MATERNAL AND NEWBORN GENETIC RISK SCORE AND DYSGLYCEMIA

THE IMPACT OF MATERNAL AND/OR NEWBORN GENETIC RISK SCORES ON MATERNAL AND NEWBORN DYSGLYCEMIA

BY JAYNEEL LIMBACHIA, BMSc

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University MASTER OF SCIENCE (2019) Hamilton, Ontario

TITLE: The impact of maternal and/or newborn genetic risk scores on maternal and newborn dysglycemia
AUTHOR: Jayneel Limbachia, BMSc
SUPERVISOR: Dr. Sonia Anand
NUMBER OF PAGES: xii, 93

Lay Abstract

Background: South Asians are approximately two times more at risk for developing gestational diabetes mellitus (GDM) compared to white Caucasians. Genetic factors may contribute to this elevated risk. Polygenic risk scores (PRSs), which combine the effects of multiple disease loci and variants associated with the disease into one variable could be useful in further understanding how GDM develops in South Asians.

Methods: Data from the South Asian Birth Cohort (START) was used to test the association of three PRSs with the outcomes of interest.

Results: The type 2 diabetes PRS was independently associated with GDM. The insulin-based maternal PRS was not associated with cord blood insulin but the insulin-based newborn PRS was independently associated with cord blood insulin. However, neither the insulin-based maternal nor newborn PRS was associated with cord blood glucose/insulin ratio.

Conclusion: The PRSs suggests a possible genetic component, which contributes to abnormal glycemic status development in South Asian mothers and their newborns.

Abstract

Background: South Asians are at an increased risk of developing dysglycemia during and after pregnancy. In pregnant women, dysglycemia often develops in the form of gestational diabetes mellitus (GDM), which may predispose their newborns to adverse health outcomes through abnormal cord blood insulin levels. However, reasons for the elevated risk of dysglycemia in South Asians have not been extensively studied. Genetic factors may contribute to the heritability of GDM and abnormal cord blood insulin levels in South Asians.

Objectives: The objectives of this thesis were to test the association of:
1) A type 2 diabetes polygenic risk score with GDM in South Asian pregnant women from the South Asian Birth Cohort (START);
2) maternal and newborn insulin-based polygenic risk scores with cord blood insulin and glucose/insulin ratio in South Asian newborns from START

Methods: Three polygenic risk scores were created to test their association with participant data (N=1012) from START. GDM was defined using cut-offs established by the Born in Bradford cohort of South Asian women. The type 2 diabetes polygenic risk score was created in 832 START mothers and included 35,274 independent variants. The maternal and newborn insulin-based polygenic risk scores were created in 604 START newborns and included 1128017 independent variants. Univariate and multiple logistic and linear regression models were used to test the associations between the polygenic risk scores and dysglycemia outcomes.

Results: The type 2 diabetes polygenic risk score was associated with GDM in both univariate (OR: 2.00, 95% CI: 1.46-2.75, $P \leq 0.001$), and multivariable models (OR: 1.81, 95% CI: 1.30-2.53, $P \leq 0.001$). The maternal insulin-based polygenic risk score was not associated with cord blood insulin or cord glucose/insulin ratio. However, the newborn insulin-based polygenic risk score was associated with cord blood insulin in a multivariable model adjusted for maternal insulin-based polygenic risk score (β = 0.036, 95% CI: 0.002 – 0.069; P=0.038 among other factors.

Conclusion: A type 2 diabetes polygenic risk score and a newborn insulin-based polygenic risk score may be associated with maternal and newborn dysglycemia.

Acknowledgements

These past two years have been nothing short of inspiring in more than one way. I was fortunate to have been part of an amazing team of individuals who made sure my time at McMaster was smooth. Their infectious energy and willingness to help has been the highlight of my masters. I would like to take this opportunity to thank each and every one of them.

To start off, I am beyond grateful to my supervisor and the charismatic leader of this team, Dr. Sonia Anand for mentoring me continuously over the last two years. I have realized the importance of mentorship through her continuous involvement in my growth, both personally and academically. Thank you for being an active supervisor and for pushing me beyond my boundaries. You have led from the front — from being a perfect role model whether it be as a mentor, leader, or as a team member in collaborative endeavours. Above all, I have learned from you how to prioritize and manage time efficiently for everything that matters in life. I have learned from watching you work how anything in life is achievable if you have the willingness to work for it. Thank you so much for that. Your persistence and belief in improving South Asian health is palpable and inspires me to continue striving to do so myself.

Second, I would like to thank my committee members, Dr. Russell de Souza and Dr. Guillaume Pare for their expert opinion and their invaluable timely feedback throughout my masters. I would like to particularly express my sincerest gratitude towards Dr. Russell de Souza who has been a constant support system, always going above and beyond as a mentor, friend and confidante. Your humble attitude and willingness to help is unparalleled. I do not think I would have been able to achieve the things I was able to achieve during my masters if it wasn't for your ability to always be present, whether it be to answer my questions or to provide me with opportunities to challenge myself. I find it amazing how you continue to maintain a kind-hearted personality and instantly make people around you feel comfortable and at peace. You have become someone I can come to discuss almost anything, and you would take out the time to mentor me through it. I am more than fortunate to have found in you a great supervisor, mentor, colleague, and friend.

Third, I would like to express my deepest gratitude towards Amel. I was so fortunate to have someone like you show me the way, right from day one, all the while literally redefining patience and passion for teaching. You are simply amazing. Your ability to give is unfathomable. Your constant availability is what made me achieve what I was able to achieve in a timely manner. Your passion for your work reflects in your character and motivates others around you, including myself. It is only your passion that gave me the confidence to achieve what I was able to in the last couple of years.

Fourth, I'd like to thank Dipika Desai, Sherry Zafar, Nora Abdalla (whom I'd like to collectively refer to as the dynamic trio that run the study) for their contribution to my journey. You all provided me with opportunities that made me go above and beyond my capabilities to develop a diverse skill set. On top of that, you all were, and continue to be my biggest supporters, cheering for me on every possible occasion. I of course cannot forget Karleen, our statistician, who gave me the best advice on how to go about my projects in the most meticulous manner. Karleen, thank you for being so attentive and interested in my progress while also challenging me throughout the process in your own unique way to ensure I learned in the right way.

A personal note of thank you to Kathy Stewart and Natalie Williams who work tirelessly to make our lives simpler and easier. It was truly a privilege to have worked with individuals like yourself who are always willing to give their time to others one way or the other.

Lastly, but certainly not the least, I would like to thank my external examiner, Dr. David Meyre for being so understanding and accommodating throughout my defence process. I am fortunate to have someone with your expertise and knowledge critique my thesis with an objective lens and provide valuable suggestions that have improved my work, immensely. Thank you for your time.

This section would be incomplete without mentioning the people that worked in the background to ensure my success. First and foremost, thank you so much mom and dad for always giving me the freedom to express myself and go about things my way, whether it be academically or personally. Your support means the world to me and I have been fortunate to have unwavering support from the both of you – whether it be in the form of weekly food parcels, to not keeping any expectations from your one and only child. I thank you for setting up a great example of parenting.

Thank you to my friends and family who have been very understanding and encouraging, while being my biggest critiques. Thank you for taking interest in my endeavours while pushing me to balance my life.

Thank you to the students on our team – Ben, Priya, Pinaz, Loshana, Sujane, Talha, and Vanessa who have been a part of my growth as a student and a person in one way or the other. They have been a constant source of companionship and positive energy throughout my masters, especially towards the end—always encouraging me to be confident. Vanessa and Sujane, you both in particular have a special place in my heart for helping me get settled in the beginning when I was the most overwhelmed.

I THANK YOU ALL, immensely!

Table of Contents

List of Figures and Tables

List of abbreviations and symbols

AUC	Area Under the Curve
BMI	Body Mass Index
CI	Confidence Interval
DIAGRAM	DIAbetes Genetics Replication and Meta-analysis
GDM	Gestational Diabetes Mellitus
GWAS	Genome-Wide Association Study
GWAMA	GWAS Meta-Analysis
LD	Linkage Disequilibrium
OGTT	Oral Glucose Tolerance Test
OR	Odds Ratio
P+T	Pruning and Thresholding
PAR	population attributable risk (PAR)
PRS	Polygenic Risk Score
QC	Quality Control
SA	South Asian
SNP	Single Nucleotide Polymorphism
START	SouTh Asian biRth cohorT
T2D	Type 2 Diabetes

Declaration of Academic Achievement

I, Jayneel Limbachia, declare this thesis entitled, “the impact of maternal and/or newborn genetic risk scores on maternal and newborn dysglycemia” to be my own work, submitted for the degree of Masters of Science (M.S.c) to McMaster University. I completed this work between September 2017 and June 2019. I am the sole author of this document and no part of this work has been submitted or published elsewhere.

To the best of my knowledge, the contents of this thesis do not infringe upon any copyrights. Permissions were acquired where needed.

My supervisor, Dr. Sonia Anand, and my supervisory committee, which comprises of Dr. Russell J de Souza and Dr. Guillaume Pare´, have provided their guidance and support throughout my graduate studies. I carried out all the primary analyses and wrote this entire document myself.

Chapter 1: Introduction – Establishing the context

1.0 Gestational Diabetes Mellitus overview

1.0.1 What is Gestational Diabetes Mellitus?

Pregnancy is a state characterized by many physiological changes. Women undergo several metabolic adaptations during pregnancy to support the developing fetus. One such adaptation is the change in insulin sensitivity over the course of the pregnancy.(1) Insulin sensitivity is higher in the first half of the pregnancy but an increase in other local and placental growth hormones such as estrogen, progesterone, leptin, cortisol, placental lactogen and placental growth hormone reduces insulin sensitivity, as pregnancy progresses.(1) Since insulin is responsible for regulating plasma glucose levels, a lower insulin sensitivity (or a higher insulin resistance) in the latter stages of the pregnancy results in elevated levels of glucose in the blood.(1)

In women affected with gestational diabetes mellitus (GDM), this insulin resistance tends to be higher than in healthy pregnant women, which results in more than normal levels of glucose in the blood.(1) Thus, GDM is defined by the World Health Organization as a physiological state characterized by impaired glucose tolerance first detected anytime during pregnancy.(2)

People originating from the South Asian continent may be more prone to developing GDM due to their elevated cardiometabolic risk. There is increasing evidence suggesting higher rates of insulin resistance, glucose intolerance, and type 2 diabetes (T2D) amongst this population, attributable to a greater amount of visceral fat in South Asians compared to white Caucasians.(3-6) South Asians find it difficult to adequately regulate

plasma glucose levels due to this increased propensity for insulin resistance.(6) Although the biology underlying GDM has not been extensively studied in a South Asian context, it may be reasonable to assume that South Asian women are more susceptible to GDM based on the pathophysiology of the disease.

1.0.2. Prevalence and burden of disease in South Asians

GDM is becoming an increasing concern worldwide. The global prevalence of the disease has been estimated to be 16.9%.(7) According to the International Diabetes Federation, 21.3 million women (representing 16.2% of live births) experienced dysglycemia in pregnancy, with 85.1% of the cases being attributable to GDM.(8) Approximately, 1 in 7 births is influenced by maternal dysglycemia in the form of GDM.(8) The prevalence of hyperglycemia during pregnancy increases with age-45.4% of women who become pregnant after the age of 45 develop a form of dysglycemia during pregnancy, although the amount of pregnancies in this age group are rare.(9)

The estimated burden of GDM varies across the globe. Women aged 20-49 years from South-East Asia have the highest prevalence of GDM (24.2%), and those in Africa have the lowest (10.4% - Table 1).(9) In absolute terms GDM affects 6.9 million live births in South East Asia, and 3.4 million live births in Africa (Table 1).(9). In general, South Asians have an estimated twofold-higher risk for GDM compared to white-Caucasians.(10, 11) Several studies conducted in India have shown that the prevalence of GDM ranges from 0 to 41.9% across regions (Table 2).(12) In addition, there is a high proportion of GDM cases among South Asian immigrants to Canada and the United Kingdom (UK) (Table 3)

as well. According to a study by Anand et al. the prevalence of GDM in South Asians residing in Ontario, Canada is 36.3%(11), while the prevalence of GDM in South Asians living in the UK is 24.2 %. (13)

Table 1: The prevalence of GDM in women aged 20-49 years of age by region in 2017(9)

	Raw Prevalence (%)	Number of live births affected
Africa	10.4	3.4 million
Europe	16.2	1.7 million
Middle East and North Africa	21.8	3.8 million
North America and Carribean	14.6	1.0 million
South America and Central America	13.1	0.9 million
South East Asia	24.2	6.9 million
Western Pacific	12.6	3.6 million

Table 2: Regions with the highest and lowest prevalence estimates of GDM across India(12)

Region	**Prevalence[a]**
Highest	
Uttar Pradesh	13.4 – 41.9%
Andhra Pradesh	17.20 - 21.81%
Lowest	
Jammu and Kashmir	3.8-11%

Maharashtra	0.5-9.5%
Assam	3.0%
Manipur	0-1%
Overall India	8.9%

[a] A majority of the prevalence estimates listed here are based on the WHO99 and IADPSG definition of GDM. The review(12), however, does not provide the exact definition used to calculate each of the prevalence estimate since multiple studies provided such information.

