR12}$
Youth	Adult	0.465	0.269	0.196 (0.072)	0.006
Youth	Birth	0.465	0.384	0.081 (0.056)	0.148
Birth	Adult	0.384	0.269	0.115 (0.055)	0.037

Table 5.9. Bootstrapping analyses for DIAGRAM 2018 T2D PS for models without adjustment for clinical covariates: AUC and differences (N = 4,770).

Cohort 1	Cohort 2	AUC1	AUC2
Youth	Adult	0.625	0.582
Youth	Birth	0.625	0.632
Birth	Adult	0.632	0.582

Table 5.10. Bootstrapping analyses for the DIAGRAM 2018 T2D PS for models adjustment for clinical variables: ΔAUC and differences (N = 4,770).

Cohort 1	Cohort 2	ΔAUC1	ΔAUC2	$diff_{\Delta AUC12}$ (SE)	p-value for $diff_{\Delta AUC12}$
Youth	Adult	0.125	0.082	0.043 (0.023)	0.061
Youth	Birth	0.125	0.132	-0.007 (0.017)	0.658
Birth	Adult	0.132	0.082	0.050 (0.019)	0.007

Taken together, the results of Tables 5.10 and 5.11 display the area under the ROC curve for the DIAGRAM 2018 T2D PS in each of the cohorts (without adjustment for clinical variables), along with the delta AUC (ΔAUC) values that denote the increase in AUC with the addition of the DIAGRAM 2018 T2D PS to clinical variables for each of the cohorts. The pairwise differences of ΔAUC between cohorts were also calculated, and that of the birth and adult cohort was significant ($diff_{\Delta AUC12}$ = 0.050, SE = 0.019, p = 0.007).

The NRI values for the DIAGRAM 2018 T2D PS were calculated for each cohort, with adjustment for clinical variables. The pairwise differences in NRI for the DIAGRAM 2018 T2D PS were calculated between cohorts.

Table 5.11. Bootstrapping analyses for the DIAGRAM 2018 T2D PS for models with adjustment for clinical variables: NRI and differences (N = 4,770).

Cohort 1	Cohort 2	NRI1	NRI2	$diff_{NRI12}$ (SE)	p-value for $diff_{NRI12}$
Youth	Adult	0.397	0.266	0.131 (0.100)	0.188
Youth	Birth	0.397	0.368	0.029 (0.093)	
Birth	Adult	0.368	0.266	0.103 (0.079)	

However, the p-values for the comparison of the NRI of the birth cohort with others were not calculated because the NRI was calculated at 30 years follow-up versus 10 years follow-up for the other cohorts. The difference in NRI values for the DIAGRAM 2018 T2D PS in the youth and adult cohorts was not found to be significant in this analysis.

5.3.7. Additional genotypic analyses

The effects of some variants strongly associated with T2D in this Indigenous study population were not captured in the DIAGRAM 2018 PS. To address this, I assessed the contribution of genotypes for *KCNQ1* rs2237895 and *ABCC8* rs1272388614 in the adult cohort. For each genotype, associations were significant ($p < 0.05$), and they contributed modestly to the model of clinical factors and the PS, as assessed by AUC and NRI analyses (Table 5.12, 5.13, 5.14).

Table 5.12. AUC analyses for combinations of clinical variables, *ABCC8* rs1272388614 (R1420H) genotype, and the DIAGRAM 2018 T2D PS, in the adult cohort (N = 2,333).

Parameters in model	AUC
age, sex, *ABCC8* genotype	0.590
age, sex, *ABCC8* genotype, PS	0.620
age, sex, *ABCC8* genotype, parental diabetes, FPG, HbA1c	0.728
age, sex, ABCC8 genotype, parental diabetes, FPG, HbA1c, PS	0.735
age, sex, *ABCC8* genotype, parental diabetes, FPG, 2hPG, HbA1c	0.760
age, sex, *ABCC8* genotype, parental diabetes, FPG, 2hPG, HbA1c, PS	0.765

Table 5.13. AUC analyses for combinations of clinical variables, *KCNQ1* rs2237895 genotype, and the DIAGRAM 2018 T2D PS, in the adult cohort (N = 2,333).

Parameters in model	AUC
age, sex, maternally-derived *KCNQ1* risk allele, paternally-derived *KCNQ1* risk allele	0.598
age, sex, maternally-derived *KCNQ1* risk allele, paternally-derived *KCNQ1* risk allele, PS	0.619
age, sex, maternally-derived *KCNQ1* risk allele, paternally-derived *KCNQ1* risk allele, parental diabetes, FPG, HbA1c	0.723
age, sex, maternally-derived *KCNQ1* risk allele, paternally-derived *KCNQ1* risk allele, parental diabetes, FPG, HbA1c, PS	0.727
age, sex, maternally-derived *KCNQ1* risk allele, paternally-derived *KCNQ1* risk allele, parental diabetes, FPG, 2hPG, HbA1c	0.756
age, sex, maternally-derived *KCNQ1* risk allele, paternally-derived *KCNQ1* risk allele, parental diabetes, FPG, 2hPG, HbA1c, PS	0.758

Table 5.14. AUC analyses for combinations of clinical variables, *ABCC8* rs1272399614 (R1420H), *KCNQ1* rs2237895 genotypes, and the DIAGRAM 2018 T2D PS, in the adult cohort (N = 2,333).

Parameters in model	AUC
age, sex, *ABCC8* genotype, maternally-derived *KCNQ1* risk allele, paternally-derived *KCNQ1* risk allele	0.598
age, sex, *ABCC8* genotype, maternally-derived *KCNQ1* risk allele, paternally-derived *KCNQ1* risk allele, PS	0.619
age, sex, *ABCC8* genotype, maternally-derived *KCNQ1* risk allele, paternally-derived *KCNQ1* risk allele, parental diabetes, FPG, HbA1c	0.730
age, sex, *ABCC8* genotype, maternally-derived *KCNQ1* risk allele, paternally-derived *KCNQ1* risk allele, parental diabetes, FPG, HbA1c, PS	0.735
age, sex, *ABCC8* genotype, maternally-derived *KCNQ1* risk allele, paternally-derived *KCNQ1* risk allele, parental diabetes, FPG, 2hPG, HbA1c	0.763
age, sex, *ABCC8* genotype, maternally-derived *KCNQ1* risk allele, paternally-derived *KCNQ1* risk allele, parental diabetes, FPG, 2hPG, HbA1c, PS[f]	0.765

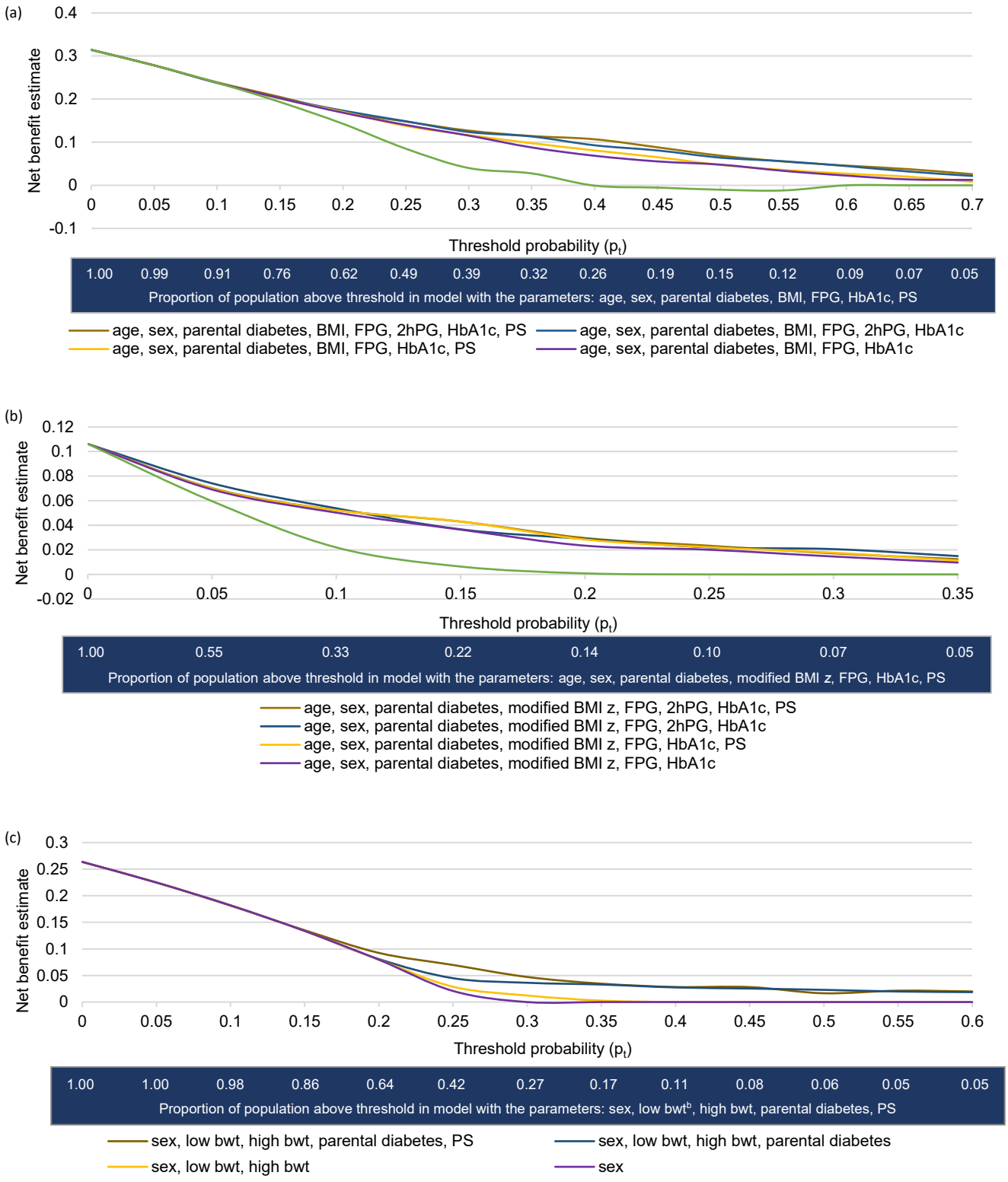
[f] NRI values for each genetic factor in the bottommost row of Table S3c: 0.021 for *ABCC8* genotype, 0.148 for the maternally-derived and paternally-derived *KCNQ1* risk alleles, and 0.250 for the PS.

The inclusion of the single *ABCC8* rs1272399614 *and KCNQ1* rs2237895 genotypes to the other parameters in each model was not found to contribute substantially to the overall AUC.

5.3.8. Decision curve analyses

Decision curve analyses were undertaken to estimate the net benefit of including the PS at a range of threshold probabilities (i.e., minimum probabilities of disease that would warrant intervention). When the costs of false positives are low (i.e., as p_t approaches 0), population-wide interventions are favoured; thus, screening by clinical or genetic means would have little net benefit. When false-positive costs are higher (i.e., higher p_t values), net clinical benefit can be increased by screening to target the intervention to higher-risk individuals.

In the adult cohort, the net benefit of including the PS in addition to clinical factors was most pronounced at p_t values 0.3-0.5 (up to 18% improvement); this corresponded to 15-40% of the highest-risk individuals selected for the intervention (Figure 5.10). In the youth cohort, the net benefit of including the PS was most pronounced at p_t values 0.05-0.35 (up to 21% improvement) (Figure 5.11). In the birth cohort, the net benefit of including the PS was most pronounced at p_t values 0.15-0.35 (up to 56% improvement) (Figures 5.10).



Figures 5.10. Net benefit of predictive models with or without the DIAGRAM 2018 PS for T2D prediction.
(a) Over 10 year follow-up time, in the adult cohort.
(b) Over 10-year follow-up time, in the youth cohort.
(c) Over 30-year follow-up time, in the birth cohort.
[b] 'Bwt' stands for birth weight.

5.4. Summary

Polygenic scores potentially have utility for identification of individuals with higher risk of T2D. Previous studies generally reported significant associations between T2D PS and T2D incidence and modest prediction improvement as measured by AUC: ΔAUC from 0.005 to 0.02 (Meigs et al., 2008; Lyssenko et al., 2008; Mars et al., 2020; Vaxillaire et al., 2014; Park et al., 2015; He et al., 2021). A limited number of studies include measures of reclassification: continuous net reclassification improvement (NRIs) ranged from 0.044 to 0.285. In the present study, the DIAGRAM 2018 PS was strongly statistically significant in predicting T2D incidence in adult, youth, and birth cohorts. Results of AUC analyses are consistent with findings of previous studies: improvement in prediction contributed by the T2D PS was statistically significant but modest. However, ΔAUC does not fully capture the contribution of a single variable to individual risk (Pencina et al., 2012). I calculated NRIs for individual variables to address this limitation. NRIs for the PS across all cohorts ranged from 0.2 to 0.3, which is considered intermediate power for identifying T2D risk, and were comparable to those of commonly measured clinical variables (e.g., HbA1c and FPG).

5.4.1. Construction of PS

While all T2D PS were significantly associated with T2D incidence across all cohorts, the DIAGRAM 2018 PS, derived from European-ancestry populations, performed slightly better than the others. While collaborators at NIDDK PECRB have previously shown modest heterogeneity in effects of established T2D variants between Europeans and this study population (Hanson et al., 2015), the DIAGRAM 2018 PS even out-performed a population-specific variant PS with a comparable number of variants, derived by two-fold cross-validation in the present population (N ≈ 3850). The expectation is that a PS derived from a GWAS in a more closely-matched ancestry group would perform better than one from a different ancestry

group, if sample sizes are equal (Conrad et al., 2006). Further research could focus upon the optimization of PS that are derived from and used within a given study population or derived from and tested within closely-matched study populations.

5.4.2. Optimal age for preventive interventions

Genetic effects of the DIAGRAM 2018 T2D PS with respect to T2D incidence were greatest in the youth and birth cohorts. This is consistent with the hypothesis that genetic effects for many chronic diseases are strongest earlier in life (Jiang et al., 2021), and consistent with the finding that familial recurrence risk of diabetes in this population is higher when it occurs at younger ages (Hanson & Knowler, 1998). The present findings could also reflect the limited availability of phenotypic data for study participants at birth or young ages. However, when analysed without any clinical covariates, HRs associated for the birth cohort and the youth cohort were higher than that for the adult cohort; tests for differences in the HRs between the adult and youth cohorts and adult and birth cohorts yielded p = 0.037 and p=0.006, respectively; differences between birth and youth cohorts were not significant (p=0.15).

The improvements in AUC and net benefit upon adding the PS to clinical variables were greatest in the birth cohort. The use of type 2 diabetes PS at birth could be particularly beneficial as phenotypic manifestations of risk (e.g., hyperglycaemia and obesity) are less apparent. Specifically with respect to the birth cohort, the use of T2D PS at birth could be particularly beneficial as phenotypic manifestations of risk (e.g., hyperglycaemia and obesity) are often less apparent. Without adjustment for demographic and clinical predictors, the T2D PS had substantially higher HR, AUC and NRI values for youth and birth cohorts compared with adult cohorts, suggesting the greater predictive utility of the PS at earlier ages. However, when clinical predictors were included in prediction models, differences in results of survival, AUC and NRI analyses were attenuated.

6. General discussion

The scope of this thesis included the use of genome-wide association studies, polygenic scores (PS) and partitioned/process-specific polygenic scores (pPS) to study the genetic epidemiology of birth weight and T2D and the complex relationship between the two traits. The main aim of this thesis was to investigate genetic and non-genetic influences on birth weight and type 2 diabetes, using methods in genetic and clinical epidemiology. Analyses were conducted within a cohort study of participants from an Indigenous community in the Southwestern United States to accomplish three main objectives. First, genetic variants that were associated with birth weight using seven different models of GWAS were identified, and the genetic relationship between birth weight and T2D was evaluated using PS for birth weight and T2D. Second, the associations between pPS for T2D with T2D and related traits were assessed. Third, the contribution of clinical variables in addition to PS for T2D to the prediction of the incidence of T2D in various age groups was evaluated within the longitudinal PECRB study. The following discussion sections discuss the findings, strengths, and limitations of the corresponding chapters.

6.1. Genetics of birth weight and relationship between birth weight and T2D

6.1.1. GWAS of own and offspring birth weight

Seven different models of GWAS of birth weight were conducted, using various combinations of covariates, imputed or directly genotyped variants, and maternal and foetal genotypes:

(1) Additive foetal GWAS of own birth weight across imputed variants, adjusted for birth year and the first five genetic principal components (PCs 1-5);

(2) Additive foetal GWAS of own birth weight across imputed variants, adjusted for birth year, PCs 1-5, and *in utero* exposure to maternal diabetes;

(3) Additive foetal GWAS of own birth weight across imputed variants, adjusted for birth year, PCs 1-5, and gestational age (by multiple imputation);

(4) Dominant-recessive foetal GWAS of own birth weight across directly genotyped variants, adjusted for birth year, PCs 1-5, and gestational age (by multiple imputation);

(5) Additive foetal GWAS of own birth weight across directly genotyped variants, adjusted for birth year, PCs 1-5, and gestational age (without multiple imputation, i.e., complete-case analysis);

(6) Additive foetal GWAS of own birth weight across directly genotyped variants, adjusted for own birth year, own PCs 1-5, and conditioned on maternal genotype; and

(7) Additive maternal GWAS of offspring birth weight across directly genotyped variants, adjusted for offspring birth year, offspring PCs 1-5, and conditioned on foetal genotype.

None of the GWAS models yielded genome-wide significant ($p < 5.0 \times 10^{-8}$) associations; a threshold of 'suggestive significance' was defined for the purposes of these analysis at $p < 5.0 \times 10^{-4}$ to maximise the proportion of SNPs that were replicated in the Early Growth Consortium 2019 GWAS meta-analysis of birth weight for given *p*-value thresholds (Warrington et al., 2019). A substantial number of SNPs met that threshold in each of the seven models. However, none of the QQ plots suggest substantial excess of significant signals; most adhere closely to null expectations. A high degree of replication in the EGG consortium GWAS suggests that some of the suggestive signals are true positives. However, it is difficult to know which SNPs are true positives without further replication data.

Results were similar across Models 1, 2, and 3; notably, in Model 3, which was adjusted for gestational age (by multiple imputation), the *SUR1/ABCC8* R1420H variant identified previously by Baier et al as associated with birth weight (β = 169.1g per copy of the H allele, $p = 1.5 \times 10^{-3}$) and T2D (OR 2.02, 95% CI 1.45-2.82; $p = 3.6 \times 10^{-5}$) had the greatest significance of association with birth weight among imputed variants. Other variants that appeared across GWAS Models 1, 2, and 3, as well as multiple other models that were conducted across directly genotyped variants, were a variant in *UBE4B*, as well as one on chromosome 19.

GWAS Model 4, the dominant-recessive foetal GWAS of own birth weight adjusted for birth year, PCs 1-5, and gestational age (with multiple imputation) was conducted in a sample of the same size as that of Model 3 (N = 3,700). In Model 4, the aforementioned R1420H variant in *SUR1/ABCC8* was also the SNP with the greatest significance of association with own birth weight, and the variant in *UBE4B* also appeared among the associations of greatest significance. GWAS Model 5, the additive foetal GWAS of own birth weight adjusted for birth year, PCs 1-5, and gestational age (without multiple imputation) was conducted in a slightly smaller sample because approximately 70% of individuals had gestational age data available. It functioned as a complete-case analysis relative to GWAS Model 3, which included gestational age (by multiple imputation) as a covariate. The correlation between the beta coefficients of Models 5 and 3 was strong and positive: the Pearson correlation coefficient, r, was 0.88 ($p < 10^{-4}$). In GWAS Model 5, the variant in *UBE4B* also appeared among the associations of greatest significance.

I also conducted GWAS Models 6 and 7, a foetal GWAS of own birth weight and a maternal GWAS of offspring birth weight, respectively, because previous studies have shown that maternal and foetal genetic effects with respect to birth weight can act in opposite directions (Tyrrell et al., 2013; Warrington et al., 2019). GWAS Model 6 was an additive foetal GWAS of

own birth weight adjusted for own birth year and own PCs 1-5 and conditioned on maternal genotype. GWAS Model 7 was an additive maternal GWAS of offspring birth weight adjusted for offspring birth year and offspring PCs 1-5 and conditioned on foetal genotype. This analysis did not identify SNPs that were associated with birth weight at a level of genome-wide significance. Further, substantial genomic inflation was exhibited in GWAS Model 7. Future analyses could explore the inclusion of maternal genetic principal components as covariates and adding an additional variance component for a shared maternal effect to address this inflation.

Taken together, the results of GWAS Models 1 through 7 did not include evidence of genome-wide significant associations between genetic variants and own and offspring birth weights. The R1420H variant in *ABCC8* was the top signal in GWAS Models 3 and 4, which supports the findings of previous analyses in this population that identified this as a potential real association with birth weight. While a great degree of genotypic (genotypes, genetic PCs, and population substructure) and phenotypic data (birth weight, likelihood of *in utero* exposure to maternal diabetes, birth year) were available for its members, the small subsample of the PECRB cohort that had birth weight and genotypic data available limited the statistical power of these GWAS to detect significant associations with birth weight. This subsample of the PECRB cohort—3,700 PECRB study participants who were singleton births and had genotypic data available—consisted of over 1,000 additional participants who had entered the study after the list of participants in the subsample had last been updated by PECRB colleagues (Lindsay et al., 2000). Given that the longitudinal study ended in 2007, this subsample comprises the largest possible study population within the PECRB cohort for GWAS of birth weight.

6.1.2. Bivariate genetic correlation analyses

Bivariate genetic correlation analyses were undertaken to assess the proportion of variance shared due to genetic causes between birth weight and T2D and glycaemic and anthropometric traits. These traits included T2D, FPG, 2hPG, fasting insulin, height, HOMA-β, HOMA-IR, log(M-value), log(Maximum BMI), and log(Acute insulin response). Pairwise environmental correlation estimates between birth weight and these traits were relatively strong, as expected. However, genetic correlations, with some exceptions, were rather modest. Maximum BMI had the highest positive genetic correlation with respect to birth weight; this appeared to be a significant correlation. Notably, T2D had an inverse phenotypic correlation and inverse environmental correlation with birth weight; however, the *p*-value for the test of whether the genetic correlation between the traits was nonzero was not significant. The results of these analyses are difficult to parse and prompt future analyses to address their limitations. The relationship between birth weight and T2D risk in the PECRB cohort has been shown to be nonlinear: it is possible that the genetic relationships also follow a quadratic relationship, which is difficult to model as a conventional linear genetic correlation.

While genetic correlation analyses in other populations have supported the role of common genetic effects in the association between low birth weight and T2D, such analyses assume a linear relationship between the two phenotypes. Additionally, maternal and foetal genotypes correlate, such that shared genetic effects may not be directly attributable to foetal genotype, but rather to the intrauterine environment (which is influenced by maternal genotype) to which the foetus has been exposed. Eaves et al developed M-GCTA (maternal effects genome-wide complex trait analysis), an extension of the program GCTA, which can estimate the relative direct effects of foetal and maternal genetic effects on complex traits (Eaves et al., 2014). A study that used M-GCTA to analyse the influence of maternal and foetal genotype on birth weight in the Avon Longitudinal Study of Parents and Children demonstrated that foetal genotype contributes far more than maternal genotype to variance in birth weight

(Horikoshi et al., 2016). However, the relative contributions of foetal and maternal genotype to the shared covariance in birth weight and T2D remain to be quantified. Thus, genetic correlation analyses that account for non-linear relationships between traits, such as a trivariate genetic correlation analysis between categories of birth weight (i.e., high and low) and T2D, could be conducted to and further analyse the relationship between the traits.

6.1.3. Associations of T2D and birth weight polygenic scores with T2D and birth weight

The birth weight PS had a significant, positive association with birth weight and the T2D PS had a significant, positive association with T2D. Results indicate that genetic variants that associate with birth weight and T2D from larger GWAS of European-ancestry study populations largely also influence birth weight and T2D within the Indigenous PECRB study population. However, the birth weight PS was significantly, inversely associated with T2D and the T2D PS was not significantly associated with birth weight. This is somewhat contrary to the bivariate genetic correlation analyses, in which genetic correlation was not significantly negative (i.e., inverse). This may be explained by the fact that the variants that associated with both low birth weight and T2D are largely captured in GWAS meta-analyses in European-ancestry study populations, while other variants that are associated with high birth weight and T2D are not. Another potential explanation is that the bivariate genetic correlation analyses reflect environmental factors that are shared in families, which mask an inverse genetic correlation.

Notably, there was no evidence for a non-linear relationship between the birth weight PS and T2D (e.g., a U-shaped curve, as in the relationship between birth weight itself and T2D in this same study population). These findings support the notion that certain variants that confer susceptibility to low birth weight also confer susceptibility to T2D. They may also align with

the foetal insulin hypothesis, which posited that the genetic susceptibility to insulin resistance predisposes offspring to low birth weight as well as subsequent diabetes (Hattersley & Tooke, 1999). Further research is needed to partition the specific contributions of maternal and foetal genotype on birth weight and subsequent diabetes. This could be done by constructing separate maternal and foetal polygenic scores for birth weight and assessing associations between those PS with birth weight and T2D.

6.2. Associations of PS and pPS with T2D and related traits

6.2.1. Construction of PS

While all T2D PS were significantly associated with T2D incidence across all cohorts, the T2D PS derived from European-ancestry populations (i.e., DIAMANTE 2020 – Multi-ancestry European T2D PS, DIAGRAM 2018 T2D PS, and DIAMANTE 2020 – Multi-ancestry composite T2D PS) performed better than the others. These PS likely performed well due to the large sample size and extensive fine-mapping in the DIAGRAM 2018 and DIAMANTE 2020 T2D meta-analyses. The heterogeneity, as measured by I^2, between SNPs in the DIAGRAM 2018 T2D PS and those in the DIAGRAM 2018 GWAS was modest: this was consistent with PECRB collaborators' previous work that incorporated a smaller number of established T2D-associated variants (Hanson et al., 2015).

While further work is needed to optimize T2D PS, the present results suggest that PS constructed using results of GWAS in larger populations may be suitable for translation across study populations in whom well-powered GWAS are not available, particularly those for which index variants derived from other populations may capture common causal variants, potentially due to high LD. Since many Indigenous communities are smaller in size than broader ancestry groups (e.g., East Asian-ancestry, EAS; South Asian, SAS; Hispanic/Latino,

MXL; etc) as defined by large consortia, obtaining suitably powered GWAS for deriving polygenic scores for study populations from these communities may pose additional challenges. Obtaining adequately powered GWAS that are specific for certain communities may be unfeasible. This could necessitate the use of summary statistics from larger GWAS meta-analyses for primary derivation.

6.2.2. Associations of T2D PS and pPS

Strong and significant cross-sectional associations of T2D PS and pPS with T2D and related traits were identified in the PECRB cohort. A strength of these analyses was the inclusion of rigorously measured phenotypic data for PFAT, logM, and logAIR that had been collected by PECRB staff from the PECRB study population (further described in Section 2.2). Results of the T2D pPS- and PS-metabolic trait association study broadly agree with the putative pathophysiological mechanisms that underlie each of the six clusters by Mahajan et al. For instance, the adiposity pPS—which contains loci in the well-known obesity-associated genes *FTO* and *MC4R*—was strongly and significantly associated with PFAT, and significantly associated with logmaxBMI and T2D, in the Indigenous study population. However, the lipids cluster, which contained 3 SNPs in both the extended Mahajan et al cluster and the cluster once merged with genotypes of the PECRB cohort, was not significantly associated with any of the traits. This prompts future large-scale GWAS and GWAS meta-analyses on T2D and related metabolic traits that could enable further extension of these pPS to explore pPS-trait associations. Additionally, analytical approaches to account for disparities in the sizes of various pPS in such association studies could also be implemented; for examples, pPS could be scaled on a per-allele basis in all scores so that odds ratios could be more similarly interpreted.

6.3. Predicting the incidence of T2D using T2D PS

6.3.1. AUC, NRI and decision curve analyses

Polygenic scores potentially have utility for identification of individuals with higher risk of T2D. This relationship was examined longitudinally, as this is more applicable to the presumed use of this score in clinical settings. The PECRB cohort was partitioned into birth, youth, and adult cohorts, based on the timing of examinations with available data. The contribution of the DIAGRAM 2018 T2D PS to the information provided by various models containing clinical predictors was assessed using survival, area under the ROC curve (AUC), net reclassification improvement (NRI), and decision curve analyses.

Results of AUC analyses in the present Indigenous study population are consistent with findings of previous studies: improvement in prediction contributed by the T2D PS was statistically significant but modest. NRIs for the PS across all cohorts showed that it had intermediate power for identifying T2D risk and were comparable to those of commonly measured clinical variables (e.g., HbA1c and FPG).

Decision curve analyses provide a method to facilitate clinical decision making by quantifying the potential net benefit of an intervention across a range of threshold probability (p_t) values. Decision curve analysis assumes that the intervention will be equally effective regardless of how risk is determined. There are limited data on how T2D PS affects response to preventive interventions. For example, a study within the Diabetes Prevention Program Outcomes Study suggested that lifestyle and metformin interventions were both effective, even in those with greater T2D PS (Hivert et al., 2011). Ultimately, clinical utility may depend on how the T2D PS affects the decision to implement preventive interventions.

Across adult, youth, and birth cohorts, results of the decision curve analyses suggest modest increases in clinical benefit for using PS at moderately stringent p_t values. There are few data

on optimal p_t values for T2D prevention: they depend upon preferences of individual patients and clinicians, health care system characteristics, and the nature of the interventions considered. Many clinicians would recommend lifestyle prevention for individuals with impaired glucose regulation; based on cumulative incidence in the adult cohort, this would correspond to p_t = 0.21-0.49. This is in the range in which the analyses suggest meaningful, albeit modest, improvement in clinical benefit from incorporating the PS.

6.3.2. Optimal age for preventive interventions

Genetic effects of the PS with respect to T2D incidence were greatest in the youth and birth cohorts. This is consistent with the hypothesis that genetic effects on T2D are strongest earlier in life. The improvements in AUC and net benefit upon adding the PS to clinical variables were greatest in the birth cohort. The use of T2D PS at birth could be particularly beneficial as phenotypic manifestations of risk (e.g., hyperglycaemia and obesity) are less apparent. Without adjustment for demographic and clinical predictors, the T2D PS had substantially higher HR, AUC and NRI values for youth and birth cohorts compared with adult cohorts, suggesting the greater predictive utility of the PS at earlier ages. However, when clinical predictors were included in prediction models, differences in results of survival, AUC and NRI analyses were attenuated.

There are limited data on how T2D PS affects response to preventive interventions. . While there is strong evidence that T2D can be prevented by lifestyle modification, pharmacologic treatment, or bariatric surgery, there are few data on the efficacy of preventive efforts initiated in youth or infancy (Crandall et al., 2008). Thus, while the analyses described in Chapter 5 suggest that the T2D PS has the strongest contribution to prediction of T2D incidence in the

birth cohort, adults may be a more appropriate target population for preventive interventions in the near term.

6.3.3. Future research

The analyses described in Chapter 5 showed that T2D polygenic scores, as currently constructed, can provide utility for assessing T2D risk. The T2D PS was strongly associated with incident diabetes in this Indigenous study population and, while the improvement in overall prediction measured by ΔAUC is modest, as measured by NRI analyses information from the T2D PS is comparable to that of widely used clinical factors (e.g., HbA1c and BMI). Furthermore, the decision curve analysis shows improvement of net benefit with the T2D PS beyond the use of clinical factors over a range of p_t values; while the increments in benefit are modest, they may be meaningful in a population basis. Further optimization of the PS could provide better prediction of the incidence of T2D.

Optimisation of T2D PS as well as of prediction models is warranted if T2D PS are to be used in clinical practice. Notably, some relevant clinical measures that may be readily obtained at birth (e.g., birth length for calculation of adiposity measures) were not available for inclusion in prediction models in the present study. Advances in laboratory methods and informatics are also required to make PS and risk algorithms available to clinicians and patients. To ensure that PS could be applied to a diverse set of patients and study participants in the future, the transferability of PS across ancestry groups is important to further assess. A recent perspective piece discussed the potential benefits of polygenic risk estimation on clinical medicine, including the selection study participants for clinical trials, such as improvements in cost and efficiency of care delivery and the conduct of research (Fahed et al., 2022). Future research could investigate the net benefit of PS for T2D and other complex traits for these purposes before such methods are implemented.This study showed that T2D

polygenic scores, as currently constructed, can provide utility for assessing T2D risk. The T2D PS was strongly associated with incident diabetes in this PECRB cohort. While the improvement in overall prediction measured by ΔAUC is modest, as measured by NRI analyses information from the T2D PS is comparable to that of widely used clinical factors (e.g., HbA1c and BMI). Furthermore, the decision curve analysis shows improvement of net benefit with the T2D PS beyond the use of clinical factors over a range of p_t values; while the increments in benefit are modest, they may be meaningful in a population basis. To ensure that PS could be applied to a diverse set of patients and study participants in the future, the transferability of PS across ancestry groups is important to assess. Health economics studies are needed to investigate which clinical settings and constructions of T2D PS would maximize net benefit for prediction of T2D incidence. With such knowledge, more informed decisions about the use of genetic information in prevention of T2D could be made.

7. Appendices

7.1. Appendix 1: Publications, abstracts, grants, and prizes

7.1.1. Publications

Wedekind L. E., Mahajan A., Hsueh W.-C., Chen P., Olaiya M. T., Kobes S., Sinha M., Baier L. J., Knowler W. C., McCarthy M. I., Hanson R. L. The utility of a type 2 diabetes polygenic score in addition to clinical variables for prediction of type 2 diabetes incidence in birth, youth and adult cohorts in an Indigenous study population. *Diabetologia* 2023 (in press).

Wedekind L. E., Mitchell C. M., Andersen C. A., Hanson R. L., Knowler W. C. Epidemiology of type 2 diabetes in Indigenous communities in the United States. *Current Diabetes Reports* 2021; **21**(11):47.

Wedekind L. E.,* Noé A.,* Mokaya J.,* Tamandjou C., Kapulu M., Ruecker A., Kestelyn E., et al. Equity for excellence in academic institutions: A manifesto for change [version 1; peer review: 2 approved]. *Wellcome Open Research* 2021; **6**:142.

Olaiya M. T., **Wedekind L. E.**, Hanson R. L., Sinha M., Kobes S., Nelson R. G., Baier L. J., Knowler WC. Birthweight and early-onset type 2 diabetes in American Indians: differential effects in adolescents and young adults and additive effects of genotype, BMI and maternal diabetes. *Diabetologia* 2019; **62**(9):1628-1637.

* Equal first-authorship

7.1.2. Published abstracts

Wedekind L. E., Mahajan A., Hsueh W.-C., Kobes S., Baier L. J., Knowler W. C., McCarthy M. I., Hanson R. L. Associations between type 2 diabetes partitioned/process-specific polygenic scores and metabolic traits. *Diabetes* 2022; 71(S1).

Wedekind L. E., Hsueh W.-C., Olaiya M. T., Kobes S., Baier L. J., Knowler W. C., Mahajan A., et al. 1132-P: Genetic relationships between birth weight and type 2 diabetes. *Diabetes* 2021; 70(S1).

Wedekind L. E., Kobes S., Hsueh W.-C., Baier L. J., Knowler W. C., Mahajan A., McCarthy M. I., Hanson R. L. 1460-P: Type 2 diabetes polygenic score in addition to clinical factors for prediction of diabetes incidence in an Indigenous American population. *Diabetes* 2020; 69(S1).

Hanson R. L., **Wedekind L. E.**, Hsueh W.-C., Kobes S., Baier L. J., Bogardus C., Knowler W. C. 1709-P: Assessing the role of 211 “established” type 2 diabetes variants in American Indians. *Diabetes* 2019; 68(S1).

Wedekind L. E., Hsueh W.-C., Kobes S., Olaiya M. T., Knowler W. C., Baier L. J., Hanson R. L. 1701-P: Genome-wide association study identifies novel potential associations with birth weight in a Southwestern American Indian population. *Diabetes* 2019; 68(S1).

Wedekind L. E, Walter M., Kobes S., Chen P., Hsueh W.-C., Nelson R. G., Baier L. J., et al. Genome-wide association study of inflammatory cytokines in Southwestern Native Americans identifies novel association between *WWOX* and serum TNF-α levels. *Diabetes* 2018; 67(S1).

7.1.3. Grants

National Institutes of Health Graduate Partnership Program Track 3 Medical Scientist Training Program Funding ($220,000 funding support for medical training). National Institute of Diabetes and Digestive and Kidney Diseases. 2022.

Poster Award ($1,000 Travel Grant). Annual National Institute of Diabetes and Digestive and Kidney Diseases Scientific Conference. 2021.

Fellows Award for Research Excellence ($1,500 Travel Grant). NIH Fellows Committee, Scientific Directors and Office of Intramural Training and Education. 2020.

National Institutes of Health (NIH)-Oxford Fellowship and NIH Scientific Directors First-Year Stipend Award ($250,000 fellowship plus $20,000 support for research lab). NIH Oxford-Cambridge Scholars Program. 2018.

7.1.4. Prizes

Building a Better Community Through Service Award. International Biomedical Research Alliance. 2020.

University of Oxford Nuffield Department of Medicine (NDM) Public Engagement Prize. University of Oxford NDM. 2019.

7.2. Appendix 2: Summary statistics for SNPs included in T2D PS

Table 7.2.1. SNPs included in the construction of the AGEN 2020 (Spracklen et al., 2020) PS.

chr[h]	position[i]	SNP	a1[j]	a2	a1f[k]	OR[l]	p-value
1	20688352	rs60573766	C	T	0.6347	1.04	4.30×10^{-10}
1	39942242	rs371894931	CA	C	0.1572	1.06	2.68×10^{-11}
1	46244900	rs562138031	C	CT	0.7256	1.06	3.99×10^{-12}
1	51191935	rs11205766	T	A	0.9031	1.09	7.49×10^{-15}
1	64107893	rs2269245	G	A	0.8153	1.06	5.40×10^{-10}
1	177878933	rs532504	A	G	0.2133	1.06	7.39×10^{-12}
1	184014593	rs1327123	C	G	0.4596	1.04	7.00×10^{-9}
1	204474581	rs201297151	CAAAAAAAA	C	0.4404	1.04	3.37×10^{-8}
1	214155398	rs12403994	C	A	0.6228	1.05	6.05×10^{-12}
1	229642499	rs238763	T	A	0.6783	1.05	5.04×10^{-11}
2	632789	rs10634531	CTTG	C	0.9064	1.1	2.37×10^{-17}
2	27730940	rs1260326	C	T	0.4562	1.07	1.01×10^{-21}
2	45192080	rs12712928	C	G	0.4022	1.06	1.83×10^{-14}
2	60586707	rs243018	G	C	0.6656	1.06	1.55×10^{-15}
2	120231070	rs3731600	C	G	0.9675	1.13	6.86×10^{-9}
2	149568261	rs200576292	CT	C	0.6603	1.05	1.29×10^{-9}
2	165381518	rs75536691	A	G	0.976	1.2	1.16×10^{-15}
2	213687103	rs75179644	T	C	0.8992	1.08	5.36×10^{-10}
2	234191103	rs117809958	A	T	0.0181	1.25	1.98×10^{-15}
3	12385357	rs3963364	C	A	0.95	1.12	3.49×10^{-11}
3	23258614	rs11926494	G	A	0.82	1.12	2.69×10^{-37}
3	63904715	rs67114627	AAG	A	0.6459	1.09	9.10×10^{-32}
3	114968018	rs6806156	T	C	0.6107	1.05	1.59×10^{-11}
3	121965199	rs9859381	G	T	0.486	1.04	2.86×10^{-9}
3	123174832	rs60054445	C	T	0.6643	1.05	5.62×10^{-12}
3	124921920	rs12497133	A	G	0.3231	1.04	1.14×10^{-8}
3	152382352	rs1850421	A	C	0.2777	1.05	1.41×10^{-9}
3	170643788	rs201018682	T	C	0.8162	1.06	1.01×10^{-11}
3	185495320	rs13092876	A	G	0.3122	1.13	1.91×10^{-66}
3	186649931	rs11332772	C	CG	0.5571	1.05	1.07×10^{-10}
3	187698333	rs13086331	T	C	0.8187	1.05	7.29×10^{-9}
3	195830310	rs9866168	T	A	0.6394	1.05	1.55×10^{-9}
4	1254535	rs7656416	C	T	0.6833	1.11	9.01×10^{-42}
4	1784605	rs6831006	G	C	0.2736	1.07	2.92×10^{-16}
4	6303731	rs147834269	G	A	0.9784	1.23	9.10×10^{-12}
4	45186139	rs10938398	A	G	0.2921	1.05	3.84×10^{-10}

4	71844118	rs28599782	A	G	0.209	1.07	4.64×10^{-16}
4	85339618	rs117624659	T	C	0.9765	1.23	1.95×10^{-16}
4	153520279	rs10011838	G	A	0.493	1.08	1.43×10^{-27}
5	14768092	rs6885132	C	G	0.556	1.04	2.59×10^{-9}
5	36257018	rs16902871	G	A	0.149	1.06	3.34×10^{-9}
5	50079603	rs74334916	C	A	0.0748	1.07	4.33×10^{-8}
5	51751574	rs12109081	T	A	0.3608	1.04	1.08×10^{-8}
5	55810305	rs256904	T	A	0.4923	1.08	3.56×10^{-29}
5	74574984	rs2126736	A	G	0.4279	1.04	1.84×10^{-8}
5	95848503	rs6556925	C	A	0.4155	1.04	3.12×10^{-9}
5	133864599	rs329122	A	G	0.3865	1.04	2.22×10^{-8}
5	176513896	rs3135911	A	C	0.4324	1.05	1.51×10^{-12}
6	7231843	rs9379084	G	A	0.7986	1.07	2.20×10^{-14}
6	20683164	rs9350271	A	G	0.4225	1.21	4.96×10^{-183}
6	31026236	rs76541615	T	G	0.7959	1.08	1.08×10^{-17}
6	34214670	rs4711389	A	G	0.8925	1.1	7.28×10^{-17}
6	39046644	rs742762	A	C	0.7118	1.08	1.79×10^{-22}
6	50787459	rs62405419	T	G	0.2682	1.05	3.79×10^{-9}
6	117996631	rs80196932	T	C	0.7859	1.06	7.57×10^{-13}
6	126964510	rs4273712	G	A	0.4688	1.05	2.56×10^{-12}
6	131954797	rs7739842	G	T	0.3564	1.05	1.55×10^{-11}
6	137294771	rs35389258	T	TA	0.45	1.05	9.49×10^{-14}
6	139205386	rs9376382	C	T	0.5989	1.04	1.50×10^{-8}
6	143056556	rs9390022	T	C	0.7997	1.05	6.35×10^{-9}
7	13886654	rs7787720	T	C	0.4209	1.06	2.26×10^{-15}
7	14898282	rs17168486	T	C	0.4181	1.07	8.23×10^{-22}
7	28219310	rs3735567	G	A	0.7776	1.06	3.05×10^{-12}
7	44174857	rs2908279	G	T	0.627	1.05	8.42×10^{-11}
7	55984953	rs565050730	GA	G	0.3341	1.04	4.43×10^{-8}
7	69189726	rs610930	A	G	0.2868	1.07	1.38×10^{-19}
7	69696905	rs12698877	G	A	0.3362	1.07	6.96×10^{-22}
7	89752238	rs62469016	C	G	0.2231	1.07	1.52×10^{-15}
7	93107093	rs2074120	A	C	0.323	1.04	8.38×10^{-9}
7	102336979	rs75990271	T	C	0.8152	1.07	3.22×10^{-11}
7	126526991	rs117737118	G	A	0.0934	1.18	3.31×10^{-31}
7	127253550	rs2233580	T	C	0.0858	1.34	2.73×10^{-132}
7	127761917	rs61342118	A	C	0.0831	1.34	4.78×10^{-105}
7	140579350	rs71170768	TA	T	0.3159	1.07	2.17×10^{-10}
7	157024510	rs1182444	G	A	0.4947	1.05	1.67×10^{-12}
8	17927609	rs34642578	T	C	0.0533	1.09	1.60×10^{-9}
8	36832310	rs56687477	A	G	0.3227	1.05	1.53×10^{-10}
8	37391203	rs4739515	G	C	0.5383	1.05	1.68×10^{-11}
8	38343012	rs328301	T	C	0.3278	1.04	4.11×10^{-8}

8	41512648	rs33981001	GGT	G	0.3888	1.08	5.27×10^{-28}
8	73503743	rs349359	C	A	0.2424	1.04	3.05×10^{-8}
8	75214398	rs149265787	G	A	0.0239	1.14	5.68×10^{-10}
8	95960886	rs896852	G	T	0.2998	1.04	6.42×10^{-9}
8	118184783	rs13266634	C	T	0.5857	1.12	3.73×10^{-67}
8	126471274	rs60089934	A	G	0.3777	1.04	3.33×10^{-9}
8	132879795	rs73708054	C	T	0.2523	1.04	4.41×10^{-8}
9	1032567	rs1016565	A	G	0.4209	1.04	2.18×10^{-8}
9	4290085	rs4237150	C	G	0.4262	1.07	4.46×10^{-27}
9	22132878	rs10965248	T	C	0.563	1.2	4.42×10^{-164}
9	81917127	rs1328412	T	C	0.9448	1.1	6.41×10^{-11}
9	84308948	rs2796441	G	A	0.3897	1.08	1.43×10^{-28}
9	98278413	rs113154802	C	T	0.8885	1.06	3.51×10^{-8}
9	107597527	rs201375651	CA	C	0.3948	1.04	2.56×10^{-8}
9	136149500	rs529565	C	T	0.4445	1.04	1.68×10^{-10}
9	139246588	rs376993806	G	A	0.8821	1.14	4.52×10^{-26}
10	12309139	rs11257657	G	C	0.4832	1.12	9.77×10^{-62}
10	23487778	rs77065181	A	G	0.0468	1.09	1.57×10^{-8}
10	63712602	rs141583966	G	GGTGT	0.9093	1.08	7.65×10^{-10}
10	64976133	rs148928116	T	TA	0.7945	1.06	2.53×10^{-13}
10	71273357	rs1955163	G	A	0.5409	1.05	1.68×10^{-11}
10	77323643	rs34907385	C	T	0.4914	1.05	5.37×10^{-12}
10	80951130	rs34204798	C	CG	0.5681	1.06	5.02×10^{-19}
10	89684214	rs1236816	A	C	0.499	1.04	4.31×10^{-10}
10	93592703	rs147689733	T	C	0.0209	1.31	1.21×10^{-27}
10	94435673	rs35906730	G	A	0.2645	1.15	1.27×10^{-71}
10	95009180	rs565236700	C	T	0.0235	1.23	5.54×10^{-19}
10	99056921	rs10736116	C	G	0.3058	1.05	9.21×10^{-11}
10	112678657	rs7895872	T	G	0.5787	1.05	1.37×10^{-11}
10	114754088	rs7901695	C	T	0.0382	1.32	8.18×10^{-62}
10	122929493	rs10886863	C	T	0.6641	1.06	5.28×10^{-17}
10	124150342	rs112820281	C	CTGGA	0.4098	1.05	1.42×10^{-10}
11	2203154	rs11043003	C	T	0.0819	1.11	1.03×10^{-16}
11	2858546	rs2237897	C	T	0.6321	1.28	1.88×10^{-245}
11	17415190	rs4148646	C	G	0.3846	1.08	1.70×10^{-26}
11	27729505	rs4922793	A	G	0.5655	1.04	1.62×10^{-10}
11	69462642	rs602652	A	G	0.8088	1.06	5.32×10^{-9}
11	72463435	rs7109575	G	A	0.9448	1.15	5.46×10^{-21}
11	92708710	rs10830963	G	C	0.421	1.04	4.49×10^{-8}
12	4381981	rs7304270	C	T	0.7614	1.07	1.04×10^{-12}
12	27963402	rs3751236	G	A	0.6717	1.07	6.58×10^{-21}
12	31441179	rs80234489	C	A	0.1699	1.11	4.33×10^{-32}
12	50269863	rs77978149	T	C	0.0898	1.08	5.70×10^{-9}

12	66232810	rs2583934	T	G	0.3396	1.06	4.95×10^{-16}
12	71449521	rs7313668	T	G	0.3742	1.05	4.91×10^{-11}
12	97850215	rs10860209	C	A	0.589	1.04	5.67×10^{-9}
12	108629780	rs1426371	G	A	0.4911	1.05	7.76×10^{-12}
12	111836771	rs149212747	A	AC	0.7948	1.07	2.07×10^{-11}
12	112736118	rs77768175	A	G	0.8068	1.07	1.39×10^{-10}
12	114123722	rs7307263	G	C	0.4272	1.04	3.64×10^{-8}
12	118400856	rs111246699	A	G	0.2572	1.06	1.54×10^{-15}
12	121363506	rs118074491	G	A	0.0323	1.21	8.83×10^{-20}
13	22589883	rs9316706	A	G	0.3505	1.04	3.33×10^{-9}
13	26781367	rs568052023	C	CA	0.5812	1.07	2.62×10^{-22}
13	33557173	rs7983505	T	A	0.1572	1.08	3.22×10^{-18}
13	51088809	rs123378	G	A	0.1951	1.05	2.22×10^{-10}
13	80707429	rs1215468	A	G	0.7185	1.09	1.32×10^{-31}
13	91949562	rs9515905	A	G	0.831	1.08	7.24×10^{-18}
14	24878370	rs12437434	C	T	0.7127	1.05	1.02×10^{-9}
14	38809661	rs61975988	A	G	0.4569	1.04	1.97×10^{-9}
14	77382503	rs58524310	G	A	0.3268	1.05	8.41×10^{-11}
14	101255172	rs73347525	A	G	0.7556	1.06	7.46×10^{-11}
14	103237952	rs55700915	A	G	0.4344	1.04	1.50×10^{-8}
15	28546173	rs76704029	T	C	0.7322	1.06	3.39×10^{-8}
15	38828140	rs8043085	T	G	0.4493	1.05	2.06×10^{-14}
15	40615872	rs12907887	C	G	0.229	1.08	1.74×10^{-20}
15	52587740	rs149336329	G	T	0.9494	1.11	1.70×10^{-9}
15	62394264	rs8037894	G	C	0.5856	1.08	7.30×10^{-33}
15	68080886	rs4776970	A	T	0.2214	1.04	3.41×10^{-8}
15	75742095	rs8038760	A	C	0.6081	1.05	1.81×10^{-11}
15	77776562	rs952472	C	A	0.3949	1.07	1.62×10^{-26}
15	90428894	rs10852123	A	C	0.201	1.06	8.38×10^{-13}
15	91522253	rs8026714	A	G	0.486	1.07	1.06×10^{-22}
15	93825384	rs61021634	A	G	0.4379	1.05	1.41×10^{-11}
15	99366409	rs79826452	A	G	0.8904	1.07	3.16×10^{-8}
16	3647098	rs2240885	A	G	0.4027	1.04	2.79×10^{-9}
16	20323168	rs117267808	A	G	0.0782	1.11	4.85×10^{-17}
16	53800954	rs1421085	C	T	0.1668	1.14	1.55×10^{-48}
16	72022534	rs12600132	T	C	0.4316	1.04	5.95×10^{-9}
16	73100308	rs6416749	C	T	0.3746	1.05	3.40×10^{-12}
16	81534790	rs2925979	T	C	0.3635	1.04	1.51×10^{-9}
17	6953781	rs186568031	T	C	0.0942	1.12	8.99×10^{-24}
17	29642430	rs7502556	T	C	0.5346	1.05	3.79×10^{-11}
17	36101586	rs8064454	A	C	0.3049	1.13	6.47×10^{-61}
17	65641651	rs2706710	T	C	0.0813	1.07	1.67×10^{-8}
17	73187031	rs35559984	CA	C	0.6519	1.05	7.88×10^{-9}

18	7076836	rs9948462	T	C	0.7048	1.05	8.70×10^{-10}
18	57852587	rs476828	C	T	0.243	1.09	4.81×10^{-27}
18	60845884	rs12454712	T	C	0.5122	1.06	1.40×10^{-15}
19	7293119	rs8101064	T	C	0.1277	1.07	3.53×10^{-8}
19	7986638	rs475002	G	C	0.5184	1.04	9.80×10^{-10}
19	12509536	rs4804181	A	C	0.5139	1.04	1.54×10^{-8}
19	21529576	rs145389767	G	A	0.9794	1.22	2.42×10^{-13}
19	22100706	rs142395395	A	G	0.9691	1.24	6.95×10^{-23}
19	33894846	rs10422861	C	T	0.5358	1.06	2.29×10^{-16}
19	46157928	rs113036890	CAAAAAAAA	C	0.7313	1.1	1.03×10^{-32}
20	22430241	rs73085586	G	A	0.6437	1.04	1.66×10^{-9}
20	42994812	rs12625671	C	T	0.442	1.07	2.25×10^{-21}
20	48830772	rs13040225	A	T	0.5132	1.05	1.64×10^{-14}
20	50155386	rs6021276	T	C	0.4101	1.04	6.66×10^{-10}
20	57477177	rs11477757	TC	TCC	0.6687	1.06	1.27×10^{-8}
22	29380119	rs147413364	T	TA	0.3571	1.04	3.38×10^{-8}
22	46313618	rs28637892	T	G	0.2153	1.05	3.66×10^{-9}
22	50356302	rs28691713	C	T	0.5552	1.07	1.79×10^{-17}

[h] Chr is chromosome.
[i] Position is in base pairs (bp) on Genome Reference Consortium Human Build 37.
[j] A1 and a2 are the alleles from AGEN 2020, with the effect allele given first.
[k] A1f is the frequency of a1 in AGEN 2020 study population.
[l] OR is reported per copy of the effect allele and is also reported with the 95% confidence interval (CI) of the OR.

Table 7.2.2. Summary statistics for SNPs included in the construction of the DIAGRAM 2018 (DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium et al., 2014) PS.

chr[m]	position[n]	SNP	a1[o]	a2	a1f[p]	OR[q]	p-value
1	40035928	rs3768321	T	G	0.2	1.09	2.6×10^{-26}
1	117532790	rs1127215	C	T	0.584	1.05	1.6×10^{-13}
1	120526982	rs1493694	T	C	0.109	1.09	2.7×10^{-16}
1	150786038	rs10305745	A	G	0.015	1.28	4.4×10^{-6}
1	177889025	rs539515	C	A	0.198	1.05	1.6×10^{-10}
1	205114873	rs12048743	G	C	0.442	1.04	3.5×10^{-9}
1	206593900	rs9430095	C	G	0.494	1.04	1.9×10^{-8}
1	214159256	rs340874	C	T	0.556	1.07	1.6×10^{-22}
1	219748818	rs2820446	C	G	0.706	1.06	3.3×10^{-16}
1	229672955	rs348330	G	A	0.361	1.05	2.7×10^{-14}
1	235690800	rs291367	G	A	0.632	1.04	4.7×10^{-10}
2	422144	rs62107261	T	C	0.954	1.12	3.8×10^{-12}
2	653575	rs35913461	C	T	0.829	1.06	1.6×10^{-11}
2	16574669	rs11680058	A	G	0.863	1.06	1.4×10^{-8}
2	25643221	rs17802463	G	T	0.731	1.04	2.9×10^{-8}
2	27730940	rs1260326	C	T	0.607	1.07	6.5×10^{-25}
2	43207872	rs28525376	G	T	0.422	1.03	2.7×10^{-6}
2	43430440	rs6708643	A	G	0.501	1.04	3.9×10^{-8}
2	58981064	rs10193538	T	G	0.61	1.04	8.9×10^{-9}
2	59307725	rs6545714	G	A	0.392	1.04	8.9×10^{-9}
2	60583665	rs243024	A	G	0.46	1.06	2.5×10^{-20}
2	65287896	rs2249105	A	G	0.634	1.1	2.2×10^{-14}
2	65355270	rs2052261	G	A	0.304	1.07	2.5×10^{-6}
2	65655012	rs2028150	C	G	0.598	1.05	2.3×10^{-12}
2	121318166	rs11688931	C	G	0.849	1.04	8.2×10^{-6}
2	121347612	rs11688682	G	C	0.728	1.05	4.2×10^{-9}
2	147861633	rs35999103	T	C	0.155	1.05	9.7×10^{-9}
2	158339550	rs13426680	A	G	0.937	1.09	6.7×10^{-10}
2	161135544	rs3772071	T	C	0.714	1.05	1.2×10^{-11}
2	165513091	rs10195252	T	C	0.586	1.07	6.0×10^{-25}
2	227101411	rs2972144	G	A	0.639	1.1	2.1×10^{-46}
3	12336507	rs11709077	G	A	0.877	1.14	1.8×10^{-36}
3	12489342	rs17819328	G	T	0.425	1.06	4.8×10^{-16}
3	23510044	rs17013314	G	A	0.031	1.11	8.4×10^{-9}
3	46925539	rs11926707	C	T	0.626	1.27	2.1×10^{-8}
3	47242923	rs75423501	G	A	0.101	1.05	7.5×10^{-6}
3	49980596	rs4688760	T	C	0.684	1.04	3.5×10^{-10}
3	53127677	rs2581787	T	G	0.563	1.04	2.4×10^{-8}
3	63962339	rs3774723	G	A	0.844	1.07	1.6×10^{-13}

3	64701146	rs9860730	A	G	0.704	1.06	4.9×10^{-15}
3	72865183	rs13085136	C	T	0.928	1.08	1.5×10^{-8}
3	77671721	rs2272163	C	A	0.618	1.04	9.6×10^{-9}
3	123065778	rs11708067	A	G	0.772	1.09	5.2×10^{-32}
3	124926637	rs649961	T	C	0.465	1.04	9.9×10^{-10}
3	152086533	rs147579559	A	G	0.6	1.05	8.1×10^{-13}
3	152417881	rs74653713	C	A	0.957	1.1	1.2×10^{-8}
3	152433628	rs35497231	C	T	0.317	1.04	7.6×10^{-8}
3	168218841	rs7629630	A	T	0.857	1.05	2.5×10^{-8}
3	170733076	rs9873618	G	A	0.71	1.07	4.8×10^{-21}
3	183738460	rs2872246	A	C	0.454	1.04	1.5×10^{-8}
3	185503456	rs6780171	A	T	0.314	1.14	9.0×10^{-56}
3	185541213	rs11717959	G	T	0.621	1.04	3.0×10^{-6}
3	185829891	rs1516728	A	T	0.759	1.03	6.3×10^{-6}
3	186665645	rs3887925	T	C	0.547	1.07	3.1×10^{-22}
3	186675277	rs7645517	A	G	0.058	1.08	2.5×10^{-8}
4	744972	rs1531583	T	G	0.046	1.13	3.5×10^{-14}
4	1010077	rs35654957	C	T	0.367	1.03	4.2×10^{-7}
4	1784403	rs56337234	C	T	0.503	1.06	8.6×10^{-18}
4	3241845	rs362307	T	C	0.077	1.08	1.1×10^{-9}
4	6302519	rs1801212	A	G	0.709	1.05	4.5×10^{-6}
4	6306763	rs10937721	C	G	0.588	1.06	1.5×10^{-8}
4	17792869	rs12640250	C	A	0.715	1.04	3.7×10^{-8}
4	45186139	rs10938398	A	G	0.429	1.05	3.6×10^{-12}
4	52818664	rs2102278	G	A	0.319	1.04	3.7×10^{-8}
4	89740894	rs1903002	G	C	0.501	1.04	2.7×10^{-8}
4	104140848	rs1580278	C	A	0.473	1.04	2.2×10^{-10}
4	137083193	rs1296328	A	C	0.446	1.04	3.5×10^{-8}
4	153513369	rs7669833	T	A	0.705	1.06	1.2×10^{-14}
4	157652753	rs28819812	C	A	0.677	1.04	2.2×10^{-8}
4	185717759	rs58730668	T	C	0.858	1.07	1.3×10^{-13}
5	14768092	rs6885132	C	G	0.904	1.07	1.7×10^{-8}
5	44682589	rs6884702	G	A	0.393	1.04	1.5×10^{-10}
5	51791225	rs17261179	T	C	0.517	1.04	1.3×10^{-8}
5	52100489	rs3811978	G	A	0.167	1.06	7.7×10^{-11}
5	52774510	rs62370480	A	G	0.22	1.04	2.0×10^{-6}
5	53271420	rs702634	A	G	0.69	1.05	7.7×10^{-14}
5	53412620	rs279744	C	A	0.691	1.04	3.1×10^{-8}
5	55808475	rs465002	T	C	0.742	1.11	6.1×10^{-38}
5	55848669	rs2431115	A	G	0.402	1.04	3.9×10^{-10}
5	55861595	rs9687832	A	G	0.198	1.08	1.7×10^{-20}
5	56196604	rs96844	G	A	0.262	1.04	5.4×10^{-8}
5	67714246	rs4976033	G	A	0.411	1.05	1.0×10^{-9}

5	75003678	rs2307111	T	C	0.605	1.05	2.1×10^{-16}
5	76424949	rs4457053	G	A	0.304	1.06	8.4×10^{-18}
5	78430607	rs1316776	C	A	0.648	1.05	2.6×10^{-12}
5	86577352	rs7719891	G	A	0.259	1.04	2.4×10^{-8}
5	133414622	rs244665	A	G	0.703	1.03	9.9×10^{-6}
5	133864599	rs329122	A	G	0.429	1.04	3.6×10^{-9}
5	157928196	rs3934712	C	T	0.206	1.05	3.2×10^{-8}
6	7035734	rs112498319	C	A	0.409	1.03	4.2×10^{-7}
6	7255650	rs9505097	C	T	0.799	1.05	8.6×10^{-10}
6	20679709	rs7756992	G	A	0.274	1.15	2.4×10^{-88}
6	32573415	rs601945	G	A	0.178	1.06	4.7×10^{-8}
6	34524698	rs2233632	T	C	0.688	1.04	5.3×10^{-7}
6	40409243	rs34298980	T	C	0.497	1.04	9.3×10^{-10}
6	43760327	rs11967262	G	C	0.486	1.04	8.8×10^{-10}
6	43814190	rs6458354	C	T	0.289	1.05	2.1×10^{-12}
6	50788778	rs3798519	C	A	0.184	1.06	2.6×10^{-12}
6	51180765	rs2465043	G	A	0.644	1.03	2.9×10^{-6}
6	107431688	rs4946812	G	A	0.674	1.04	8.2×10^{-9}
6	126792095	rs11759026	G	A	0.232	1.07	2.4×10^{-18}
6	127416930	rs2800733	A	G	0.717	1.05	6.0×10^{-11}
6	137300960	rs9494624	A	G	0.29	1.04	6.1×10^{-9}
6	140249466	rs616279	A	G	0.738	1.04	6.7×10^{-7}
6	160770312	rs474513	A	G	0.517	1.04	8.1×10^{-10}
6	164133001	rs4709746	C	T	0.868	1.06	5.8×10^{-9}
7	14898282	rs17168486	T	C	0.181	1.07	2.3×10^{-17}
7	15063569	rs10228066	T	C	0.537	1.07	1.1×10^{-28}
7	15206239	rs2908334	T	C	0.631	1.03	5.9×10^{-6}
7	23434606	rs78840640	G	C	0.022	1.11	2.8×10^{-6}
7	23512896	rs4279506	G	C	0.61	1.06	4.8×10^{-8}
7	28198677	rs1708302	C	T	0.512	1.1	1.1×10^{-48}
7	30728452	rs917195	C	T	0.77	1.05	4.2×10^{-11}
7	44255643	rs878521	A	G	0.245	1.06	1.9×10^{-13}
7	102486254	rs11496066	T	C	0.818	1.08	1.1×10^{-8}
7	102987583	rs62482405	G	T	0.082	1.05	6.9×10^{-6}
7	103444978	rs39328	T	C	0.433	1.04	3.7×10^{-8}
7	117495667	rs6976111	A	C	0.313	1.04	1.2×10^{-8}
7	130027037	rs2268382	C	A	0.327	1.03	7.4×10^{-6}
7	130457914	rs1562396	G	A	0.319	1.06	9.9×10^{-18}
7	150537635	rs62492368	A	G	0.308	1.05	1.1×10^{-10}
7	156930550	rs6459733	G	C	0.673	1.06	2.4×10^{-17}
8	9974824	rs17689007	G	A	0.533	1.04	2.5×10^{-9}
8	10808687	rs57327348	A	T	0.782	1.04	4.5×10^{-8}
8	30863938	rs10954772	T	C	0.314	1.04	1.8×10^{-9}

8	41509915	rs4736819	T	C	0.554	1.04	5.4×10^{-7}
8	110123183	rs12680028	C	G	0.534	1.04	2.5×10^{-8}
8	118185025	rs3802177	G	A	0.685	1.11	1.1×10^{-55}
8	128711742	rs17772814	G	A	0.915	1.08	5.4×10^{-10}
8	129568078	rs1561927	C	T	0.269	1.04	1.5×10^{-9}
8	145507304	rs4977213	C	T	0.375	1.05	9.1×10^{-14}
8	145879883	rs12719778	T	C	0.538	1.04	5.0×10^{-9}
9	3965689	rs510807	A	C	0.491	1.03	1.4×10^{-6}
9	4291928	rs10974438	C	A	0.357	1.05	1.5×10^{-14}
9	19067833	rs7022807	G	A	0.401	1.04	2.7×10^{-10}
9	20241069	rs7867635	C	T	0.412	1.04	4.0×10^{-8}
9	20662703	rs7847880	C	T	0.843	1.04	2.1×10^{-6}
9	22133773	rs76011118	A	G	0.034	1.11	1.4×10^{-7}
9	22134068	rs10811660	G	A	0.828	1.27	1.4×10^{-115}
9	22134172	rs10757283	T	C	0.43	1.11	1.7×10^{-41}
9	22157908	rs1333052	A	C	0.66	1.03	6.3×10^{-7}
9	28410683	rs1412234	C	T	0.323	1.04	1.9×10^{-10}
9	34074476	rs12001437	C	T	0.372	1.04	2.8×10^{-10}
9	81359113	rs11137820	C	G	0.575	1.04	2.9×10^{-8}
9	81905590	rs17791513	A	G	0.932	1.1	3.1×10^{-14}
9	84308948	rs2796441	G	A	0.592	1.07	4.4×10^{-24}
9	97001682	rs55653563	A	C	0.732	1.04	2.2×10^{-9}
9	136149229	rs505922	C	T	0.332	1.05	3.9×10^{-12}
9	139235606	rs78403475	G	C	0.896	1.06	1.2×10^{-6}
9	139241030	rs28505901	G	A	0.752	1.09	6.7×10^{-26}
9	139507212	rs11793035	C	T	0.331	1.04	2.2×10^{-7}
10	12307894	rs11257655	T	C	0.218	1.09	1.5×10^{-32}
10	71321279	rs177045	G	A	0.316	1.07	6.6×10^{-18}
10	71321658	rs61850200	C	G	0.277	1.04	7.3×10^{-6}
10	71466578	rs2642588	G	T	0.702	1.05	2.2×10^{-14}
10	80952826	rs703972	G	C	0.533	1.07	1.7×10^{-29}
10	81096589	rs1317617	G	A	0.798	1.04	1.8×10^{-6}
10	89769340	rs11202627	T	C	0.152	1.06	4.7×10^{-8}
10	93924663	rs7078559	T	C	0.578	1.03	4.1×10^{-7}
10	94462427	rs10882101	T	C	0.587	1.06	1.4×10^{-8}
10	94479107	rs1112718	A	G	0.598	1.06	5.0×10^{-7}
10	114703136	rs7918400	T	C	0.476	1.06	2.0×10^{-15}
10	114757956	rs78025551	C	G	0.851	1.05	1.6×10^{-7}
10	114758349	rs7903146	T	C	0.295	1.37	5.8×10^{-447}
10	114871594	rs34855922	A	G	0.716	1.05	5.5×10^{-12}
10	122915345	rs72631105	A	G	0.19	1.06	3.7×10^{-9}
10	124193181	rs2280141	T	G	0.516	1.05	1.4×10^{-13}
11	1704596	rs12802972	A	G	0.428	1.03	1.5×10^{-6}