Table 3: Prevalence of GDM among South Asians outside the Indian subcontinent(11, 13-15)

Region	**Prevalence**	**Diagnostic Criteria**	**Study Design**
Canada	36.3%	Born in Bradford OGTT cut-offs	Prospective cohort
United Kingdom	24.2%	Born in Bradford OGTT cut-offs	Prospective cohort
Australia	11.5%	NA	Cross-sectional study
USA	9.7%	Self-reported	Cross sectional analysis of a prospective cohort study

1.0.2. Diagnostic criteria for GDM

Differences in prevalence estimates are possibly due to the wide range of diagnostic criteria used in these studies.(12) Identifying criteria for GDM have been highly contested in recent times. One set of criteria is the threshold recently established by the International

Association of Diabetes and Pregnancy Study groups (IADPSG) to diagnose the disease by identifying children who were large for gestational age, highly adipose at birth, and had high cord blood C-peptide levels .(13) The IADPSG criteria was developed to elucidate whether there were any complications associated with hyperglycemia in pregnancy that is not completely characterized as overt diabetes. Researchers from the Hyperglyemia and Adverse Pregnancy Outcomes (HAPO) group had 25,505 pregnant women from nine countries undergo 75-g oral glucose tolerance tests between 24 to 32 weeks of gestation.(16) They assessed if participants who had 1-hour fasting plasma glucose levels of 5.8mmol/L or less and 2-hour plasma glucose levels of 11.1 mmol/L or less had any adverse pregnancy complications such as primary caesarean delivery, their newborns being born with a birth weight above the 90th percentile for gestational age, clinically diagnosed neonatal hypoglycemia, or cord-blood serum C-peptide level being above the 90th percentile. They also assessed secondary outcomes such as delivery before 37 weeks of gestation, shoulder dystocia or birth injury, need for intensive neonatal care, hyperbilirubinemia, and preeclampsia. The HAPO group found that the lower than clinically used diagnostic thresholds of overt diabetes were in fact associated with adverse pregnancy outcomes such as increased birth weight (OR: 1.38, 95% CI: 1.32 – 1.44) and increased cord blood C-peptide levels (OR: 1.55, 95% CI: 1.47 – 1.64).(16) The IADPSG criteria increased the amount of women diagnosed with GDM as compared to the more restrictive definitions such as the WHO or Canadian Diabetes Association criteria.(7, 13)

However, the applicability of IADPSG criteria across different ethnic groups, such as South Asians, is questionable given the differences in glucose regulations as discussed

previously. It is unclear whether the IADPSG criteria is appropriate for diagnosing GDM in South Asians, who are at a higher risk for GDM compared to White-Caucasians.(13) The Born in Bradford cohort study, conducted in South Asians of predominantly Pakistani origin residing in the UK, thus established new cut points to diagnose GDM in South Asian pregnant women: a fasting glucose level of 5.2 mmol/L or higher, or a 2-hour post-load level of 7.2 mmol/l or higher. These criteria were selected based on the specificity for high infant birthweight (>90th percentile for gestational age) and adiposity (sum of skinfold measurements > 90th percentile for gestational age) in 5408 infants born to South Asian women from this cohort.(13)

1.0.3 Fetal programming

GDM often increases the odds of large for gestational age (LGA) babies and results in other pregnancy related complications.(11, 17, 18) These adverse health outcomes in mothers and their newborns can be attributed to a concept known as "fetal programming".

Fetal programming, an idea first proposed by a British epidemiologist, David Barker, suggests that the environmental insults a fetus undergoes *in utero* will have long-term consequences on its health as an adult.(18, 19) The effects on long-term health are exacerbated when the post-natal energy environment, characterized by excess nutrition, does not match the energy conditions *in utero*.(5)

GDM may influence fetal programming by i) contributing to the "thin-fat" phenotype: a characteristic feature of babies born to South Asian women who are low in birthweight but have high visceral adiposity, ii) contributing to macrosomia - large for

gestational age babies, and/or iii) by causing beta cell dysfunction/stress in the fetal pancreas.(3, 20) In mothers with GDM, an excess amount of glucose and other growth factors such as some amino acids and lipids are passed onto the developing fetus (Figure 1), exposing the latter to a hyperglycemic environment. Insulin, however, is unable to crossover to the placenta due to its large size. (21, 22) The fetal pancreas must overwork to produce enough insulin to meet the increased glucose, which results in the development of excess fetal adipose tissue, and an increased fetal size.(22)

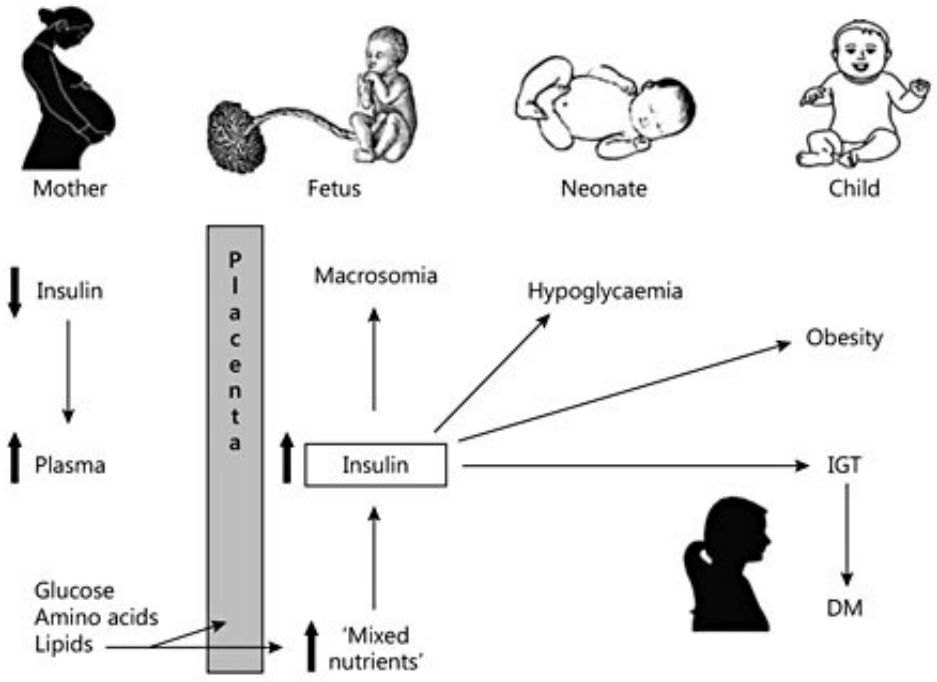


Figure 1: Exchange of nutrients at the fetal-placental membrane and the eventual progression to dysglycemia in newborns and young children.
IGT = Impaired Glucose Tolerance, DM = Diabetes Mellitus (23)
Image reproduced with permission from the publisher: S Karger AG, Basel, Switzerland

1.0.4. Future risk for disease

Based on the "fetal programming" theory, prenatal development is an important time period that defines the mother and their newborn's life trajectory and adult health. Women with GDM have an increased risk of developing cardiometabolic conditions such as type 2 diabetes (T2D), atherosclerosis, and cardiovascular disease.(24-26) A recent meta-analysis of 30 cohort studies with 2,626,905 pregnant women worldwide showed that women with GDM were almost eight times more likely (OR: 7.76, 95%CI: 5.10-11.81; P=-0.0091) to develop T2D in comparison to women with no GDM.(27) On the other hand, offspring born to mothers who had GDM during pregnancy were two-to four times more likely to have increased body fat percentage, insulin resistance in childhood, and were more likely to develop metabolic diseases such as obesity and T2D in childhood and adulthood. (17, 18, 22, 28) These findings have been replicated in South Asian women and their offspring. (22, 29) In particular, South Asian offspring born to mothers with GDM are more likely to have an increased amount of adipose tissue at birth.(22, 29)

Understanding the risk factors that eventually lead to the development of GDM and subsequent adult diseases is important in preventing and controlling the disease outcomes.

1.1 Determinants of Gestational Diabetes Mellitus (GDM)

1.1.1 Modifiable and non-modifiable risk factors

Reasons for increased GDM risk among South Asians are unclear. Observational studies conducted in multiethnic populations from different parts of the world have identified both modifiable and non-modifiable risk factors. Some of the established risk

factors include modifiable factors such as increased maternal age, living in an urbanized habitat, maternal overweight, and cigarette smoking.(30-33), and non-modifiable risk factors such as family history of T2D, low maternal height, and ethnicity.(30-33) Studies conducted in South Asian pregnant women from different parts of India and UK have also showed the role of parity (3 or more children) and vitamin B12 levels in the onset of GDM.(30, 34-36) A recent study conducted in South Asian pregnant women from Canada established another novel risk factor in the form of maternal diet quality, which is also implicated in the risk of GDM.(11)

Furthermore, apart from the environmental risk factors discussed previously, there is evidence which supports a genetic association with GDM. Genetic factors such as family history of T2D have been shown to be moderately associated with GDM (OR = 1.65, 95% CI: 1.01-1.04).(11) However, there are few large scale studies using genome wide data investigating the genetic association with GDM. To our knowledge, there are no genome wide association studies (GWASs) or genome wide association meta-analyses (GWAS-MAs) on the outcome of GDM which have been conducted in South Asians. Furthermore, to our knowledge no prior studies have estimated the heritability of GDM.

1.1.2 Studying the genetics of GDM

Several glucose- and T2D-related traits have been extensively studied and documented through GWASs and GWAS-MAs. In fact, some candidate gene studies have identified associations between T2D related genes and GDM. Independent testing of some T2D associated loci (e.g. *TCF7L2, PPARG, CDKN2A/B, KCNQ1, GCK*) has revealed an

association with GDM, when tested separately or as genetic risk scores.(37-41) The few GDM GWASs which have been conducted show signals near/within genes (*CDKAL1, MTNR1B, GCKR, PCSK1, PPP1R3B* and *G6PC2*) that have previously been shown to be associated with T2D/glucose related traits.(38, 42) South Asian specific studies have also revealed associations with T2D which have been reported to be associated with GDM in South Asians, including *HMG20A* (rs7178572), *HNF4A* (rs4812829), and *CDKAL1* (rs7754840 and rs7756992).(43, 44) Thus, it is reasonable to study T2D specific genetic variants to assess the heritability of GDM in South Asians due to shared risk factors between the two states and their similar pathophysiology.

Recently, researchers have opted to use polygenic risk scores (PRSs), derived from capturing genetic information from multiple loci, to quantify the risk of a disease.(45, 46) This approach combines multiple single variants associated with risk across the genome into a single genetic risk score. When used in combination, gene scores are more useful for assessing the joint effects of multiple genes on disease susceptibility. (46) Like other complex diseases, i.e. those caused by several genetic and non-genetic exposures, such as cardiovascular disease or T2D, GDM is unlikely to have a singular genetic etiology, and thus lends itself to the PRS approach. Use of a PRS using expanded GWASs genotype information may be particularly useful in South Asians given their higher propensity for developing GDM and its strong association with the family history of T2D, which is suggestive of an underlying genetic association.