11	2118860	rs11042596	G	T	0.665	1.04	2.0×10^{-8}
11	2197286	rs4929965	A	G	0.383	1.07	4.0×10^{-26}
11	2372356	rs4930091	C	T	0.759	1.04	3.7×10^{-6}
11	2579163	rs2283164	A	G	0.947	1.08	1.2×10^{-7}
11	2691500	rs231361	A	G	0.256	1.08	5.0×10^{-25}
11	2755548	rs2283220	A	G	0.69	1.05	1.4×10^{-9}
11	2850828	rs234853	G	A	0.248	1.08	6.8×10^{-16}
11	2857194	rs2237895	C	A	0.426	1.12	6.0×10^{-52}
11	2858546	rs2237897	C	T	0.954	1.23	8.4×10^{-32}
11	2908754	rs445084	G	A	0.361	1.03	1.7×10^{-6}
11	17408404	rs5213	C	T	0.362	1.07	3.5×10^{-27}
11	28534898	rs4923543	A	G	0.332	1.04	4.5×10^{-8}
11	32927778	rs145678014	G	T	0.957	1.11	2.0×10^{-10}
11	34642668	rs286925	A	G	0.182	1.04	5.0×10^{-6}
11	34982148	rs2767036	C	A	0.291	1.04	3.3×10^{-8}
11	43877934	rs1061810	A	C	0.288	1.05	6.0×10^{-13}
11	45912013	rs7115753	A	G	0.449	1.04	3.8×10^{-9}
11	47529947	rs7124681	A	C	0.41	1.04	5.1×10^{-9}
11	65294799	rs1783541	T	C	0.204	1.06	2.0×10^{-14}
11	68997225	rs61881115	G	A	0.838	1.05	4.1×10^{-7}
11	72460398	rs77464186	A	C	0.836	1.11	4.7×10^{-33}
11	92708710	rs10830963	G	C	0.277	1.1	4.8×10^{-43}
11	128042575	rs10893829	T	C	0.853	1.06	1.3×10^{-10}
11	128234144	rs10750397	A	G	0.282	1.05	8.3×10^{-13}
11	128398938	rs67232546	T	C	0.207	1.06	1.3×10^{-11}
12	4031104	rs10848958	C	T	0.804	1.04	1.5×10^{-7}
12	4300172	rs11063028	C	T	0.18	1.06	8.5×10^{-11}
12	4376089	rs4238013	C	T	0.209	1.06	3.2×10^{-11}
12	4399050	rs3217860	G	A	0.258	1.05	3.9×10^{-9}
12	26453283	rs718314	G	A	0.253	1.05	8.4×10^{-11}
12	27965150	rs10842994	C	T	0.805	1.08	4.1×10^{-20}
12	66221060	rs2258238	T	A	0.104	1.1	4.5×10^{-21}
12	66358347	rs1042725	T	C	0.49	1.05	1.8×10^{-13}
12	71522953	rs1796330	G	C	0.571	1.05	2.2×10^{-14}
12	97848775	rs77864822	A	G	0.932	1.08	1.1×10^{-8}
12	108629780	rs1426371	G	A	0.739	1.05	8.2×10^{-12}
12	118412373	rs34965774	A	G	0.144	1.06	2.0×10^{-9}
12	118489636	rs12578639	A	T	0.828	1.04	2.2×10^{-6}
12	121297815	rs11065299	A	G	0.075	1.06	5.8×10^{-7}
12	121432117	rs56348580	G	C	0.689	1.05	2.3×10^{-13}
12	123450765	rs4148856	C	G	0.781	1.05	1.7×10^{-10}
12	124468572	rs7978610	G	C	0.666	1.27	2.0×10^{-8}
12	124509177	rs825452	A	G	0.603	1.04	2.4×10^{-6}

12	133069698	rs12811407	A	G	0.331	1.05	1.7×10^{-12}
13	26776999	rs34584161	A	G	0.76	1.05	2.2×10^{-10}
13	31042452	rs11842871	G	T	0.735	1.04	1.2×10^{-8}
13	33554302	rs576674	G	A	0.169	1.05	8.3×10^{-10}
13	51096095	rs963740	A	T	0.713	1.04	2.1×10^{-8}
13	58366634	rs9537803	C	T	0.277	1.04	4.6×10^{-8}
13	58965435	rs9569864	C	T	0.825	1.05	8.7×10^{-8}
13	59077406	rs9563615	A	T	0.71	1.05	6.4×10^{-11}
13	80717156	rs1359790	G	A	0.72	1.09	2.4×10^{-31}
13	109947213	rs7987740	T	C	0.609	1.04	4.0×10^{-8}
13	110431626	rs4771648	G	A	0.669	1.04	8.9×10^{-8}
14	23288935	rs17122772	G	C	0.228	1.04	1.6×10^{-8}
14	33302882	rs17522122	T	G	0.474	1.04	3.2×10^{-9}
14	38848419	rs8017808	G	T	0.743	1.04	2.1×10^{-8}
14	79932041	rs17836088	C	G	0.217	1.06	6.7×10^{-14}
14	103894071	rs62007683	G	T	0.653	1.04	3.1×10^{-8}
15	38834033	rs8032939	C	T	0.246	1.06	3.5×10^{-14}
15	41809205	rs11070332	A	G	0.358	1.05	1.1×10^{-13}
15	53091553	rs2456530	T	C	0.127	1.06	5.4×10^{-9}
15	57456802	rs117483894	G	A	0.037	1.1	3.9×10^{-8}
15	62394264	rs8037894	G	C	0.566	1.05	2.6×10^{-13}
15	63871292	rs7178762	C	T	0.46	1.04	5.4×10^{-10}
15	68080886	rs4776970	A	T	0.641	1.04	5.0×10^{-9}
15	75932129	rs13737	G	T	0.759	1.05	5.6×10^{-10}
15	77818128	rs1005752	A	C	0.715	1.08	2.5×10^{-29}
15	90423293	rs4932265	T	C	0.267	1.07	4.2×10^{-20}
15	91511260	rs12910825	G	A	0.361	1.05	1.6×10^{-15}
16	295795	rs6600191	T	C	0.825	1.06	9.3×10^{-13}
16	3583173	rs3751837	T	C	0.22	1.04	1.4×10^{-8}
16	28915217	rs8046545	G	A	0.359	1.04	1.9×10^{-8}
16	53800954	rs1421085	C	T	0.415	1.13	3.1×10^{-84}
16	69651866	rs862320	C	T	0.578	1.04	3.9×10^{-11}
16	75234872	rs72802342	C	A	0.923	1.17	$4.0 \times 10^{-3}2$
16	75516534	rs3115960	G	C	0.37	1.03	2.8×10^{-6}
16	81534790	rs2925979	T	C	0.3	1.05	1.4×10^{-14}
17	3828086	rs1043246	G	C	0.157	1.05	7.9×10^{-7}
17	3860356	rs3826482	A	T	0.576	1.03	2.1×10^{-7}
17	4045440	rs1377807	C	G	0.312	1.05	4.2×10^{-13}
17	7549681	rs1641523	C	T	0.428	1.05	1.2×10^{-10}
17	9785187	rs7222481	C	G	0.324	1.04	1.4×10^{-8}
17	17661802	rs4925109	A	G	0.316	1.05	2.8×10^{-12}
17	36046451	rs10962	C	G	0.226	1.05	9.9×10^{-10}
17	36099952	rs10908278	T	A	0.481	1.08	6.4×10^{-36}

17	40731411	rs34855406	C	G	0.277	1.05	2.3×10^{-12}
17	47060322	rs35895680	C	A	0.678	1.06	2.5×10^{-15}
17	61965043	rs2727301	T	C	0.754	1.04	1.3×10^{-6}
17	62203304	rs60276348	T	C	0.14	1.05	2.6×10^{-8}
17	65892507	rs61676547	C	G	0.192	1.06	2.9×10^{-11}
18	7070642	rs7240767	C	T	0.376	1.04	1.6×10^{-8}
18	53452144	rs28719468	C	T	0.159	1.04	1.9×10^{-6}
18	54675384	rs17684074	G	C	0.74	1.04	2.9×10^{-8}
18	56876228	rs9957145	G	A	0.829	1.05	8.1×10^{-9}
18	57848369	rs523288	T	A	0.238	1.05	7.6×10^{-13}
18	58056566	rs74452128	C	A	0.976	1.15	1.0×10^{-9}
18	60668270	rs10469140	G	A	0.485	1.03	6.6×10^{-6}
18	60845884	rs12454712	T	C	0.614	1.05	4.6×10^{-13}
19	4948862	rs7249758	A	G	0.204	1.05	3.4×10^{-9}
19	7240848	rs75253922	C	T	0.191	1.05	2.7×10^{-8}
19	7970635	rs4804833	A	G	0.39	1.05	7.7×10^{-13}
19	19388500	rs8107974	T	A	0.077	1.1	3.3×10^{-15}
19	33890838	rs10406327	C	G	0.523	1.04	3.8×10^{-8}
19	44938870		A	G	0.001	1.61	8.3×10^{-6}
19	45411941	rs429358	T	C	0.846	1.08	2.6×10^{-18}
19	46157019	rs10406431	A	G	0.563	1.05	9.6×10^{-14}
19	46178661	rs2238689	C	T	0.418	1.04	5.4×10^{-9}
19	47569003	rs3810291	A	G	0.673	1.05	8.9×10^{-12}
20	21466795	rs13041756	C	T	0.107	1.06	1.4×10^{-8}
20	32596704	rs2268078	A	G	0.657	1.04	2.3×10^{-10}
20	43001721	rs4810426	T	C	0.106	1.09	3.1×10^{-17}
20	43042364	rs1800961	T	C	0.035	1.18	2.3×10^{-22}
20	43233649	rs11696357	A	G	0.934	1.06	9.9×10^{-6}
20	45598564	rs6063048	G	A	0.725	1.05	2.2×10^{-11}
20	48832135	rs11699802	C	T	0.536	1.04	1.8×10^{-11}
20	51223594	rs34454109	A	T	0.771	1.04	7.1×10^{-9}
20	57394628	rs6070625	G	C	0.517	1.05	5.3×10^{-14}
20	62450664	rs6011155	T	C	0.63	1.04	6.3×10^{-6}
20	62693175	rs59944054	A	G	0.238	1.06	1.5×10^{-8}
22	30609554	rs6518681	G	A	0.914	1.09	1.1×10^{-12}
22	32348841	rs117001013	C	T	0.912	1.07	1.7×10^{-8}
22	41489920	rs5758223	A	G	0.717	1.04	3.8×10^{-8}
22	44324730	rs738408	T	C	0.226	1.05	1.4×10^{-10}
22	50356850	rs1801645	C	T	0.275	1.04	1.5×10^{-8}

[m] Chr is chromosome.
[n] Position is in base pairs (bp) on Genome Reference Consortium Human Build 37.
[o] A1 and a2 are the alleles from AGEN 2020, with the effect allele given first.
[p] A1f is the frequency of a1 in the DIAGRAM 2018 study population.
[q] OR is reported per copy of the effect allele.

Table 7.2.3. SNPs included in the PS derived from various ancestry-specific GWAS meta-analyses in DIAMANTE 2020, as well as the including the PECRB study population-specific weight PS.

chr[r]	position[s]	SNP	a1/a2[t]	OR_comp[u]	OR_afr	OR_eas	OR_eur	OR_his	OR_sas	freq_pop[v]	OR_pop[w]	p-value
1	20729451	rs10916784	G/C	1.03	1.03	1.03	1.03	1.02	1.06	0.49	1.05	3.1×10^{-1}
1	39870793	rs3768301	T/C	1.07	1.08	1.06	1.08	1.04	1.05	0.30	1.06	2.6×10^{-1}
1	46358862	rs34444543	G/A	1.04	1.02	1.06	1.03	1.01	1.02	0.61	1.05	3.0×10^{-1}
1	64114429	rs11576729	G/T	1.05	1.08	1.06	1.04	1.09	1.03	0.50	1.04	3.6×10^{-1}
1	117532790	rs1127215	C/T	1.04	1.02	1.02	1.05	1.06	1.05	0.21	0.96	4.7×10^{-1}
1	120455586	rs835576	C/T	1.07	1.03	1.09	1.08	1.09	1.03	0.10	0.94	4.6×10^{-1}
1	177889025	rs539515	C/A	1.06	1.04	1.06	1.06	1.06	1.03	0.20	1.03	5.4×10^{-1}
1	200416099	rs10919928	A/G	1.03	0.77	1.04	1.15	1.14	0.98	0.13	1.13	9.1×10^{-2}
1	204539291	rs6689629	A/G	1.04	1.05	1.04	1.03	1.03	1.04	0.98	0.93	6.7×10^{-1}
1	205107793	rs12039805	A/G	1.03	1.03	1.02	1.04	1.06	1.01	0.39	1.03	6.0×10^{-1}
1	206600992	rs9429893	A/G	1.03	1.08	1.02	1.04	1.02	1.03	0.48	1.03	5.5×10^{-1}
1	214159256	rs340874	C/T	1.05	1.01		1.06	1.04	1.04	0.35	1.04	4.2×10^{-1}
1	219748818	rs2820446	C/G	1.05	1.06	1.03	1.06	1.02	1.03	0.38	1.06	1.8×10^{-1}
1	229672955	rs348330	G/A	1.05	1.02	1.06	1.05	1.03	1.02	0.68	1.01	8.0×10^{-1}
2	422144	rs62107261	T/C	1.11	1.11		1.11	1.12	0.97	0.99	1.35	1.9×10^{-1}
2	653874	rs10188334	C/T	1.06	1.11		1.05	1.07	1.10	0.85	1.07	3.2×10^{-1}
2	25533568	rs55928417	G/T	1.03	0.99	1.03	1.04	1.04	1.05	0.84	1.06	4.1×10^{-1}
2	27730940	rs1260326	C/T	1.06	1.07	1.06	1.07	1.03	1.08	0.90	1.07	5.3×10^{-1}
2	45192080	rs12712928	C/G	1.01	0.95	1.07	0.97	0.95	1.00	0.35	0.91	8.8×10^{-2}
2	60586707	rs243018	G/C	1.06	1.09	1.06	1.06	1.06	1.05	0.74	1.02	6.8×10^{-1}
2	65284231	rs2540949	A/T	1.03	1.03	1.02	1.04	1.03	1.02	0.79	1.10	1.3×10^{-1}
2	65666674	rs6752053	T/C	1.05	1.04	1.05	1.05	1.05	1.05	0.73	0.93	1.9×10^{-1}
2	121347612	rs11688682	G/C	1.06	1.02	1.01	1.06	1.05	1.06	0.99	1.47	3.0×10^{-1}
2	121440218	rs10864859	T/G	1.06	1.03	1.04	1.06	1.05	1.13	0.65	1.02	7.1×10^{-1}
2	158390468	rs7594480	T/C	1.08	1.07	0.99	1.08	1.10	1.09	0.84	0.98	7.7×10^{-1}
2	161144055	rs1020731	A/G	1.03	1.04	1.02	1.04	1.06	0.98	0.94	1.14	1.6×10^{-1}
2	163649480	rs12614955	T/C	1.03	1.05	1.03	1.03	1.01	1.04	0.60	0.96	3.5×10^{-1}

2	165508389	rs10184004	C/T	1.07	1.11	1.05	1.06	1.08	1.09	0.97	1.14	3.8×10^{-1}
2	213818731	rs16849467	T/C	1.04	1.06	1.04	1.03	1.02	1.05	0.97	0.98	8.5×10^{-1}
2	227100490	rs2943648	G/A	1.09	1.08	1.05	1.10	1.07	1.08	0.99	1.08	7.3×10^{-1}
3	12329783	rs17036160	C/T	1.12	1.09	1.12	1.12	1.12	1.12	0.91	1.18	4.6×10^{-2} *
3	12490951	rs4684855	T/C	1.04	1.11	1.03	1.05	1.03	1.01	0.40	1.07	1.5×10^{-1}
3	23457080	rs13094957	T/C	1.07	1.02	1.11	1.07	1.05	1.08	0.95	0.99	8.2×10^{-1}
3	23632174	rs76435632	G/C	1.10	1.21	1.11	1.11	1.08	1.08	0.12	1.07	2.9×10^{-1}
3	50174197	rs2624847	G/T	1.03	0.97	1.04	1.04	1.04	1.00	0.13	1.07	3.6×10^{-1}
3	58338809	rs12629058	T/C	1.04	1.13	1.04	1.03	1.05	1.06	0.16	1.04	6.6×10^{-1}
3	63897215	rs2292662	C/T	1.06	1.02	1.09	1.07	1.01	1.01	0.98	1.01	9.0×10^{-1}
3	64703394	rs66815886	G/T	1.04	1.05	1.01	1.05	1.03	1.03	0.96	0.84	2.7×10^{-1}
3	114960798	rs1459513	C/A	1.05	0.98	1.06	1.06	1.06	1.01	0.16	1.00	6.3×10^{-1}
3	123065778	rs11708067	A/G	1.10	1.13	1.18	1.09	1.13	1.11	0.54	1.03	6.5×10^{-1}
3	124921457	rs9873519	T/C	1.04	1.05	1.04	1.04	1.01	1.05	0.29	1.01	8.4×10^{-1}
3	151998053	rs1426385	A/G	1.04	1.05	1.02	1.04	1.02	1.05	0.42	1.03	5.0×10^{-1}
3	152399693	rs10935897	A/G	1.03	1.04	1.03	1.04	1.06	1.01	0.18	0.93	2.2×10^{-1}
3	152530027	rs75417759	C/T	1.07	1.07	0.96	1.11	1.11	1.06	0.98	1.07	7.1×10^{-1}
3	170724883	rs8192675	T/C	1.06	0.97	1.06	1.07	1.01	1.05	0.76	1.08	1.7×10^{-1}
3	185510613	rs7633675	G/T	1.12	1.13	1.14	1.12	1.16	1.12	0.18	1.07	2.8×10^{-1}
3	186665645	rs3887925	T/C	1.05	1.03	1.04	1.06	1.05	1.05	0.14	1.09	2.0×10^{-1}
3	186676455	rs9799068	A/C	1.04	1.03	1.03	1.06	1.03	1.03	0.15	1.01	6.6×10^{-1}
3	195825077	rs74289356	T/C	1.05	1.07	1.06	1.04	1.06	1.01	0.41	0.96	5.4×10^{-1}
4	1240299	rs730831	T/G	1.09	1.09	1.11	1.08	1.11	1.04	0.53	1.02	7.6×10^{-1}
4	1784605	rs6831006	G/C	1.06	1.02	1.07	1.06	1.04	1.04	0.31	0.99	7.4×10^{-1}
4	6293237	rs9998835	G/C	1.08	0.99	1.12	1.09	1.13	1.06	0.94	1.05	6.5×10^{-1}
4	18047401	rs6855926	A/G	1.04	1.07	1.03	1.04	1.02	1.06	0.66	1.05	4.1×10^{-1}
4	45175691	rs13130484	T/C	1.04	1.08	1.03	1.04	1.04	1.02	0.30	1.13	1.8×10^{-2} *
4	71835822	rs7674402	A/G	1.06	1.04	1.07	1.06	1.06	1.06	0.15	1.03	6.2×10^{-1}
4	83587562	rs10471048	G/C	1.03	1.05	1.02	1.04	1.01	1.04	0.23	0.96	5.1×10^{-1}
4	102135363	rs2659518	A/G	1.04	1.06	1.03	1.05	1.07	1.03	0.40	1.07	1.7×10^{-1}
4	103725894	rs223423	G/A	1.02	1.04	1.00	1.04	1.07	0.98	0.30	1.03	5.6×10^{-1}

4	106048291	rs17035289	C/T	1.04	1.07	1.04	1.04	1.10	1.00	0.22	1.03	5.9×10^{-1}
4	153520475	rs6813195	C/T	1.06	1.08	1.07	1.06	1.08	1.06	0.53	1.10	5.3×10^{-2}
4	157725916	rs1425482	T/C	1.03	1.06	1.00	1.04	1.09	1.02	0.69	1.06	2.2×10^{-1}
5	14780521	rs30614	A/G	1.04	1.07		1.04	1.07	1.05	0.23	0.95	3.5×10^{-1}
5	44682589	rs6884702	G/A	1.03	0.99	1.02	1.04	1.06	1.02	0.52	1.00	8.4×10^{-1}
5	51791225	rs17261179	T/C	1.03	1.05	1.03	1.04	1.03	1.02	0.77	0.99	7.8×10^{-1}
5	53271420	rs702634	A/G	1.04	1.04	1.04	1.04	1.03	1.02	0.96	0.89	3.9×10^{-1}
5	53303595	rs6876198	C/T	1.04	1.07	1.02	1.04	1.07	1.04	0.24	1.00	8.9×10^{-1}
5	55810305	rs256904	T/A	1.08	1.11	1.09	1.09	1.07	1.05	0.77	1.17	8.7×10^{-3} *
5	55840633	rs42251	A/G	1.03	1.00	1.03	1.04	1.06	1.03	0.25	1.13	2.6×10^{-2} *
5	55860866	rs3936510	T/G	1.08	1.02	1.08	1.09	1.13	1.07	0.08	1.08	4.4×10^{-1}
5	67716793	rs57634870	G/T	1.04	1.04	1.04	1.04	1.04	1.05	0.23	0.99	7.3×10^{-1}
5	75003678	rs2307111	T/C	1.04	1.12	1.00	1.05	1.05	1.02	0.76	1.03	6.4×10^{-1}
5	76435004	rs7732130	G/A	1.06	1.15	1.08	1.06	1.04	1.03	0.35	1.11	4.1×10^{-2} *
5	78472599	rs10052346	G/T	1.04	1.08	1.02	1.04	1.07	1.02	0.47	0.96	4.5×10^{-1}
5	122704342	rs4267865	G/T	1.08	1.18	1.06	1.08	1.07	1.12	0.86	1.13	7.2×10^{-2}
5	133864599	rs329122	A/G	1.04	1.01	1.04	1.04	1.07	1.06	0.80	1.06	3.4×10^{-1}
5	176589585	rs244708	G/A	1.03	1.02	1.04	1.02	1.06	1.01	0.45	1.11	4.9×10^{-2} *
6	20680678	rs9348441	A/T	1.17	1.08	1.23	1.15	1.09	1.12	0.28	1.13	1.7×10^{-2} *
6	31139452	rs879882	C/T	1.04	0.97	1.05	1.04	1.07		0.60	1.02	6.2×10^{-1}
6	32373378	rs3806155	T/A	1.19	0.99	1.19	1.19	0.78		0.01	0.56	1.1×10^{-1}
6	32439077	rs7452864	C/T	1.04	1.14	1.00	1.05	1.02		0.97	1.10	6.6×10^{-1}
6	33524820	rs62405954	T/C	1.10	1.17	1.05	1.11	1.08		0.81	0.98	7.5×10^{-1}
6	34214670	rs4711389	A/G	1.08	1.30	1.13	1.05	1.07		0.51	1.06	2.9×10^{-1}
6	38992668	rs2281342	T/C	1.03	1.08	1.03	1.03	1.09	1.01	0.80	1.03	7.0×10^{-1}
6	39046644	rs742762	A/C	1.04	1.02	1.08	1.01	1.08	0.98	0.73	1.01	7.6×10^{-1}
6	39284184	rs3734618	G/A	1.04	1.05	1.07	1.03	1.05	1.05	0.15	1.17	2.6×10^{-2} *
6	40409243	rs34298980	T/C	1.03	1.01	1.02	1.04	1.07	1.02	0.41	1.06	2.8×10^{-1}
6	43758873	rs6905288	A/G	1.04	1.03	1.03	1.04	1.04	1.04	0.83	1.13	5.9×10^{-2}
6	43814190	rs6458354	C/T	1.04	1.03	1.04	1.05	1.05	0.99	0.06	1.01	8.2×10^{-1}
6	50788778	rs3798519	C/A	1.06	1.04	1.05	1.06	1.07	1.05	0.56	1.07	1.3×10^{-1}

6	107433400	rs1665901	A/T	1.04	1.00	1.04	1.04	1.02	1.03	0.66	0.98	6.7×10^{-1}
6	118011723	rs72951506	C/T	1.04	1.11	1.05	1.04	1.00	1.03	0.73	1.12	3.9×10^{-2} *
6	126792095	rs11759026	G/A	1.07	1.12	1.06	1.07	1.11	1.07	0.77	1.03	6.1×10^{-1}
6	127416930	rs2800733	A/G	1.05	1.04	1.07	1.05	1.05	1.04	0.84	1.03	6.5×10^{-1}
6	131954797	rs7739842	G/T	1.04	0.98	1.05	1.03	1.07	1.04	0.40	1.06	2.7×10^{-1}
6	137291281	rs6937795	A/C	1.04	0.98	1.04	1.05	1.01	1.02	0.63	1.02	6.6×10^{-1}
6	138855975	rs9376353	A/T	1.03	1.03	1.03	1.03	1.00	1.07	0.57	0.99	7.7×10^{-1}
6	143058692	rs6570526	G/C	1.03	1.03	1.03	1.03	1.04	1.02	0.49	0.98	7.1×10^{-1}
6	153438573	rs6932473	T/A	1.04	1.01	1.04	1.04	1.06	1.03	0.28	1.07	2.1×10^{-1}
6	160770360	rs539298	A/G	1.04	1.05	1.03	1.04	1.02	1.07	0.63	0.95	2.8×10^{-1}
6	164133001	rs4709746	C/T	1.05	1.02	1.05	1.06	1.04	0.99	0.50	1.09	6.1×10^{-2}
7	13887008	rs12154701	A/C	1.03	1.03	1.05	1.02	1.02	1.02	0.26	1.07	2.6×10^{-1}
7	14898282	rs17168486	T/C	1.06	1.01	1.07	1.07	1.05	1.05	0.83	1.05	5.1×10^{-1}
7	15062983	rs2215383	C/T	1.07	1.08	1.07	1.07	1.06	1.08	0.24	0.98	8.2×10^{-1}
7	28192280	rs849133	C/T	1.05	1.08		1.04	1.11	1.06	0.76	1.05	3.5×10^{-1}
7	28205303	rs552707	T/C	1.03	1.04		1.04	1.01	1.00	0.01	0.76	2.1×10^{-1}
7	30728452	rs917195	C/T	1.05	1.04	1.04	1.05	1.08	1.05	0.59	0.98	6.5×10^{-1}
7	44178829	rs882019	G/A	1.03	1.07	1.04	1.03	1.01	1.01	0.52	1.02	7.6×10^{-1}
7	44255643	rs878521	A/G	1.04	1.05	1.03	1.06	1.04	1.04	0.59	0.94	2.1×10^{-1}
7	50809085	rs13236710	G/A	1.05	1.05	1.06	1.05	1.09	0.97	0.93	1.02	8.8×10^{-1}
7	69055951	rs2533457	G/A	1.04	1.06	1.06	1.03	1.03	1.07	0.27	1.03	6.3×10^{-1}
7	89800241	rs6978118	A/T	1.03	1.05	1.07	1.02	1.04	1.01	0.39	1.07	1.9×10^{-1}
7	102481891	rs7781557	C/T	1.05	1.10	1.19	1.05	1.06	1.02	0.99	0.94	7.9×10^{-1}
7	130457914	rs1562396	G/A	1.04	1.01	1.03	1.06	1.04	1.00	0.33	1.06	2.6×10^{-1}
7	140631823	rs11983228	C/G	1.05	1.04	1.07	1.05	1.06	1.06	0.01	1.78	5.5×10^{-2}
7	150537635	rs62492368	A/G	1.03	1.00	1.03	1.04	1.03	1.01	0.53	0.97	6.0×10^{-1}
7	156794983	rs887609	A/G	1.03	1.01	1.06	1.02	1.10	0.99	0.25	1.01	7.8×10^{-1}
7	156992461	rs2366214	A/G	1.05	1.06	1.04	1.06	1.05	1.03	0.48	1.05	3.2×10^{-1}
8	10787612	rs4240673	T/C	1.04	1.03	1.00	1.04	1.02	1.05	0.89	0.99	8.5×10^{-1}
8	12618225	rs12680692	A/T	1.03	1.05	1.01	1.04	1.02	1.06	0.65	1.04	4.7×10^{-1}
8	36854711	rs10092900	G/T	1.04	1.06	1.04	1.03	1.03	1.06	0.16	1.05	4.3×10^{-1}

8	37397803	rs12680217	T/C	1.05	1.09	1.06	1.05	0.98	1.05	0.53	1.02	8.3×10^{-1}
8	41510260	rs12550613	C/G	1.04	1.07	1.05	1.03	1.01	1.04	0.72	0.96	4.6×10^{-1}
8	41522991	rs508419	G/A	1.05	1.03	1.07	1.05	1.06	1.06	0.92	0.98	7.7×10^{-1}
8	95965695	rs13257021	A/G	1.04	0.99	1.05	1.05	1.06	1.02	0.40	0.98	7.4×10^{-1}
8	116497173	rs800909	T/C	1.03	1.08	1.02	1.03	1.06	1.02	0.26	1.06	2.7×10^{-1}
8	118184783	rs13266634	C/T	1.12	1.08	1.13	1.11	1.10	1.11	0.87	1.15	5.9×10^{-2}
8	129569999	rs4733612	G/A	1.04	1.02	1.08	1.04	1.08	1.03	0.03	1.24	2.3×10^{-1}
8	145544720	rs3890400	A/G	1.04	1.04	1.04	1.05	1.04	1.04	0.11	1.03	6.8×10^{-1}
8	145972670	rs7014773	T/C	1.03	1.06	1.04	1.03	1.03	0.99	0.65	0.98	6.0×10^{-1}
9	4290085	rs4237150	C/G	1.05	1.01	1.06	1.04	1.09	1.07	0.68	1.12	2.8×10^{-2} *
9	4297892	rs4258054	T/C	1.03	1.00	1.03	1.04	1.03	1.01	0.24	0.91	9.6×10^{-2}
9	19074538	rs12380322	G/A	1.03	0.99	1.03	1.04	1.00	1.03	0.23	1.04	4.4×10^{-1}
9	21840834	rs7856455	G/T	1.04	0.98	1.04	1.10	1.01	1.03	0.76	1.03	6.8×10^{-1}
9	22133984	rs10757282	C/T	1.08	1.05	1.06	1.09	1.08	1.06	0.40	1.07	1.4×10^{-1}
9	22134094	rs10811661	T/C	1.17	1.08	1.16	1.19	1.15	1.20	0.92	1.22	3.2×10^{-2} *
9	28410683	rs1412234	C/T	1.03	1.07	1.01	1.04	1.02	1.06	0.22	0.98	6.5×10^{-1}
9	34074476	rs12001437	C/T	1.03	1.00	1.03	1.04	1.03	1.00	0.47	1.03	4.5×10^{-1}
9	81914978	rs13290396	C/T	1.10	1.12	1.12	1.10	1.09	1.10	0.70	1.09	8.6×10^{-2}
9	83998346	rs9332453	C/T	1.03	0.98	1.05	1.02	1.02	1.06	0.73	1.02	7.4×10^{-1}
9	84308948	rs2796441	G/A	1.07	1.01	1.08	1.07	1.06	1.05	0.45	1.03	4.9×10^{-1}
9	96971175	rs12345069	C/T	1.04	1.05	1.01	1.04	1.07	1.04	0.98	1.00	8.9×10^{-1}
9	98278413	rs113154802	C/T	1.05	1.05	1.07	1.03	1.24	1.08	0.97	1.12	3.8×10^{-1}
9	126015103	rs2416899	T/G	1.03	0.97	1.03	1.05	0.99	1.08	0.28	0.97	5.6×10^{-1}
9	136149229	rs505922	C/T	1.05	1.02	1.05	1.05	1.09	1.02	0.11	0.94	4.5×10^{-1}
9	139243334	rs28429551	A/T	1.09	1.08	1.16	1.08	1.08	1.10	0.72	1.14	1.9×10^{-2} *
9	139247229	rs74604683	C/T	1.05	1.23	1.06	1.05	1.08	1.03	0.87	0.97	7.8×10^{-1}
10	12307894	rs11257655	T/C	1.11	1.09	1.13	1.09	1.15	1.10	0.41	1.08	1.2×10^{-1}
10	26497704	rs7923442	A/G	1.04	1.04	1.04	1.03	1.03	1.05	0.81	1.07	2.2×10^{-1}
10	64974380	rs41274074	G/C	1.06	0.99	1.07	1.08	1.00	0.98	0.78	1.03	6.2×10^{-1}
10	71320943	rs190925	A/G	1.04	0.95	1.06	1.05	1.03	1.06	0.36	0.92	9.9×10^{-2}
10	71466578	rs2642588	G/T	1.05	0.99	1.16	1.05	1.01	1.07	0.70	1.00	8.6×10^{-1}

10	77244336	rs3012060	T/A	1.05	1.20	1.05	1.05	1.05	1.01	0.24	0.99	8.1×10^{-1}
10	80943841	rs703980	G/A	1.06	1.04	1.06	1.07	1.04	1.04	0.60	1.06	2.0×10^{-1}
10	89766368	rs10887775	A/G	1.04	1.01	1.04	1.04	1.02	1.07	0.29	1.02	6.9×10^{-1}
10	94460650	rs10882099	T/C	1.02	1.03	1.09	1.02	1.08	1.00	0.43	1.04	3.9×10^{-1}
10	94479107	rs1112718	A/G	1.02	0.99	1.07	1.01	1.01	1.02	0.40	1.04	5.0×10^{-1}
10	99056190	rs10748694	A/T	1.04	1.03	1.06	1.03	0.99	1.08	0.39	1.15	4.0×10^{-3} *
10	112621837	rs7067540	C/T	1.04	1.03	1.06	1.04	1.03	1.03	0.55	0.96	3.9×10^{-1}
10	114344288	rs12243296	G/A	1.04		1.22	1.03	1.09	1.00	0.00	0.69	5.6×10^{-1}
10	114381965	rs7100404	C/T	1.05		1.37	1.05	1.08	0.99	1.00	0.71	4.9×10^{-1}
10	114428364	rs2859885	C/T	1.05		1.01	1.07	0.97	1.00	0.60	1.04	4.5×10^{-1}
10	114552267	rs10787461	G/A	1.04		0.86	1.05	1.00	1.01	0.45	1.00	8.3×10^{-1}
10	114715598	rs2104598	G/A	1.04		1.21	1.04	1.11	1.00	0.67	1.07	2.0×10^{-1}
10	114758349	rs7903146	T/C	1.33		1.18	1.35	1.29	1.26	0.09	1.02	6.5×10^{-1}
10	114797893	rs7076754	G/A	1.07		1.14	1.06	1.14	1.04	0.97	0.92	6.2×10^{-1}
10	114859416	rs7081841	G/C	1.05		1.00	1.05	1.07	1.00	0.43	1.00	8.6×10^{-1}
10	115016408	rs12257761	T/C	1.06		1.03	1.09	1.00	1.02	0.70	1.10	5.5×10^{-2}
10	115069951	rs11196296	T/C	1.02		0.91	1.16	0.83	0.98	0.17	0.94	3.8×10^{-1}
10	115247447	rs11596522	T/G	1.05			1.07	0.86	1.01	0.02	1.03	8.2×10^{-1}
10	122834572	rs11199753	G/T	1.05	1.13	1.07	1.04	1.06	0.99	0.87	0.99	8.4×10^{-1}
10	122909625	rs2172073	A/C	1.05	1.06	1.04	1.07	1.07	1.04	0.72	1.06	3.7×10^{-1}
10	122968964	rs11592107	A/G	1.03	1.05	1.04	1.03	1.05	1.03	0.01	1.53	1.0×10^{-1}
10	124167512	rs2421016	C/T	1.04	0.96	1.04	1.05	1.07	1.04	0.09	1.02	8.3×10^{-1}
11	2077271	rs76547628	T/C	1.04	1.01	0.95	1.06	1.10	1.00	0.12	1.18	2.0×10^{-2} *
11	2194420	rs10770142	G/C	1.07	1.14	1.10	1.07	1.07	1.08	0.46	1.28	1.8 × 10-7 *
11	2235129	rs4930050	G/A	1.06	1.15	0.89	1.02	1.15	1.00	0.46	0.89	1.8×10^{-2} *
11	2364549	rs800125	A/C	1.01	0.91	0.99	1.03	0.94	1.00	0.52	1.04	4.6×10^{-1}
11	2375458	rs79495865	G/A	1.02	1.13	0.94	1.04	1.07	1.02	0.66	0.96	3.5×10^{-1}
11	2579163	rs2283164	A/G	1.09	1.16	1.08	1.11	1.01	1.10	0.99	0.86	6.8×10^{-1}
11	2681072	rs151215	G/A	1.04	1.04	0.96	1.06	1.06	1.05	0.05	1.08	4.6×10^{-1}
11	2691500	rs231361	A/G	1.09	1.12	1.04	1.10	1.12	1.08	0.66	0.91	9.7×10^{-2}
11	2799679	rs2237884	T/C	1.04	1.03	1.07	1.04	1.10	0.96	0.55	1.05	3.7×10^{-1}

11	2857897	rs234866	G/A	1.05	1.10	1.03	1.05	1.14	1.08	0.96	1.06	7.1×10^{-1}
11	2858546	rs2237897	C/T	1.15	1.15	1.12	1.21	1.23	1.20	0.55	1.32	2.2 × 10-8 *
11	2908754	rs445084	G/A	1.04	1.07	1.03	1.04	1.12	1.04	0.41	1.05	2.8×10^{-1}
11	8654528	rs10769936	C/T	1.03	1.02	1.03	1.04	1.06	1.00	0.79	1.00	8.4×10^{-1}
11	17408630	rs5215	C/T	1.08	1.05	1.09	1.07	1.07	1.07	0.39	1.06	2.2×10^{-1}
11	27683618	rs4923464	C/T	1.03	1.13	1.05	1.02	0.99	1.02	0.84	0.98	7.9×10^{-1}
11	32927778	rs145678014	G/T	1.10	1.20	0.51	1.11	1.07	0.95	0.99	1.98	4.4×10^{-2} *
11	43816200	rs6485462	C/T	1.03	1.02	1.00	1.05	1.04	1.02	0.53	1.00	8.5×10^{-1}
11	49477266	rs6485981	T/C	1.04	0.98	1.03	1.06	1.08	1.07	0.14	1.00	8.2×10^{-1}
11	65326154	rs12789028	A/G	1.06	1.06	1.04	1.06	1.03	1.08	0.10	1.22	1.6×10^{-2} *
11	72460398	rs77464186	A/C	1.11	1.07	1.15	1.11	1.09	1.09	0.95	1.13	2.8×10^{-1}
11	76156973	rs61894507	G/A	1.04	1.00	1.05	1.03	1.05	1.07	0.94	0.86	1.9×10^{-1}
11	92708710	rs10830963	G/C	1.08	1.15	1.06	1.11	1.11	1.09	0.14	1.06	4.5×10^{-1}
11	93131667	rs11020308	A/C	1.04	1.02	1.03	1.04	0.98	1.06	0.26	1.02	7.4×10^{-1}
11	128040810	rs10893827	A/G	1.04	1.03	1.03	1.06	1.06	1.05	0.65	1.03	6.4×10^{-1}
11	128235252	rs7104712	C/A	1.04	1.03	1.04	1.04	1.06	1.00	0.56	1.01	8.2×10^{-1}
11	128389391	rs11819995	T/C	1.05	1.03	1.04	1.05	1.02	1.05	0.15	1.12	1.2×10^{-1}
12	4033222	rs10848960	G/C	1.04	0.98	0.60	1.05	0.99	1.02	0.98	1.29	1.7×10^{-1}
12	4382324	rs3812821	G/C	1.05	0.96	1.21	1.05	1.07	1.04	0.59	1.14	3.6×10^{-3} *
12	26474867	rs10842708	G/A	1.04	1.07	1.03	1.05	1.01	1.04	0.69	1.01	8.5×10^{-1}
12	27964996	rs12578595	C/T	1.07	1.02	1.07	1.08	1.05	1.05	0.75	0.92	1.2×10^{-1}
12	33370406	rs6488140	A/G	1.04	1.09	1.01	1.05	1.06	1.03	0.69	1.08	1.4×10^{-1}
12	50263148	rs7132908	A/G	1.03	1.00	1.04	1.03	1.03	1.00	0.08	1.02	7.8×10^{-1}
12	66255005	rs343093	G/C	1.07	1.12	1.07	1.08	1.09	1.05	0.23	1.10	9.3×10^{-2}
12	66360164	rs7970350	T/C	1.04	1.00	1.04	1.05	1.01	1.03	0.82	0.99	7.7×10^{-1}
12	71449521	rs7313668	T/G	1.04	0.97	1.05	1.05	1.03	1.02	0.04	1.20	1.5×10^{-1}
12	97851611	rs7972074	C/T	1.04	1.06	1.04	1.02	1.10	1.01	0.82	1.06	2.9×10^{-1}
12	108629780	rs1426371	G/A	1.05	1.04	1.05	1.05	1.04	1.04	0.82	1.01	8.4×10^{-1}
12	118412373	rs34965774	A/G	1.06	1.05	1.07	1.06	1.06	1.05	0.32	1.03	6.5×10^{-1}
12	121456616	rs61953351	G/T	1.06	0.98	1.06	1.07	1.12	0.99	0.98	0.91	5.8×10^{-1}
12	123618544	rs1790116	T/G	1.04	1.03	1.03	1.04	1.03	1.07	0.55	1.00	7.1×10^{-1}

12	124545435	rs2451321	C/G	1.03	1.06	1.03	1.03	1.06	1.04	0.75	1.03	6.2×10^{-1}
12	133069698	rs12811407	A/G	1.05	1.04	1.06	1.05	1.07	1.01	0.10	1.17	4.6×10^{-2} *
13	23309382	rs314879	C/T	1.04	1.08	1.04	1.04	1.05	1.04	0.02	1.27	2.0×10^{-1}
13	26776999	rs34584161	A/G	1.05	1.00	1.06	1.05	1.02	1.04	0.34	1.06	2.8×10^{-1}
13	33554587	rs2858980	G/A	1.06	1.02	1.07	1.06	1.06	1.07	0.31	1.12	2.4×10^{-2} *
13	51096095	rs963740	A/T	1.04	1.03	1.04	1.04	1.04	1.02	0.90	1.12	1.8×10^{-1}
13	54107583	rs9568868	T/G	1.05	1.02	1.05	1.04	1.06	1.07	0.62	0.99	8.4×10^{-1}
13	80707429	rs1215468	A/G	1.08	1.06	1.09	1.08	1.11	1.08	0.61	1.07	1.6×10^{-1}
13	91942919	rs34165267	C/T	1.05	1.04	1.10	1.04	1.06	1.02	0.99	1.40	2.1×10^{-1}
14	33303540	rs12883788	T/C	1.04	1.00	1.04	1.04	1.04	1.03	0.39	1.03	5.5×10^{-1}
14	38803756	rs2183237	G/A	1.04	1.03	1.04	1.04	1.10	1.03	0.45	1.01	8.1×10^{-1}
14	79944099	rs8008910	A/G	1.06	1.06	1.26	1.06	1.07	1.03	0.31	1.02	6.8×10^{-1}
14	101124721	rs12878003	G/A	1.04	1.06	1.01	1.04	1.14	1.04	0.94	1.18	1.1×10^{-1}
14	101255172	rs73347525	A/G	1.05	1.05	1.07	1.05	1.07	1.02	0.63	1.10	7.9×10^{-2}
14	101301866	rs1053900	C/T	1.03	1.03	1.02	1.03	1.08	1.01	0.14	0.94	3.8×10^{-1}
14	103252270	rs11160699	A/G	1.04	1.03	1.04	1.04	1.08	1.05	0.54	1.09	8.2×10^{-2}
15	38843887	rs28582094	G/A	1.04	1.05	1.03	1.05	1.06	1.01	0.56	1.10	5.7×10^{-2}
15	40616742	rs3743140	A/G	1.05	1.04	1.09	1.03	1.07	1.02	0.26	1.00	8.3×10^{-1}
15	41818917	rs1473781	A/G	1.03	1.10	1.00	1.05	1.02	1.02	0.61	1.02	5.9×10^{-1}
15	52517714	rs3825801	C/T	1.05	1.00	1.04	1.05	0.95	1.08	0.96	0.92	4.7×10^{-1}
15	62391608	rs7163757	C/T	1.06	1.03	1.09	1.05	1.05	1.05	0.38	0.93	1.3×10^{-1}
15	63871292	rs7178762	C/T	1.04	1.02	1.04	1.04	1.06	1.03	0.79	0.95	3.9×10^{-1}
15	68080886	rs4776970	A/T	1.04	1.00	1.05	1.04	1.03	1.03	0.10	1.07	4.7×10^{-1}
15	75815758	rs11636031	T/C	1.05	1.04	1.05	1.05	1.07	1.03	0.86	1.13	9.1×10^{-2}
15	77776562	rs952472	C/A	1.08	1.06	1.08	1.08	1.08	1.06	0.65	1.07	2.4×10^{-1}
15	90379632	rs6496609	C/A	1.07	1.13	1.06	1.07	1.01	1.06	0.08	1.19	6.6×10^{-2}
15	91513157	rs2890156	A/T	1.07	1.06	1.08	1.07	1.05	1.08	0.39	1.01	7.7×10^{-1}
15	93832067	rs7167984	G/A	1.04	1.04	1.06	1.03	1.05	1.01	0.53	1.05	2.9×10^{-1}
16	295795	rs6600191	T/C	1.04	1.04	1.03	1.06	1.02	1.04	0.47	1.04	4.8×10^{-1}
16	3613126	rs12445430	T/C	1.04	1.06	1.04	1.04	1.06	1.03	0.18	1.10	1.6×10^{-1}
16	53809123	rs55872725	T/C	1.13	1.11	1.15	1.13	1.15	1.07	0.15	1.15	4.8×10^{-2} *

16	69651866	rs862320	C/T	1.04	1.02	1.03	1.04	1.02	1.01	0.88	0.88	7.3×10^{-2}
16	73100308	rs6416749	C/T	1.04	1.01	1.06	1.03	1.05	1.01	0.21	1.01	8.4×10^{-1}
16	75243657	rs72802358	G/C	1.09	0.99	1.10	1.12	1.07	1.10	0.95	1.21	8.3×10^{-2}
16	81534790	rs2925979	T/C	1.05	1.03	1.04	1.06	1.02	1.08	0.07	0.97	6.8×10^{-1}
16	88554480	rs9937296	C/T	1.04	0.99	1.04	1.05	1.07	1.02	0.53	1.02	8.0×10^{-1}
17	3828086	rs1043246	G/C	1.05	1.00	1.04	1.05	1.04	1.08	0.57	1.00	7.9×10^{-1}
17	3988451	rs8071043	C/T	1.04	1.07	0.99	1.05	1.03	1.02	0.66	0.97	5.7×10^{-1}
17	6953155	rs113748381	A/G	1.11	1.08	1.13	1.08	1.13	0.81	0.40	1.02	7.0×10^{-1}
17	17751478	rs1108646	A/G	1.04	0.99	1.04	1.04	1.05	1.04	0.64	1.01	8.4×10^{-1}
17	29704002	rs1048317	T/C	1.04	1.04	1.05	1.03	1.06	1.02	0.83	1.15	2.3×10^{-2} *
17	36043653	rs3094515	C/T	1.04	1.06	1.03	1.05	1.04	1.02	0.75	1.11	5.1×10^{-2}
17	36056076	rs12449654	C/G	1.05	1.04	1.08	1.04	1.12	1.01	0.45	1.05	4.0×10^{-1}
17	36099952	rs10908278	T/A	1.09	1.00	1.14	1.08	1.04	1.08	0.33	1.09	9.6×10^{-2}
17	40696915	rs684214	T/C	1.04	1.08	1.03	1.05	1.04	1.02	0.05	1.07	5.3×10^{-1}
17	47060322	rs35895680	C/A	1.05	1.04	1.04	1.06	1.01	1.02	0.74	1.02	7.3×10^{-1}
17	62203128	rs57676627	T/C	1.06	1.02	1.41	1.06	1.07	1.09	0.02	0.99	8.6×10^{-1}
17	65957568	rs9899520	A/G	1.04	0.97	1.04	1.05	1.01	1.04	0.38	1.06	2.7×10^{-1}
17	76792179	rs1044486	G/A	1.04	1.07	1.04	1.03	1.06	1.04	0.56	1.08	1.4×10^{-1}
18	7076836	rs9948462	T/C	1.04	0.99	1.05	1.04	1.03	1.05	0.72	0.98	7.2×10^{-1}
18	56876430	rs9957320	G/T	1.05	1.00	1.02	1.06	1.09	1.04	0.58	1.05	3.8×10^{-1}
18	57829135	rs6567160	C/T	1.07	1.08	1.09	1.06	1.13	1.06	0.03	0.84	3.3×10^{-1}
18	60845884	rs12454712	T/C	1.05	1.02	1.05	1.05	1.06	1.03	0.64	1.05	3.0×10^{-1}
19	4951064	rs262549	G/C	1.05	1.03	1.05	1.05	1.05	1.04	0.09	1.10	2.9×10^{-1}
19	7968168	rs2115107	A/G	1.05	1.07	1.05	1.05	1.04	1.02	0.33	1.01	8.2×10^{-1}
19	12509536	rs4804181	A/C	1.04	1.01	1.04	1.04	1.04	1.03	0.66	1.02	6.9×10^{-1}
19	19379549	rs58542926	T/C	1.07	1.05	1.03	1.09	1.12	1.04	0.02	1.07	7.2×10^{-1}
19	33890838	rs10406327	C/G	1.04	1.02	1.07	1.04	1.00	1.03	0.76	1.02	6.8×10^{-1}
19	45326768	rs1871045	T/C	1.03	1.04	1.04	1.03	1.04	1.05	0.38	1.08	1.1×10^{-1}
19	45411941	rs429358	T/C	1.06	1.04	1.03	1.07	1.05	1.04	0.85	0.88	7.1×10^{-2}
19	46157019	rs10406431	A/G	1.05	1.02	1.07	1.05	1.02	1.05	0.46	1.00	9.2×10^{-1}
19	46178661	rs2238689	C/T	1.04	1.00	1.04	1.04	1.03	1.03	0.42	1.06	2.3×10^{-1}

19	47569003	rs3810291	A/G	1.05	1.05	1.04	1.04	1.09	1.04	0.63	1.05	4.4×10^{-1}
20	22427370	rs2181063	C/G	1.03	1.03	1.05	1.03	0.98	1.02	0.83	1.05	4.5×10^{-1}
20	32674967	rs4911405	T/C	1.04	1.08	1.01	1.04	1.07	1.03	0.65	1.04	4.2×10^{-1}
20	42994812	rs12625671	C/T	1.08	1.14	1.08	1.08	1.08	1.11	0.81	1.05	5.1×10^{-1}
20	43042364	rs1800961	T/C	1.16	1.10	1.15	1.18	1.14	1.07	0.03	1.39	1.4×10^{-2} *
20	45596378	rs6063046	A/G	1.04	1.07	1.03	1.05	1.04	1.03	0.84	1.16	2.9×10^{-2} *
20	48832020	rs6091115	T/C	1.05	1.04	1.05	1.04	1.06	1.05	0.52	1.06	2.3×10^{-1}
20	57387352	rs736266	T/A	1.03	1.03	1.03	1.04	1.02	0.99	0.70	1.07	2.4×10^{-1}
22	30205572	rs36575	C/T	1.08	1.08	0.80	1.08	1.08	1.07	0.99	0.96	8.1×10^{-1}
22	32203334	rs75307421	A/G	1.10	1.06	1.06	1.11	1.16	1.08	0.21	1.15	2.3×10^{-2} *
22	44324730	rs738408	T/C	1.03	1.03	1.02	1.04	1.06	1.02	0.81	0.96	4.5×10^{-1}
22	50356302	rs28691713	C/T	1.05	1.06	1.08	1.04	1.00	1.03	0.60	1.03	5.6×10^{-1}

[r] Chr is chromosome.
[s] Position is in base pairs (bp) on Genome Reference Consortium Human Build 37.
[t] A1/a2 are the alleles from DIAMANTE 2020, with the effect allele given first.
[u] OR is reported per copy of the effect allele, with all but OR_pop calculated for participants within each ancestry group included in the DIAMANTE 2020 GWAS meta-analysis: OR_comp for the multi-ancestry – composite PS, OR_afr for the multi-ancestry – African PS, OR_eas for the multi-ancestry – East Asian PS, OR_eur for the multi-ancestry – European PS, OR_his for the multi-ancestry – Latino/Hispanic PS, OR_sas for the multi-ancestry – South Asian PS. OR is not reported for all SNPs across analyses for various ancestry groups: see DIAMANTE 2020 methodologies for details (Mahajan et al., 2020).
[v] Freq_pop is the frequency of a1 in the present Indigenous study population.
[w] OR_pop is the OR calculated and cross-validated for the population-specific weight score using the same set of variants as the multi-ancestry composite PS.
*Indicates nominally significant ($p < 0.05$), and directionally consistent with multi-ancestry result, association in the current study population.

Table S4. SNPs included in the population-specific variant PS.

chr[x]	position[y]	SNP	subsets[z]	a1/a2[aa]	a1f[ab]	OR[ac]	p-value
1	3637812	rs2181487	1 > 2	C/A	0.41	1.28	9.9×10^{-5}
1	3675959	rs12131045	1 > 2	G/T	0.32	1.37	5.0×10^{-6}
1	7532704	rs1474950912	1 > 2	C/G	0.94	1.62	2.5×10^{-4}
1	10201624	rs77567553	1 > 2	A/G	0.85	1.36	5.5×10^{-4}
1	16322282	rs1024537880	2 > 1	G/A	0.08	1.65	2.0×10^{-5}
1	18793386	rs12041192	2 > 1	A/G	0.07	1.55	4.4×10^{-4}
1	20610244	rs55928751	2 > 1	A/G	0.26	1.28	6.0×10^{-4}
1	38590813	rs898985	1 > 2	C/T	0.76	1.29	5.5×10^{-4}
1	39130160	rs4474214	1 > 2	G/T	0.84	1.36	4.1×10^{-4}
1	58273678	rs72664686	2 > 1	T/G	0.49	1.25	4.2×10^{-4}
1	63673963	rs147237461	2 > 1	A/G	0.25	1.33	8.0×10^{-5}
1	63722576	rs141607445	2 > 1	G/A	0.25	1.35	4.0×10^{-5}
1	66602990	rs551000	1 > 2	C/T	0.12	1.43	2.8×10^{-4}
1	67826180	rs114967702	2 > 1	A/G	0.76	1.35	3.4×10^{-5}
1	67983243	rs10789236	1 > 2	C/T	0.47	1.28	1.7×10^{-4}
1	68225671	rs554355	1 > 2	A/G	0.20	1.32	4.6×10^{-4}
1	69459021	rs199961703	1 > 2	I/D	0.86	1.38	5.6×10^{-4}
1	71093805	rs8179355	1 > 2	C/A	0.63	1.33	2.2×10^{-5}
1	71172305	rs76562719	1 > 2	G/A	0.66	1.31	5.7×10^{-5}
1	89948844	rs9427984	1 > 2	A/G	0.59	1.26	4.7×10^{-4}
1	91402075	rs358694	1 > 2	C/T	0.36	1.28	2.2×10^{-4}
1	91526198	rs2625760	1 > 2	A/G	0.40	1.32	2.3×10^{-5}
1	98141714	rs189939377	2 > 1	C/T	0.07	1.52	4.5×10^{-4}
1	111727851	chr1:111727851	2 > 1	A/C	0.96	1.86	6.1×10^{-5}
1	115721655	rs10858066	2 > 1	C/T	0.55	1.29	9.8×10^{-5}
1	116403499	rs975972	1 > 2	C/T	0.20	1.34	2.3×10^{-4}
1	164492700	rs6677872	1 > 2	T/G	0.93	1.60	1.8×10^{-4}
1	169455435	rs2056926	2 > 1	C/G	0.20	1.39	1.3×10^{-5}
1	169529973	rs6029	2 > 1	T/C	0.19	1.39	3.0×10^{-5}
1	183709624	rs12407737	1 > 2	C/A	0.95	1.67	3.5×10^{-4}
1	183747518	rs78151551	1 > 2	A/G	0.95	1.67	3.3×10^{-4}
1	187743426	rs115672901	1 > 2	T/G	0.90	1.45	4.3×10^{-4}
1	193504097	rs12034778	2 > 1	T/A	0.84	1.36	3.1×10^{-4}
1	194775338	rs6656153	2 > 1	T/C	0.08	1.47	5.2×10^{-4}
1	202114469	rs11581162	1 > 2	G/A	0.10	1.48	3.3×10^{-4}
1	221303867	rs10863588	1 > 2	C/T	0.12	1.44	2.7×10^{-4}
1	230011750	rs149517808	1 > 2	T/C	0.14	1.38	5.3×10^{-4}
1	230486870	rs74634891	1 > 2	G/A	0.88	1.55	6.6×10^{-6}
1	230619350	rs12048353	1 > 2	C/T	0.34	1.31	9.2×10^{-5}
1	237215582	chr1:237215582	1 > 2	G/T	0.04	1.78	4.7×10^{-4}

1	237326254	rs111689687	1 > 2	G/T	0.05	1.78	9.3×10^{-5}
1	247363038	rs12130823	1 > 2	G/T	0.25	1.29	4.1×10^{-4}
2	3457459	rs12619324	1 > 2	G/A	0.83	1.34	5.4×10^{-4}
2	5689487	rs2882275	1 > 2	T/C	0.28	1.28	4.1×10^{-4}
2	16277283	rs138932039	2 > 1	C/T	0.26	1.33	7.8×10^{-5}
2	16336621	rs144177256	2 > 1	C/G	0.21	1.37	6.3×10^{-5}
2	28819516	rs10194430	1 > 2	A/G	0.34	1.26	6.4×10^{-4}
2	33243179	rs75886200	1 > 2	G/C	0.07	1.64	8.6×10^{-5}
2	33267614	rs13393464	2 > 1	T/C	0.88	1.42	1.3×10^{-4}
2	33626885	rs78541119	1 > 2	G/A	0.04	1.90	2.2×10^{-4}
2	33626957	rs609277	2 > 1	T/G	0.38	1.25	6.2×10^{-4}
2	42114002	rs13028888	2 > 1	C/T	0.39	1.25	4.6×10^{-4}
2	45572181	rs13036196	1 > 2	T/C	0.76	1.31	2.5×10^{-4}
2	45668873	rs3770252	1 > 2	G/T	0.81	1.32	5.7×10^{-4}
2	49641556	rs17835319	2 > 1	A/G	0.20	1.34	2.4×10^{-4}
2	55090468	rs35459932	2 > 1	A/G	0.84	1.37	2.2×10^{-4}
2	98481711	rs139613671	1 > 2	C/T	0.23	1.33	1.8×10^{-4}
2	98647080	rs143198148	1 > 2	C/G	0.19	1.33	5.2×10^{-4}
2	100862903	rs6737502	1 > 2	C/T	0.78	1.35	1.9×10^{-4}
2	106401187	chr2:106401187	2 > 1	D/I	0.01	2.79	5.1×10^{-4}
2	110282530	chr2:110282530	2 > 1	T/G	0.95	1.71	5.7×10^{-4}
2	111413674	chr2:111413674	2 > 1	G/A	0.96	1.82	2.9×10^{-4}
2	111881704	chr2:111881704	2 > 1	C/T	0.96	1.83	2.2×10^{-4}
2	121084993	rs73951213	2 > 1	A/G	0.92	1.50	2.8×10^{-4}
2	129213748	rs10175448	2 > 1	T/C	0.41	1.24	5.7×10^{-4}
2	130178361	rs66638135	2 > 1	I/D	0.67	1.26	4.8×10^{-4}
2	170981031	rs4668198	1 > 2	C/T	0.62	1.26	4.5×10^{-4}
2	174267408	rs10194659	1 > 2	C/G	0.31	1.34	3.1×10^{-5}
2	180454209	rs79903308	1 > 2	A/G	0.07	1.63	1.2×10^{-4}
2	180543780	rs1515286	1 > 2	A/C	0.18	1.42	2.2×10^{-5}
2	198169130	rs12464356	1 > 2	T/C	0.88	1.46	1.3×10^{-4}
2	198298448	rs150021501	1 > 2	I/D	0.88	1.41	4.5×10^{-4}
2	198515254	rs116121900	1 > 2	G/A	0.86	1.40	2.7×10^{-4}
2	199679110	rs10189905	1 > 2	G/T	0.24	1.31	4.3×10^{-4}
2	207412722	rs3217266	2 > 1	I/D	0.80	1.37	4.2×10^{-5}
2	207518184	rs10186535	2 > 1	T/C	0.78	1.29	5.9×10^{-4}
2	208333760	chr2:208333760	2 > 1	C/T	0.95	1.67	4.3×10^{-4}
2	215243598	rs6705674	1 > 2	C/T	0.41	1.26	3.6×10^{-4}
2	215401934	rs77091556	1 > 2	T/G	0.27	1.33	8.1×10^{-5}
2	216023682	rs7419736	1 > 2	G/A	0.75	1.35	6.1×10^{-5}
2	216138939	rs112349663	1 > 2	I/D	0.80	1.39	4.6×10^{-5}
2	220821732	rs80349138	1 > 2	A/C	0.84	1.39	1.4×10^{-4}
2	221639854	rs6755421	2 > 1	A/G	0.64	1.26	4.3×10^{-4}

2	222027666	rs2019083	1 > 2	C/T	0.59	1.25	5.8×10^{-4}
2	225105625	chr2:225105625	1 > 2	G/C	0.96	1.79	3.2×10^{-4}
2	226360771	rs1963400	2 > 1	T/G	0.80	1.30	5.7×10^{-4}
2	226857138	rs138124232	1 > 2	C/A	0.87	1.39	5.9×10^{-4}
2	226979415	rs139975434	1 > 2	A/G	0.83	1.35	6.2×10^{-4}
2	227131174	rs41504645	1 > 2	T/C	0.86	1.39	3.8×10^{-4}
2	230284815	rs2396663	2 > 1	T/C	0.72	1.28	4.5×10^{-4}
2	232190769	rs56678553	1 > 2	A/G	0.65	1.32	3.8×10^{-5}
2	239775835	rs150530918	1 > 2	G/A	0.12	1.43	3.1×10^{-4}
2	242490658	rs13408032	1 > 2	T/C	0.47	1.25	4.6×10^{-4}
2	242577950	rs17140248	1 > 2	A/G	0.39	1.25	4.8×10^{-4}
3	1293504	rs3772339	1 > 2	A/C	0.29	1.29	3.3×10^{-4}
3	7589639	rs117534611	2 > 1	A/G	0.14	1.42	1.0×10^{-4}
3	9977244	rs3894571	2 > 1	T/C	0.30	1.30	9.0×10^{-5}
3	15382870	rs2103079	1 > 2	T/C	0.77	1.32	1.7×10^{-4}
3	39844963	rs62261638	1 > 2	G/A	0.33	1.31	1.0×10^{-4}
3	39966748	rs11717034	1 > 2	G/A	0.30	1.30	1.8×10^{-4}
3	57038055	rs183677269	2 > 1	C/A	0.86	1.36	4.8×10^{-4}
3	57151483	rs189430546	2 > 1	T/C	0.86	1.37	5.0×10^{-4}
3	58822664	rs79407372	1 > 2	G/A	0.87	1.40	4.6×10^{-4}
3	59777427	rs67777617	1 > 2	G/T	0.56	1.25	4.0×10^{-4}
3	60742058	rs13063937	1 > 2	C/A	0.73	1.31	1.8×10^{-4}
3	60833504	rs76349782	1 > 2	T/A	0.86	1.43	9.2×10^{-5}
3	62531625	rs17695565	1 > 2	T/C	0.82	1.38	1.3×10^{-4}
3	64835496	rs73119726	2 > 1	C/T	0.87	1.45	7.1×10^{-5}
3	65023127	rs4688534	2 > 1	C/A	0.24	1.29	4.8×10^{-4}
3	73701944	rs6781766	1 > 2	T/C	0.33	1.27	3.8×10^{-4}
3	73860347	rs61663855	1 > 2	D/I	0.13	1.39	5.4×10^{-4}
3	89994954	chr3:89994954	2 > 1	C/T	0.07	1.53	4.6×10^{-4}
3	96503627	rs75849815	1 > 2	T/A	0.10	1.45	5.8×10^{-4}
3	98562718	rs79834212	1 > 2	G/C	0.41	1.28	1.8×10^{-4}
3	98736435	rs80040807	1 > 2	A/C	0.38	1.27	3.3×10^{-4}
3	105005089	rs16851027	1 > 2	C/T	0.30	1.29	3.6×10^{-4}
3	105047616	rs10575418	1 > 2	I/D	0.63	1.26	5.0×10^{-4}
3	108652222	rs5851624	1 > 2	D/I	0.31	1.27	5.5×10^{-4}
3	112556904	rs6795521	2 > 1	G/C	0.31	1.30	9.0×10^{-5}
3	112914615	rs972554	2 > 1	A/G	0.32	1.25	6.1×10^{-4}
3	116317395	rs1518327	1 > 2	T/G	0.33	1.29	1.7×10^{-4}
3	131374802	rs9861011	1 > 2	G/A	0.73	1.29	4.6×10^{-4}
3	131835569	rs13074860	2 > 1	C/T	0.15	1.37	3.6×10^{-4}
3	133258597	rs4533654	1 > 2	T/G	0.73	1.32	8.2×10^{-5}
3	140296782	rs75476344	2 > 1	A/G	0.20	1.30	5.9×10^{-4}
3	141031298	rs79507063	2 > 1	T/G	0.82	1.34	3.7×10^{-4}

3	143806538	rs16854922	1 > 2	G/C	0.88	1.41	5.2×10^{-4}
3	148122713	rs4321499	2 > 1	G/T	0.21	1.32	3.9×10^{-4}
3	149050925	rs1526786	2 > 1	T/C	0.24	1.29	3.1×10^{-4}
3	167082105	rs79573605	2 > 1	G/C	0.10	1.44	4.4×10^{-4}
3	172707771	rs76050108	1 > 2	A/T	0.92	1.58	1.2×10^{-4}
3	174762060	rs9863629	2 > 1	T/C	0.86	1.40	1.6×10^{-4}
3	176927519	rs62296619	2 > 1	C/T	0.78	1.30	3.4×10^{-4}
3	181080833	rs118012541	2 > 1	G/A	0.84	1.37	2.5×10^{-4}
3	181505245	rs74909304	1 > 2	C/A	0.46	1.31	3.9×10^{-5}
3	184118600	rs7648852	1 > 2	G/A	0.40	1.30	6.9×10^{-5}
3	185680144	rs10937233	1 > 2	T/C	0.30	1.27	6.0×10^{-4}
3	185877745	rs138925570	1 > 2	G/A	0.93	1.61	1.0×10^{-4}
3	186346272	rs73061438	1 > 2	G/A	0.42	1.28	1.8×10^{-4}
3	188149586	rs7649407	2 > 1	A/G	0.87	1.38	4.8×10^{-4}
3	189048167	rs4687038	2 > 1	C/T	0.34	1.32	2.9×10^{-5}
3	194400812	rs9820108	1 > 2	A/G	0.33	1.27	5.7×10^{-4}
4	3682614	chr4:3682614	2 > 1	T/C	0.90	1.42	6.4×10^{-4}
4	8755903	rs140893109	2 > 1	G/A	0.94	1.62	2.4×10^{-4}
4	9902230	rs4235344	2 > 1	A/G	0.13	1.38	4.8×10^{-4}
4	10535135	rs117906214	2 > 1	G/A	0.20	1.30	5.4×10^{-4}
4	11572999	rs2191065	2 > 1	G/A	0.89	1.42	4.1×10^{-4}
4	14865904	rs2047203	2 > 1	A/T	0.19	1.32	6.6×10^{-4}
4	24280616	rs59580288	2 > 1	D/I	0.22	1.32	2.0×10^{-4}
4	33764638	rs144917101	1 > 2	G/A	0.88	1.44	2.1×10^{-4}
4	34075738	rs116018934	1 > 2	A/C	0.89	1.42	3.8×10^{-4}
4	42123860	rs137880547	1 > 2	T/C	0.91	1.50	5.0×10^{-4}
4	68091816	rs28528999	2 > 1	T/C	0.72	1.27	4.3×10^{-4}
4	78552172	rs28884780	1 > 2	A/G	0.34	1.32	4.3×10^{-5}
4	84052440	rs41332145	1 > 2	C/T	0.70	1.27	4.7×10^{-4}
4	90095095	rs180910763	1 > 2	C/T	0.80	1.34	2.4×10^{-4}
4	91765642	rs78416641	1 > 2	A/C	0.05	1.68	2.2×10^{-4}
4	121486977	rs36072791	1 > 2	A/G	0.82	1.32	6.2×10^{-4}
4	126112363	rs186374925	1 > 2	G/T	0.95	1.74	1.3×10^{-4}
4	136750399	rs13139177	2 > 1	G/A	0.59	1.24	6.3×10^{-4}
4	138486135	rs188206537	2 > 1	A/G	0.17	1.50	1.6×10^{-6}
4	138534031	rs141998354	2 > 1	T/C	0.19	1.45	4.6×10^{-6}
4	148175509	rs181321827	1 > 2	A/G	0.88	1.49	7.7×10^{-5}
4	166861524	rs28669377	2 > 1	T/A	0.90	1.50	2.0×10^{-4}
4	168716957	rs191684316	2 > 1	C/T	0.06	1.80	1.5×10^{-5}
4	178830486	rs188918690	2 > 1	G/C	0.11	1.41	5.9×10^{-4}
4	184835242	rs10866265	1 > 2	C/T	0.31	1.27	5.8×10^{-4}
5	2089792	rs139754007	2 > 1	D/I	0.38	1.24	6.3×10^{-4}
5	2902999	rs2860294	2 > 1	A/G	0.50	1.25	2.9×10^{-4}