1.2 Issue of ethnic transferability

GDM is a complex disease. Most of the GWASs conducted to date on GDM or T2D have been conducted in white/European populations. Although most of the results can be transferred or applied to other ethnicities, the predictive ability of such GWASs may be limited in non-white populations due to population differences in allelic frequencies and linkage disequilibrium structures.(47)

T2D GWASs are one of the few GWAS-MAs which include variants from both white Caucasians as well as multiethnic populations.(47) These studies suggest a cautionary approach when interpreting results from a single-ancestry GWAS and applying those results across other populations. A study conducted by Martin et al. shows that the orders of magnitude of the PRSs derived from these T2D GWASs depend on whether the data were derived from a white/European or a multiethnic population.(47) They found directional inconsistencies in all the PRSs which were developed from white/European data, when applied to other populations. For example, their height-based PRS predicted a decrease in height in populations that were genetically different than Europeans despite the empirical evidence that shows that West Africans are equally as tall as Europeans, on average.(47) Such directional inconsistencies in the PRSs developed from white/European data may not be controlled for by simply accounting for the observed vs expected bias this approach generates using an analytic technique alone.(47) The transferability of single-ancestry GWAS studies therefore requires the inclusion of diverse populations. Finally, there is a possibility that differences in genetic background or environmental exposures may lead to the same locus having a differential effect across ethnically diverse populations

when assessing the association between common variants and T2D.(48) This may explain why novel variants are often identified in non-white/European populations. (43) For example, variants in *KCNQ1,* were associated with the risk of developing T2D in South East Asian populations,(48) while variants of HMG20A and HNF4A have been shown to be associated with GDM in South Asians through GWASs. Conducting more ethnically diverse GWASs may be necessary to identify other unique variants that may be implicated in disease risk and further understand disease pathology.

The utility of a PRS developed using GWAS summary statistics depends on genetic similarity between the group in which the GWAS was conducted and the target group to which PRS is being applied.(47) Conducting genetic studies in ethnic groups like South Asians is important to further understand the biological and genetic mechanisms behind T2D and GDM, since they share the largest burden of such metabolic diseases and are therefore more likely to harbour rare genetic variants exclusive to South Asians.

Chapter 2: Genetic Contribution to Gestational Diabetes in South Asian women: Analysis from the START-Canada Birth Cohort study - Paper

2.0 Abstract

Background: Women of South Asian (SA) ancestry are at increased risk of developing gestational diabetes mellitus (GDM). Few studies have investigated the contribution of the maternal genetic profile to GDM risk. We built a type 2 diabetes (T2D) polygenic risk score (PRS) based on genotypes from the genome-wide SNP genotyping array and large consortium data and investigated whether the maternal genetic load is associated with GDM.

Methods: As part of the Canadian South Asian Birth Cohort (START) prospective birth cohort study we recruited 1,012 SA pregnant women and assessed them in the second trimester. 832 women had a PRS for T2D based on a multi-ethnic GWAS meta-analysis, which included 35,274 independent variants. GDM was defined based on glucose values established by the Born in Bradford cohort of SA women.

Results: 301 (36.2%) women were classified as having GDM. The mean PRS was significantly higher in women without GDM, $P \leq 0.001$. The tertiles of the PRS (tertiles 2 and 3 versus 1) were associated with GDM in both univariate (OR: 2.00, 95% CI: 1.46-2.75, $P \leq 0.001$), and multivariable models (OR: 1.81, 95% CI: 1.30-2.53, $P \leq 0.001$) including other known predictors of GDM: maternal age, pre-pregnancy weight, family history of T2D, low quality diet and height. The population attributable risk of the PRS

tertiles 2 and 3 in the univariate model was 38.0% and in the multivariable model was 35.3%.

Discussion/Conclusion: A PRS including 35,274 independent variants for T2D is strongly associated with GDM in SA women living in Canada, independent of a reported family history of T2D, maternal age, pre-pregnancy weight, height and low diet quality.

Abbreviations

AUC	Area Under the Curve
BMI	Body Mass Index
CI	Confidence Interval
DIAGRAM	DIAbetes Genetics Replication and Meta-analysis
GDM	Gestational Diabetes Mellitus
GWAS	Genome-Wide Association Study
GWAMA	GWAS Meta-Analysis
LD	Linkage Disequilibrium
OGTT	Oral Glucose Tolerance Test
OR	Odds Ratio
P+T	Pruning and Thresholding
PAR	population attributable risk (PAR)
PRS	Polygenic Risk Score
QC	Quality Control
SA	South Asian
SNP	Single Nucleotide Polymorphism
START	SouTh Asian biRth cohorT
T2D	Type 2 Diabetes

2.1 Introduction

Gestational diabetes mellitus (GDM) is defined as diabetes first diagnosed during pregnancy. This abnormal increase in blood glucose levels is associated with an increased risk of adverse health outcomes for both mother and their fetus/child during pregnancy, and later in life.(49) It is estimated that in 2017, 1 in 7 live births were affected by gestational diabetes (GDM) worldwide.(9) Furthermore, the prevalence of GDM has been shown to vary widely between ethnicities, countries/regions.(50) For example, South Asian women (those who originate from the Indian subcontinent) have nearly a 2-fold increased odds of developing GDM, compared to white Caucasian women.(9, 13, 51-53) Reasons for this disproportionate risk are not well understood and have not been fully examined in a South Asian context. In a prior analysis, we reported that pre-pregnancy weight, low diet quality in pregnancy, advanced maternal age, maternal height, and family history of type 2 diabetes (T2D) are associated with the risk of developing GDM in South Asian women living in Canada.(11)

The contribution of genetic factors to the development of GDM is not well understood. GDM is a complex disorder, which is strongly influenced by maternal characteristics such as body weight, dietary intake, smoking status, and family history of T2D.(32, 33) It is generally accepted that GDM and T2D share a common genetic background. This is illustrated by the fact that top genetic signals from genome wide association studies (GWAS) of GDM and dysglycemia during pregnancy are located within/near genes/loci previously known for their association with glucose related traits in non-expectant populations.(38, 42) Hence, data from large T2D consortia can be used to

estimate a polygenic risk score (PRS) of T2D/GDM for each individual in studies with available genome wide genetic data. In the absence of genetic information, clinicians/researchers use self-reported family history of T2D as a proxy of the genetic risk in people at risk for GDM.

In this analysis, the PRS for each mother part of the South Asian Birth Cohort (START) was calculated using their genotypes from the genome-wide SNP genotyping array and data from a large multi-ethnic GWAS meta-analysis.(54, 55) Our aim is to test the association of this T2D PRS with GDM in South Asian pregnant women, and to assess the independent association of the PRS with GDM in a multivariable model which includes family history of T2D.

2.2 Methods

Study design and participants

The South Asian Birth Cohort (START) is a prospective cohort study designed to evaluate the environmental and genetic determinants of cardio-metabolic traits of South Asian women and their offspring, living in Canada. The rationale and study design are described elsewhere.(56) In brief, 1,012 South Asian pregnant women, aged between 18 and 40 years old, were recruited during their second trimester of pregnancy from Peel Region (Ontario, Canada) through physician referrals between July 2011 and November 2015.

All START participants provided informed consent for participation and genetic testing and the study is approved by local ethics committees (Hamilton Integrated Research Ethics Board, William Osler Health System, and Trillium Health Partners, March 3rd, 2011). A detailed description of the maternal measurements has been published previously.(11) Briefly, weight and height were measured using standard procedures, and collected information about family and personal medical history using questionnaires. A validated ethnic-specific food frequency questionnaire (57) was used to collected dietary information. Each participant without a pre-pregnancy history of T2DM had a 75-gram oral glucose tolerance test (OGTT).

Exposures and Outcomes in this analysis: Family history of T2D is defined based on maternal reported parental history of T2D. GDM status is defined using the South Asian specific cutoffs defined in the Born in Bradford study (fasting glucose level of 5.2 mmol/L or higher, or a 2-hour post load level of 7.2 mmol/L or higher).(13) Self-reported GDM status is used if these measures were unavailable. Women with pre-existing T2D were excluded from this analysis. Diet quality is coded as a dichotomous variable (low *vs* medium + high) as previously described.(11)

DNA extraction, Genotyping, Imputation and Filtering:

DNA was extracted and genotyped from a total of 867 samples (START mothers) using the Illumina Human CoreExome-24 and Infinium CoreExome-24 arrays (Illumina, San-Digeo, CA, USA). Data has been cleaned using standard quality control (QC) procedures (58) and 837 samples passed the QC. Genotypes have been subsequently phased

using SHAPEIT v2.12 (59), and imputed with the IMPUTE v2.3.2 software (60) using the 1000 Genomes (phase 3, all ethnic groups) data as a reference panel.(61) Variants with an info score ≥ 0.7 have been kept for analysis. Using these criteria, 832 START participants with known GDM status (301 cases and 531 controls) and available genotypes have been included in the analysis.

Building the PRS

The procedure used to build the PRS is described elsewhere (54). In brief, the PRS has been built based on data from the DIAGRAM consortium (55), using a pruning and thresholding method (p-value cutoff 0.2). In this analysis, we standardized the continuous PRS to a mean of 0 and a standard deviation of 1. The PRS was divided into tertiles to calculate the population attributable risk (PAR).

Statistical Analysis

We calculated means (and standard deviations [SDs]) and counts to summarize continuous variables and categorical data, respectively, and means (and standard errors [SEs]) for adjusted continuous results. First, a univariate logistic regression model was used to assess if maternal PRS was associated with GDM. We then built upon the GDM predictive model constructed in our previously published analysis (11) by adding the maternal PRS component. The effect of the addition of PRS was compared to the model without PRS using a General Likelihood Ratio test. Diet quality is coded as a dichotomous variable (low *vs* medium + high) as previously described.(11) These analyses were

conducted using SPSS v.25.(62) The Interactive Risk Attributable Program (IRAP v2.2, US National Cancer Institute) was used to calculate the PAR by considering the frequency of the exposure in the population and the relation of the exposure to GDM. For calculation of PAR, continuous variables were recoded into categorical variables: age was divided into categories [(29-31, 32-43) vs 19-28], body mass index was divided into 3 categories (<18.5, 18.5-23, and >23) and tertiles of the PRS, whereby Tertile 2 and 3 of the PRS is compared to Tertile 1.

2.3 Results

Table 5 shows the characteristics of the START cohort. The women enrolled into START originate from North India (69.5%), and Pakistan (21.5%), followed by South India and/or Sri Lankan (7.5%), and other countries (1.6%). About half the women identify as Sikh, one-quarter as Muslim, one-fifth as Hindu, and less than 5% as Christian or other faiths. Women with GDM were older, have higher pre-pregnancy weight, lower height, lower quality dietary intake, and are more likely to report a family history of T2D compared to pregnant women without GDM. (Table 5)

The standardized final PRS includes data from 35,274 variants and ranges from -2.99 to 3.21 (mean = 0, SD = 1). Women with GDM have a higher mean polygenic score compared to women without GDM (.26 [SD=.99] vs -.17 [SD=.99], P≤0.001). Similarly, women with GDM are more likely to have a PRS that is categorized in tertile 2 or 3 compared to tertile 1 (tertile 2 and 3: [76.1%] *vs*. tertile 1: [61.4%], P≤0.001).

The PRS is significantly associated with the risk of GDM in a univariate model (OR: 1.57, 95% CI: 1.35-1.82 [P≤0.001 per 1 unit increase in the PRS]). The risk of GDM increases progressively comparing tertile 2 to 1 (OR: 1.55 95% CI: 1.08-2.23, P=0.002), and tertile 3 vs 1 (OR: 2.56 95% CI: 1.79-3.65, P≤0.001). When tertile 2 and 3 are combined and compared to tertile 1, and the increase odds of GDM is OR: 2.00 (95% CI: 1.46-2.75, P≤0.001). Results for the pooled PRS categories will be presented in the rest of the analysis.