5	18835357	rs114710558	1 > 2	T/C	0.12	1.40	5.4×10^{-4}
5	19488982	rs13170473	2 > 1	G/A	0.44	1.24	5.3×10^{-4}
5	27367354	rs59371379	1 > 2	I/D	0.87	1.42	1.4×10^{-4}
5	27432606	rs4367265	1 > 2	A/G	0.86	1.42	1.3×10^{-4}
5	27609865	rs55668326	1 > 2	T/G	0.85	1.45	4.0×10^{-5}
5	31660006	rs6861360	1 > 2	G/A	0.36	1.28	2.0×10^{-4}
5	67169005	rs141412809	2 > 1	C/G	0.10	1.45	4.6×10^{-4}
5	67950535	rs10076830	2 > 1	T/C	0.94	1.61	1.9×10^{-4}
5	73221549	rs6453032	2 > 1	G/T	0.47	1.27	1.1×10^{-4}
5	98388026	rs59910509	2 > 1	A/G	0.12	1.38	6.4×10^{-4}
5	109204107	rs2241690	2 > 1	T/G	0.66	1.26	4.4×10^{-4}
5	109947984	rs117569584	2 > 1	G/A	0.94	1.57	4.7×10^{-4}
5	118705545	rs32652	2 > 1	T/G	0.38	1.27	2.0×10^{-4}
5	128130002	rs62391237	2 > 1	C/T	0.93	1.61	6.9×10^{-5}
5	128313080	rs2577423	2 > 1	T/C	0.71	1.32	6.5×10^{-5}
5	129700024	rs73239262	2 > 1	T/C	0.95	1.66	1.6×10^{-4}
5	130810992	rs149151967	2 > 1	G/T	0.95	1.64	6.5×10^{-4}
5	135215554	rs2269927	2 > 1	C/T	0.92	1.49	4.1×10^{-4}
5	136932292	chr5:136932292	2 > 1	C/T	0.05	1.73	2.1×10^{-4}
5	150881537	rs58610768	2 > 1	G/A	0.86	1.36	5.6×10^{-4}
5	163661781	rs73799291	2 > 1	G/C	0.06	1.59	3.7×10^{-4}
5	173273681	rs3849720	2 > 1	T/C	0.23	1.39	1.6×10^{-5}
6	3030410	rs6596934	1 > 2	C/G	0.29	1.28	5.5×10^{-4}
6	9512507	rs2327189	1 > 2	T/C	0.29	1.29	3.4×10^{-4}
6	15936733	rs6936104	2 > 1	G/A	0.91	1.43	6.2×10^{-4}
6	17850142	rs676296	1 > 2	A/G	0.79	1.31	5.3×10^{-4}
6	21506530	rs6927607	2 > 1	T/C	0.75	1.29	4.1×10^{-4}
6	24270120	rs117831339	1 > 2	G/A	0.19	1.32	6.0×10^{-4}
6	31481299	rs2516399	1 > 2	A/G	0.84	1.35	4.6×10^{-4}
6	37649437	rs62398397	1 > 2	G/A	0.55	1.27	2.6×10^{-4}
6	40547677	rs115546979	1 > 2	T/C	0.94	1.64	2.6×10^{-4}
6	43233482	chr6:43233482	2 > 1	T/C	0.99	3.27	3.9×10^{-5}
6	43413575	rs369686699	2 > 1	G/A	0.99	3.26	4.5×10^{-5}
6	45404230	rs1997992	1 > 2	T/C	0.29	1.31	1.2×10^{-4}
6	45922402	rs3777591	1 > 2	G/A	0.23	1.30	5.2×10^{-4}
6	64732017	rs9344660	1 > 2	G/A	0.09	1.48	4.9×10^{-4}
6	70767320	rs3805997	1 > 2	G/A	0.25	1.31	3.2×10^{-4}
6	79912072	chr6:79912072	1 > 2	G/T	0.03	2.03	4.7×10^{-4}
6	80178863	rs2803183	1 > 2	T/C	0.42	1.26	4.6×10^{-4}
6	93467089	chr6:93467089	1 > 2	A/C	0.05	1.64	5.3×10^{-4}
6	101940762	rs2579944	1 > 2	A/G	0.74	1.32	1.5×10^{-4}
6	103295786	chr6:103295786	1 > 2	G/A	0.96	1.85	2.7×10^{-4}
6	117832836	rs9387485	1 > 2	C/T	0.10	1.52	8.2×10^{-5}

6	139039029	rs9389610	1 > 2	G/A	0.57	1.26	3.8×10^{-4}
6	139759781	rs3010312	2 > 1	C/A	0.63	1.25	4.4×10^{-4}
6	151353805	rs12183135	1 > 2	C/G	0.06	1.63	4.7×10^{-4}
6	154668204	rs4870278	2 > 1	T/C	0.82	1.34	2.8×10^{-4}
6	155315556	rs13199643	1 > 2	G/A	0.37	1.33	1.2×10^{-5}
6	156494674	rs7774516	1 > 2	C/G	0.46	1.30	3.5×10^{-5}
6	156549813	rs9384441	1 > 2	A/G	0.24	1.40	8.8×10^{-6}
6	158295488	rs181880963	2 > 1	T/A	0.94	1.57	4.5×10^{-4}
6	164360807	rs12193589	1 > 2	A/G	0.40	1.27	2.4×10^{-4}
6	166896330	rs6929010	1 > 2	T/C	0.46	1.25	5.4×10^{-4}
7	1632513	rs13235561	1 > 2	A/C	0.16	1.38	1.7×10^{-4}
7	3078035	rs35049983	1 > 2	A/G	0.44	1.25	4.4×10^{-4}
7	3252555	rs34932678	2 > 1	A/G	0.67	1.32	2.7×10^{-5}
7	4224116	chr7:4224116	2 > 1	A/G	0.02	2.35	4.4×10^{-4}
7	7187816	rs150130341	2 > 1	T/C	0.92	1.52	4.2×10^{-4}
7	8189620	rs3757524	1 > 2	C/T	0.21	1.32	3.7×10^{-4}
7	9619380	rs2259306	2 > 1	C/T	0.58	1.24	5.5×10^{-4}
7	13907827	rs251994	2 > 1	C/T	0.53	1.24	3.7×10^{-4}
7	13945127	rs2237295	2 > 1	C/T	0.41	1.25	3.8×10^{-4}
7	20021609	rs6956048	1 > 2	C/A	0.52	1.25	4.7×10^{-4}
7	23997775	rs57329175	2 > 1	C/T	0.93	1.52	5.8×10^{-4}
7	24958042	rs120	2 > 1	A/G	0.25	1.30	1.9×10^{-4}
7	26122034	rs17153193	2 > 1	T/C	0.70	1.30	1.3×10^{-4}
7	28485676	rs17156696	2 > 1	T/A	0.32	1.26	5.3×10^{-4}
7	28547048	rs12333835	1 > 2	C/T	0.23	1.33	1.8×10^{-4}
7	29668519	rs16875105	2 > 1	A/G	0.19	1.31	4.6×10^{-4}
7	38673833	rs9918553	2 > 1	A/G	0.67	1.25	6.7×10^{-4}
7	43272084	rs181157662	2 > 1	T/C	0.18	1.35	1.8×10^{-4}
7	45982233	chr7:45982233	2 > 1	A/G	0.06	1.59	4.2×10^{-4}
7	51682401	rs4947627	1 > 2	A/G	0.16	1.36	3.9×10^{-4}
7	73141464	rs2030922	1 > 2	C/T	0.47	1.29	5.8×10^{-5}
7	93816686	rs80145019	1 > 2	C/A	0.29	1.30	2.0×10^{-4}
7	95897827	rs142857687	1 > 2	A/T	0.18	1.33	5.5×10^{-4}
7	103023537	rs10268808	1 > 2	A/C	0.34	1.26	5.2×10^{-4}
7	103193651	rs2711879	1 > 2	C/T	0.40	1.29	7.2×10^{-5}
7	121304227	rs187383253	2 > 1	C/A	0.91	1.48	2.2×10^{-4}
7	130404776	rs4067228	1 > 2	A/G	0.50	1.25	4.7×10^{-4}
7	135575461	rs62489095	2 > 1	C/A	0.07	1.53	3.4×10^{-4}
7	143602779	rs6947163	1 > 2	G/T	0.15	1.38	3.7×10^{-4}
7	149988544	rs147486244	2 > 1	G/A	0.63	1.28	1.4×10^{-4}
8	2838945	rs656466	2 > 1	C/G	0.12	1.39	5.5×10^{-4}
8	3026578	rs10503198	2 > 1	C/G	0.28	1.32	6.4×10^{-5}
8	3106370	rs79693896	2 > 1	T/A	0.88	1.40	5.5×10^{-4}

8	3290516	rs2406295	2 > 1	C/G	0.90	1.45	2.4×10^{-4}
8	3341226	rs79915250	2 > 1	C/T	0.93	1.52	4.2×10^{-4}
8	3904192	rs3849818	1 > 2	T/G	0.20	1.32	6.2×10^{-4}
8	6042400	rs1834203	2 > 1	T/C	0.50	1.27	1.3×10^{-4}
8	6709922	rs2951864	2 > 1	C/G	0.46	1.25	2.8×10^{-4}
8	8328109	rs2921039	1 > 2	G/C	0.11	1.44	3.3×10^{-4}
8	11439961	rs11250149	1 > 2	C/G	0.25	1.35	6.0×10^{-5}
8	12786947	chr8:12786947	1 > 2	T/C	0.05	1.75	2.1×10^{-4}
8	13209750	rs76989997	1 > 2	C/T	0.52	1.25	5.4×10^{-4}
8	13243084	rs1481679	1 > 2	A/G	0.55	1.26	4.2×10^{-4}
8	14443650	rs10086949	1 > 2	T/G	0.41	1.28	1.7×10^{-4}
8	27769687	rs35857051	1 > 2	T/C	0.39	1.25	6.0×10^{-4}
8	35956929	chr8:35956929	1 > 2	A/T	0.06	1.61	4.9×10^{-4}
8	55308198	rs12546306	1 > 2	C/T	0.68	1.27	6.2×10^{-4}
8	62495263	rs6471961	1 > 2	T/C	0.77	1.30	3.8×10^{-4}
8	62634563	rs149447795	1 > 2	T/C	0.19	1.37	1.7×10^{-4}
8	62663272	rs199622822	1 > 2	D/I	0.20	1.37	1.1×10^{-4}
8	70653945	rs182136050	2 > 1	T/G	0.16	1.38	1.3×10^{-4}
8	70692868	rs138718121	2 > 1	T/A	0.17	1.37	1.1×10^{-4}
8	81365072	chr8:81365072	1 > 2	A/G	0.93	1.67	1.1×10^{-4}
8	87327440	rs148141053	1 > 2	T/G	0.12	1.42	3.8×10^{-4}
8	89008404	rs180694224	2 > 1	A/G	0.92	1.48	5.1×10^{-4}
8	93586384	rs2595612	1 > 2	T/C	0.06	1.63	2.8×10^{-4}
8	99455167	rs201599771	1 > 2	D/I	0.88	1.47	7.2×10^{-5}
8	100002032	rs79434662	1 > 2	G/A	0.88	1.49	3.5×10^{-5}
8	100709154	rs140502344	1 > 2	A/G	0.87	1.41	2.9×10^{-4}
8	109912830	rs4734184	2 > 1	A/G	0.83	1.32	6.6×10^{-4}
8	111025983	rs55893742	1 > 2	A/G	0.03	2.08	5.5×10^{-5}
8	124987478	rs7465584	2 > 1	C/T	0.37	1.29	8.5×10^{-5}
8	136671060	rs4243847	2 > 1	A/G	0.41	1.25	4.8×10^{-4}
8	136682479	rs4075568	2 > 1	C/T	0.54	1.30	2.7×10^{-5}
9	538401	rs10975042	2 > 1	C/T	0.19	1.35	1.2×10^{-4}
9	4375454	rs2039194	1 > 2	G/C	0.82	1.33	6.0×10^{-4}
9	5498562	rs34719356	2 > 1	D/I	0.34	1.25	6.5×10^{-4}
9	8726153	rs1473822	2 > 1	C/T	0.66	1.28	1.6×10^{-4}
9	8924538	rs1368684	2 > 1	C/T	0.26	1.34	3.0×10^{-5}
9	9028285	rs324496	2 > 1	T/C	0.74	1.28	5.5×10^{-4}
9	9173304	rs968079	2 > 1	G/A	0.23	1.31	2.6×10^{-4}
9	9191639	rs62529522	2 > 1	T/C	0.68	1.30	7.5×10^{-5}
9	11876599	rs192161934	1 > 2	T/G	0.08	1.51	4.5×10^{-4}
9	25335125	rs113501731	1 > 2	A/C	0.15	1.42	7.5×10^{-5}
9	28393919	rs143327236	2 > 1	G/A	0.04	1.76	1.8×10^{-4}
9	29789814	rs950316	2 > 1	G/A	0.52	1.25	3.0×10^{-4}

9	32194564	rs188065424	2 > 1	G/A	0.75	1.31	2.2×10^{-4}
9	32320727	chr9:32320727	2 > 1	T/C	0.96	1.72	4.9×10^{-4}
9	80709044	rs62573096	2 > 1	G/A	0.26	1.29	4.5×10^{-4}
9	81158113	rs7027911	1 > 2	G/A	0.55	1.27	1.8×10^{-4}
9	87132349	rs35342431	2 > 1	I/D	0.78	1.30	4.8×10^{-4}
9	90719044	rs56257218	1 > 2	G/A	0.07	1.54	5.7×10^{-4}
9	98252899	chr9:98252899	1 > 2	D/I	0.86	1.38	5.0×10^{-4}
9	101490087	rs337570	2 > 1	C/T	0.62	1.25	4.8×10^{-4}
9	105695436	rs79931339	1 > 2	A/C	0.79	1.32	5.2×10^{-4}
9	113085232	rs17807186	2 > 1	G/A	0.35	1.28	1.5×10^{-4}
9	129468586	rs13299536	2 > 1	A/G	0.32	1.29	1.1×10^{-4}
9	135277259	rs186468471	1 > 2	T/C	0.99	2.78	3.1×10^{-4}
9	137514801	rs57187152	1 > 2	C/T	0.08	1.54	1.8×10^{-4}
9	138995207	rs59623632	1 > 2	A/G	0.28	1.31	1.4×10^{-4}
9	140143728	rs201094187	2 > 1	D/I	0.14	1.40	2.7×10^{-4}
10	8277297	rs192871141	2 > 1	A/C	0.91	1.44	5.0×10^{-4}
10	9034011	rs12776100	2 > 1	C/A	0.73	1.28	4.9×10^{-4}
10	13861399	rs72771044	1 > 2	A/T	0.55	1.29	1.0×10^{-4}
10	14134826	rs789767	1 > 2	T/G	0.68	1.30	1.3×10^{-4}
10	14818870	rs12256790	1 > 2	G/A	0.72	1.35	3.1×10^{-5}
10	14991430	rs7906967	1 > 2	C/A	0.32	1.33	5.1×10^{-5}
10	23217727	rs10828375	2 > 1	G/T	0.30	1.28	2.9×10^{-4}
10	23342159	rs4237359	2 > 1	G/A	0.60	1.28	1.0×10^{-4}
10	25528562	rs4749015	2 > 1	A/G	0.82	1.34	2.6×10^{-4}
10	32549517	rs76220129	1 > 2	C/G	0.07	1.59	2.8×10^{-4}
10	72276664	chr10:72276664	2 > 1	C/G	0.01	2.47	5.2×10^{-4}
10	72494629	rs4747092	1 > 2	G/A	0.45	1.31	2.4×10^{-5}
10	73334219	rs144584204	1 > 2	I/D	0.10	1.46	4.2×10^{-4}
10	73398722	rs61650587	1 > 2	G/A	0.78	1.32	2.3×10^{-4}
10	73483645	rs3802713	1 > 2	T/C	0.18	1.35	3.8×10^{-4}
10	78670661	rs1907746	1 > 2	G/A	0.81	1.34	2.4×10^{-4}
10	78773064	rs74948619	1 > 2	G/A	0.77	1.32	2.3×10^{-4}
10	87300091	rs1912299	2 > 1	T/C	0.74	1.28	4.7×10^{-4}
10	89474072	rs2302404	1 > 2	C/T	0.90	1.44	4.9×10^{-4}
10	89593380	rs151154388	1 > 2	T/C	0.87	1.44	1.2×10^{-4}
10	92756381	rs35945428	2 > 1	G/T	0.64	1.26	3.3×10^{-4}
10	95982217	rs4918159	2 > 1	A/G	0.48	1.24	4.1×10^{-4}
10	97484887	rs34826225	2 > 1	I/D	0.70	1.31	6.3×10^{-5}
10	97603178	rs3181129	2 > 1	T/G	0.72	1.28	2.8×10^{-4}
10	97734022	rs10509696	2 > 1	T/C	0.65	1.25	4.4×10^{-4}
10	101366135	rs5787355	1 > 2	D/I	0.39	1.28	1.7×10^{-4}
10	101433770	rs55714089	1 > 2	D/I	0.40	1.25	4.4×10^{-4}
10	101558746	rs2756109	1 > 2	G/T	0.55	1.27	2.2×10^{-4}

10	114272761	rs12780297	1 > 2	T/G	0.19	1.32	6.3×10^{-4}
10	121029044	rs61874482	2 > 1	A/G	0.10	1.46	2.8×10^{-4}
10	123804177	rs192412818	2 > 1	G/A	0.90	1.47	2.2×10^{-4}
10	131059102	rs12573583	2 > 1	C/G	0.92	1.61	2.4×10^{-5}
11	2189185	rs4074905	2 > 1	G/A	0.21	1.31	4.1×10^{-4}
11	2206387	rs11564707	1 > 2	G/C	0.43	1.39	2.4×10^{-7}
11	2286243	rs739543	1 > 2	A/G	0.44	1.26	3.3×10^{-4}
11	2857194	rs2237895	1 > 2	C/A	0.49	1.40	1.9×10^{-7}
11	2868868	rs79788804	2 > 1	A/G	0.36	1.30	6.2×10^{-5}
11	2931981	rs450208	1 > 2	T/G	0.69	1.33	3.3×10^{-5}
11	5463425	rs16931041	2 > 1	T/C	0.81	1.32	3.0×10^{-4}
11	11205233	rs2722772	2 > 1	A/G	0.50	1.28	6.5×10^{-5}
11	12309002	rs3736304	1 > 2	C/T	0.57	1.28	1.5×10^{-4}
11	20388097	rs7928503	1 > 2	G/A	0.73	1.29	3.8×10^{-4}
11	22781497	rs12422017	2 > 1	T/C	0.63	1.26	3.8×10^{-4}
11	27526304	rs114388342	2 > 1	T/C	0.04	1.72	5.2×10^{-4}
11	43255059	rs140092128	1 > 2	G/C	0.13	1.40	4.6×10^{-4}
11	61183329	rs17156014	1 > 2	C/T	0.25	1.30	4.9×10^{-4}
11	62528410	chr11:62528410	1 > 2	I/D	0.98	2.18	3.8×10^{-4}
11	63017480	rs118020373	1 > 2	C/A	0.73	1.31	2.4×10^{-4}
11	63215235	rs116986544	1 > 2	G/T	0.81	1.32	5.9×10^{-4}
11	68833565	rs189122876	2 > 1	C/T	0.14	1.39	2.6×10^{-4}
11	69448373	rs654240	2 > 1	T/C	0.40	1.27	1.6×10^{-4}
11	70518864	rs7104178	1 > 2	A/G	0.73	1.30	4.6×10^{-4}
11	70968452	rs117835161	1 > 2	A/G	0.22	1.46	1.3×10^{-6}
11	71019766	rs1660861	1 > 2	T/C	0.61	1.26	3.7×10^{-4}
11	77999865	chr11:77999865	2 > 1	G/A	0.05	1.78	7.2×10^{-5}
11	79415407	rs2105527	2 > 1	T/C	0.07	1.62	8.0×10^{-5}
11	81582466	rs149635064	2 > 1	C/T	0.07	1.53	5.9×10^{-4}
11	86628005	rs4944640	2 > 1	C/T	0.36	1.31	3.5×10^{-5}
11	110693356	rs184443	1 > 2	G/A	0.33	1.26	5.9×10^{-4}
11	112353922	chr11:112353922	1 > 2	A/T	0.96	1.75	2.9×10^{-4}
11	112476540	rs4291707	1 > 2	G/T	0.94	1.71	4.5×10^{-5}
11	113964043	rs77660877	2 > 1	C/T	0.85	1.35	5.4×10^{-4}
11	115658805	rs74369070	1 > 2	G/A	0.86	1.43	9.5×10^{-5}
11	119239689	rs680268	2 > 1	T/C	0.86	1.37	4.0×10^{-4}
11	126285301	rs7949566	1 > 2	G/A	0.53	1.28	1.1×10^{-4}
11	126601540	rs10893557	1 > 2	G/T	0.31	1.28	3.6×10^{-4}
12	3352183	rs77587	2 > 1	G/C	0.53	1.25	2.9×10^{-4}
12	4150894	rs78555230	2 > 1	C/T	0.06	1.65	8.5×10^{-5}
12	12823025	rs143606772	2 > 1	C/T	0.63	1.25	4.7×10^{-4}
12	17571193	rs183764686	1 > 2	G/A	0.05	1.71	3.6×10^{-4}
12	43990838	rs60167591	2 > 1	I/D	0.61	1.25	4.2×10^{-4}

12	44023884	rs1478618	2 > 1	A/G	0.58	1.24	5.2×10^{-4}
12	57856493	rs78179158	1 > 2	G/A	0.19	1.37	1.3×10^{-4}
12	58197231	rs60780489	1 > 2	T/G	0.24	1.30	5.2×10^{-4}
12	63450335	rs3913041	2 > 1	A/G	0.60	1.26	2.8×10^{-4}
12	81141539	rs1603216	1 > 2	G/A	0.86	1.38	4.7×10^{-4}
12	107743313	rs76454120	1 > 2	C/T	0.69	1.27	6.2×10^{-4}
12	117815549	rs526159	2 > 1	G/A	0.86	1.37	2.6×10^{-4}
12	128537287	rs7975535	1 > 2	T/C	0.34	1.28	2.3×10^{-4}
12	128709513	rs11059588	2 > 1	G/T	0.55	1.33	9.1×10^{-6}
12	129704082	rs78134285	2 > 1	T/G	0.06	1.56	5.7×10^{-4}
13	32569815	rs121046	2 > 1	G/A	0.46	1.24	5.5×10^{-4}
13	39279459	rs56255874	2 > 1	A/G	0.64	1.27	1.9×10^{-4}
13	40803985	rs71425795	1 > 2	C/T	0.32	1.28	2.8×10^{-4}
13	46342822	rs4942430	2 > 1	C/A	0.82	1.33	4.0×10^{-4}
13	54432416	rs1841058	1 > 2	C/T	0.79	1.32	4.0×10^{-4}
13	92982848	chr13:92982848	2 > 1	D/I	0.08	1.54	2.9×10^{-4}
13	93363027	chr13:93363027	2 > 1	G/C	0.08	1.60	6.9×10^{-5}
13	93586124	chr13:93586124	2 > 1	C/T	0.07	1.59	1.1×10^{-4}
13	93955178	chr13:93955178	2 > 1	T/C	0.08	1.58	1.0×10^{-4}
13	94070190	chr13:94070190	2 > 1	C/T	0.08	1.52	3.5×10^{-4}
13	94634966	chr13:94634966	2 > 1	A/G	0.08	1.55	1.5×10^{-4}
13	98833517	rs7986669	2 > 1	C/T	0.93	1.55	3.2×10^{-4}
13	100390236	rs17577153	2 > 1	A/G	0.88	1.47	5.4×10^{-5}
13	100456122	rs2390357	2 > 1	A/G	0.16	1.37	2.2×10^{-4}
13	101064282	rs12429310	1 > 2	G/A	0.85	1.38	2.8×10^{-4}
13	105339826	rs701565	1 > 2	A/C	0.44	1.25	5.2×10^{-4}
13	105898262	rs16966166	2 > 1	C/T	0.62	1.26	2.3×10^{-4}
14	33131189	rs12895513	1 > 2	A/G	0.58	1.24	5.9×10^{-4}
14	33152745	rs11156756	1 > 2	A/G	0.45	1.26	2.8×10^{-4}
14	41767938	rs1778367	1 > 2	T/G	0.11	1.43	3.4×10^{-4}
14	77396193	rs112854307	2 > 1	D/I	0.20	1.32	3.4×10^{-4}
14	78205694	rs185044034	1 > 2	C/T	0.91	1.47	5.2×10^{-4}
14	80185987	rs178388	2 > 1	A/C	0.25	1.34	5.6×10^{-5}
14	82431921	rs74963973	2 > 1	C/T	0.08	1.50	2.4×10^{-4}
14	103463561	rs7157813	1 > 2	A/G	0.87	1.44	2.0×10^{-4}
15	26153832	rs1877249	2 > 1	A/G	0.39	1.24	5.9×10^{-4}
15	26186978	rs71463472	2 > 1	A/C	0.10	1.44	4.6×10^{-4}
15	26273281	rs72625124	1 > 2	T/G	0.89	1.44	3.9×10^{-4}
15	32337237	rs4779563	2 > 1	T/C	0.63	1.29	7.7×10^{-5}
15	39163687	rs11632328	1 > 2	T/C	0.49	1.27	1.6×10^{-4}
15	39641049	rs74634178	2 > 1	C/T	0.96	1.87	8.3×10^{-5}
15	53879272	rs4447369	1 > 2	C/T	0.72	1.29	2.8×10^{-4}
15	54757794	rs57385376	2 > 1	A/C	0.88	1.39	4.9×10^{-4}

15	58132503	rs6493957	1 > 2	G/T	0.58	1.26	2.7×10^{-4}
15	58838440	rs1973024	1 > 2	T/C	0.37	1.26	4.0×10^{-4}
15	63595545	rs72212605	1 > 2	D/I	0.71	1.28	4.8×10^{-4}
15	74408580	rs117223627	1 > 2	C/T	0.90	1.48	1.6×10^{-4}
15	74903614	rs140284162	1 > 2	G/C	0.96	1.81	1.5×10^{-4}
15	79488593	rs7170146	1 > 2	A/G	0.56	1.30	4.9×10^{-5}
15	80818730	rs147006180	1 > 2	G/C	0.15	1.38	3.3×10^{-4}
15	82391732	rs11073033	1 > 2	A/G	0.14	1.39	3.6×10^{-4}
15	86103176	rs16941432	1 > 2	G/A	0.29	1.32	7.4×10^{-5}
15	87333293	rs16977987	1 > 2	G/A	0.74	1.28	5.7×10^{-4}
15	87422428	rs57116355	1 > 2	C/T	0.77	1.30	4.6×10^{-4}
15	88044080	rs34297582	2 > 1	A/G	0.76	1.30	2.9×10^{-4}
15	89180915	rs10152159	1 > 2	C/T	0.45	1.26	3.2×10^{-4}
15	91990598	rs34901612	2 > 1	G/A	0.16	1.34	5.9×10^{-4}
15	98701022	rs977832	2 > 1	T/C	0.89	1.48	9.2×10^{-5}
16	322345	rs74003728	2 > 1	C/T	0.95	1.71	1.9×10^{-4}
16	370443	rs117116749	1 > 2	A/G	0.02	2.45	3.0×10^{-4}
16	4060661	rs2239310	2 > 1	A/G	0.26	1.28	4.6×10^{-4}
16	6328022	rs2160166	2 > 1	T/C	0.17	1.36	2.1×10^{-4}
16	6916942	chr16:6916942	2 > 1	C/A	0.04	1.81	3.1×10^{-4}
16	7700265	rs184543795	1 > 2	G/A	0.12	1.43	3.7×10^{-4}
16	13355924	rs78178451	1 > 2	C/T	0.10	1.45	4.3×10^{-4}
16	17842398	rs12598358	2 > 1	G/A	0.89	1.44	2.4×10^{-4}
16	18863882	rs74337241	1 > 2	A/G	0.95	1.72	4.8×10^{-4}
16	24325877	chr16:24325877	2 > 1	T/A	0.96	1.71	5.2×10^{-4}
16	51106209	rs4785498	2 > 1	T/C	0.65	1.33	1.6×10^{-5}
16	52065724	rs1345322	2 > 1	G/A	0.57	1.25	3.8×10^{-4}
16	52169755	rs4785094	2 > 1	G/A	0.69	1.31	4.4×10^{-5}
16	52261784	rs13337467	2 > 1	G/T	0.65	1.33	1.2×10^{-5}
16	52424157	rs187680	2 > 1	A/C	0.46	1.30	2.4×10^{-5}
16	53606724	chr16:53606724	1 > 2	G/C	0.04	1.78	5.9×10^{-4}
16	53987525	chr16:53987525	1 > 2	T/A	0.04	2.00	2.4×10^{-5}
16	56223793	rs3859112	1 > 2	C/T	0.17	1.36	3.8×10^{-4}
16	56308431	rs12600108	1 > 2	A/C	0.27	1.31	1.3×10^{-4}
16	56944947	rs75109156	1 > 2	A/G	0.76	1.28	6.3×10^{-4}
16	61291839	rs16963175	1 > 2	A/G	0.86	1.44	6.1×10^{-5}
16	68867456	rs1801026	2 > 1	T/C	0.26	1.32	9.1×10^{-5}
16	73375719	rs4346228	2 > 1	G/A	0.54	1.24	5.5×10^{-4}
16	74062203	rs2716591	2 > 1	C/T	0.87	1.50	1.3×10^{-5}
16	79061586	chr16:79061586	2 > 1	T/C	0.95	1.67	5.8×10^{-4}
16	85822447	chr16:85822447	1 > 2	T/C	0.95	1.76	3.4×10^{-4}
16	86119528	rs9935120	2 > 1	T/C	0.17	1.33	4.2×10^{-4}
16	86120025	rs4843903	1 > 2	T/C	0.44	1.25	5.6×10^{-4}

17	2295633	rs2429919	2 > 1	C/A	0.70	1.27	4.7×10^{-4}
17	5846281	rs967787	2 > 1	A/G	0.41	1.24	6.6×10^{-4}
17	12562530	rs12603774	1 > 2	T/C	0.43	1.29	7.8×10^{-5}
17	48457722	rs77309724	2 > 1	G/C	0.90	1.42	6.7×10^{-4}
17	52943191	rs16955263	2 > 1	A/G	0.61	1.26	5.0×10^{-4}
17	55604097	rs74401513	2 > 1	T/C	0.95	1.65	5.8×10^{-4}
17	57556414	rs9890999	2 > 1	C/T	0.51	1.25	3.8×10^{-4}
17	58149745	rs3744375	2 > 1	A/G	0.67	1.34	1.1×10^{-5}
17	58212091	rs345171	2 > 1	T/C	0.77	1.31	3.5×10^{-4}
17	58749017	chr17:58749017	2 > 1	C/G	0.96	1.80	3.1×10^{-4}
17	58965766	chr17:58965766	2 > 1	C/T	0.96	1.81	3.0×10^{-4}
17	59214727	chr17:59214727	2 > 1	A/G	0.95	1.74	3.1×10^{-4}
17	63477337	rs187812044	1 > 2	G/A	0.15	1.37	3.4×10^{-4}
17	63722528	rs4346239	1 > 2	G/A	0.53	1.28	9.0×10^{-5}
17	71943095	rs145834533	2 > 1	C/T	0.36	1.25	4.9×10^{-4}
17	74336545	rs11077813	1 > 2	T/C	0.82	1.37	1.7×10^{-4}
17	75215103	rs7215278	1 > 2	G/T	0.64	1.27	2.6×10^{-4}
17	77831802	chr17:77831802	2 > 1	C/T	0.02	2.04	6.7×10^{-4}
18	5377643	rs34657168	1 > 2	T/G	0.85	1.36	4.3×10^{-4}
18	7127805	rs9965559	1 > 2	A/G	0.17	1.42	2.7×10^{-5}
18	7689006	rs4121621	2 > 1	G/A	0.24	1.31	1.9×10^{-4}
18	11694187	rs8090294	1 > 2	G/T	0.10	1.48	4.5×10^{-4}
18	26291497	rs16945199	2 > 1	G/A	0.87	1.43	1.1×10^{-4}
18	27562746	rs8097943	1 > 2	C/T	0.20	1.32	5.9×10^{-4}
18	32076237	rs8092794	1 > 2	A/G	0.73	1.31	1.5×10^{-4}
18	37496313	rs138230935	2 > 1	D/I	0.15	1.34	5.8×10^{-4}
18	43856724	rs77784260	2 > 1	G/A	0.77	1.40	4.9×10^{-6}
18	45034793	rs7506616	2 > 1	T/C	0.37	1.25	5.3×10^{-4}
18	45059898	rs12604167	2 > 1	A/G	0.76	1.28	6.0×10^{-4}
18	45546445	rs12955249	1 > 2	C/T	0.49	1.30	3.5×10^{-5}
18	60111809	rs11152345	1 > 2	C/A	0.15	1.42	9.6×10^{-5}
18	61540993	rs184075750	2 > 1	A/G	0.94	1.60	3.6×10^{-4}
18	64832417	rs7235347	1 > 2	C/T	0.52	1.25	5.5×10^{-4}
18	69308225	rs144134528	2 > 1	D/I	0.10	1.55	3.4×10^{-5}
19	1027797	rs7253254	1 > 2	C/A	0.15	1.38	3.4×10^{-4}
19	1138540	rs189976016	1 > 2	T/C	0.16	1.41	6.8×10^{-5}
19	10552211	chr19:10552211	2 > 1	T/C	0.96	1.82	4.0×10^{-4}
19	11411889	chr19:11411889	2 > 1	G/A	0.98	2.11	3.3×10^{-4}
19	11558451	rs111813645	2 > 1	A/G	0.95	1.69	1.2×10^{-4}
19	13876841	chr19:13876841	1 > 2	C/T	0.98	2.54	9.9×10^{-5}
19	13949434	chr19:13949434	1 > 2	G/A	0.98	2.56	8.4×10^{-5}
19	15529989	rs145008437	2 > 1	G/T	0.07	1.55	3.6×10^{-4}
19	29790947	rs79231989	2 > 1	T/C	0.65	1.27	2.7×10^{-4}

19	29998033	rs34950765	2 > 1	I/D	0.86	1.36	6.2×10^{-4}
19	30897578	rs79725917	2 > 1	C/T	0.82	1.34	3.6×10^{-4}
19	35757250	rs916147	1 > 2	A/G	0.30	1.28	2.7×10^{-4}
19	40094528	rs1014206	2 > 1	G/A	0.57	1.24	6.6×10^{-4}
19	41172997	chr19:41172997	2 > 1	G/C	0.98	2.32	3.1×10^{-4}
19	42985831	rs16975834	2 > 1	G/A	0.09	1.49	1.7×10^{-4}
19	44953196	rs73935677	2 > 1	A/G	0.90	1.43	5.9×10^{-4}
19	45202027	rs12150984	2 > 1	A/G	0.22	1.29	6.7×10^{-4}
19	45379516	rs412776	2 > 1	A/G	0.18	1.33	3.8×10^{-4}
19	46262673	chr19:46262673	1 > 2	A/G	0.01	5.22	4.3×10^{-4}
19	47774666	rs138370983	1 > 2	D/I	0.36	1.26	4.3×10^{-4}
19	49263886	rs78627671	1 > 2	A/G	0.10	1.52	5.6×10^{-5}
19	53482311	rs145685675	1 > 2	I/D	0.20	1.34	2.7×10^{-4}
19	54333563	rs71363373	1 > 2	A/C	0.20	1.32	4.6×10^{-4}
20	292327	rs6076049	2 > 1	C/T	0.34	1.25	4.5×10^{-4}
20	364747	rs7265169	2 > 1	C/A	0.54	1.24	6.4×10^{-4}
20	550815	rs2665782	2 > 1	A/C	0.57	1.26	2.3×10^{-4}
20	2375262	rs2076405	1 > 2	A/G	0.03	1.98	2.4×10^{-4}
20	3780962	chr20:3780962	2 > 1	D/I	0.01	3.94	1.6×10^{-4}
20	22732845	rs17827832	2 > 1	G/T	0.17	1.34	4.6×10^{-4}
20	44577244	rs189615511	2 > 1	C/T	0.96	1.81	1.1×10^{-4}
20	46410015	rs6012265	1 > 2	C/T	0.24	1.36	5.3×10^{-5}
20	52468966	rs4811476	2 > 1	T/C	0.42	1.25	3.7×10^{-4}
20	57155037	rs185637455	2 > 1	C/T	0.96	1.77	3.3×10^{-4}
20	62772637	rs2983435	1 > 2	C/T	0.54	1.25	5.3×10^{-4}
21	22436984	rs62207566	1 > 2	G/A	0.89	1.50	7.6×10^{-5}
21	25536983	rs186705241	1 > 2	C/T	0.07	1.52	5.7×10^{-4}
21	27397380	rs149868148	2 > 1	C/T	0.07	1.53	5.0×10^{-4}
21	33782887	rs2211789	2 > 1	T/C	0.54	1.27	1.1×10^{-4}
21	34423899	rs2834078	2 > 1	T/G	0.65	1.25	6.5×10^{-4}
21	36469037	rs984657	1 > 2	C/T	0.63	1.26	4.2×10^{-4}
21	39738517	rs34613855	2 > 1	A/G	0.85	1.36	3.3×10^{-4}
21	39788503	rs2836379	2 > 1	A/C	0.91	1.48	2.9×10^{-4}
21	41580503	rs117371768	2 > 1	G/A	0.06	1.61	3.2×10^{-4}
21	41808578	rs76272754	2 > 1	T/G	0.08	1.46	6.2×10^{-4}
21	41877887	rs7282270	2 > 1	A/G	0.18	1.31	6.1×10^{-4}
21	43142343	rs55711022	1 > 2	G/C	0.55	1.24	6.0×10^{-4}
21	43144340	rs7283002	2 > 1	C/T	0.75	1.34	5.8×10^{-5}
21	43530575	rs220124	2 > 1	T/C	0.32	1.31	5.7×10^{-5}
22	17428896	rs148856575	1 > 2	T/G	0.10	1.50	2.4×10^{-4}
22	26469229	rs9613096	2 > 1	A/G	0.77	1.30	4.6×10^{-4}
22	27424318	rs12483907	1 > 2	A/G	0.13	1.41	3.2×10^{-4}
22	36133739	rs5001172	2 > 1	A/G	0.70	1.30	8.0×10^{-5}

22	39939913	rs12628516	2 > 1	A/G	0.90	1.41	6.4×10^{-4}
22	43006541	rs182783110	2 > 1	T/C	0.08	1.54	2.0×10^{-4}
22	43867177	rs8142313	1 > 2	G/C	0.51	1.27	2.5×10^{-4}
22	46266351	rs117192130	1 > 2	A/G	0.05	1.69	1.8×10^{-4}
22	47372330	rs4439	1 > 2	T/C	0.38	1.27	4.4×10^{-4}
X	16675963	rs6632861	2 > 1	G/A	0.73	1.22	5.7×10^{-4}
X	19785191	rs148543908	2 > 1	G/A	0.05	1.48	5.0×10^{-4}
X	19942966	rs7060796	2 > 1	A/G	0.06	1.45	3.6×10^{-4}
X	30119777	rs6628506	1 > 2	G/A	0.54	1.21	2.3×10^{-4}
X	32934766	rs5928113	2 > 1	A/G	0.09	1.38	3.6×10^{-4}
X	35847989	rs1536848	2 > 1	G/A	0.25	1.23	3.3×10^{-4}
X	44226172	rs5953389	2 > 1	A/G	0.40	1.23	5.5×10^{-5}
X	84534059	rs6617041	1 > 2	A/C	0.18	1.29	1.7×10^{-4}
X	111659846	rs6643034	2 > 1	T/C	0.21	1.26	2.4×10^{-4}
X	125060265	rs5932947	2 > 1	C/T	0.66	1.20	4.6×10^{-4}
X	125583629	rs5931011	2 > 1	G/A	0.62	1.22	1.2×10^{-4}
X	135769607	rs141122592	2 > 1	T/C	0.93	1.43	2.8×10^{-4}
X	150396570	rs760105	2 > 1	A/G	0.24	1.23	5.4×10^{-4}

[x] Chr is chromosome.
[y] Position is in base pairs (bp) on Genome Reference Consortium Human Build 37.
[z] Subsets represent, which of the two randomly assigned groups is the discovery group (listed before the arrow) or the target group (listed after the arrow). The frequency, OR and p-value is reported for the discovery group.
[aa] A1/a2 are the alleles with the effect allele given first.
[ab] A1f is the frequency of the effect allele.
[ac] OR is reported per copy of the effect allele.

8. Bibliography

Ahlqvist, E., Storm, P., Käräjämäki, A., Martinell, M., Dorkhan, M., Carlsson, A., Vikman, P., Prasad, R. B., Aly, D. M., Almgren, P., Wessman, Y., Shaat, N., Spégel, P., Mulder, H., Lindholm, E., Melander, O., Hansson, O., Malmqvist, U., Lernmark, Å., … Groop, L. (2018). Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *The Lancet Diabetes & Endocrinology*, *6*(5), 361–369. https://doi.org/10.1016/S2213-8587(18)30051-2

Alfred, G. T. (2009). Colonialism and State Dependency. *International Journal of Indigenous Health*, *5*(2), 42–60. https://doi.org/10.3138/IJIH.V5I2.28982

Ali-Khan, S. E., Krakowski, T., Tahir, R., & Daar, A. S. (2011). The use of race, ethnicity and ancestry in human genetic research. *The HUGO Journal*, *5*(1–4), 47–63. https://doi.org/10.1007/S11568-011-9154-5

Ali, O. (2013). Genetics of type 2 diabetes. *World Journal of Diabetes*, *4*(4), 114. https://doi.org/10.4239/wjd.v4.i4.114

Allison, P. D. (2000). Multiple imputation for missing data: A cautionary tale. *Sociological Methods and Research*, *28*(3), 301–309. https://doi.org/10.1177/0049124100028003003

Almasy, L., & Blangero, J. (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. *American Journal of Human Genetics*, *62*(5), 1198–1211. https://doi.org/10.1086/301844

Almgren, P., Lehtovirta, M., Isomaa, B., Sarelin, L., Taskinen, M. R., Lyssenko, V., Tuomi, T., & Groop, L. (2011). Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia*, *54*(11), 2811–2819. https://doi.org/10.1007/S00125-011-2267-5

American Diabetes Association. (2014). Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*, *37*(Supplement 1), S81–S90. https://doi.org/10.2337/DC14-S081

American Diabetes Association. (2020). 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2020. *Diabetes Care*, *43*(Supplement_1), S14–S31. https://doi.org/10.2337/DC20-S002

American Diabetes Association. (2021). 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2021. *Diabetes Care*, *44*(Supplement 1), S15–S33. https://doi.org/10.2337/DC21-S002

Andersen, M. K., Pedersen, C. E. T., Moltke, I., Hansen, T., Albrechtsen, A., & Grarup, N. (2016). Genetics of type 2 diabetes: the power of isolated populations. *Current Diabetes Reports*, *16*(7), 1–9. https://doi.org/10.1007/S11892-016-0757-Z

Ang, G. Y. (2020). Age of onset of diabetes and all-cause mortality. *World Journal of Diabetes*, *11*(4), 95. https://doi.org/10.4239/WJD.V11.I4.95

Baier, L. J., & Hanson, R. L. (2004). Genetic Studies of the Etiology of Type 2 Diabetes in Pima Indians: Hunting for Pieces to a Complicated Puzzle. *Diabetes*, *53*(5), 1181–1186. https://doi.org/10.2337/diabetes.53.5.1181

Baier, L. J., Muller, Y. L., Remedi, M. S., Traurig, M., Piaggi, P., Wiessner, G., Huang, K.,

Stacy, A., Kobes, S., Krakoff, J., Bennett, P. H., Nelson, R. G., Knowler, W. C., Hanson, R. L., Nichols, C. G., & Bogardus, C. (2015). ABCC8 R1420H loss-of-function variant in a southwest American Indian community: Association with increased birth weight and doubled risk of type 2 diabetes. *Diabetes*, *64*(12), 4322–4332. https://doi.org/10.2337/db15-0459

Beaumont, R. N., Warrington, N. M., Cavadino, A., Tyrrell, J., Nodzenski, M., Horikoshi, M., Geller, F., Myhre, R., Richmond, R. C., Paternoster, L., Bradfield, J. P., Kreiner-Møller, E., Huikari, V., Metrustry, S., Lunetta, K. L., Painter, J. N., Hottenga, J.-J., Allard, C., Barton, S. J., … Freathy, R. M. (2018). Genome-wide association study of offspring birth weight in 86 577 women identifies five novel loci and highlights maternal genetic effects that are independent of fetal genetics. *Human Molecular Genetics*, *27*(4), 742–756. https://doi.org/10.1093/hmg/ddx429

Bellou, V., Belbasis, L., Tzoulaki, I., & Evangelou, E. (2018). Risk factors for type 2 diabetes mellitus: An exposure-wide umbrella review of meta-analyses. *PLoS ONE*, *13*(3), e0194127. https://doi.org/10.1371/JOURNAL.PONE.0194127

Belmont, J. W., Hardenbol, P., Willis, T. D., Yu, F., Yang, H., Ch'Ang, L. Y., Huang, W., Liu, B., Shen, Y., Tam, P. K. H., Tsui, L. C., Waye, M. M. Y., Wong, J. T. F., Zeng, C., Zhang, Q., Chee, M. S., Galver, L. M., Kruglyak, S., Murray, S. S., … Tanaka, T. (2003). The International HapMap Project. *Nature*, *426*(6968), 789–796. https://doi.org/10.1038/nature02168

Berinstein, D. M., Stahn, R. M., Welty, T. K., Leonardson, G. R., & Herlihy, J. J. (1997). The prevalence of diabetic retinopathy and associated risk factors among Sioux Indians. *Diabetes Care*, *20*(5), 757–759. https://doi.org/10.2337/diacare.20.5.757

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, *10*(2–3), 191–203. https://doi.org/10.1016/0098-3004(84)90020-7

Blangero, J., Diego, V. P., Dyer, T. D., Almeida, M., Peralta, J., Kent, J. W., Williams, J. T., Almasy, L., & GÖring, H. H. H. (2013). A kernel of truth: statistical advances in polygenic variance component models for complex human pedigrees. *Advances in Genetics*, *81*, 1–31. https://doi.org/10.1016/B978-0-12-407677-8.00001-4

Blangero, J., Diego, V. P., Dyer, T. D., Almeida, M., Peralta, J., Kent, J. W., Williams, J. T., Almasy, L., Göring, H. H. H., & Göring, H. H. H. (2013). A kernel of truth: statistical advances in polygenic variance component models for complex human pedigrees. *Advances in Genetics*, *81*, 1–31. https://doi.org/10.1016/B978-0-12-407677-8.00001-4

Blom, G. (1958). *Statistical estimates and transformed beta-variables* [Almqvist & Wiksell]. http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-75457

Bonferroni, C. E. (1935). Il calcolo delle assicurazioni su gruppi di teste. *Studi in Onore Del Professore Salvatore Ortu Carboni*, 13–60. https://cir.nii.ac.jp/crid/1571980074213687424

Braun, K. L., Browne, C. V., Ka'Opua, L. S., Kim, B. J., & Mokuau, N. (2014). Research on indigenous elders: From positivistic to decolonizing methodologies. *Gerontologist*, *54*(1), 117–126. https://doi.org/10.1093/geront/gnt067

Browning, B. L., & Browning, S. R. (2011). A fast, powerful method for detecting identity by descent. *American Journal of Human Genetics*, *88*(2), 173–182.

https://doi.org/10.1016/j.ajhg.2011.01.010

Bullock, A., Sheff, K., Hora, I., Rios Burrows, N., Benoit, S. R., Saydah, S. H., Hardin, C. L., & Gregg, E. W. (2020). Prevalence of diagnosed diabetes in American Indian and Alaska Native adults, 2006-2017 What are the new findings? *BMJ Open Diab Res Care*, *8*, 1218. https://doi.org/10.1136/bmjdrc-2020-001218

Bunt, J. C., Krakoff, J., Ortega, E., Knowler, W. C., & Bogardus, C. (2007). Acute insulin response is an independent predictor of type 2 diabetes mellitus in individuals with both normal fasting and 2-h plasma glucose concentrations. *Diabetes/Metabolism Research and Reviews*, *23*(4), 304–310. https://doi.org/10.1002/dmrr.686

Bursell, S. E., Fonda, S. J., Lewis, D. G., & Horton, M. B. (2018). Prevalence of diabetic retinopathy and diabetic macular edema in a primary care-based teleophthalmology program for American indians and alaskan natives. *PLoS ONE*, *13*(6). https://doi.org/10.1371/journal.pone.0198551

Butler, A. E., Janson, J., Bonner-Weir, S., Ritzel, R., Rizza, R. A., & Butler, P. C. (2003). Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. *Diabetes*, *52*(1), 102–110. https://doi.org/10.2337/DIABETES.52.1.102

Callaghan, W. M., & Dietz, P. M. (2010). Differences in Birth Weight for Gestational Age Distributions According to the Measures Used to Assign Gestational Age. *American Journal of Epidemiology*, *171*(7), 826–836. https://doi.org/10.1093/AJE/KWP468

Campbell, I. W. (2000). Epidemiology and Clinical Presentation of Type 2 Diabetes. *Value in Health*, *3*(S1), 3–6. https://doi.org/10.1046/j.1524-4733.2000.36014.x

Campbell, J. A., Walker, R. J., Smalls, B. L., & Egede, L. E. (2012). Glucose control in diabetes: the impact of racial differences onmonitoring and outcomes. *Endocrine*, *42*(3), 471. https://doi.org/10.1007/S12020-012-9744-6

Carlson, C. S., Matise, T. C., North, K. E., Haiman, C. A., Fesinmeyer, M. D., Buyske, S., Schumacher, F. R., Peters, U., Franceschini, N., Ritchie, M. D., Duggan, D. J., Spencer, K. L., Dumitrescu, L., Eaton, C. B., Thomas, F., Young, A., Carty, C., Heiss, G., Le Marchand, L., … Kooperberg, C. L. (2013). Generalization and Dilution of Association Results from European GWAS in Populations of Non-European Ancestry: The PAGE Study. *PLoS Biology*, *11*(9), 1001661. https://doi.org/10.1371/journal.pbio.1001661

Carpenter, J. R., & Kenward, M. G. (2007). *Missing data in randomised controlled trials-a practical guide*.

Carter-Pokras, O. D., Offutt-Powell, T. N., Kaufman, J. S., Giles, W. H., & Mays, V. M. (2012). Epidemiology, Policy, and Racial/Ethnic Minority Health Disparities. *Annals of Epidemiology*, *22*(6), 446. https://doi.org/10.1016/J.ANNEPIDEM.2012.04.018

Centers for Disease Control and Prevention. (2018). *Summary Health Statistics: National Health Interview Survey, 2018*. http://www.cdc.gov/nchs/nhis/SHS/tables.htm.

Centers for Disease Control and Prevention. (2020). *National Diabetes Statistics Report, 2020*.

Cerf, M. E. (2013). Beta Cell Dysfunction and Insulin Resistance. *Frontiers in Endocrinology*, *4*(MAR). https://doi.org/10.3389/FENDO.2013.00037

Chambers, M., Tanamas, S. K., Clark, E. J., Dunnigan, D. L., Kapadia, C. R., Hanson, R. L.,

Nelson, R. G., Knowler, W. C., & Sinha, M. (2017). Growth tracking in severely obese or underweight children. *Pediatrics*, *140*(6). https://doi.org/10.1542/peds.2017-2248

Chambless, L. E., Cummiskey, C. P., & Cui, G. (2011). Several methods to assess improvement in risk prediction models: Extension to survival analysis. *Statistics in Medicine*, *30*(1), 22–38. https://doi.org/10.1002/sim.4026

Chande, A. T., Rishishwar, L., Conley, A. B., Valderrama-Aguirre, A., Medina-Rivas, M. A., & Jordan, I. K. (2020). Ancestry effects on type 2 diabetes genetic risk inference in Hispanic/Latino populations. *BMC Medical Genetics*, *21*(Suppl 2). https://doi.org/10.1186/s12881-020-01068-0

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1). https://doi.org/10.1186/S13742-015-0047-8

Chen, R., Corona, E., Sikora, M., Dudley, J. T., Morgan, A. A., Moreno-Estrada, A., Nilsen, G. B., Ruau, D., Lincoln, S. E., Bustamante, C. D., & Butte, A. J. (2012). Type 2 diabetes risk alleles demonstrate extreme directional differentiation among human populations, compared to other diseases. *PLoS Genetics*, *8*(4). https://doi.org/10.1371/journal.pgen.1002621

Chikowore, T., Ekoru, K., Vujkovi, M., Gill, D., Pirie, F., Young, E., Sandhu, M. S., McCarthy, M., Rotimi, C., Adeyemo, A., Motala, A., & Fatumo, S. (2022). Polygenic Prediction of Type 2 Diabetes in Africa. *Diabetes Care*, *45*(3), 717–723. https://doi.org/10.2337/DC21-0365

Cho, N. H., Shaw, J. E., Karuranga, S., Huang, Y., da Rocha Fernandes, J. D., Ohlrogge, A. W., & Malanda, B. (2018). IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Research and Clinical Practice*, *138*, 271–281. https://doi.org/10.1016/J.DIABRES.2018.02.023

Choi, S. W., & O'Reilly, P. F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience*, *8*(7), 1–6. https://doi.org/10.1093/GIGASCIENCE/GIZ082

Claw, K. G., Anderson, M. Z., Begay, R. L., Tsosie, K. S., Fox, K., Garrison, N. A., Bader, A. C. C., Bardill, J., Bolnick, D. A. A., Brooks, J., Cordova, A., Malhi, R. S., Nakatsuka, N., Neller, A., Raff, J. A. A., Singson, J., TallBear, K., Vargas, T., & Yracheta, J. M. (2018). A framework for enhancing ethical genomic research with Indigenous communities. *Nature Communications*, *9*(1), 1–7. https://doi.org/10.1038/S41467-018-05188-3

Concannon, P., Erlich, H. A., Julier, C., Morahan, G., Nerup, J., Pociot, F., Todd, J. A., Rich, S. S., & Type 1 Diabetes Genetics Consortium. (2005). Type 1 diabetes: evidence for susceptibility loci from four genome-wide linkage scans in 1,435 multiplex families. *Diabetes*, *54*(10), 1997–2002. https://doi.org/10.2337/DIABETES.54.10.2995

Conrad, D. F., Jakobsson, M., Coop, G., Wen, X., Wall, J. D., Rosenberg, N. A., & Pritchard, J. K. (2006). A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nature Genetics*, *38*(11), 1251–1260. https://doi.org/10.1038/ng1911

Cowie, C. C., Casagrande, S. S., & Geiss, L. S. (2018). Prevalence and Incidence of Type 2 Diabetes and Prediabetes. In C. C. Cowie, S. S. Casagrande, & A. Menke (Eds.), *Diabetes in America* (3rd ed.). National Institute of Diabetes and Digestive and Kidney Diseases (US). http://www.ncbi.nlm.nih.gov/pubmed/33651562

Craig Venter, J., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., … Zhu, X. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304–1351. https://doi.org/10.1126/SCIENCE.1058040/SUPPL_FILE/1058040S3-3_MED.GIF

Crandall, J. P., Knowler, W. C., Kahn, S. E., Marrero, D., Florez, J. C., Bray, G. A., Haffner, S. M., Hoskin, M., & Nathan, D. M. (2008). The prevention of type 2 diabetes. In *Nature Clinical Practice Endocrinology and Metabolism* (Vol. 4, Issue 7, pp. 382–393). Nat Clin Pract Endocrinol Metab. https://doi.org/10.1038/ncpendmet0843

Dabelea, D. (2007). The predisposition to obesity and diabetes in offspring of diabetic mothers. *Diabetes Care*, *30 Suppl 2*(Supplement 2), S169-74. https://doi.org/10.2337/dc07-s211

Dabelea, D., Pettitt, D. J., Jones, K. L., & Arslanian, S. A. (1999). Type 2 diabetes mellitus in minority children and adolescents: An emerging problem. *Endocrinology and Metabolism Clinics of North America*, *28*(4), 709–729. https://doi.org/10.1016/S0889-8529(05)70098-0

Dabelea, D, Hanson, R. L., Lindsay, R. S., Pettitt, D. J., Imperatore, G., Gabir, M. M., Roumain, J., Bennett, P. H., & Knowler, W. C. (2000). Intrauterine exposure to diabetes conveys risks for type 2 diabetes and obesity: a study of discordant sibships. *Diabetes*, *49*(12), 2208–2211. http://www.ncbi.nlm.nih.gov/pubmed/11118027

Dabelea, D, Knowler, W. C., & Pettitt, D. J. (2000). Effect of diabetes in pregnancy on offspring: follow-up research in the Pima Indians. *The Journal of Maternal-Fetal Medicine*, *9*(1), 83–88. https://doi.org/10.1002/(SICI)1520-6661(200001/02)9:1<83::AID-MFM17>3.0.CO;2-O

Dabelea, Dana, & Crume, T. (2011). Maternal Environment and the Transgenerational Cycle of Obesity and Diabetes. *Diabetes*, *60*(7), 1849. https://doi.org/10.2337/DB11-0400

Dabelea, Dana, Mayer-Davis, E. J., Saydah, S., Imperatore, G., Linder, B., Divers, J., Bell, R., Badaru, A., Talton, J. W., Crume, T., Liese, A. D., Merchant, A. T., Lawrence, J. M., Reynolds, K., Dolan, L., Liu, L. L., & Hamman, R. F. (2014). Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009. *JAMA - Journal of the American Medical Association*, *311*(17), 1778–1786. https://doi.org/10.1001/jama.2014.3201

Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J., & Lander, E. S. (2001). High-resolution haplotype structure in the human genome. *Nature Genetics*, *29*(2), 229–232. https://doi.org/10.1038/ng1001-229

Davis-Delano, L. R., Gone, J. P., & Fryberg, S. A. (2020). The psychosocial effects of Native American mascots: a comprehensive review of empirical research findings. *Race Ethnicity and Education*, *23*(5), 613–633. https://doi.org/10.1080/13613324.2020.1772221

DeForest, N., & Majithia, A. R. (2022). Genetics of Type 2 Diabetes: Implications from Large-Scale Studies. *Current Diabetes Reports*, *1*, 1–9. https://doi.org/10.1007/S11892-022-01462-3/FIGURES/1

Del Guerra, S., Lupi, R., Marselli, L., Masini, M., Bugliani, M., Sbrana, S., Torri, S., Pollera, M., Boggi, U., Mosca, F., Del Prato, S., & Marchetti, P. (2005). Functional and Molecular

Defects of Pancreatic Islets in Human Type 2 Diabetes. *Diabetes*, *54*(3), 727–735. https://doi.org/10.2337/DIABETES.54.3.727

Delaneau, O., Howie, B., Cox, A. J., Zagury, J. F., & Marchini, J. (2013). Haplotype estimation using sequencing reads. *American Journal of Human Genetics*, *93*(4), 687–696. https://doi.org/10.1016/J.AJHG.2013.09.002/ATTACHMENT/7A8C37C3-5B9B-4EC3-B043-D0A76046ACA4/MMC1.PDF

Dennis, J. M., Shields, B. M., Henley, W. E., Jones, A. G., & Hattersley, A. T. (2019). Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. *The Lancet Diabetes and Endocrinology*, *7*(6), 442–451. https://doi.org/10.1016/S2213-8587(19)30087-7/ATTACHMENT/FB46E195-2766-47A9-A7CD-7825A4DD116B/MMC1.PDF

DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium, South Asian Type 2 Diabetes (SAT2D) Consortium, Mexican American Type 2 Diabetes (MAT2D) Consortium, Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples (T2D-GENES) Consortium, Mahajan, A., Go, M. J., Zhang, W., Below, J. E., Gaulton, K. J., Ferreira, T., Horikoshi, M., Johnson, A. D., Ng, M. C. Y., Prokopenko, I., Saleheen, D., Wang, X., Zeggini, E., Abecasis, G. R., … Morris, A. P. (2014). Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature Genetics*, *46*(3), 234–244. https://doi.org/10.1038/ng.2897

DiCorpo, D., LeClair, J., Cole, J. B., Sarnowski, C., Ahmadizar, F., Bielak, L. F., Blokstra, A., Bottinger, E. P., Chaker, L., Chen, Y.-D. I., Chen, Y., de Vries, P. S., Faquih, T., Ghanbari, M., Gudmundsdottir, V., Guo, X., Hasbani, N. R., Ibi, D., Ikram, M. A., … Udler, M. S. (2022). Type 2 Diabetes Partitioned Polygenic Scores Associate With Disease Outcomes in 454,193 Individuals Across 13 Cohorts. *Diabetes Care*, *45*(3), 674–683. https://doi.org/10.2337/DC21-1395

Dimas, A. S., Lagou, V., Barker, A., Knowles, J. W., Mägi, R., Hivert, M. F., Benazzo, A., Rybin, D., Jackson, A. U., Stringham, H. M., Song, C., Fischer-Rosinsky, A., Welløv Boesgaard, T., Grarup, N., Abbasi, F. A., Assimes, T. L., Hao, K., Yang, X., Lecoeur, C., … Prokopenko, I. (2014). Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes*, *63*(6), 2158–2171. https://doi.org/10.2337/DB13-0949

Drong, A. W., Lindgren, C. M., & McCarthy, M. I. (2012). The genetic and epigenetic basis of type 2 diabetes and obesity. *Clinical Pharmacology and Therapeutics*, *92*(6), 707–715. https://doi.org/10.1038/clpt.2012.149

Dudbridge, F., & Gusnanto, A. (2008). Estimation of significance thresholds for genomewide association scans. *Genetic Epidemiology*, *32*(3), 227. https://doi.org/10.1002/GEPI.20297

Eaves, L. J., Pourcain, B. S. t., Smith, G. D. ave., York, T. P., & Evans, D. M. (2014). Resolving the effects of maternal and offspring genotype on dyadic outcomes in genome wide complex trait analysis (“M-GCTA”). *Behavior Genetics*, *44*(5), 445–455. https://doi.org/10.1007/S10519-014-9666-6

Edwards, M. (2017). The Barker Hypothesis. *Handbook of Famine, Starvation, and Nutrient*

Deprivation, 1–21. https://doi.org/10.1007/978-3-319-40007-5_71-1

Egeland, G. M., Skjærven, R., & Irgens, L. M. (2000). Birth characteristics of women who develop gestational diabetes: population based study. *BMJ*, *321*(7260), 546–547. https://doi.org/10.1136/BMJ.321.7260.546

Eneanya, N. D., Yang, W., & Reese, P. P. (2019). Reconsidering the consequences of using race to estimate kidney function. *JAMA*, *322*(2), 113–114. https://doi.org/10.1001/JAMA.2019.5774

Epstein, A. M., Ayanian, J. Z., Keogh, J. H., Noonan, S. J., Armistead, N., Cleary, P. D., Weissman, J. S., David-Kasdan, J. A., Carlson, D., Fuller, J., Marsh, D., & Conti, R. M. (2009). Racial disparities in access to renal transplantation — clinically appropriate or due to underuse or overuse? *Http://Dx.Doi.Org/10.1056/NEJM200011233432106*, *343*(21), 1537–1544. https://doi.org/10.1056/NEJM200011233432106

Esparza, J., Fox, C., Harper, I. T., Bennett, P. H., Schulz, L. O., Valencia, M. E., & Ravussin, E. (2000). Daily energy expenditure in Mexican and USA Pima Indians: Low physical activity as a possible cause of obesity. *International Journal of Obesity*, *24*(1), 55–59. https://doi.org/10.1038/sj.ijo.0801085

Færch, K., Witte, D. R., Tabák, A. G., Perreault, L., Herder, C., Brunner, E. J., Kivimäki, M., & Vistisen, D. (2013). Trajectories of cardiometabolic risk factors before diagnosis of three subtypes of type 2 diabetes: a post-hoc analysis of the longitudinal Whitehall II cohort study. *The Lancet Diabetes & Endocrinology*, *1*(1), 43–51. https://doi.org/10.1016/S2213-8587(13)70008-1

Fahed, A. C., Philippakis, A. A., & Khera, A. V. (2022). The potential of polygenic scores to improve cost and efficiency of clinical trials. *Nature Communications*, *13*(1), 1–4. https://doi.org/10.1038/s41467-022-30675-z

Flannick, J., Mercader, J. M., Fuchsberger, C., Udler, M. S., Mahajan, A., Wessel, J., Teslovich, T. M., Caulkins, L., Koesterer, R., Barajas-Olmos, F., Blackwell, T. W., Boerwinkle, E., Brody, J. A., Centeno-Cruz, F., Chen, L., Chen, S., Contreras-Cubas, C., Córdova, E., Correa, A., … Boehnke, M. (2019). Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature*, *570*(7759), 71–76. https://doi.org/10.1038/s41586-019-1231-2

Fonseca, V. A. (2009). Defining and Characterizing the Progression of Type 2 Diabetes. *Diabetes Care*, *32*(suppl 2), S151–S156. https://doi.org/10.2337/DC09-S301

Forouhi, N. G., Balkau, B., Borch-Johnsen, K., Dekker, J., Glumer, C., Qiao, Q., Spijkerman, A., Stolk, R., Tabac, A., & Wareham, N. J. (2006). The threshold for diagnosing impaired fasting glucose: a position statement by the European Diabetes Epidemiology Group. *Diabetologia*, *49*(5), 822–827. https://doi.org/10.1007/S00125-006-0189-4

Fowler, M. J. (2011). Microvascular and Macrovascular Complications of Diabetes. *Clinical Diabetes*, *29*(3), 116–122. https://doi.org/10.2337/DIACLIN.29.3.116

Franks, P. W., & McCarthy, M. I. (2016). Exposing the exposures responsible for type 2 diabetes and obesity. *Science*, *354*(6308), 69–73. https://doi.org/10.1126/SCIENCE.AAF5094

Freathy, R. M., Mook-Kanamori, D. O., Sovio, U., Prokopenko, I., Timpson, N. J., Berry, D. J., Warrington, N. M., Widen, E., Hottenga, J. J., Kaakinen, M., Lange, L. A., Bradfield, J.

P., Kerkhof, M., Marsh, J. A., Mägi, R., Chen, C.-M., Lyon, H. N., Kirin, M., Adair, L. S., … McCarthy, M. I. (2010). Variants in ADCY5 and near CCNL1 are associated with fetal growth and birth weight. *Nature Genetics*, *42*(5), 430–435. https://doi.org/10.1038/ng.567

Freidlin, B., Zheng, G., Li, Z., & Gastwirth, J. L. (2002). Trend Tests for Case-Control Studies of Genetic Markers: Power, Sample Size and Robustness. *Human Heredity*, *53*(3), 146–152. https://doi.org/10.1159/000064976

Fuchsberger, C., Flannick, J., Teslovich, T. M., Mahajan, A., Agarwala, V., Gaulton, K. J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D. J., Rivas, M. A., Perry, J. R. B., Sim, X., Blackwell, T. W., Robertson, N. R., Rayner, N. W., Cingolani, P., Locke, A. E., Tajes, J. F., … McCarthy, M. I. (2016). The genetic architecture of type 2 diabetes. *Nature*, *536*(7614), 41–47. https://doi.org/10.1038/nature18642

Garrison, N. A., Hudson, M., Ballantyne, L. L., Garba, I., Martinez, A., Taualii, M., Arbour, L., Caron, N. R., & Rainie, S. C. (2019). Genomic research through an Indigenous lens: understanding the expectations. *Https://Doi.Org/10.1146/Annurev-Genom-083118-015434*, *20*, 495–517. https://doi.org/10.1146/ANNUREV-GENOM-083118-015434

Ge, T., Chen, C. Y., Ni, Y., Feng, Y. C. A., & Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-09718-5

Ge, T., Nichols, T. E., Lee, P. H., Holmes, A. J., Roffman, J. L., Buckner, R. L., Sabuncu, M. R., & Smoller, J. W. (2015). Massively expedited genome-wide heritability analysis (MEGHA). *Proceedings of the National Academy of Sciences of the United States of America*, *112*(8), 2479–2484. https://doi.org/10.1073/pnas.1415603112

Ge, T., Patki, A., Srinivasasainagendra, V., Lin, Y.-F., Irvin, M. R., Tiwari, H. K., Armstrong, N., Davis, B. H., Perez, E., Gainer, V., Benoit, B., O'Connor, M. J., Narayan, R., Etheridge, B., Stamou, M., Leong, A., Udler, M. S., Choi, K. W., Miles, A. D., … Karlson, E. W. (2021). Validation of a Trans-Ancestry Polygenic Risk Score for Type 2 Diabetes in Diverse Populations. *MedRxiv*, *15*, 2021.09.11.21263413. https://doi.org/10.1101/2021.09.11.21263413

Gerstein, H. C., Santaguida, P., Raina, P., Morrison, K. M., Balion, C., Hunt, D., Yazdi, H., & Booker, L. (2007). Annual incidence and relative risk of diabetes in people with various categories of dysglycemia: A systematic overview and meta-analysis of prospective studies. *Diabetes Research and Clinical Practice*, *78*(3), 305–312. https://doi.org/10.1016/J.DIABRES.2007.05.004

Gloyn, A. L., Weedon, M. N., Owen, K. R., Turner, M. J., Knight, B. A., Hitman, G., Walker, M., Levy, J. C., Sampson, M., Halford, S., McCarthy, M. I., Hattersley, A. T., & Frayling, T. M. (2003). Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes*, *52*(2), 568–572. https://doi.org/10.2337/DIABETES.52.2.568

Godfrey, K. M., & Barker, D. J. (2000). Fetal nutrition and adult disease. *The American Journal of Clinical Nutrition*, *71*(5), 1344S-1352S. https://doi.org/10.1093/ajcn/71.5.1344s

González, J. R., Carrasco, J. L., Dudbridge, F., Armengol, L., Estivill, X., & Moreno, V.

(2008). Maximizing association statistics over genetic models. *Genetic Epidemiology*, *32*(3), 246–254. https://doi.org/10.1002/GEPI.20299

Greenwood, T. A., Rana, B. K., & Schork, N. J. (2004). Human Haplotype Block Sizes Are Negatively Correlated With Recombination Rates. *Genome Research*, *14*(7), 1358. https://doi.org/10.1101/GR.1540404

Gu, S., Pakstis, A. J., Li, H., Speed, W. C., Kidd, J. R., & Kidd, K. K. (2007). Significant variation in haplotype block structure but conservation in tagSNP patterns among global populations. *European Journal of Human Genetics*, *15*(3), 302–312. https://doi.org/10.1038/sj.ejhg.5201751

Guillory, J. D. (1968). The pro-slavery arguments of Dr. Samuel A. Cartwright. *Louisiana History: The Journal of the Louisiana Historical Association*, *9*(3), 209–227. https://www.jstor.org/stable/4231017

Guo, T., Hanson, R. L., Traurig, M., Muller, Y. L., Ma, L., Mack, J., Kobes, S., Knowler, W. C., Bogardus, C., & Baier, L. J. (2007). TCF7L2 is not a major susceptibility gene for type 2 diabetes in Pima Indians: analysis of 3,501 individuals. *Diabetes*, *56*(12), 3082–3088. https://doi.org/10.2337/db07-0621

Haggstrom, G. W. (1983). Logistic Regression and Discriminant Analysis by Ordinary Least Squares. *Journal of Business & Economic Statistics*, *1*(3), 229. https://doi.org/10.2307/1391344

Haines, L., Wan, K. C., Lynn, R., Barrett, T. G., & Shield, J. P. H. (2007). Rising incidence of type 2 diabetes in children in the U.K. *Diabetes Care*, *30*(5), 1097–1101. https://doi.org/10.2337/dc06-1813

Hakonarson, H., & Grant, S. F. A. (2011). Genome-wide association studies (GWAS): impact on elucidating the aetiology of diabetes. *Diabetes/Metabolism Research and Reviews*, *27*(7), 685–696. https://doi.org/10.1002/DMRR.1221

Hales, C. N., & Barker, D. J. P. (1992). Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis*. *Diabetologia*, *35*, 595–601.