Independent predictors of GDM

In a multivariable model including maternal age, pre-pregnancy weight, height, diet quality and family history of T2D, being in the top 2 tertiles of polygenic score is strongly and independently associated with GDM, with an odds ratio (OR) of 1.81 [95% CI: 1.30-2.53, P<0.001] (Table 6). The addition of the PRS to the model reduced the effect size of the family history of T2D; however, the effect remains significant OR: 1.62 (95% CI: 1.20-2.20, P=0.002; change in models: X^2 (1) = 12.77, P=0.0004).

Population attributable risks:

The PAR of all independent predictors of GDM including categorical maternal PRS was calculated using their multivariable ORs and the exposure frequency. Results are shown in Table 7 and Figure 2. When all factors are considered, the collective PAR is 74.7% (95% CI: 65.1%-84.3%), which is a notable increase from the multivariable model without genetic information included (total PAR = 62.7%, 95% CI: 54.6% – 74.10%). The maternal PRS independently accounts for 35.3% (95% CI: 19.8% - 50.7%) of the PAR for GDM. The total inherited component of GDM, which includes maternal PRS and family history of T2D, accounts for 47.9% (95% CI: 33.8% - 62.0%) of the PAR.

2.4 Discussion

Using a T2D polygenic risk score derived from GWAS significant SNPs, we demonstrate that there is a strong genetic association with GDM in South Asian women, which accounts for a substantial proportion (35.3%) of the PAR of GDM. This association is independent and additive to other known factors associated with GDM which include family history of T2D, maternal age, pre-pregnancy weight and height, and low diet quality. Together these factors including the PRS account for a PAR of almost 75%.

The addition of PRS reduces but does not negate the impact of family history of T2D on GDM, which has an odds ratio for GDM of 1.62 and a PAR of 20.4%. This could reflect that the PRS does not capture information from all of the genetic variants associated with GDM, and/or that family history also represents non-genetic shared lifestyle factors.(63)

There has only been one published GWAS of GDM(38), and all GDM associated genes/loci known to date are also associated with T2D. However, genetic variants that are exclusively associated to GDM are yet to be discovered. Genes/loci that are significantly associated to T2D at a GWAS significance level (P-value < 5×10^{-8}) are largely common between South Asians and European populations, but prior South Asian-specific T2D GWAS have also yielded some unique variants.(64, 65) In addition, the analysis of whole genome sequencing data in South Asians performed by Chambers *et al.,* also shows a 1.5-fold enrichment for stratified SNPs at T2D loci in South Asians genomes compared to Europeans. This enrichment could underlie the increased risk of T2D in SA populations.(66) Prior lines of evidence include the following explanations: frequent

endogamous unions due to socio-cultural constructs like the caste system among South Asians, and within regions of India (i.e. North *vs* South) resulted in a reduction in heterozygosity rates, which in turn could have favored the increase in frequency of genetic variants associated with T2D/GDM.(67, 68) The caste system which was enforced in the Indian Subcontinent for several thousand years encouraged marriages among 2nd or 3rd degree relatives and/or within the same caste, village, or region. This system was declared illegal in 1950 in India, and discrimination based on caste categories is condemned by local South Asian governments. However, these cultural beliefs and practices, along with promoting marriages within Hindus, Sikh, Muslim, Christian and other subgroups are still deeply rooted in South Asian culture. Increasing the awareness that greater genetic admixture will likely lower the population's risk to develop diseases with a strong genetic component may reduce the prevalence of GDM in South Asians over time.

Our estimations indicate that GDM has a strong genetic component in START (heritability $h^2_{SWG\_SNPs}$ in START = 0.55, SE = 0.42).(54) This could explain why the PAR of our PRS (35.3%) is substantial. This observation is also in line with our additional observation that the proportion of the variance explained by the SNPs included in our top P+T PRS ($h^2_{SNPs\ in\ PRS}$) approximates 0.15, SE = 0.13. Furthermore, our results show that all independent predictors of GDM together have a PAR of 74.7%. Prevention of GDM reduces both the mother's future risk of T2D and cardiovascular disease, and lowers the offspring's risk of future obesity, insulin resistance and T2D.(24, 50, 69-72) Future interventions to reduce GDM should focus on reducing modifiable risk factors of GDM such as pre-pregnancy weight and diet quality. Genetic factors are typically considered non-

modifiable, and indeed this is true for the pregnant women at risk of GDM.

The implications of our findings should be considered. Currently, inquiring about the family history of T2D is an easy and informative way of identifying an inherited predisposition to T2DM and GDM. However, the use of genetic information in the clinical setting is expensive as it includes counselling, genotyping and analytic costs. Since studies thus far have not consistently shown that knowledge of genetic predisposition for selected health conditions serves as a motivator to health behavior change or improved treatment(73-75), whether genetic testing for GDM should be implemented in a clinical setting in is still open to debate. Hence, for now taking a thorough personal, family and dietary history along with measuring pre-pregnancy BMI and performing an OGTT is the most cost-efficient clinical risk assessment for GDM. This may change in the future as the predictive value of the PRS is refined and genotyping costs are reduced.

The strengths of our study include: i. the PRS is optimized in order to target South Asian population by using data from a multi-ethnic GWAMA and by restricting the list of variants to those tested in South Asians(54, 55), ii. the PRS is also based on genotypes from the genome-wide SNP genotyping array which allows us to capture more genetic information and improve the predictive power of our PRS, and iii. GDM status is determined using validated South Asian-specific cutoffs and objective measures (OGTT test). There are some limitations of our data: i. the PRS is based on a multi-ethnic GWAS meta-analysis of primarily white Caucasians as only ~ 20% of the study sample were South Asian, ii. the genetic variants included in the PRS represent T2D loci as there were too few GDM GWAS variants previously reported, iii. our observations are made in South Asian

pregnant women living in Canada who are predominantly of North Indian origin; and generalizability to other South Asians i.e. South Indian, or other ethnic groups should be made cautiously. (55)

Conclusion: A polygenic risk score including 35,274 independent variants for T2D is strongly associated with GDM in South Asian women, independent of a reported family history of T2DM, maternal age, pre-pregnancy weight, and low diet quality.

*Note: Manuscript to be submitted

Genetic contribution to Gestational Diabetes in South Asian women: Analysis from the START-Canada Birth Cohort study

Amel Lamri[1,2*], Jayneel Limbachia[1,2*], Karleen M. Schulze[2], Dipika Desai[2], Russell J. de Souza [2,3,] Milan Gupta[1], Guillaume Paré[2-4], Sonia S. Anand[1-3] for the START Canada Investigators

*A.L. and J.L. equally contributed to this work.

1. Department of Medicine, McMaster University and Hamilton Health Sciences, Hamilton, Ontario, Canada
2. Population Health Research Institute, McMaster University and Hamilton Health Sciences, Hamilton, Ontario, Canada
3. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada
4. Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada

Corresponding Author: Dr. Sonia S Anand,

McMaster University, MDCL 3202,

1280 Main St W, Hamilton,

Ontario L8S 4K1, Canada.

Tel.: +1-905-525-9140, ext. 21523;

Fax: +1-905-528-2814.

e-mail: anands@mcmaster.ca

Keywords: Polygenic risk score (PRS), gestational diabetes mellitus (GDM), type 2 diabetes (T2D), family history of type 2 diabetes, South Asian

WORD COUNT: 2415

Chapter 3: Genetic contribution of a maternal and newborn insulin polygenic risk score to newborn dysglycemia

3.0 Physiology of fetal glucose and insulin

Insulin is a peptide hormone responsible for regulating plasma and intracellular glucose levels across different tissues. In peripheral tissues, insulin is responsible for regulating glucose uptake by activating a signalling cascade through the GLUT-4 transporter.(76) Maternal glucose, but not insulin, crosses the placenta due to its smaller size. A specialized glucose transporter, GLUT-1, facilitates the movement of glucose across the placenta, which is independent of maternal insulin activity.(76) Umbilical cord blood insulin is produced by the fetal pancreas in response to maternal glucose levels.(77) It is an important marker of glucose homeostasis in the newborn and a potential indicator of the intrauterine environment of the developing fetus.(78)

Newborns whose mothers have GDM or impaired glucose tolerance during pregnancy are more likely to have elevated levels of cord blood insulin.(78-81) There is an increased glucose transfer from the mother to the fetus through the placenta due to the high concentration of maternal plasma glucose. The fetal pancreas responds to the increased glucose by increasing its insulin production which can lead to fetal hyperinsulinemia or insulin dysregulation at birth.(22, 80) The compensation by the fetal pancreas may affect the development and long-term function of the pancreatic beta cells, eventually leading to insulin resistance and increased risk of future metabolic diseases in the newborn (82).

3.0.1 Genetic determinants of insulin regulation and response in adults

Elevated plasma insulin in adults is a marker for subsequent GDM and T2D. Genetic variants associated with reduced beta cell function and independently with insulin resistance have been identified in prior GWASs.

Some of the main genetic loci implicated in T2D operate through their effect on insulin secretion. For example, variants associated with loci such as *TCF7L2* and *MTNR1B* reduce insulin secretion by altering the gene expression in adult pancreatic beta cells.(83) Similarly, a variant (rs1552224) at another loci in the *CENTD2* gene has been suggested to impair the function of beta cells and insulin production.(84) On the contrary, the lead variant of *KLF14* has an opposite effect on insulin levels—it has been suggested that the variant increases insulin production via an insulin resistance-like effect in peripheral tissues.(84) Based on the GWAS catalog, 9 genetic variants associated with beta cell function (measured by HOMA B) and 24 associated with peripheral insulin resistance (measured by HOMA IR, insulin resistance, and insulin sensitivity measurement) have been identified to date (at the GWAS significance level: $P \leq 5 \times 10^{-8}$).(85)

Accordingly, there are several genetic determinants of insulin levels in adults. Although the variants differ in the way they regulate insulin activity and affect subsequent disease development, it is important to note that the association of these variants to fasting insulin levels is weak. In addition, there are unique genetic variants which may be specific to a certain population or limited to the pregnancy period. There may be other nongenetic determinants of plasma insulin levels in adults, especially pregnant women.(86) Further

research is needed to elucidate if there are unique genetic variants implicated in the long term glycemic changes which occur as a result of a pregnancy.(86)

3.0.2 Genetic determinants of cord blood insulin regulation

Cord blood insulin is a marker for intrauterine glucose homeostasis. Abnormal cord blood insulin regulation may lead to a series of downstream consequences in growth, which may eventually lead to adult health problems. Thus, it is important to understand the underlying genetic factors, in addition to the more obvious environmental factors such as maternal glucose levels, which regulate cord blood insulin. There is some evidence that genome-wide epigenetic modifications could explain some of the mechanisms behind this fetal programming, which has led other studies to explore how pregnancy may affect the genetic underpinnings of future disease risk.(87)

Interestingly, the risk of T2D in newborns and the growing offspring may be mediated by genetic risk variants associated with cord blood insulin in a similar way that plasma insulin is in adults. Genetic risk variants may directly affect the ability of the fetal pancreas to secrete insulin.(81, 88) In fact, a study by Dungar et al. identified a common variant in the maternally expressed gene, *H19*, to be associated with *IGF2* (insulin regulating gene) levels in cord blood.(89) However, more evidence supporting such an association is needed—no GWASs to date have identified genetic variants that are directly associated with cord blood insulin levels. It is reasonable to consider that the genetic variants which regulate plasma insulin levels in adults may also regulate cord blood insulin levels in the developing fetus.

In this analysis, we investigated: i. the association of an insulin-based maternal PRS, and ii. insulin-based newborn PRS with cord blood insulin and cord blood glucose/insulin ratio in South Asian newborns.

3.1 Methods

3.1.1. Study design and participants

South Asians in Canada have a unique risk profile. They are at twofold greater risk for T2D and cardiovascular disease (CVD) compared to white Europeans.(90) They also have higher amounts of fasting insulin and glucose levels compared to their Chinese and Europeans counterparts.(90)

The study population for this analysis comes from START.(56) As described previously (Chapter 2), START was designed to evaluate the environmental and genetic determinants of adverse metabolic conditions among South Asian women and their offspring, living in Canada. This birth cohort aims to further understand the risk factors in pregnancy and early life that are associated with the elevated risk of T2D, atherosclerosis and coronary artery disease in the South Asian population.(56) More information on rationale, study design and participants can be found elsewhere.(56)

START participants were approached for consent to collect and analyze their DNA as described before (Chapter 2). 92.9% of pregnant women provided consent for their sample, and 91.6% for their offspring. Details regarding maternal measurements have been published elsewhere(11) and briefly described before (Chapter 2). With regards to START infants, their weight (to the nearest 10g) and length were measured at birth, 6 months and

12 months using standard procedures. Newborn adiposity was measured using skin-fold thickness at birth, 1 year, 3 and 5 years. Fetal sonograms were used to determine gestational age and assess other fetal growth characteristics.

Exposures and outcomes in this analysis: Cord blood was collected and processed at the time of birth from the umbilical vein at the local hospital following a standard protocol and transported to Hamilton for storage in liquid nitrogen. These samples were batched and subsequently insulin and glucose were analyzed using the electrochemiluminescence immunoassay (on the Roche Elecsys® 2010 immunoassay analyzer) and the Bectron Dickenson Unicell DxC 600 Synchron Clinical System assay, respectively.(56) With regards to maternal exposures, GDM status is defined using the Born in Bradford criteria (see Chapter 2).

3.1.2. DNA extraction, Genotyping, Imputation and Filtering:

Information on DNA extraction, genotyping, quality control, phasing, imputation, and filtering has been presented before (Chapter 2). For this analysis, PRSs were developed for 1441 participants in START (837 mothers and 604 infants). Cord blood insulin was available for 638 START newborns (Note: from the 1000 START newborns, 777 provided their blood for analysis. We received 739 samples back from the lab analysis, out of which 638 had insulin values available, while the remaining samples [N=98] could not be analyzed due to insulin hemolysis or miscellaneous sample problems, [N=3]). From the 638 START newborns, 515 had available DNA to be included in this analysis.

3.1.3. Building the maternal and infant PRS

The maternal and infant PRS used in this analysis were developed using a pruning and thresholding (P+T) method. The P+T method is one of the many methods used to create a PRS, which involves using empirically derived p-value thresholds to select the SNP variants that make up the PRS. The process involves clumping linked SNP variants into groups from which variants with the highest p-values (lowest significance) are selected against while the ones with the lowest p-values (highest significance) are included in the risk score.(91)

Consortium data: In brief, summary statistics (P-value, effect size) for plasma insulin levels from the Meta-analyses of Glucose and Insulin-related traits Consortium (MAGIC) were used to build the maternal and infant PRS using R v3.5(92) and Plink v1.9(93, 94) (https://www.cog-genomics.org/plink1.9). Initially, two studies from MAGIC were identified as relevant to the study objectives. The study conducted by Dupuis et al.(95) collected information from a GWAS-MA in which fasting glucose, fasting insulin (log-transformed) and indices of B-cell function (HOMA-B) and insulin resistance (HOMA-IR) in 46,186 non-diabetic participants were measured. The second study, conducted by Lagou et al., (2018 *manuscript in preparation)* is a GWAS-MA measuring fasting plasma glucose (mmol/L) and log transformed fasting plasma insulin (pmol/L concentrations at study level in 160692 European, non-diabetic individuals. The study has sex-specific summary statistics on fasting glucose in 67,506 men and 73,809 women and fasting insulin in 47,806 men and 50,404 women, collected separately. The Lagou et al. study, (2018 *manuscript in preparation)* was used for this analysis due to its larger sample size and sex specific

summary statistics. Results from this study were downloaded from MAGIC's main website (https://www.magicinvestigators.org/downloads/).

The genomic position of the variants included in the consortium were extracted from the Ensembl database using R v3.5(92) (biomaRt package)(96, 97). Both male and female effect estimates (with P-value) from Lagou data were extracted separately to create different male and female PRSs. From START, 1449 genotyped samples were used. SNPs with a minor allele frequency ≥ 0.01 in START were considered. The pruning was performed using PLINK's `--clump` function for variants with a P-value ≤ 0.05 in the consortium only. Genotypes of European participants from the 1000 genome study were used to estimate the linkage disequilibrium (LD) between variants. Other parameters were kept at default.

The associations between maternal and newborn PRS and fasting plasma insulin levels collected from START pregnant women at the time of the OGTT were tested. This was conducted to validate the PRSs since the data from Lagou et al., (2018 *manuscript in preparation)* was based on fasting plasma insulin levels in men and women. The PRSs (both maternal and newborn) are made up of 1128017 variants in total. The continuous PRSs were standardized to a mean of 0 and a standard deviation of 1.

3.1.4. Literature search informing the analytical approach for maternal and newborn PRSs

Mothers and their newborns share half of their genotypes (r = 0.5). Thus, in this analysis, it was important to delineate the separate effects of maternal and newborn PRS

on our outcome of interest in newborn's cord blood insulin. Genetic studies that consider both maternal and newborn genetic contribution to a phenotype of interest are often unable to identify whether the genetic associations reflect the effects of the newborn's own genotype or their mother's genotype or a combination of both.(98) There is also a possibility that maternal (through the intrauterine environment) and newborn genotypes (direct effect) have opposing effects (98) on the outcome of interest. Hence, a literature search was necessary to identify the methodologic approaches which capture the true effect of maternal and newborn genotypes on cord blood insulin while controlling for their correlation.

A literature search of current evidence/common practices was conducted using PUBMED. The search strategy included terms like [maternal genotypes] AND [fetal genotypes] AND [regression]", with a date range limit of the last 5 years. Sixty studies were identified. Out of the 60, only 3 were deemed relevant to the query.(98-100) In addition, two more relevant studies were reviewed based on suggestions from a colleague and content expert on this topic (Table 3).(101, 102)

Table 4: Relevant studies identified through the literature search

Study	Objective	Analysis approach	Covariates
Warrington et al., 2018	To determine whether maternal or newborn genotype or both contribute to birthweight	Structural equation modelling	Not applicable

Stordal et al., 2017	To assess the strength of association between vitamin D genetic variants in mothers and their newborns with vitamin D levels in newborn cord blood	Univariate model and multivariable regression models i) mutually adjusting for maternal and newborn SNPs and ii) newborn gene score with known covariates	Season of birth, maternal dietary intake, maternal age, pre-pregnancy BMI, geographic region of living
Tyrell et al., 2016	To assess if maternal glycemic traits-based gene scores are associated with newborn birth weight	Univariate and multivariable regression models adjusting for both maternal and newborn gene score in one model	Offspring sex, gestational age
Li et. al., 2017	To assess the association between maternal BMI based gene risk score on maternal and newborn obesity traits	Multivariable regression adjusted for both maternal and newborn genetic risk score in one model	Offspring gestational age at birth, sex, gestational weight gain, parity, gestational diabetes mellitus, smoking, and ethnicity
Li et al., 2014	To determine the (maternal and/or fetal) genetic effects associated with the development of nonsyndromic conotruncal heart defects (CTD)	Penalized Logistic Regression Approach (LASSO)	Not applicable

The most recent study, conducted by Warrington et al.(98), investigated novel ways of determining the effect of maternal and fetal genotypes independently of each other on

newborns' birthweight. The study highlights the importance of an approach that is unbiased and beyond simple univariable (one for maternal and one for fetal) regression that often fail to capture the correlation between the two. Through their analysis, Warrington et al. demonstrated the need to use structural equation modelling when genotype information is available in mothers only but phenotype (i.e. birthweight in their case) information was available in both mothers and newborns. However, it must be noted that the study does justify the use of a multivariable regression model if there are enough data to adjust for the fetal genotype in a regression model of maternal genotype on birthweight and vice versa.(98) The authors suggest that this approach yields an unbiased estimate of the effects of maternal and/or fetal genotype on birthweight by capturing the correlation between maternal and fetal genotypes as two sources of variation of equal importance.

The second study by Stordal et al.(99) was conducted to assess the influence of both maternal and fetal vitamin D genotypes on cord blood vitamin D levels. First, they used simple (univariate) linear regression analyses to study the effects of maternal and fetal genotypes individually and measured the proportion of variance explained by each. They then added maternal and fetal genotypes in a single model to test the association between the genotypes and birth weight by mutually adjusting for each other. In the model with maternal and fetal genotypes as independent predictors, adjusted for one another, only the fetal genotypes were associated with cord blood vitamin D levels.(99) Finally, they combined the fetal genotypes in a genetic risk score and tested its association with birthweight in a multivariable model with other known predictors of cord blood vitamin D levels adjusted as covariates. There were no issues with collinearity in the model with both

maternal and fetal genotypes adjusted for each other. However, the authors suggest that the the lack of association between maternal genotypes and vitamin D levels may be due to a small sample size since such relationship has been shown previously.(99)

The study by Tyrell et al.(100) aimed to assess the genetic evidence which supports the association between different maternal traits (such as BMI) and newborn birth weight. Tyrell et al. were also interested in understanding if confounders such as maternal smoking and dietary habits affect the proposed association. SNPs associated with maternal traits such as BMI, fasting plasma glucose, and systolic blood pressure were selected and a weighted genetic risk score was developed for each trait to test its association with birth weight. First, they tested different maternal genetic risk scores with newborn birth weight in univariate analyses and then repeated the analyses after adjusting for fetal genotypes in multivariable models. The maternal BMI, fasting plasma glucose and systolic blood pressure genetic risk scores were found to be associated with newborn birth weight, but the amount of variance explained was modest (0.2-5%). Replication of these results is warranted to understand the biology behind these results.

Finally, from the suggested studies, the more recent one by Li et al.(101) followed a similar statistical modelling approach as previously outlined. One of their objectives was to assess the effect of maternal BMI genetic risk score on maternal (such as pre-pregnancy BMI, postpartum weight retention at 5 years) and offspring (BMI Z-score at birth and from birth to 5 years old) traits.(101). To assess the association between the maternal BMI genetic risk score and newborn birth weight, they used a multivariable model regressing maternal BMI gene score on BMI Z-score at birth, which was then repeated by adjusting

for the newborn genetic risk score. This corrected the possible confounding effects of newborn genotype on birthweight and allowed Li et al. to study the independent association between maternal BMI gene score and newborn birth weight—which became non-significant after adjusting for the newborn genetic risk score. Once again, the lack of effect after adjustment may be attributable to a lack of power.(101)

However, the older study by Li et al. (102) used a unique approach to differentiate the effects of maternal and newborn genotype when studying the candidate genes associated with congenital heart defects. They proposed a haplotype-based analysis with a penalized regression framework (Least Absolute Shrinkage and Selection Operator) to delineate the genetic effects when mother-newborn pair data is available. The basis behind this machine learning technique requires parental genotypes to clearly differentiate the effects of maternal and newborn genotypes and also their interaction. Their technique allowed Li et al. to identify seven genes that were associated with their phenotype of interest and whether each of them corresponded to maternal main effect, newborn main effect or their interaction.(102) However, the study was powered to detect interactions between maternal and newborn genotypes from the same genomic region only. Studying multiple regions would significantly increase the statistical tests to be conducted, making the study not feasible.(102)

Based on the review, most of the studies (three(99-101) out of five studies) support the use of multivariable regression analyses conditioning maternal genotypes on newborn genotypes and vice versa to account for the correlation between the two. From the three studies, two of them are particularly relevant to our study's aim. The two studies conducted

by Tyrell et al.(100) and Li et al.(101) aimed to study a broader genetic contribution to birthweight and BMI, respectively. They were interested in validating the hypothesized causal associations between genetic variants and their outcomes of interests, whereas Stordal et al.(99) were specifically interested in knowing which of the maternal or fetal genome played a bigger role in predicting cord blood vitamin D levels. The analysis conducted in this study aims to understand the broad genetic architecture behind newborn cord blood insulin levels. There is no literature that supports the role of either the maternal or newborn genotypes in predicting cord blood insulin levels. Thus, using univariate linear regression models followed by multivariable models conditioning the effects of maternal and newborn PRSs on each other was used in the START analysis. This is justified since we had access to complete genotype and phenotypes, while the LASSO technique used by Li et al.(102) is cumbersome and limited to information derived from one genomic region only. In fact, even the Warrington et al. study, (98), which used structural equation modeling to address the correlation issue suggests that simply adjusting for maternal and newborn genotypes in one model if the data is adequate and complete with both maternal and newborn genotypes and phenotypes can provide unbiased results.