Hamman, R. F., Bell, R. A., Dabelea, D., D'Agostino, R. B., Dolan, L., Imperatore, G., Lawrence, J. M., Linder, B., Marcovina, S. M., Mayer-Davis, E. J., Pihoker, C., Rodriguez, B. L., & Saydah, S. (2014). The SEARCH for diabetes in youth study: Rationale, findings, and future directions. *Diabetes Care*, *37*(12), 3336–3344. https://doi.org/10.2337/DC14-0574/-/DC1

Hanson, R. L., Guo, T., Muller, Y. L., Fleming, J., Knowler, W. C., Kobes, S., Bogardus, C., & Baier, L. J. (2013). Strong parent-of-origin effects in the association of KCNQ1 variants with type 2 diabetes in American Indians. *Diabetes*, *62*(8), 2984–2991. https://doi.org/10.2337/db12-1767

Hanson, R. L., & Knowler, W. C. (1998). Analytic Strategies to Detect Linkage to a Common Disorder With Genetically Determined Age of Onset: Diabetes Mellitus in Pima Indians. *Genetic Epidemiology*, *15*, 299–315. https://doi.org/10.1002/(SICI)1098-2272(1998)15:3

Hanson, R. L., Rong, R., Kobes, S., Muller, Y. L., Weil, E. J., Curtis, J. M., Nelson, R. G., & Baier, L. J. (2015). Role of Established Type 2 Diabetes-Susceptibility Genetic Variants in a High Prevalence American Indian Population. *Diabetes*, *64*(7), 2646–2657. https://doi.org/10.2337/db14-1715

Harder, T., Rodekamp, E., Schellong, K., Dudenhausen, J. W., & Plagemann, A. (2007). Birth Weight and Subsequent Risk of Type 2 Diabetes: A Meta-Analysis. *American Journal of Epidemiology*, *165*(8), 849–857. https://doi.org/10.1093/aje/kwk071

Hattersley, A. T., Beards, F., Ballantyne, E., Appleton, M., Harvey, R., & Ellard, S. (1998). Mutations in the glucokinase gene of the fetus result in reduced birth weight. *Nature Genetics 1998 19:3*, *19*(3), 268–270. https://doi.org/10.1038/953

Hattersley, A. T., & Tooke, J. E. (1999). The fetal insulin hypothesis: an alternative explanation of the association of low bir thweight with diabetes and vascular disease. *The Lancet*, *353*(9166), 1789–1792. https://doi.org/10.1016/S0140-6736(98)07546-1

He, Y., Lakhani, C. M., Rasooly, D., Manrai, A. K., Tzoulaki, I., & Patel, C. J. (2021). Comparisons of Polyexposure, Polygenic, and Clinical Risk Scores in Risk Prediction of Type 2 Diabetes. *Diabetes Care*, dc202049. https://doi.org/10.2337/dc20-2049

Hegele, R. A., Cao, H., Harris, S. B., Hanley, A. J. G., & Zinman, B. (1999). The hepatic nuclear factor-1α G319S variant is associated with early-onset type 2 diabetes in Canadian Oji-Cree. *Journal of Clinical Endocrinology and Metabolism*, *84*(3), 1077–1082. https://doi.org/10.1210/jcem.84.3.5528

Heinitz, S., Piaggi, P., Bogardus, C., & Krakoff, J. (2018). Decline in the acute insulin response in relationship to plasma glucose concentrations. *Diabetes/Metabolism Research and Reviews*, *34*(2), e2953. https://doi.org/10.1002/DMRR.2953

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/SIM.1186

Hill, D. J., & Milner, R. D. G. (1985). Insulin as a Growth Factor. *Pediatric Research*, *19*(9), 879–886. https://doi.org/10.1203/00006450-198509000-00001

Hivert, M. F., Jablonski, K. A., Perreault, L., Saxena, R., McAteer, J. B., Franks, P. W., Hamman, R. F., Kahn, S. E., Haffner, S., Meigs, J. B., Altshuler, D., Knowler, W. C., & Florez, J. C. (2011). Updated genetic score based on 34 confirmed type 2 diabetes loci is associated with diabetes incidence and regression to normoglycemia in the Diabetes Prevention Program. *Diabetes*, *60*(4), 1340–1348. https://doi.org/10.2337/db10-1119

Hoet, J. J., & Hanson, M. A. (1999). Intrauterine nutrition: its importance during critical periods for cardiovascular and endocrine development. *The Journal of Physiology*, *514*(Pt 3), 617. https://doi.org/10.1111/J.1469-7793.1999.617AD.X

Horikoshi, M., Beaumont, R. N., Day, F. R., Warrington, N. M., Kooijman, M. N., Fernandez-Tajes, J., Feenstra, B., Van Zuydam, N. R., Gaulton, K. J., Grarup, N., Bradfield, J. P., Strachan, D. P., Li-Gao, R., Ahluwalia, T. S., Kreiner, E., Rueedi, R., Lyytikaïnen, L. P., Cousminer, D. L., Wu, Y., … Freathy, R. M. (2016). Genome-wide associations for birth weight and correlations with adult disease. *Nature*, *538*(7624), 248–252. https://doi.org/10.1038/nature19806

Horikoshi, M., Yaghootkar, H., Mook-Kanamori, D. O., Sovio, U., Taal, H. R., Hennig, B. J., Bradfield, J. P., St Pourcain, B., Evans, D. M., Charoen, P., Kaakinen, M., Cousminer, D. L., Lehtimäki, T., Kreiner-Møller, E., Warrington, N. M., Bustamante, M., Feenstra, B., Berry, D. J., Thiering, E., … Freathy, R. M. (2013). New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nature Genetics*, *45*(1), 76–82. https://doi.org/10.1038/ng.2477

Horton, N. J., & Lipsitz, S. R. (2001). Multiple Imputation in Practice. *The American Statistician*, *55*(3), 244–254. https://doi.org/10.1198/000313001317098266

Howie, B. N., Donnelly, P., & Marchini, J. (2009). A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genetics*, *5*(6), e1000529. https://doi.org/10.1371/journal.pgen.1000529

Hsueh, W. C., Nair, A. K., Kobes, S., Chen, P., Göring, H. H. H., Pollin, T. I., Malhotra, A., Knowler, W. C., Baier, L. J., & Hanson, R. L. (2017). Identity-by-Descent Mapping Identifies Major Locus for Serum Triglycerides in Amerindians Largely Explained by an APOC3 Founder Mutation. *Circulation: Cardiovascular Genetics*, *10*(6). https://doi.org/10.1161/CIRCGENETICS.117.001809/-/DC1

Huang, T., Wang, T., Zheng, Y., Ellervik, C., Li, X., Gao, M., Fang, Z., Chai, J. F., Ahluwalia, T. V. S., Wang, Y., Voortman, T., Noordam, R., Frazier-Wood, A., Scholz, M., Sonestedt, E., Akiyama, M., Dorajoo, R., Zhou, A., Kilpeläinen, T. O., … Qi, L. (2019a). Association of Birth Weight With Type 2 Diabetes and Glycemic Traits: A Mendelian Randomization Study. *JAMA Network Open*, *2*(9), e1910915–e1910915. https://doi.org/10.1001/JAMANETWORKOPEN.2019.10915

Huang, T., Wang, T., Zheng, Y., Ellervik, C., Li, X., Gao, M., Fang, Z., Chai, J. F., Ahluwalia, T. V. S., Wang, Y., Voortman, T., Noordam, R., Frazier-Wood, A., Scholz, M., Sonestedt, E., Akiyama, M., Dorajoo, R., Zhou, A., Kilpeläinen, T. O., … Qi, L. (2019b). Association of Birth Weight With Type 2 Diabetes and Glycemic Traits: A Mendelian Randomization Study. *JAMA Network Open*, *2*(9), e1910915. https://doi.org/10.1001/JAMANETWORKOPEN.2019.10915

Hughes, A. E., Hattersley, A. T., Flanagan, S. E., & Freathy, R. M. (2021). Two decades since the fetal insulin hypothesis: what have we learned from genetics? *Diabetologia*, *64*(4), 717–726. https://doi.org/10.1007/S00125-021-05386-7

Hughes, A. E., Nodzenski, M., Beaumont, R. N., Talbot, O., Shields, B. M., Scholtens, D. M., Knight, B. A., >William Lowe, W. L. L., Hattersley, A. T., & Freathy, R. M. (2018). Fetal genotype and maternal glucose have independent and additive effects on birth weight. *Diabetes*, *67*(5), 1024–1029. https://doi.org/10.2337/DB17-1188/-/DC1

Huvinen, E., Koivusalo, S. B., Meinilä, J., Valkama, A., Tiitinen, A., Rönö, K., Stach-Lempinen, B., & Eriksson, J. G. (2018). Effects of a Lifestyle Intervention During Pregnancy and First Postpartum Year: Findings From the RADIEL Study. *The Journal of Clinical Endocrinology & Metabolism*, *103*(4), 1669–1677. https://doi.org/10.1210/JC.2017-02477

Huxley, R., Neil, A., & Collins, R. (2002). Unravelling the fetal origins hypothesis: is there really an inverse association between birthweight and subsequent blood pressure? *Lancet (London, England)*, *360*(9334), 659–665. https://doi.org/10.1016/S0140-6736(02)09834-3

Innes, K. E., Byers, T. E., Marshall, J. A., Barón, A., Orleans, M., & Hamman, R. F. (2002). Association of a Woman's Own Birth Weight With Subsequent Risk for Gestational Diabetes. *JAMA*, *287*(19), 2534–2541. https://doi.org/10.1001/JAMA.287.19.2534

International Diabetes Federation. (2000). IDF Diabetes Atlas (1st ed.). In *International Diabetes Federation*.

International Diabetes Federation. (2003). IDF Diabetes Atlas (2nd ed.). In *International*

Diabetes Federation.

International Diabetes Federation. (2006). IDF Diabetes Atlas (3rd ed.). In *International Diabetes Federation*.

International Diabetes Federation. (2009). IDF Diabetes Atlas (4th ed.). In *International Diabetes Federation*.

International Diabetes Federation. (2011). IDF Diabetes Atlas (5th ed.). In *International Diabetes Federation*.

International Diabetes Federation. (2013). IDF Diabetes Atlas (6th ed.). In *International Diabetes Federation*.

Janssen, R. C., Bogardus, C., Takeda, J., Knowler, W. C., & Thompson, D. B. (1994). Linkage Analysis of Acute Insulin Secretion With GLUT2 and Glucokinase in Pima Indians and the Identification of a Missense Mutation in GLUT2. *Diabetes*, *43*(4), 558–563. https://doi.org/10.2337/DIAB.43.4.558

Jiang, X., Holmes, C., & McVean, G. (2021). The impact of age on genetic risk for common diseases. *PLOS Genetics*, *17*(8), e1009723. https://doi.org/10.1371/JOURNAL.PGEN.1009723

Johansson, B. B., Irgens, H. U., Molnes, J., Sztromwasser, P., Aukrust, I., Juliusson, P. B., Søvik, O., Levy, S., Skrivarhaug, T., Joner, G., Molven, A., Johansson, S., & Njølstad, P. R. (2017). Targeted next-generation sequencing reveals MODY in up to 6.5% of antibody-negative diabetes cases listed in the Norwegian Childhood Diabetes Registry. *Diabetologia*, *60*(4), 625–635. https://doi.org/10.1007/S00125-016-4167-1/TABLES/4

Jun, G., Manning, A., Almeida, M., Zawistowski, M., Wood, A. R., Teslovich, T. M., Fuchsberger, C., Feng, S., Cingolani, P., Gaulton, K. J., Dyer, T., Blackwell, T. W., Chen, H., Chines, P. S., Choi, S., Churchhouse, C., Fontanillas, P., King, R., Lee, S. Y., … Blangero, J. (2017). Evaluating the contribution of rare variants to type 2 diabetes and related traits using pedigrees. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(2), 379–384. https://doi.org/10.1073/PNAS.1705859115/SUPPL_FILE/PNAS.1705859115.SAPP.PDF

Kahn, R. (1997). Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*, *20*(7), 1183–1197. https://doi.org/10.2337/diacare.20.7.1183

Kapoor, R. R., Flanagan, S. E., Arya, V. B., Shield, J. P., Ellard, S., & Hussain, K. (2013). Clinical and molecular characterisation of 300 patients with congenital hyperinsulinism. *European Journal of Endocrinology*, *168*(4), 557–564. https://doi.org/10.1530/EJE-12-0673

Kemp, J. P., Morris, J. A., Medina-Gomez, C., Forgetta, V., Warrington, N. M., Youlten, S. E., Zheng, J., Gregson, C. L., Grundberg, E., Trajanoska, K., Logan, J. G., Pollard, A. S., Sparkes, P. C., Ghirardello, E. J., Allen, R., Leitch, V. D., Butterfield, N. C., Komla-Ebri, D., Adoum, A. T., … Evans, D. M. (2017). Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nature Genetics*, *49*(10), 1468–1475. https://doi.org/10.1038/ng.3949

Khan, M. A. B., Hashim, M. J., King, J. K., Govender, R. D., Mustafa, H., & Kaabi, J. Al. (2020). Epidemiology of Type 2 Diabetes – Global Burden of Disease and Forecasted

Trends. *Journal of Epidemiology and Global Health*, *10*(1), 107. https://doi.org/10.2991/JEGH.K.191028.001

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T., & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, *50*(9), 1219. https://doi.org/10.1038/S41588-018-0183-Z

Kirk, J. K., Passmore, L. V., Bell, R. A., Narayan, K. M. V., D'Agostino, R. B., Arcury, T. A., & Quandt, S. A. (2008). Disparities in A1C levels between Hispanic and non-Hispanic white adults with diabetes: a meta-analysis. *Diabetes Care*, *31*(2), 240–246. https://doi.org/10.2337/DC07-0382

Klebanoff, M. A., & Cole, S. R. (2008). Use of Multiple Imputation in the Epidemiologic Literature. *American Journal of Epidemiology*, *168*(4), 355–357. https://doi.org/10.1093/AJE/KWN071

Klimentidis, Y. C., Zhou, J., & Wineinger, N. E. (2014). Identification of Allelic Heterogeneity at Type-2 Diabetes Loci and Impact on Prediction. *PLoS ONE*, *9*(11), 113072. https://doi.org/10.1371/JOURNAL.PONE.0113072

Knowler, W. C., Bennett, P. H., Bottazzo, G. F., & Doniach, D. (1979). Islet cell antibodies and diabetes mellitus in Pima Indians. *Diabetologia*, *17*(3), 161–164. https://doi.org/10.1007/BF01219743

Knowler, W. C., Pettitt, D. J., Saad, M. F., & Bennett, P. H. (1990). Diabetes mellitus in the Pima Indians: incidence, risk factors and pathogenesis. *Diabetes/Metabolism Reviews*, *6*(1), 1–27. http://www.ncbi.nlm.nih.gov/pubmed/2192853

Knowler, W. C., Bennett, P. H., Hamman, R. F., & Miller, M. (1978). Diabetes incidence and prevalence in Pima Indians: A 19-fold greater incidence than in Rochester, Minnesota. *American Journal of Epidemiology*, *108*(6), 497–505. https://doi.org/10.1093/oxfordjournals.aje.a112648

Knowler, W. C., Pettitt, D. J., Savage, P. J., & Bennett, P. H. (1981). Diabetes incidence in Pima Indians: Contributions of obesity and parental diabetes. *American Journal of Epidemiology*, *113*(2), 144–156. https://doi.org/10.1093/oxfordjournals.aje.a113079

Koyama, S., Ito, K., Terao, C., Akiyama, M., Horikoshi, M., Momozawa, Y., Matsunaga, H., Ieki, H., Ozaki, K., Onouchi, Y., Takahashi, A., Nomura, S., Morita, H., Akazawa, H., Kim, C., Seo, J. sun, Higasa, K., Iwasaki, M., Yamaji, T., … Komuro, I. (2020). Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nature Genetics*, *52*(11), 1169–1177. https://doi.org/10.1038/s41588-020-0705-3

Krischer, J. P., Liu, X., Lernmark, Å., Hagopian, W. A., Rewers, M. J., She, J. X., Toppari, J., Ziegler, A. G., & Akolkar, B. (2017). The Influence of Type 1 Diabetes Genetic Susceptibility Regions, Age, Sex, and Family History on the Progression From Multiple Autoantibodies to Type 1 Diabetes: A TEDDY Study Report. *Diabetes*, *66*(12), 3122–3129. https://doi.org/10.2337/DB17-0261

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., … Morgan, M. J. (2001).

Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Langothe, H., Palmer, C. N. A., Morris, A. D., Zeggini, E., Hattersley, A. T., McCarthy, M. I., Frayling, T. M., & Weedon, M. N. (2008). Assessing the Combined Impact of 18 Common Genetic Variants of Modest Effect Sizes on Type 2 Diabetes Risk. *Diabetes*, *57*(11), 3129–3135. https://doi.org/10.2337/DB08-0504

Lawlor, D. A., Timpson, N. J., Harbord, R. M., Leary, S., Ness, A., McCarthy, M. I., Frayling, T. M., Hattersley, A. T., & Smith, G. D. (2008). Exploring the Developmental Overnutrition Hypothesis Using Parental–Offspring Associations and FTO as an Instrumental Variable. *PLOS Medicine*, *5*(3), e33. https://doi.org/10.1371/JOURNAL.PMED.0050033

Le, D. S. N. T., Brookshire, T., Krakoff, J., & Bunt, J. C. (2009). Repeatability and reproducibility of the hyperinsulinemic-euglycemic clamp and the tracer dilution technique in a controlled inpatient setting. *Metabolism: Clinical and Experimental*, *58*(3), 304–310. https://doi.org/10.1016/J.METABOL.2008.09.029

Lee, J. J., & Chow, C. C. (2014). Conditions for the validity of SNP-based heritability estimation. *Human Genetics*, *133*(8), 1011–1022. https://doi.org/10.1007/S00439-014-1441-5

Lee, J. J., McGue, M., Iacono, W. G., & Chow, C. C. (2018). The accuracy of LD Score regression as an estimator of confounding and genetic correlations in genome-wide association studies. *Genetic Epidemiology*, *42*(8), 783. https://doi.org/10.1002/GEPI.22161

Leiserson, M. D., Eldridge, J. V, Ramachandran, S., & Raphael, B. J. (2013). Network analysis of GWAS data. *Current Opinion in Genetics & Development*, *23*(6), 602–610. https://doi.org/10.1016/J.GDE.2013.09.003

Lewis, C. M., & Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine*, *12*(1), 1–11. https://doi.org/10.1186/S13073-020-00742-5

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. https://doi.org/10.1093/BIOINFORMATICS/BTP324

Li, L., Cheng, W. Y., Glicksberg, B. S., Gottesman, O., Tamler, R., Chen, R., Bottinger, E. P., & Dudley, J. T. (2015). Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Science Translational Medicine*, *7*(311). https://doi.org/10.1126/SCITRANSLMED.AAA9364/SUPPL_FILE/7-311RA174_SM.PDF

Li, Y. R., & Keating, B. J. (2014). Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. *Genome Medicine*, *6*(10), 91. https://doi.org/10.1186/s13073-014-0091-5

Lillioja, S., Mott, D. M., Spraul, M., Ferraro, R., Foley, J. E., Ravussin, E., Knowler, W. C., Bennett, P. H., & Bogardus, C. (1993). Insulin Resistance and Insulin Secretory Dysfunction as Precursors of Non-Insulin-Dependent Diabetes Mellitus: Prospective Studies of Pima Indians. *Http://Dx.Doi.Org/10.1056/NEJM199312303292703*, *329*(27), 1988–1992. https://doi.org/10.1056/NEJM199312303292703

Lindsay, R., Dabelea, D., Roumain, J., Hanson, R. L., Bennett, P. H., & Knowler, W. C. (2000). Type 2 diabetes and low birth weight: the role of paternal inheritance in the association of low birth weight and diabetes. *Diabetes*, *49*(3), 445–449. https://doi.org/10.2337/diabetes.49.3.445

Lindsay, R. S., Hanson, R. L., Bennett, P. H., & Knowler, W. C. (2000). Secular trends in birth weight, BMI, and diabetes in the offspring of diabetic mothers. *Diabetes Care*, *23*(9), 1249–1254. https://doi.org/10.2337/DIACARE.23.9.1249

Liu, N., Sawyer, S. L., Mukherjee, N., Pakstis, A. J., Kidd, J. R., Kidd, K. K., Brookes, A. J., & Zhao, H. (2004). Haplotype block structures show significant variation among populations. *Genetic Epidemiology*, *27*(4), 385–400. https://doi.org/10.1002/GEPI.20026

Lloyd-Jones, L. R., Zeng, J., Sidorenko, J., Yengo, L., Moser, G., Kemper, K. E., Wang, H., Zheng, Z., Magi, R., Esko, T., Metspalu, A., Wray, N. R., Goddard, M. E., Yang, J., & Visscher, P. M. (2019). Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-12653-0

Lunde, A., Melve, K. K., Gjessing, H. K., Skjærven, R., & Irgens, L. M. (2007). Genetic and Environmental Influences on Birth Weight, Birth Length, Head Circumference, and Gestational Age by Use of Population-based Parent-Offspring Data. *American Journal of Epidemiology*, *165*(7), 734–741. https://doi.org/10.1093/AJE/KWK107

Lyssenko, V., Jonsson, A., Almgren, P., Pulizzi, N., Isomaa, B., Tuomi, T., Berglund, G., Altshuler, D., Nilsson, P., & Groop, L. (2008). Clinical Risk Factors, DNA Variants, and the Development of Type 2 Diabetes. *New England Journal of Medicine*, *359*(21), 2220–2232. https://doi.org/10.1056/NEJMoa0801869

Mägi, R., Horikoshi, M., Sofer, T., Mahajan, A., Kitajima, H., Franceschini, N., McCarthy, M. I., Morris, A. P., & Morris, A. P. (2017). Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Human Molecular Genetics*, *26*(18), 3639–3650. https://doi.org/10.1093/hmg/ddx280

Mahajan, A., Go, M. J., Zhang, W., Below, J. E., Gaulton, K. J., Ferreira, T., Horikoshi, M., Johnson, A. D., Ng, M. C. Y., Prokopenko, I., Saleheen, D., Wang, X., Zeggini, E., Abecasis, G. R., Adair, L. S., Almgren, P., Atalay, M., Aung, T., Baldassarre, D., … Morris, A. P. (2014). Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature Genetics*, *46*(3), 234–244. https://doi.org/10.1038/NG.2897

Mahajan, A., Spracklen, C. N., Zhang, W., Ng, M. C., Petty, L. E., Kitajima, H., Yu, G. Z., Rüeger, S., Speidel, L., Kim, Y. J., Horikoshi, M., Mercader, J. M., Taliun, D., Moon, S., Kwak, S.-H., Robertson, N. R., Rayner, N. W., Loh, M., Kim, B.-J., … Lindgren, C. M. (2020). Trans-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. In *medRxiv*. Abel N Kho 142. https://doi.org/10.1101/2020.09.22.20198937

Mahajan, A., Spracklen, C. N., Zhang, W., Ng, M. C. Y., Petty, L. E., Kitajima, H., Yu, G. Z., Rüeger, S., Speidel, L., Kim, Y. J., Horikoshi, M., Mercader, J. M., Taliun, D., Moon, S., Kwak, S.-H., Robertson, N. R., Rayner, N. W., Loh, M., Kim, B.-J., … Morris, A. P. (2022). Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nature Genetics*, *54*(5), 560–572.

https://doi.org/10.1038/s41588-022-01058-3

Mahajan, A., Taliun, D., Thurner, M., Robertson, N. R., Torres, J. M., Rayner, N. W., Payne, A. J., Steinthorsdottir, V., Scott, R. A., Grarup, N., Cook, J. P., Schmidt, E. M., Wuttke, M., Sarnowski, C., Mägi, R., Nano, J., Gieger, C., Trompet, S., Lecoeur, C., … McCarthy, M. I. (2018a). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature Genetics*, *50*(11), 1505–1513. https://doi.org/10.1038/s41588-018-0241-6

Mahajan, A., Taliun, D., Thurner, M., Robertson, N. R., Torres, J. M., Rayner, N. W., Payne, A. J., Steinthorsdottir, V., Scott, R. A., Grarup, N., Cook, J. P., Schmidt, E. M., Wuttke, M., Sarnowski, C., Mägi, R., Nano, J., Gieger, C., Trompet, S., Lecoeur, C., … McCarthy, M. I. (2018b). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature Genetics*, *50*(11), 1505–1513. https://doi.org/10.1038/s41588-018-0241-6

Mahajan, A., Wessel, J., Willems, S. M., Zhao, W., Robertson, N. R., Chu, A. Y., Gan, W., Kitajima, H., Taliun, D., Rayner, N. W., Guo, X., Lu, Y., Li, M., Jensen, R. A., Hu, Y., Huo, S., Lohman, K. K., Zhang, W., Cook, J. P., … McCarthy, M. I. (2018). Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nature Genetics*, *50*(4), 559–571. https://doi.org/10.1038/s41588-018-0084-1

Malhotra, A., Kobes, S., Bogardus, C., Knowler, W. C., Baier, L. J., & Hanson, R. L. (2014). Assessing Accuracy of Genotype Imputation in American Indians. *PLOS ONE*, *9*(7), e102544. https://doi.org/10.1371/JOURNAL.PONE.0102544

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., … Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747–753. https://doi.org/10.1038/nature08494

Marigorta, U. M., & Navarro, A. (2013). High Trans-ethnic Replicability of GWAS Results Implies Common Causal Variants. *PLoS Genetics*, *9*(6), e1003566. https://doi.org/10.1371/journal.pgen.1003566

Marley, T. L., & Metzger, M. W. (2015). A longitudinal study of structural risk factors for obesity and diabetes among American Indian young adults, 1994-2008. *Preventing Chronic Disease*, *12*(5). https://doi.org/10.5888/pcd12.140469

Mars, N., Koskela, J. T., Ripatti, P., Kiiskinen, T. T. J., Havulinna, A. S., Lindbohm, J. V., Ahola-Olli, A., Kurki, M., Karjalainen, J., Palta, P., Neale, B. M., Daly, M., Salomaa, V., Palotie, A., Widén, E., & Ripatti, S. (2020). Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nature Medicine*, *26*(4), 549–557. https://doi.org/10.1038/s41591-020-0800-0

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019a). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, *51*(4), 584–591. https://doi.org/10.1038/s41588-019-0379-x

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019b). Current clinical use of polygenic scores will risk exacerbating health disparities. *Nature Genetics*, *51*(4), 584. https://doi.org/10.1038/S41588-019-0379-X

Mayhew, A. J., & Meyre, D. (2017). Assessing the Heritability of Complex Traits in Humans:

Methodological Challenges and Opportunities. *Current Genomics*, *18*(4), 332. https://doi.org/10.2174/1389202918666170307161450

Mayosi B. M., Keavney B., Kardos A., Davies C. H., Ratcliffe P. J., Farall M., Watkins H., & Davies CH. (2002). Electrocardiographic measures of left ventricular hypertrophy show greater heritability than echocardiographic left ventricular mass: a family study. *European Heart Journal*, *23*, 1963–1971. https://watermark.silverchair.com/1963.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAnIwggJuBgkqhkiG9w0BBwagggJfMIICWwIBADCCAlQGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMj50Amth-vDEKY7NMAgEQgIICJfSeECvAspdUPiN72756u0MUeJXfGr5jhTpOUmkSQSXVbEFY_Q

McCance, D. R., Pettitt, D. J., Hanson, R. L., Jacobsson, L. T. H., Knowler, W. C., & Bennett, P. H. (1994). Birth weight and non-insulin dependent diabetes: Thrifty genotype, thrifty phenotype, or surviving small baby genotype? *BMJ*, *308*(6934), 942. https://doi.org/10.1136/bmj.308.6934.942

McCarthy, M. I. (2017). Painting a new picture of personalised medicine for diabetes. *Diabetologia*, *60*(5), 793–799. https://doi.org/10.1007/S00125-017-4210-X

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L. J., Zhang, H., Mahajan, A., … Marchini, J. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, *48*(10), 1279–1283. https://doi.org/10.1038/NG.3643

Meigs, J. B., Shrader, P., Sullivan, L. M., McAteer, J. B., Fox, C. S., Dupuis, J., Manning, A. K., Florez, J. C., Wilson, P. W. F., D'Agostino, R. B., & Cupples, L. A. (2008). Genotype score in addition to common risk factors for prediction of type 2 diabetes. *New England Journal of Medicine*, *359*(21), 2208–2219. https://doi.org/10.1056/NEJMoa0804742

Mission, J. F., Catov, J., Deihl, T. E., Feghali, M., & Scifres, C. (2017). Early pregnancy diabetes screening and diagnosis: Prevalence, rates of abnormal test results, and associated factors. *Obstetrics and Gynecology*, *130*(5), 1136–1142. https://doi.org/10.1097/AOG.0000000000002277

Mohatt, J., Gilliam, L. K., Bekris, L., Ebbesson, S., & Lernmark, A. (2002). Type 1 diabetes-related autoantibodies are rare in Alaska native populations. *International Journal of Circumpolar Health*, *61*(1), 21–31. https://doi.org/10.3402/ijch.v61i1.17402

Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segrè, A. V., Steinthorsdottir, V., Strawbridge, R. J., Khan, H., Grallert, H., Mahajan, A., Prokopenko, I., Kang, H. M., Dina, C., Esko, T., Fraser, R. M., Kanoni, S., Kumar, A., Lagou, V., Langenberg, C., … McCarthy, M. I. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics*, *44*(9), 981–990. https://doi.org/10.1038/NG.2383

Naghavi, M., Wang, H., Lozano, R., Davis, A., Liang, X., Zhou, M., Vollset, S. E., Abbasoglu Ozgoren, A., Abdalla, S., Abd-Allah, F., Abdel Aziz, M. I., Abera, S. F., Aboyans, V., Abraham, B., Abraham, J. P., Abuabara, K. E., Abubakar, I., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., … Temesgen, A. M. (2015). Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*,

385(9963), 117–171. https://doi.org/10.1016/S0140-6736(14)61682-2/ATTACHMENT/5238BC3B-701B-4AF5-A9A1-6779A1791F9B/MMC4.PDF

Nagi, D. K. K., Pettitt, D. J. J., Bennett, P. H. M., Klein, R., & Knowler, W. C. C. (1997). Diabetic retinopathy assessed by fundus photography in Pima Indians with impaired glucose tolerance and NIDDM. *Diabetic Medicine*, *14*(6), 449–456. https://doi.org/10.1002/(SICI)1096-9136(199706)14:6<449::AID-DIA367>3.0.CO;2-D

Narayan, K. M. V., Hanson, R. L., Pettitt, D. J., Bennett, P. H., & Knowler, W. C. (1996). A two-step strategy for identification of high-risk subjects for a clinical trial of prevention of NIDDM. *Diabetes Care*, *19*(9), 972–978. https://doi.org/10.2337/DIACARE.19.9.972

Ng, M. C. Y., Shriner, D., Chen, B. H., Li, J., Chen, W. M., Guo, X., Liu, J., Bielinski, S. J., Yanek, L. R., Nalls, M. A., Comeau, M. E., Rasmussen-Torvik, L., Jensen, R. A., Evans, D. S., Sun, Y. V., An, P., Patel, S. R., Lu, Y., Long, J., … Surdulescu, G. (2014). Meta-Analysis of Genome-Wide Association Studies in African Americans Provides Insights into the Genetic Architecture of Type 2 Diabetes. *PLOS Genetics*, *10*(8), e1004517. https://doi.org/10.1371/JOURNAL.PGEN.1004517

Ni, G., Moser, G., Schizophrenia Working Group of the Psychiatric Genomics Consortium, S. W. G. of the P. G., Wray, N. R., & Lee, S. H. (2018). Estimation of Genetic Correlation via Linkage Disequilibrium Score Regression and Genomic Restricted Maximum Likelihood. *American Journal of Human Genetics*, *102*(6), 1185–1194. https://doi.org/10.1016/j.ajhg.2018.03.021

North, K. E., Williams, J. T., Welty, T. K., Best, L. G., Lee, E. T., Fabsitz, R. R., Howard, B. V., & MacCluer, J. W. (2003). Evidence for joint action of genes on diabetes status and CVD risk factors in American Indians: The strong heart family study. *International Journal of Obesity*, *27*(4), 491–497. https://doi.org/10.1038/sj.ijo.0802261

Ntzani, E. E., Liberopoulos, G., Manolio, T. A., & Ioannidis, J. P. A. (2012). Consistency of genome-wide associations across major ancestral groups. *Human Genetics*, *131*(7), 1057–1071. https://doi.org/10.1007/s00439-011-1124-4

O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., Traglia, M., Huang, J., Huffman, J. E., Rudan, I., McQuillan, R., Fraser, R. M., Campbell, H., Polasek, O., Asiki, G., Ekoru, K., Hayward, C., Wright, A. F., Vitart, V., … Marchini, J. (2014). A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *PLOS Genetics*, *10*(4), e1004234. https://doi.org/10.1371/JOURNAL.PGEN.1004234

O'Connor, M. J., Schroeder, P., Huerta-Chagoya, A., Cortés-Sánchez, P., Bonàs-Guarch, S., Guindo-Martínez, M., Cole, J. B., Kaur, V., Torrents, D., Veerapen, K., Grarup, N., Kurki, M., Rundsten, C. F., Pedersen, O., Brandslund, I., Linneberg, A., Hansen, T., Leong, A., Florez, J. C., & Mercader, J. M. (2022). Recessive Genome-Wide Meta-analysis Illuminates Genetic Architecture of Type 2 Diabetes. *Diabetes*, *71*(3), 554–565. https://doi.org/10.2337/DB21-0545

Ogurtsova, K., da Rocha Fernandes, J. D., Huang, Y., Linnenkamp, U., Guariguata, L., Cho, N. H., Cavan, D., Shaw, J. E., & Makaroff, L. E. (2017). IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Research and Clinical Practice*, *128*, 40–50. https://doi.org/10.1016/J.DIABRES.2017.03.024

Orban, T., Sosenko, J. M., Cuthbertson, D., Krischer, J. P., Skyler, J. S., Jackson, R., Yu, L.,