3.1.5 Statistical analysis

All analyses were conducted using SPSS v.25.(62) Two continuous outcomes were assessed – i) cord blood insulin and ii) glucose/insulin ratio (as a surrogate for insulin sensitivity). All main predictors (maternal and newborn PRSs) and outcomes (i. cord blood insulin and ii. cord blood glucose/insulin ratio) were assessed for normality (based on graphical distribution and Shapiro Wilk test $P>0.05$). Log transformations were applied to

variables that did not follow a normal distribution (i. cord blood insulin, and ii. glucose/insulin ratio).

First, a univariate model was used to assess the effects of insulin-based maternal PRS and newborn PRS (separately) on i) cord blood insulin. Second, the univariate models were built upon by adding covariates (deemed relevant based on literature(99, 101, 103) and biological plausibility) in a multivariable model. The univariate model with maternal PRS was built upon by adding maternal related covariates, such as maternal age (years), pre-pregnancy BMI (kg/m^2), parity (n, continuous), gestational weight gain during pregnancy (kg), and GDM status (yes/no). Finally, the multivariable model was tested again but with the addition of the newborn PRS as a confounder. Likewise, the univariate model with the newborn PRS was built upon by adding newborn related covariates such as newborn age (days), gestational age (weeks), newborn weight (g), newborn length (cm) and sex (M/F) in the initial multivariable model, which was then tested again by adjusting for maternal PRS in the final multivariable model.

Furthermore, a combined multivariable model regressing maternal and newborn PRS with both maternal and newborn covariates on cord blood insulin was also tested. Finally, a stratification analysis based on GDM status of mothers was conducted if one of the two PRSs (main predictors) were significantly associated with cord blood insulin in a multivariable model. The stratification analysis was complemented with an interaction term to further assess the differential effect of PRSs on cord-blood insulin based on GDM status. A similar analysis plan was followed for the ii) glucose/insulin ratio outcome.

3.2. Results

Descriptive characteristics for participants from the START cohort have been presented in Chapter 2 (Table 3).

The standardized final PRSs (both maternal and newborn) includes data from 1,128,017 SNP variants. The maternal PRS ranges from -3.10 to 3.56 (mean = 0, SD = 1), while the newborn PRS ranges from -2.60 to 2.95 (mean = 0, SD = 1). Both continuous and categorical (tertiles) maternal and newborn PRS have been assessed. Results for the continuous variables (maternal and newborn PRS) will be presented in the rest of the analysis.

i) cord blood insulin: Neither the maternal nor the newborn PRS is significantly associated with cord blood insulin in a univariate model (Table 8: P>0.05 per 1 unit increase in the PRS). The non-significant effect persists in the final multivariable model with newborn PRS adjusted for newborn covariates such as newborn age, gestational age, newborn weight, newborn length, and sex, and conditioning on maternal PRS (Table 9). In a multivariable model which includes maternal PRS adjusted for maternal covariates such as maternal age, pre-pregnancy BMI, parity, gestational weight gain, and GDM status, conditioning on newborn PRS, the newborn PRS is significantly associated with cord blood insulin (Table 10: β = 0.035, 95% CI: 0.002 – 0.069; P=0.039). However, in the combined multivariable model with both maternal and newborn covariates, including maternal and newborn PRS, the newborn PRS does not remain significantly associated with cord blood insulin (Table 11: β = 0.022, 95% CI: -0.010 – 0.055; P=0.17). The significant effect seen in the multivariable model with newborn PRS and maternal covariates also does not persist

across quartiles of cord blood insulin (P-value for linear trend=0.29)—there is a general increasing linear trend for pre-pregnancy BMI (P=0.022), gestational weight gain (P=0.008) and number of GDM cases (P=<0.001) across cord blood insulin quartiles, but not for newborn PRS (Table 12).

Moreover, in the stratification analysis, newborn PRS is nominally associated with cord blood insulin in the univariate (β = 0.041, 95% CI: <0.001 – 0.082, P=0.047) model but not in the multivariable models (β = 0.042, 95% CI: <0.001 – 0.084, P=0.051) in non-GDM participants (Table 13). No such effect was seen in either the univariate or the multivariable models of the GDM participants stratum (P>0.05; Table 14). The interaction model with newborn-based insulin PRS and GDM was also not significant ($P_{interaction}$ = 0.40).

ii) cord blood glucose/insulin ratio: Neither the maternal nor the newborn PRS is significantly associated with cord blood glucose/insulin ratio in the univariate (Table 15) or multivariable analysis (Table 16 and 17: P>0.05 per 1 unit increase in the PRS).

3.3. Summary and implications

In summary, an insulin-based newborn polygenic risk score derived from GWAS significant SNPs is nominally associated with cord blood insulin levels in newborns born to South Asian mothers from the START cohort. This association is independent of other known maternal factors such as the insulin-based maternal polygenic risk score, maternal age, pre-pregnancy BMI, parity, gestational weight gain, maternal smoking, GDM status and the insulin-based maternal polygenic risk score. However, it is important to note that

the insulin-based newborn polygenic risk score is not associated with cord blood insulin levels when the maternal factors are not held constant. In addition, the insulin-based newborn polygenic risk score is not independently associated with cord blood insulin when controlled for newborn factors such as newborn age, gestational age, newborn weight, newborn length and sex plus insulin-based maternal polygenic risk score. The sequential assessment of maternal and newborn factors as covariates in the multivariable models further elucidates how newborn genetic variants influence cord blood insulin. Based on this analysis, the newborn genotype influences cord blood insulin independent of maternal factors but not newborn factors. This may be because the newborn polygenic gene score is in itself a newborn related factor, and thus may be heavily correlated with other newborn factors included in the multivariable model. Furthermore, the maternal insulin-based polygenic risk score derived from significant SNPs is not independently associated with cord blood insulin levels in either of the adjusted models with maternal or newborn factors. The maternal genotype may not play any role in influencing cord blood insulin levels. Finally, neither the newborn nor the maternal insulin-based polygenic risk score is associated with cord blood glucose/insulin ratio in children from START.

The implications of these findings are broad. Firstly, this is a one of its kind study to establish genetic variants associated with cord blood insulin levels in South Asians. Based on our findings, the newborn's own genotype may be responsible for regulating cord blood insulin levels, independent of their mother's genotype (mediated through the intra-uterine environment). This is in line with evidence from GWASs that have assessed the role of maternal and newborn genotypes on birthweight,(104, 105) which is known to be

affected by cord blood insulin levels. In those studies, newborn genotypes explain 24% to 69% of the variance in birth weight while the maternal genetic contribution is as low as 3% to 22%.(106) The role of the newborn genotype in influencing cord blood insulin may be attributable to the fetus's response to an increased glucose demand; a common characteristic of South Asian pregnancies. Intuitively, the fact that maternal genotype does not play a role in influencing cord blood insulin is justified since maternally produced insulin does not cross the placenta. Cord blood insulin production is exclusive to the fetal pancreas. However, it must be noted that the effects of the newborn genotype are not completely independent and are in part mediated by maternal environmental factors. This is particularly important from a public health perspective as controlling the modifiable factors such as gestational weight gain or GDM during pregnancy may reduce the burden of issues against the genetic predisposition to abnormal cord blood insulin levels in South Asian newborns. Further research is warranted to determine how exactly newborn genetic variants influence cord blood insulin. Secondly, since cord blood insulin levels can be a marker of adult diseases in children,(107) this information presents a unique approach of using newborn specific PRS to evaluate and stratify at-risk children. Newborns with high insulin PRS are likely to have overworking pancreas at the time of birth which may make a case for closer monitoring of their post-natal environment, based on the "thrifty phenotype" hypothesis. Excess energy post-natally may cause a further exacerbation in pancreatic function and predispose such high-risk newborns to diseases such as T2D in the future. (20, 108, 109) Further research and GWAS assessing glycemic traits such as cord

blood insulin in multiethnic populations are needed to further understand the intra-uterine pathways that may mediate the effect of insulin-based newborn PRS on cord blood insulin.

This study has several strengths. First, the maternal and newborn PRS capture a large amount of genetic information, being adapted from studies looking at whole-genome genotypes. This adds to the predictive capability of the PRS in determining the outcome of interest. Second, the Lagou et al. (2018 *manuscript in preparation*) study had information collected separately in both males and females which adds more power and an ability to control for sex-specific differences in our analysis. Finally, we used a less extreme GWAS significant threshold to ensure a high predictive power of the PRS, without removing too many SNPs by using the established GWAS significance threshold ($P \leq 5 \times 10^{-8}$).

Nonetheless, there are some limitations inherent in the data informing the analysis. First, the null effect of the insulin-based maternal PRS may be due to a lack of power. The study did not have an adequate sample size (number of START participants with cord insulin or cord glucose/insulin ratio measurements) to rule out "no difference" (Appendix 1). Second, the genetic variants measured in the GWAS (Lagou et. al., 2018 *Manuscript in preparation*) were entirely collected in a European population. This is in contrast to the fully South Asian population from the START cohort for whom we have phenotypic information available. Ethnic differences may underestimate the true effect of the PRS and cord blood insulin or glucose/insulin ratio in our population. Third, the GWAS represents plasma insulin levels measured in an adult population while our outcomes of interest, cord blood insulin and glucose/insulin ratio are measured in offspring born to START mothers. There may be some inconsistencies in the associations between the genotypes and

phenotypes since the genetic variants are based on slightly different loci than the outcomes. The lack of effect between PRS and glucose/insulin ratio may particularly be attributed to the inconsistency between the genotypes (insulin-based) and the phenotype, which represents insulin sensitivity in this case. Finally, the generalizability of our analysis (with regards to the newborn PRS association with cord blood insulin) is limited to South Asians from Canada, mainly originating from Northern India.

Overall, the results from this analysis may support the use of an insulin-based newborn PRS made up of 1128017 variants to predict cord blood insulin levels in South Asian newborns, independent of other maternal factors such as maternal age, pre-pregnancy BMI, parity, gestational weight gain and GDM status. If true, this information can be useful for early detection and monitoring of abnormal insulin levels in newborns with a high genetic risk score.

Chapter 4: Discussion, Conclusions, and Future Directions

4.0. Discussion

Maternal and newborn dysglycemia are important predictors of future disease. In mothers, dysglycemia may take the form of gestational diabetes mellitus (GDM) during pregnancy, and eventually result in newborn dysglycemia in their offspring. Having an understanding of modifiable and non-modifiable risk factors that are implicated in the risk of maternal and newborn dysglycemia is important for prevention and better management of future diseases. This is particularly important in a South Asian population who have some of the highest rates of dysglycemic events such as GDM —with future risk of disease for both mothers and their newborns being "programmed" *in utero.*

Based on the study, a maternal T2D PRS is a significant predictor of GDM in South Asian pregnant women, while an insulin-based newborn PRS is nominally associated with cord blood insulin levels in South Asian newborns, given other maternal factors such as maternal age, pre-pregnancy BMI, parity, gestational weight gain, maternal smoking history, GDM status, and insulin-based maternal PRS are held constant.

The lack of association between the insulin-based maternal PRS and cord blood insulin may suggest a lack of maternal genetic contribution to newborn insulin levels. However, it is also reasonable to assume that the lack of effect is because the insulin-based maternal PRS does not capture enough information from all of the genetic variants associated with glycemic traits such as cord blood insulin levels in South Asians and/or that the amount of samples from START are not adequate to measure a noticeable effect of the maternal PRS on cord blood insulin.

Finally, the lack of association between the insulin-based maternal and/or newborn PRS and cord blood glucose/insulin ratio (a marker of insulin sensitivity) might be attributable to the non-specificity of the genotype data compared to the outcome and/or low study power. The same insulin data from Lagou et al., 2018 (*Manuscript in preparation*) that informed the maternal and newborn PRS was also used to predict insulin sensitivity since no other studies from the MAGIC consortium had calculated insulin sensitivity similar to how it was done in the START cohort—i.e. estimated using the glucose/insulin ratio.

4.0.1 Clinical implications

Calculating the cumulative genetic risk based on an additive effect of SNP variants in the form of a PRS has been used to investigate the underlying biology behind several diseases now, including T2D, CVD, breast cancer, schizophrenia and other psychiatric disorders.(110) Accordingly, findings from this study may have some clinical implications. Firstly, the PRS can be used to study gene-environment interactions and develop better disease prevention approaches.(111, 112) An understanding of both genetic and environmental factors can inform more personalized and tailored approaches to alleviating risk factors that contribute to disease development. The association of maternal T2D PRS with GDM in a South Asians population can be useful in further evaluating the PRS's interaction with other environmental risk factors such as pre-pregnancy weight, and/or diet quality that have also been shown to be implicated in the risk for GDM. Likewise, the nominal association of an insulin-based newborn PRS with cord blood insulin can be

further contextualized if it interacts with other maternal factors such as gestational weight gain or GDM status to modify the risk of abnormal insulin regulation in newborns. In addition, the gene-environmental interaction can elucidate which environmental factor is a stronger contributor in regulating newborn insulin levels and subsequent dysglycemia associated with abnormal insulin regulation. This idea has been supported by a study conducted by Nakamura et al.(111) assessing gene-environmental interactions in obesity. Nakamura et al. found that their genetic risk score was associated with obesity and interacted with other environmental factors such as BMI. They propose that their gene score can be useful in selecting optimal lifestyle factors (determined by the highest interaction with their gene risk score) that can be considered intervention targets for personalized and targeted obesity prevention.(111). A similar approach could be used based on the results from this study to assess optimal environmental factors which can serve as intervention targets to prevent maternal and newborn dysglycemia characterized by GDM and abnormal cord blood insulin levels.

Secondly, PRSs can not only be used to stratify participants at risk but can also inform their management plan based on their genetic risk category. For example, a study conducted by Natarajan et al.(113) identified that patients categorized with high genetic risk of coronary heart disease fared significantly better on statin therapy compared to those categorized as low genetic risk. In our case, participants who belong to the high genetic risk for GDM category (tertiles 2 and 3) could be assessed to see if they are the ones more likely to benefit from a tailored diet intervention compared to the low genetic risk group.

Lastly, PRSs can be useful in determining genetic overlap between multiple diseases in clinic.(110) Studies have shown that a PRS derived from one disease (e.g. Schizophrenia-specific PRS) can be used to ascertain the risk or onset of another disease (e.g. bipolar disorder).(110) This has led to an emerging field of using multi-polygenic risk scores (MPSs), derived from multiple GWASs, to predict outcomes by enhancing the gene score's predictive power.(114) Perhaps combining genetic variants contributing to the maternal T2D PRS and insulin-based newborn PRS in a MPS could be useful in reducing some of the costs related to utilizing gene scores while increasing the predictive capacity of the PRSs and identifying both maternal and newborn dysglycemia in a clinical setting.

Overall, the positive results from this study may indicate some uses for the PRSs in a clinical setting. The maternal T2D PRS and the insulin-based newborn PRS highlight the genetic burden of GDM (maternal and newborn dysglycemia) and adverse cord insulin profiles (may indicate subsequent newborn dysglycemia) in South Asians (primary prevention), while stratifying high risk participants who may require constant screening and monitoring (secondary prevention) or treatment optimization (tertiary prevention).(110)

4.0.2 Limitations and Challenges

There are some limitations to this study that may affect the generalizability of the results. As discussed previously, one of the biggest challenges was that the three PRS developed in this study were based on T2D and plasma insulin levels, which are different than the outcomes measured in the analysis. Although there are reasons supporting the use of T2D genetic variants to predict GDM and plasma insulin genetic variants from adults to

predict cord blood insulin in newborns based on similarity and lack of outcome specific GWASs, there is a possibility that the true effect is over or underrepresented. Nonetheless, there were some other technical challenges that were apparent when conducting this study, such as i) identifying the most appropriate methods to create the PRS, and ii) recognizing false positives and false negatives.

Appropriateness of methods used to create PRS

In epidemiology studies, ethnicity can confound the relationship between the exposure and the outcome because of its multidimensionality—ethnicity includes cultural, geographical, and biological aspects.(115) One of the main technical challenges in conducting this study was a lack of multiethnic GWASs. There is a general paucity of genetic variants related to our outcomes of interest (GDM and cord blood insulin) which have been extensively studied and collected in a South Asian population. In fact, only 20% of the Mahajan et al.(55) data which informed the T2D PRS was representative of a South Asian population while the majority of it was white/European. The Lagou et al., 2018 (*Manuscript in Preparation)* data, which informed the insulin-based PRS, was entirely collected in a white/European population. This may have led to spurious associations between the PRSs and our outcomes of interest since the allelic frequencies of the variants and beta estimates of the outcomes may vary based on ethnicity.(115) Usually, a stratified analysis is useful to delineate the true association between the genotypes and phenotypes in these cases. However, this was not possible in our study since the phenotype data has been collected from a South Asian population. A lack of diversity in GWASs can introduce

significant bias when using these PRS to understand the biology of diseases such as GDM and T2D in a clinical setting.(116)

Furthermore, another technical point of consideration when building the PRS was deciding the GWAS significance threshold. In this study, a less conservative threshold ($P \leq 0.2$ for T2D PRS and $P \leq 0.05$ for insulin PRS) than the commonly accepted ($P \leq 5 \times 10^{-8}$) threshold was used. While the higher threshold is more conservative and ensures only the SNP variants that are truly associated with the outcome of interest are included, a less conservative estimate was used to avoid missing information from variants that may be marginally associated with the outcome and to improve the overall predictive capability of the PRS.(54) A less conservative estimate is useful in avoiding the winner's curse or over inflated estimation of the true effect.(110) On the other hand, a higher than $P \leq 0.2$ or $P < 0.05$, threshold respectively, may result in poor tagging, low coverage and a smaller heritability estimate despite the reduction in number of SNPs included with that approach.(54) An appropriate threshold is required to maintain a balance between having too many and too few SNPs informing the PRS.

False positives and false negatives

One of the biggest challenges in genetic association studies is determining whether the association seen between the exposure (the PRS) and the phenotypic traits is the true effect (i.e. a non-biased estimate). There is a possibility that the genetic variants informing the PRSs have small effect sizes that can result in false positive or false negative associations through random or systematic error.(115)

In PRSs analyses, population structure or the difference in geographic location between the genetic variants and tested traits may be a source of confounding and partially contribute to false positive results.(117) For this study, while the base data informing the T2D PRS is derived from a multiethnic DIAGRAM consortium (~20% South Asian), the data informing the insulin-based maternal and newborn PRS is derived from a completely white/European MAGIC consortium. The target data or the phenotypic information, in both cases, has been measured in a South Asian population. There is a high possibility that the allele frequencies and environmental risk factors for the outcomes of interest systematically differ between the base and target data due to genetic drift or ascertainment of genotyped variants.(117) As such, the observed associations between PRS and the outcomes of interest may have resulted from differences at null SNPs,(117) creating false positive results. Although deriving base and target samples from the same population or the one that is genetically similar can control for this confounding, this was not possible in this study due to a lack of genetic data on South Asians.

Furthermore, the lack of association between insulin-based maternal PRS and cord blood insulin may be a false negative—i.e. a type 2 error due to a lack of specificity of genetic variants and power. Prior studies have showed a potential role of maternal genotypes on newborn glycemic traits.(100, 118) It is reasonable that the null effect of insulin-based maternal PRS on cord blood insulin is partially due to the differences in the genetic backgrounds of the base and target populations.(119) However, the lack of replication in studies assessing the same association (i.e. the effect of maternal genotype on cord blood insulin) makes this finding uncertain. Additionally, and perhaps more

importantly, the lack of effect may be due to the study being underpowered (50% - Appendix 1). Genetic variants that are included in the maternal PRS have a modest effect size. Combing these variants and pooling them together in a PRS further reduces their effect size.(119) Thus, association studies aimed at assessing such genetic associations would require a very large sample size, in the range of 1000-10000.(119) Although we cannot calculate the required sample size to achieve 80% power in our study, our sample size (N=522) is well under that recommended range, and so it is reasonable to assume that the lack of effect may be a false negative. This notion has been supported by a study conducted by Dudbridge which showed that negative results in polygenic risk score association studies are more likely to be due to low sample sizes and that increasing sample sizes may result in a more successful analysis.(120)

4.1. Conclusions and Future Directions

In conclusion, a type 2 diabetes (T2D) polygenic risk score (PRS) is strongly associated with GDM in South Asian women, independent of other risk factors for GDM including family history of T2D, maternal age, pre-pregnancy weight and low diet quality; whereas an insulin-based newborn PRS is nominally associated with cord blood insulin levels in South Asian newborns, when adjusted for maternal factors such as age, pre-pregnancy BMI, gestational weight gain, parity, smoking and GDM status. This study shows that PRS may be useful in predicting maternal and newborn dysglycemia, and newborn insulin levels to some extent in a South Asian population despite the lack of South Asian specific genetic data. The positive association between maternal T2D PRS and GDM

adds further evidence to the notion that GDM and T2D share common trajectories, while making a case for better risk stratification strategies during pregnancy to reduce the burden of future disease in mothers and their newborns. The positive association between insulin-based newborn PRS and cord blood insulin when adjusted for maternal factors may suggest that newborn insulin levels are not primarily driven by their own genetics but also by the *in-utero* environment, making a case for controlling environmental factors during pregnancy to reduce disease burden in newborns.

Future research should consider gene-environment interactions to further delineate how the polygenic risk scores affect each other and glycemic traits. It is important to identify how modifiable risk factors for GDM such as pre-pregnancy weight and diet quality (separately) interact with maternal T2D PRS to modify the risk of GDM. This will allow us to determine how environmental modifiable factors that contribute to the risk of GDM in South Asians alter the genetic susceptibility to the disease. If the gene-environment interaction is strong, interventions aimed at reducing pre-pregnancy weight and improving diet quality are warranted to reduce the genetic burden of GDM among South Asians. In addition, such analysis could highlight which of the two environmental factors should be targeted more for optimal effect in South Asian pregnant women based on the strength of their interaction with the PRS.(111) The use of a T2D maternal PRS that is associated with GDM may allow for more specific targeting of modifiable factors (that interact with the PRS) in a cost effective manner compared to generalized interventions. Moreover, although we aimed to test the association of our PRS in a South Asian population, future analyses could consider using data from multiethnic cohorts to improve

the generalizability of our findings. Perhaps a next step would be using cohorts with phenotypic data from other high-risk groups for GDM and subsequent T2D such as Africans and Native Americans when developing the gene score. This would be especially important to increase the accuracy and predictive power of our PRS in understanding disease biology and risk stratification.