Palmer, J. P., Schatz, D., & Eisenbarth, G. (2009). Pancreatic islet autoantibodies as predictors of type 1 diabetes in the Diabetes Prevention Trial-Type 1. *Diabetes Care*, *32*(12), 2269–2274. https://doi.org/10.2337/DC09-0934

Ørskou, J., Henriksen, T. B., Kesmodel, U., & Secher, N. J. (2003). Maternal characteristics and lifestyle factors and the risk of delivering high birth weight infants. *Obstetrics and Gynecology*, *102*(1), 115–120. https://doi.org/10.1016/S0029-7844(03)00402-2

Osabe, D., Tanahashi, T., Nomura, K., Shinohara, S., Nakamura, N., Yoshikawa, T., Shiota, H., Keshavarz, P., Yamaguchi, Y., Kunika, K., Moritani, M., Inoue, H., & Itakura, M. (2007). Evaluation of sample size effect on the identification of haplotype blocks. *BMC Bioinformatics*, *8*, 200. https://doi.org/10.1186/1471-2105-8-200

Park, H. Y., Choi, H. J., & Hong, Y. C. (2015). Utilizing Genetic Predisposition Score in Predicting Risk of Type 2 Diabetes Mellitus Incidence: A Community-based Cohort Study on Middle-aged Koreans. *Journal of Korean Medical Science*, *30*(8), 1101–1109. https://doi.org/10.3346/JKMS.2015.30.8.1101

Pavkov, M. E., Bennett, P. H., Knowler, W. C., Krakoff, J., Sievers, M. L., & Nelson, R. G. (2006). Effect of youth-onset type 2 diabetes mellitus on incidence of end-stage renal disease and mortality in young and middle-aged Pima Indians. *Journal of the American Medical Association*, *296*(4), 421–426. https://doi.org/10.1001/jama.296.4.421

Pavkov, M. E., Hanson, R. L., Knowler, W. C., Bennett, P. H., Krakoff, J., & Nelson, R. G. (2007). Changing patterns of type 2 diabetes incidence among Pima Indians. *Diabetes Care*, *30*(7), 1758–1763. https://doi.org/10.2337/dc06-2010

Pavkov, M. E., Knowler, W. C., Hanson, R. L., & Nelson, R. G. (2008). Diabetic nephropathy in American Indians, with a special emphasis on the Pima Indians. In *Current Diabetes Reports* (Vol. 8, Issue 6, pp. 486–493). NIH Public Access. https://doi.org/10.1007/s11892-008-0083-1

Pe'er, I., Yelensky, R., Altshuler, D., & Daly, M. J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genetic Epidemiology*, *32*(4), 381–385. https://doi.org/10.1002/GEPI.20303

Pearson, E. R. (2019). Type 2 diabetes: a multifaceted disease. *Diabetologia*, *62*(7), 1107. https://doi.org/10.1007/S00125-019-4909-Y

Pedersen, A. B., Mikkelsen, E. M., Cronin-Fenton, D., Kristensen, N. R., Pham, T. M., Pedersen, L., & Petersen, I. (2017). Missing data and multiple imputation in clinical epidemiological research. *Clinical Epidemiology*, *9*, 157. https://doi.org/10.2147/CLEP.S129785

Pencina, M. J., & D'Agostino, R. B. (2004). Overall C as a measure of discrimination in survival analysis: Model specific population value and confidence interval estimation. *Statistics in Medicine*, *23*(13), 2109–2123. https://doi.org/10.1002/sim.1802

Pencina, M. J., D'Agostino, R. B., Pencina, K. M., Janssens, A. C. J. W., & Greenland, P. (2012). Interpreting incremental value of markers added to risk prediction models. *American Journal of Epidemiology*, *176*(6), 473–481. https://doi.org/10.1093/aje/kws207

Pencina, M. J., D'Agostino, R. B., D'Agostino, R. B., & Vasan, R. S. (2008). Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine*, *27*(2), 157–172.

https://doi.org/10.1002/sim.2929

Perng, W., Oken, E., & Dabelea, D. (2019). Developmental overnutrition and obesity and type 2 diabetes in offspring. *Diabetologia*, *62*(10), 1779–1788. https://doi.org/10.1007/S00125-019-4914-1

Peterson, R. E., Kuchenbaecker, K., Walters, R. K., Chen, C. Y., Popejoy, A. B., Periyasamy, S., Lam, M., Iyegbe, C., Strawbridge, R. J., Brick, L., Carey, C. E., Martin, A. R., Meyers, J. L., Su, J., Chen, J., Edwards, A. C., Kalungi, A., Koen, N., Majara, L., … Duncan, L. E. (2019). Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell*, *179*(3), 589. https://doi.org/10.1016/J.CELL.2019.08.051

Pettitt, D. J., & Jovanovic, L. (2001). Birth weight as a predictor of type 2 diabetes mellitus: the U-shaped curve. *Current Diabetes Reports*, *1*(1), 78–81. https://doi.org/10.1007/s11892-001-0014-x

Pettitt, D. J., Aleck, K. A., Baird, H. R., Carraher, M. J., Bennett, P. H., & Knowler, W. C. (1988). Congenital susceptibility to NIDDM. Role of intrauterine environment. *Diabetes*, *37*(5), 622–628. https://doi.org/10.2337/DIAB.37.5.622

Pettitt, D. J., Baird, H. R., Aleck, K. A., Bennett, P. H., & Knowler, W. C. (1983). Excessive Obesity in Offspring of Pima Indian Women with Diabetes during Pregnancy. *New England Journal of Medicine*, *308*(5), 242–245. https://doi.org/10.1056/nejm198302033080502

Pettitt, D. J, Nelson, R. G., Saad, M. F., Bennett, P. H., & Knowler, W. C. (1993). Diabetes and Obesity in the Offspring of Pima Indian Women With Diabetes During Pregnancy. *Diabetes Care*, *16*(1), 310–314. https://doi.org/10.2337/DIACARE.16.1.310

Piaggi, P., Masindova, I., Muller, Y. L., Mercader, J., Wiessner, G. B., Chen, P., Kobes, S., Hsueh, W. C., Mongalo, M., Knowler, W. C., Krakoff, J., Hanson, R. L., Bogardus, C., & Baier, L. J. (2017). A genome-wide association study using a custom genotyping array identifies variants in GPR158 associated with reduced energy expenditure in American Indians. *Diabetes*, *66*(8), 2284–2295. https://doi.org/10.2337/db16-1565

Pierce, B. L., & Burgess, S. (2013). Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *American Journal of Epidemiology*, *178*(7), 1177–1184. https://doi.org/10.1093/AJE/KWT084

Pigeyre, M., Hess, S., Gomez, M. F., Asplund, O., Groop, L., Paré, G., & Gerstein, H. (2021). Validation of the classification for type 2 diabetes into five subgroups: a report from the ORIGIN trial. *Diabetologia 2021*, 1–10. https://doi.org/10.1007/S00125-021-05567-4

Polfus, L. M., Darst, B. F., Highland, H., Sheng, X., Ng, M. C. Y., Below, J. E., Petty, L., Bien, S., Sim, X., Wang, W., Fontanillas, P., Patel, Y., Preuss, M., Schurmann, C., Du, Z., Lu, Y., Rhie, S. K., Mercader, J. M., Tusie-Luna, T., … Cade, B. E. (2021). *Genetic discovery and risk characterization in type 2 diabetes across diverse populations: The 23andMe Research Team, DIAMANTE Hispanic/Latino Consortium, MEta-analysis of type 2 DIabetes in African Americans Consortium*. https://doi.org/10.1016/j.xhgg.2021.100029

Popejoy, A. B., Crooks, K. R., Fullerton, S. M., Hindorff, L. A., Hooker, G. W., Koenig, B. A., Pino, N., Ramos, E. M., Ritter, D. I., Wand, H., Wright, M. W., Yudell, M., Zou, J. Y., Plon, S. E., Bustamante, C. D., & Ormond, K. E. (2020). Clinical Genetics Lacks

Standard Definitions and Protocols for the Collection and Use of Diversity Measures. *American Journal of Human Genetics*, *107*(1), 72–82. https://doi.org/10.1016/J.AJHG.2020.05.005

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Auwera, G. A. Van der, Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., & Banks, E. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178. https://doi.org/10.1101/201178

Pyle, L., & Kelsey, M. M. (2021). Youth-onset type 2 diabetes: translating epidemiology into clinical trials. *Diabetologia*, *64*(8), 1709–1716. https://doi.org/10.1007/S00125-021-05480-W

Roberts, H., Jiles, R., Mokdad, A., Beckles, G., & Rios-Burrows, N. (2009). Trend analysis of diagnosed diabetes prevalence among American Indian/Alaska Native Young Adults - United States, 1994-2007. *Ethnicity and Disease*, *19*(3), 276–279. http://apt.allenpress.com/aptonline/?request=get-document&issn=1049-510X&volume=019&issue=03&page=0276

Rode, L., Hegaard, H. K., Kjærgaard, H., Møller, L. F., Tabor, A., & Ottesen, B. (2007). Association between maternal weight gain and birth weight. *Obstetrics and Gynecology*, *109*(6), 1309–1315. https://doi.org/10.1097/01.AOG.0000266556.69952.DE

Rodriguez-Lonebear, D. (2021). The blood line: Racialized boundary making and citizenship among Native nations. *Https://Doi.Org/10.1177/2332649220981589*, *7*(4), 527–542. https://doi.org/10.1177/2332649220981589

Roseboom, T., de Rooij, S., & Painter, R. (2006). The Dutch famine and its long-term consequences for adult health. *Early Human Development*, *82*(8), 485–491. https://doi.org/10.1016/J.EARLHUMDEV.2006.07.001

Rosenberg, N. A., Huang, L., Jewett, E. M., Szpiech, Z. A., Jankovic, I., & Boehnke, M. (2010). Genome-wide association studies in diverse populations. *Nature Reviews Genetics*, *11*(5), 356–366. https://doi.org/10.1038/nrg2760

Ruan, Y., Lin, Y. F., Feng, Y. C. A., Chen, C. Y., Lam, M., Guo, Z., Ahn, Y. M., Akiyama, K., Arai, M., Baek, J. H., Chen, W. J., Chung, Y. C., Feng, G., Fujii, K., Glatt, S. J., Ha, K., Hattori, K., Higuchi, T., Hishimoto, A., … Ge, T. (2022). Improving polygenic prediction in ancestrally diverse populations. *Nature Genetics*, *54*(5), 573–580. https://doi.org/10.1038/s41588-022-01054-7

Rubin, D. B. (1996). Multiple Imputation After 18+ Years. *Journal of the American Statistical Association*, *91*(434), 473. https://doi.org/10.2307/2291635

Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N., Colagiuri, S., Guariguata, L., Motala, A. A., Ogurtsova, K., Shaw, J. E., Bright, D., & Williams, R. (2019). Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Research and Clinical Practice*, *157*, 107843. https://doi.org/10.1016/J.DIABRES.2019.107843

Saran, R., Robinson, B., Abbott, K. C., Agodoa, L. Y. C., Bragg-Gresham, J., Balkrishnan, R., Bhave, N., Dietrich, X., Ding, Z., Eggers, P. W., Gaipov, A., Gillen, D., Gipson, D., Gu, H., Guro, P., Haggerty, D., Han, Y., He, K., Herman, W., … Shahinian, V. (2019). US

Renal Data System 2018 Annual Data Report: Epidemiology of Kidney Disease in the United States. *American Journal of Kidney Diseases*, *73*(3), A7–A8. https://doi.org/10.1053/J.AJKD.2019.01.001

Saran, R., Robinson, B., Abbott, K. C., Bragg-Gresham, J., Chen, X., Gipson, D., Gu, H., Hirth, R. A., Hutton, D., Jin, Y., Kapke, A., Kurtz, V., Li, Y., McCullough, K., Modi, Z., Morgenstern, H., Mukhopadhyay, P., Pearson, J., Pisoni, R., … Shahinian, V. (2020). US Renal Data System 2019 Annual Data Report: Epidemiology of Kidney Disease in the United States. *American Journal of Kidney Diseases*, *75*(1), A6–A7. https://doi.org/10.1053/j.ajkd.2019.09.003

Sattar, N., Rawshani, A., Franzén, S., Rawshani, A., Svensson, A.-M., Rosengren, A., McGuire, D. K., Eliasson, B., & Gudbjörnsdottir, S. (2019). Age at Diagnosis of Type 2 Diabetes Mellitus and Associations With Cardiovascular and Mortality Risks. *Circulation*, *139*(19), 2228–2237. https://doi.org/10.1161/CIRCULATIONAHA.118.037885

Saxton, S. N., Clark, B. J., Withers, S. B., Eringa, E. C., & Heagerty, A. M. (2019). Mechanistic links between obesity, diabetes, and blood pressure: Role of perivascular adipose tissue. *Physiological Reviews*, *99*(4), 1701–1763. https://doi.org/10.1152/physrev.00034.2018

Scheinost, D., Noble, S., Horien, C., Greene, A. S., Lake, E. M., Salehi, M., Gao, S., Shen, X., O'Connor, D., Barron, D. S., Yip, S. W., Rosenberg, M. D., & Constable, R. T. (2019). Ten simple rules for predictive modeling of individual differences in neuroimaging. In *NeuroImage* (Vol. 193, pp. 35–45). Academic Press Inc. https://doi.org/10.1016/j.neuroimage.2019.02.057

Schluchter, M. D. (2003). Publication bias and heterogeneity in the relationship between systolic blood pressure, birth weight, and catch-up growth--a meta analysis. *Journal of Hypertension*, *21*(2), 273–279. https://doi.org/10.1097/00004872-200302000-00017

Schulz, L. O., Bennett, P. H., Ravussin, E., Kidd, J. R., Kidd, K. K., Esparza, J., & Valencia, M. E. (2006). Effects of traditional and western environments on prevalence of type 2 diabetes in Pima Indians in Mexico and the U.S. *Diabetes Care*, *29*(8), 1866–1871. https://doi.org/10.2337/dc06-0138

Schwartz, S. S., Epstein, S., Corkey, B. E., Grant, S. F. A., Gavin, J. R., & Aguilar, R. B. (2016). The Time Is Right for a New Classification System for Diabetes: Rationale and Implications of the β-Cell–Centric Classification Schema. *Diabetes Care*, *39*(2), 179–186. https://doi.org/10.2337/DC15-1585

Scott, R. A., Scott, L. J., Mägi, R., Marullo, L., Gaulton, K. J., Kaakinen, M., Pervjakova, N., Pers, T. H., Johnson, A. D., Eicher, J. D., Jackson, A. U., Ferreira, T., Lee, Y., Ma, C., Steinthorsdottir, V., Thorleifsson, G., Qi, L., Van Zuydam, N. R., Mahajan, A., … Prokopenko, I. (2017). An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes*, *66*(11), 2888–2902. https://doi.org/10.2337/DB16-1253/-/DC1

Seidlerová, J., Bochud, M., Staessen, J. A., Cwynar, M., Dolejsová, M., Kuznetsova, T., Nawrot, T., Olszanecka, A., Stolarz, K., Thijs, L., Wojciechowska, W., Struijker-Boudier, H. A., Kawecka-Jaszcz, K., Elston, R. C., Fagard, R., Filipovský, J., & EPOGH investigators. (2008). Heritability and intrafamilial aggregation of arterial characteristics. *Journal of Hypertension*, *26*(4), 721–728. https://doi.org/10.1097/HJH.0b013e3282f4d1e7

Shields, B. M., Shepherd, M., Hudson, M., McDonald, T. J., Colclough, K., Peters, J., Knight, B., Hyde, C., Ellard, S., Pearson, E. R., & Hattersley, A. T. (2017). Population-based assessment of a biomarker-based screening pathway to aid diagnosis of monogenic diabetes in young-onset patients. *Diabetes Care*, *40*(8), 1017. https://doi.org/10.2337/DC17-0224

Sluiter, W. J., Erkelens, D. W., Terpstra, P., Reitsma, W. D., & Doorenbos, H. (1976). Glucose Tolerance and Insulin Release, A Mathematical Approach: II. Approximation of the Peripheral Insulin Resistance After Oral Glucose Loading. *Diabetes*, *25*(4), 245–249. https://doi.org/10.2337/DIAB.25.4.245

Sokal, R. R., & Rohlf, F. J. (1981). *Biometry : the principles and practice of statistics in biological research*. W.H. Freeman.

Spracklen, C. N., Horikoshi, M., Kim, Y. J., Lin, K., Bragg, F., Moon, S., Suzuki, K., Tam, C. H. T., Tabara, Y., Kwak, S. H., Takeuchi, F., Long, J., Lim, V. J. Y., Chai, J. F., Chen, C. H., Nakatochi, M., Yao, J., Choi, H. S., Iyengar, A. K., … Sim, X. (2020). Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature*, *582*(7811), 240–245. https://doi.org/10.1038/s41586-020-2263-3

Stein, C. E., Fall, C. H. D., Kumaran, K., Osmond, C., Cox, V., & Barker, D. J. P. (1996). Fetal growth and coronary heart disease in south India. *Lancet (London, England)*, *348*(9037), 1269–1273. https://doi.org/10.1016/S0140-6736(96)04547-3

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*, *338*(7713), 157–160. https://doi.org/10.1136/BMJ.B2393

Strmic-Pawl, H. V., Jackson, B. A., & Garner, S. (2017). Race Counts: Racial and Ethnic Data on the U.S. Census and the Implications for Tracking Inequality. *Sociology of Race and Ethnicity*, *4*(1), 1–13. https://doi.org/10.1177/2332649217742869

Suzuki, K., Akiyama, M., Ishigaki, K., Kanai, M., Hosoe, J., Shojima, N., Hozawa, A., Kadota, A., Kuriki, K., Naito, M., Tanno, K., Ishigaki, Y., Hirata, M., Matsuda, K., Iwata, N., Ikeda, M., Sawada, N., Yamaji, T., Iwasaki, M., … Kadowaki, T. (2019). Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. *Nature Genetics*, *51*(3), 379–386. https://doi.org/10.1038/s41588-018-0332-4

TallBear, K. (2013). *Native American DNA: tribal belonging and the false promise of genetic science*.

Tan, V. Y. F., & Févotte, C. (2013). Automatic relevance determination in nonnegative matrix factorization with the β-divergence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(7), 1592–1605. https://doi.org/10.1109/TPAMI.2012.240

Tataranni, P. A., & Ravussin, E. (1995). Use of dual-energy X-ray absorptiometry in obese individuals. *The American Journal of Clinical Nutrition*, *62*(4), 730–734. https://doi.org/10.1093/AJCN/62.4.730

The SIGMA Type 2 Diabetes Consortium, Williams Amy, A. L., Jacobs Suzanne, S. B. R., Moreno-Macías, H., Huerta-Chagoya, A., Churchhouse, C., Márquez-Luna, C., Gómez-Vázquez, M. J., Burtt Noël, N. P., Aguilar-Salinas, C. A., González-Villalpando, C., Florez, J. C., Orozco, L., Tusié-Luna, T., Altshuler, D., Ripke, S., Manning, A. K., García-Ortíz, H., Neale, B., … Orozco Lorena, L. (2014). Sequence variants in

SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature*, *506*(7486), 97–101. https://doi.org/10.1038/NATURE12828

Thompson, D. B., Janssen, R. C., Ossowski, V. M., Prochazka, M., Knowler, W. C., & Bogardus, C. (1995). Evidence for Linkage Between a Region on Chromosome 1p and the Acute Insulin Response in Pima Indians. *Diabetes*, *44*(4), 478–481. https://doi.org/10.2337/DIAB.44.4.478

Thomsen, S. K., Ceroni, A., Van De Bunt, M., Burrows, C., Barrett, A., Scharfmann, R., Ebner, D., McCarthy, M. I., & Gloyn, A. L. (2016). Systematic Functional Characterization of Candidate Causal Genes for Type 2 Diabetes Risk Variants. *Diabetes*, *65*(12), 3805–3811. https://doi.org/10.2337/DB16-0361

Torres, J. M., Abdalla, M., Payne, A., Fernandez-Tajes, J., Thurner, M., Nylander, V., Gloyn, A. L., Mahajan, A., & McCarthy, M. I. (2020). A Multi-omic Integrative Scheme Characterizes Tissues of Action at Loci Associated with Type 2 Diabetes. *American Journal of Human Genetics*, *107*(6), 1011–1028. https://doi.org/10.1016/J.AJHG.2020.10.009

Tuttle, K. R., Bakris, G. L., Bilous, R. W., Chiang, J. L., De Boer, I. H., Goldstein-Fuchs, J., Hirsch, I. B., Kalantar-Zadeh, K., Narva, A. S., Navaneethan, S. D., Neumiller, J. J., Patel, U. D., Ratner, R. E., Whaley-Connell, A. T., & Molitch, M. E. (2014). Diabetic Kidney Disease: A Report From an ADA Consensus Conference. *Diabetes Care*, *37*(10), 2864–2883. https://doi.org/10.2337/DC14-1296

Tyrrell, J. S., Yaghootkar, H., Freathy, R. M., Hattersley, A. T., & Frayling, T. M. (2013). Parental diabetes and birthweight in 236 030 individuals in the UK biobank study. *International Journal of Epidemiology*, *42*(6), 1714–1723. https://doi.org/10.1093/IJE/DYT220

Udler, M. S., Kim, J., von Grotthuss, M., Bonàs-Guarch, S., Cole, J. B., Chiou, J., Boehnke, M., Laakso, M., Atzmon, G., Glaser, B., Mercader, J. M., Gaulton, K., Flannick, J., Getz, G., Florez, J. C., & Florez, J. C. (2018). Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. *PLOS Medicine*, *15*(9), e1002654. https://doi.org/10.1371/JOURNAL.PMED.1002654

Udler, M. S., McCarthy, M. I., Florez, J. C., & Mahajan, A. (2019). Genetic risk scores for diabetes diagnosis and precision medicine. *Endocrine Reviews*, *40*(6), 1500–1520. https://doi.org/10.1210/ER.2019-00088

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1), 1–21. https://doi.org/10.1038/s43586-021-00056-9

Van Baal, C. G. M., & Boomsma, D. I. (1998). Etiology of individual differences in birth weight of twins as a function of maternal smoking during pregnancy. *Twin Research and Human Genetics*, *1*(3), 123–130. https://doi.org/10.1375/TWIN.1.3.123

Van der Auwera, G., & O'Connor, B. (2020). *Genomics in the Cloud*. https://www.oreilly.com/library/view/genomics-in-the/9781491975183/

Vassy, J. L., Hivert, M. F., Porneala, B., Dauriz, M., Florez, J. C., Dupuis, J., Siscovick, D. S., Fornage, M., Rasmussen-Torvik, L. J., Bouchard, C., & Meigs, J. B. (2014). Polygenic type 2 diabetes prediction at the limit of common variant detection. *Diabetes*, *63*(6), 2172–2182. https://doi.org/10.2337/db13-1663

Vaxillaire, M., Yengo, L., Lobbens, S., Rocheleau, G., Eury, E., Lantieri, O., Marre, M., Balkau, B., Bonnefond, A., & Froguel, P. (2014). Type 2 diabetes-related genetic risk scores associated with variations in fasting plasma glucose and development of impaired glucose homeostasis in the prospective DESIR study. *Diabetologia*, *57*(8), 1601–1610. https://doi.org/10.1007/s00125-014-3277-x

Vickers, A. J., Cronin, A. M., Elkin, E. B., & Gonen, M. (2008). Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Medical Informatics and Decision Making*, *8*(1), 53. https://doi.org/10.1186/1472-6947-8-53

Vickers, A. J., & Elkin, E. B. (2006). Decision curve analysis: A novel method for evaluating prediction models. *Medical Decision Making*, *26*(6), 565–574. https://doi.org/10.1177/0272989X06295361

Vijayakumar, P., Nelson, R. G., Hanson, R. L., Knowler, W. C., & Sinha, M. (2017). HbA1c and the prediction of type 2 diabetes in children and adults. *Diabetes Care*, *40*(1), 16–21. https://doi.org/10.2337/dc16-1358

Vilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., Genovese, G., Loh, P.-R., Bhatia, G., Do, R., Hayeck, T., Won, H.-H., Ripke, S., Neale, B. M., Corvin, A., Walters, J. T. R., Farh, K.-H., Holmans, P. A., Lee, P., … Price, A. L. (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *The American Journal of Human Genetics*, *97*(4), 576–592. https://doi.org/10.1016/J.AJHG.2015.09.001

Voight, B. F., Scott, L. J., Steinthorsdottir, V., Morris, A. P., Dina, C., Welch, R. P., Zeggini, E., Huth, C., Aulchenko, Y. S., Thorleifsson, G., McCulloch, L. J., Ferreira, T., Grallert, H., Amin, N., Wu, G., Willer, C. J., Raychaudhuri, S., McCarroll, S. A., Langenberg, C., … McCarthy, M. I. (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genetics*, *42*(7), 579–589. https://doi.org/10.1038/NG.609

Vujkovic, M., Keaton, J. M., Lynch, J. A., Miller, D. R., Zhou, J., Tcheandjieu, C., Huffman, J. E., Assimes, T. L., Lorenz, K., Zhu, X., Hilliard, A. T., Judy, R. L., Huang, J., Lee, K. M., Klarin, D., Pyarajan, S., Danesh, J., Melander, O., Rasheed, A., … Saleheen, D. (2020). Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nature Genetics*, *52*(7), 680–691. https://doi.org/10.1038/s41588-020-0637-y

Wand, H., Lambert, S. A., Tamburro, C., Iacocca, M. A., O'Sullivan, J. W., Sillari, C., Kullo, I. J., Rowley, R., Dron, J. S., Brockman, D., Venner, E., McCarthy, M. I., Antoniou, A. C., Easton, D. F., Hegele, R. A., Khera, A. V., Chatterjee, N., Kooperberg, C., Edwards, K., … Wojcik, G. L. (2021). Improving reporting standards for polygenic scores in risk prediction studies. *Nature*, *591*(7849), 211–219. https://doi.org/10.1038/s41586-021-03243-6

Wang, K., Gaitsch, H., Poon, H., Cox, N. J., & Rzhetsky, A. (2017). Classification of common human diseases derived from shared genetic and environmental determinants. *Nature Genetics*, *49*(9), 1319–1325. https://doi.org/10.1038/NG.3931

Wang, T., Huang, T., Li, Y., Zheng, Y., Manson, J. A. E., Hu, F. B., & Qi, L. (2016). Low birthweight and risk of type 2 diabetes: a Mendelian randomisation study. *Diabetologia*, *59*(9), 1920–1927. https://doi.org/10.1007/S00125-016-4019-Z/FIGURES/3

Wang, X., Liu, X., Sim, X., Xu, H., Khor, C.-C., Ong, R. T.-H., Tay, W.-T., Suo, C., Poh, W.-T., Ng, D. P.-K., Liu, J., Aung, T., Chia, K.-S., Wong, T.-Y., Tai, E.-S., & Teo, Y.-Y. (2012). A statistical method for region-based meta-analysis of genome-wide association studies in genetically diverse populations. *European Journal of Human Genetics*, *20*(4), 469. https://doi.org/10.1038/EJHG.2011.219

Wang, Y., Guo, J., Ni, G., Yang, J., Visscher, P. M., & Yengo, L. (2020). Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nature Communications*, *11*(1), 1–9. https://doi.org/10.1038/s41467-020-17719-y

Warrington, N. M., Beaumont, R. N., Horikoshi, M., Day, F. R., Helgeland, Ø., Laurin, C., Bacelis, J., Peng, S., Hao, K., Feenstra, B., Wood, A. R., Mahajan, A., Tyrrell, J., Robertson, N. R., Rayner, N. W., Qiao, Z., Moen, G.-H., Vaudel, M., Marsit, C. J., … Freathy, R. M. (2019). Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nature Genetics*, *51*(5), 804–814. https://doi.org/10.1038/s41588-019-0403-1

Warrington, N. M., Freathy, R. M., Neale, M. C., & Evans, D. M. (2018). Using structural equation modelling to jointly estimate maternal and fetal effects on birthweight in the UK Biobank. *International Journal of Epidemiology*, *47*(4), 1229–1241. https://doi.org/10.1093/IJE/DYY015

Wedekind, L. E., Mitchell, C. M., Andersen, C. C., Knowler, W. C., & Hanson, R. L. (2021). Epidemiology of Type 2 Diabetes in Indigenous Communities in the United States. *Current Diabetes Reports*, *21*(11), 1–11. https://doi.org/10.1007/S11892-021-01406-3/FIGURES/1

Weedon, M. N., McCarthy, M. I., Hitman, G., Walker, M., Groves, C. J., Zeggini, E., Rayner, N. W., Shields, B., Owen, K. R., Hattersley, A. T., & Frayling, T. M. (2006). Combining Information from Common Type 2 Diabetes Risk Polymorphisms Improves Disease Prediction. *PLoS Medicine*, *3*(10), 1877–1882. https://doi.org/10.1371/JOURNAL.PMED.0030374

Whincup, P. H., Kaye, S. J., Owen, C. G., Huxley, R., Cook, D. G., Anazawa, S., Barrett-Connor, E., Bhargava, S. K., Birgisdottir, B. E., Carlsson, S., Rooij, S. R. de, Dyck, R. F., Eriksson, J. G., Falkner, B., Fall, C., Forsén, T., Grill, V., Gudnason, V., Hulman, S., … Yarbrough, D. E. (2008). Birth Weight and Risk of Type 2 Diabetes. *JAMA*, *300*(24), 2886. https://doi.org/10.1001/jama.2008.886

Wild, S., Roglic, G., Green, A., Sicree, R., & King, H. (2004). Global Prevalence of Diabetes. *Diabetes Care*, *27*(5), 1047–1053. https://doi.org/10.2337/DIACARE.27.5.1047

Will, J. C., Strauss, K. F., Mendlein, J. M., Ballew, C., White, L. L., & Peter, D. G. (1997). Diabetes mellitus among Navajo Indians: Findings from the Navajo Health and Nutrition Survey. *Journal of Nutrition*, *127*(10 SUPPL.). https://doi.org/10.1093/jn/127.10.2106s

Willemsen, G., Ward, K. J., Bell, C. G., Christensen, K., Bowden, J., Dalgård, C., Harris, J. R., Kaprio, J., Lyle, R., Magnusson, P. K. E., Mather, K. A., Ordoňana, J. R., Perez-Riquelme, F., Pedersen, N. L., Pietiläinen, K. H., Sachdev, P. S., Boomsma, D. I., & Spector, T. (2015). The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium. *Twin Research and Human Genetics*, *18*(6), 762–771. https://doi.org/10.1017/thg.2015.83

World Health Organization, & International Diabetes Federation. (2006). *Definition and Diagnosis of Diabetes Mellitus and Intermediate Hyperglycaemia*.

Xu, G., Liu, B., Sun, Y., Du, Y., Snetselaar, L. G., Hu, F. B., & Bao, W. (2018). Prevalence of diagnosed type 1 and type 2 diabetes among US adults in 2016 and 2017: Population based study. *BMJ (Online)*, *362*, 1497. https://doi.org/10.1136/bmj.k1497

Yaghootkar, H., Lotta, L. A., Tyrrell, J., Smit, R. A. J., Jones, S. E., Donnelly, L., Beaumont, R., Campbell, A., Tuke, M. A., Hayward, C., Ruth, K. S., Padmanabhan, S., Jukema, J. W., Palmer, C. C., Hattersley, A., Freathy, R. M., Langenberg, C., Wareham, N. J., Wood, A. R., … Frayling, T. M. (2016). Genetic Evidence for a Link Between Favorable Adiposity and Lower Risk of Type 2 Diabetes, Hypertension, and Heart Disease. *Diabetes*, *65*(8), 2448–2460. https://doi.org/10.2337/DB15-1671

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A. A. E., Lee, S. H., Robinson, M. R., Perry, J. R. B., Nolte, I. M., Van Vliet-Ostaptchouk, J. V., Snieder, H., Esko, T., Milani, L., Mägi, R., Metspalu, A., Hamsten, A., Magnusson, P. K. E., Pedersen, N. L., Ingelsson, E., … Visscher, P. M. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature Genetics*, *47*(10), 1114–1120. https://doi.org/10.1038/ng.3390

Yang, J., Ferreira, T., Morris, A. P., Medland, S. E., Madden, P. A. F., Heath, A. C., Martin, N. G., Montgomery, G. W., Weedon, M. N., Loos, R. J., Frayling, T. M., McCarthy, M. I., Hirschhorn, J. N., Goddard, M. E., & Visscher, P. M. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics*, *44*(4), 369–375. https://doi.org/10.1038/ng.2213

Yang, J., Weedon, M. N., Purcell, S., Lettre, G., Estrada, K., Willer, C. J., Smith, A. V., Ingelsson, E., O'Connell, J. R., Mangino, M., Mägi, R., Madden, P. A., Heath, A. C., Nyholt, D. R., Martin, N. G., Montgomery, G. W., Frayling, T. M., Hirschhorn, J. N., McCarthy, M. I., … Visscher, P. M. (2011). Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics 2011 19:7*, *19*(7), 807–812. https://doi.org/10.1038/ejhg.2011.39

Zanetti, D., Tikkanen, E., Gustafsson, S., Priest, J. R., Burgess, S., & Ingelsson, E. (2018). Birthweight, Type 2 Diabetes Mellitus, and Cardiovascular Disease: Addressing the Barker Hypothesis With Mendelian Randomization. *Circulation. Genomic and Precision Medicine*, *11*(6), e002054. https://doi.org/10.1161/CIRCGEN.117.002054

Zeng, C.-P., Chen, Y.-C., Lin, X., Greenbaum, J., Chen, Y.-P., Peng, C., Wang, X.-F., Zhou, R., Deng, W.-M., Shen, J., & Deng, H.-W. (2017). Increased identification of novel variants in type 2 diabetes, birth weight and their pleiotropic loci. *Journal of Diabetes*, *9*(10), 898–907. https://doi.org/10.1111/1753-0407.12510

Zhang, X., Gregg, E. W., Williamson, D. F., Barker, L. E., Thomas, W., Bullard, K. M. K., Imperatore, G., Williams, D. E., & Albright, A. L. (2010). A1C Level and Future Risk of Diabetes: A Systematic Review. *Diabetes Care*, *33*(7), 1665–1673. https://doi.org/10.2337/DC09-1939

Zheng, G., Freidlin, B., & Gastwirth, J. L. (2006). Robust Genomic Control for Association Studies. *American Journal of Human Genetics*, *78*(2), 350. https://doi.org/10.1086/500054

Zhu, Y., Sidell, M. A., Arterburn, D., Daley, M. F., Desai, J., Fitzpatrick, S. L., Horberg, M. A., Koebnick, C., McCormick, E., Oshiro, C., Young, D. R., & Ferrara, A. (2019). Racial/Ethnic Disparities in the Prevalence of Diabetes and Prediabetes by BMI: Patient Outcomes Research To Advance Learning (PORTAL) Multisite Cohort of Adults in the U.S. *Diabetes Care*, *42*(12), 2211–2219. https://doi.org/10.2337/DC19-0532