References

1. Plows JF, Stanley JL, Baker PN, Reynolds CM, Vickers MH. The Pathophysiology of Gestational Diabetes Mellitus. Int J Mol Sci. 2018;19(11).
2. Diagnostic criteria and classification of hyperglycaemia first detected in pregnancy: a World Health Organization Guideline. Diabetes Res Clin Pract. 2014;103(3):341-63.
3. Yajnik CS. Early life origins of insulin resistance and type 2 diabetes in India and other Asian countries. J Nutr. 2004;134(1):205-10.
4. Volgman AS, Palaniappan LS, Aggarwal NT, Gupta M, Khandelwal A, Krishnan AV, et al. Atherosclerotic Cardiovascular Disease in South Asians in the United States: Epidemiology, Risk Factors, and Treatments: A Scientific Statement From the American Heart Association. Circulation. 2018;138(1):e1-e34.
5. Krishnaveni GV, Yajnik CS. Developmental origins of diabetes-an Indian perspective. Eur J Clin Nutr. 2017;71(7):865-9.
6. Gupta Y, Kapoor D, Josyula LK, Praveen D, Naheed A, Desai AK, et al. A lifestyle intervention programme for the prevention of Type 2 diabetes mellitus among South Asian women with gestational diabetes mellitus [LIVING study]: protocol for a randomized trial. Diabet Med. 2019;36(2):243-51.
7. Diabetes Canada Clinical Practice Guidelines Expert C, Feig DS, Berger H, Donovan L, Godbout A, Kader T, et al. Diabetes and Pregnancy. Can J Diabetes. 2018;42 Suppl 1:S255-S82.
8. Gestational Diabetes Brussels, Belgium International Diabetes Federation; 2017 [8th:[Available from: https://www.idf.org/our-activities/care-prevention/gdm.
9. International Diabetes Federation. IDF Diabetes Atlas, 8th edn. Brussels, Belgium: International Diabetes Federation. . 2017.
10. Jenum AK, Morkrid K, Sletner L, Vangen S, Torper JL, Nakstad B, et al. Impact of ethnicity on gestational diabetes identified with the WHO and the modified International Association of Diabetes and Pregnancy Study Groups criteria: a population-based cohort study. Eur J Endocrinol. 2012;166(2):317-24.
11. Anand SS, Gupta M, Teo KK, Schulze KM, Desai D, Abdalla N, et al. Causes and consequences of gestational diabetes in South Asians living in Canada: results from a prospective cohort study. CMAJ Open. 2017;5(3):E604-E11.
12. Li KT, Naik S, Alexander M, Mathad JS. Screening and diagnosis of gestational diabetes in India: a systematic review and meta-analysis. Acta Diabetol. 2018;55(6):613-25.
13. Farrar D, Fairley L, Santorelli G, Tuffnell D, Sheldon TA, Wright J, et al. Association between hyperglycaemia and adverse perinatal outcomes in south Asian and white British women: analysis of data from the Born in Bradford cohort. Lancet Diabetes Endocrinol. 2015;3(10):795-804.
14. Gadgil MD, Oza-Frank R, Kandula NR, Kanaya AM. Type 2 diabetes after gestational diabetes mellitus in South Asian women in the United States. Diabetes Metab Res Rev. 2017;33(5).

15. Carolan M, Davey MA, Biro MA, Kealy M. Maternal age, ethnicity and gestational diabetes mellitus. Midwifery. 2012;28(6):778-83.
16. Group HSCR, Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, et al. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med. 2008;358(19):1991-2002.
17. Arabin B, Baschat AA. Pregnancy: An Underutilized Window of Opportunity to Improve Long-term Maternal and Infant Health-An Appeal for Continuous Family Care and Interdisciplinary Communication. Front Pediatr. 2017;5:69.
18. Sletner L, Jenum AK, Yajnik CS, Morkrid K, Nakstad B, Rognerud-Jensen OH, et al. Fetal growth trajectories in pregnancies of European and South Asian mothers with and without gestational diabetes, a population-based cohort study. PLoS One. 2017;12(3):e0172946.
19. Marciniak A, Patro-Malysza J, Kimber-Trojnar Z, Marciniak B, Oleszczuk J, Leszczynska-Gorzelak B. Fetal programming of the metabolic syndrome. Taiwan J Obstet Gynecol. 2017;56(2):133-8.
20. Yajnik CS, Fall CH, Coyaji KJ, Hirve SS, Rao S, Barker DJ, et al. Neonatal anthropometry: the thin-fat Indian baby. The Pune Maternal Nutrition Study. Int J Obes Relat Metab Disord. 2003;27(2):173-80.
21. Krishnaveni GVY, C.S. Foetal programming in a diabetic pregnancy: long-term implications for the offspring CURRENT SCIENCE 2017;113(7):1321-6.
22. Venkataraman H, Ram U, Craik S, Arungunasekaran A, Seshadri S, Saravanan P. Increased fetal adiposity prior to diagnosis of gestational diabetes in South Asians: more evidence for the 'thin-fat' baby. Diabetologia. 2017;60(3):399-405.
23. Yajnik CS. Transmission of obesity-adiposity and related disorders from the mother to the baby. Ann Nutr Metab. 2014;64 Suppl 1:8-17.
24. Bellamy L, Casas JP, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. Lancet. 2009;373(9677):1773-9.
25. Kramer CK, Campbell S, Retnakaran R. Gestational diabetes and the risk of cardiovascular disease in women: a systematic review and meta-analysis. Diabetologia. 2019.
26. Retnakaran R. Hyperglycemia in pregnancy and its implications for a woman's future risk of cardiovascular disease. Diabetes Res Clin Pract. 2018;145:193-9.
27. Song C, Lyu Y, Li C, Liu P, Li J, Ma RC, et al. Long-term risk of diabetes in women at varying durations after gestational diabetes: a systematic review and meta-analysis with more than 2 million women. Obes Rev. 2018;19(3):421-9.
28. Kubo A, Ferrara A, Windham GC, Greenspan LC, Deardorff J, Hiatt RA, et al. Maternal hyperglycemia during pregnancy predicts adiposity of the offspring. Diabetes Care. 2014;37(11):2996-3002.
29. Krishnaveni GV, Veena SR, Hill JC, Kehoe S, Karat SC, Fall CH. Intrauterine exposure to maternal diabetes is associated with higher adiposity and insulin resistance and clustering of cardiovascular risk markers in Indian children. Diabetes Care. 2010;33(2):402-4.

30. Arora GP, Thaman RG, Prasad RB, Almgren P, Brons C, Groop LC, et al. Prevalence and risk factors of gestational diabetes in Punjab, North India: results from a population screening program. Eur J Endocrinol. 2015;173(2):257-67.
31. Berkowitz GS, Lapinski RH, Wein R, Lee D. Race/ethnicity and other risk factors for gestational diabetes. Am J Epidemiol. 1992;135(9):965-73.
32. Hedderson MM, Darbinian JA, Quesenberry CP, Ferrara A. Pregravid cardiometabolic risk profile and risk for gestational diabetes mellitus. Am J Obstet Gynecol. 2011;205(1):55 e1-7.
33. Solomon CG, Willett WC, Carey VJ, Rich-Edwards J, Hunter DJ, Colditz GA, et al. A prospective study of pregravid determinants of gestational diabetes mellitus. JAMA. 1997;278(13):1078-83.
34. Krishnaveni GV, Hill JC, Veena SR, Bhat DS, Wills AK, Karat CL, et al. Low plasma vitamin B12 in pregnancy is associated with gestational 'diabesity' and later diabetes. Diabetologia. 2009;52(11):2350-8.
35. Zargar AH, Sheikh MI, Bashir MI, Masoodi SR, Laway BA, Wani AI, et al. Prevalence of gestational diabetes mellitus in Kashmiri women from the Indian subcontinent. Diabetes Res Clin Pract. 2004;66(2):139-45.
36. Sukumar N, Venkataraman H, Wilson S, Goljan I, Selvamoni S, Patel V, et al. Vitamin B12 Status among Pregnant Women in the UK and Its Association with Obesity and Gestational Diabetes. Nutrients. 2016;8(12).
37. Cho YM, Kim TH, Lim S, Choi SH, Shin HD, Lee HK, et al. Type 2 diabetes-associated genetic variants discovered in the recent genome-wide association studies are related to gestational diabetes mellitus in the Korean population. Diabetologia. 2009;52(2):253-61.
38. Kwak SH, Kim SH, Cho YM, Go MJ, Cho YS, Choi SH, et al. A genome-wide association study of gestational diabetes mellitus in Korean women. Diabetes. 2012;61(2):531-41.
39. Mao H, Li Q, Gao S. Meta-analysis of the relationship between common type 2 diabetes risk gene variants with gestational diabetes mellitus. PLoS One. 2012;7(9):e45882.
40. Wang Y, Nie M, Li W, Ping F, Hu Y, Ma L, et al. Association of six single nucleotide polymorphisms with gestational diabetes mellitus in a Chinese population. PLoS One. 2011;6(11):e26953.
41. Zhang C, Bao W, Rong Y, Yang H, Bowers K, Yeung E, et al. Genetic variants and the risk of gestational diabetes mellitus: a systematic review. Hum Reprod Update. 2013;19(4):376-90.
42. Hayes MG, Urbanek M, Hivert MF, Armstrong LL, Morrison J, Guo C, et al. Identification of HKDC1 and BACE2 as genes influencing glycemic traits during pregnancy through genome-wide association studies. Diabetes. 2013;62(9):3282-91.
43. Kanthimathi S, Chidambaram M, Bodhini D, Liju S, Bhavatharini A, Uma R, et al. Association of recently identified type 2 diabetes gene variants with Gestational Diabetes in Asian Indian population. Mol Genet Genomics. 2017;292(3):585-91.

44. Kanthimathi S, Chidambaram M, Liju S, Bhavadharini B, Bodhini D, Prakash VG, et al. Identification of Genetic Variants of Gestational Diabetes in South Indians. Diabetes Technol Ther. 2015;17(7):462-7.
45. Belsky DW, Moffitt TE, Sugden K, Williams B, Houts R, McCarthy J, et al. Development and evaluation of a genetic risk score for obesity. Biodemography Soc Biol. 2013;59(1):85-100.
46. Cooke Bailey JN, Igo RP, Jr. Genetic Risk Scores. Curr Protoc Hum Genet. 2016;91:1 29 1-1 9.
47. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am J Hum Genet. 2017;100(4):635-49.
48. Sim X, Ong RT, Suo C, Tay WT, Liu J, Ng DP, et al. Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. PLoS Genet. 2011;7(4):e1001363.
49. Farrar D, Simmonds M, Bryant M, Sheldon TA, Tuffnell D, Golder S, et al. Hyperglycaemia and risk of adverse perinatal outcomes: systematic review and meta-analysis. BMJ. 2016;354:i4694.
50. Archambault C, Arel R, Filion KB. Gestational diabetes and risk of cardiovascular disease: a scoping review. Open Med. 2014;8(1):e1-9.
51. Anand SS, Gupta MK, Schulze KM, Desai D, Abdalla N, Wahi G, et al. What accounts for ethnic differences in newborn skinfold thickness comparing South Asians and White Caucasians? Findings from the START and FAMILY Birth Cohorts. Int J Obes (Lond). 2016;40(2):239-44.
52. Cosson E, Cussac-Pillegand C, Benbara A, Pharisien I, Jaber Y, Banu I, et al. The diagnostic and prognostic performance of a selective screening strategy for gestational diabetes mellitus according to ethnicity in Europe. J Clin Endocrinol Metab. 2014;99(3):996-1005.
53. Dornhorst A, Paterson CM, Nicholls JS, Wadsworth J, Chiu DC, Elkeles RS, et al. High prevalence of gestational diabetes in women from ethnic minority groups. Diabet Med. 1992;9(9):820-5.
54. Lamri A, Mao S, Gupta M, Paré G, Anand SS. Genome-Wide Polygenic Risk Scores and prediction of Gestational Diabetes in South Asian Women. bioRxiv. 2018.
55. Mahajan A, Go MJ, Zhang W, Below JE, Gaulton KJ, Ferreira T, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet. 2014;46(3):234-44.
56. Anand SS, Vasudevan A, Gupta M, Morrison K, Kurpad A, Teo KK, et al. Rationale and design of South Asian Birth Cohort (START): a Canada-India collaborative study. BMC Public Health. 2013;13:79.
57. Kelemen LE, Anand SS, Vuksan V, Yi Q, Teo KK, Devanesen S, et al. Development and evaluation of cultural food frequency questionnaires for South Asians, Chinese, and Europeans in North America. J Am Diet Assoc. 2003;103(9):1178-84.
58. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc. 2010;5(9):1564-73.

59. Delaneau O, Marchini J, Genomes Project C, Genomes Project C. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat Commun. 2014;5:3934.
60. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5(6):e1000529.
61. Consortium TGP, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.
62. Corp. I. IBM SPSS Statistics for Macintosh, Version 24.0. 2016.
63. Chow CK, Islam S, Bautista L, Rumboldt Z, Yusufali A, Xie C, et al. Parental history and myocardial infarction risk across the world: the INTERHEART Study. J Am Coll Cardiol. 2011;57(5):619-27.
64. Sharma V, Sharma I, Sethi I, Mahajan A, Singh G, Angural A, et al. Replication of newly identified type 2 diabetes susceptible loci in Northwest Indian population. Diabetes Res Clin Pract. 2017;126:160-3.
65. Qi Q, Wang X, Strizich G, Wang T. Genetic Determinants of Type 2 Diabetes in Asians. Int J Diabetol Vasc Dis Res. 2015;2015(Suppl 1).
66. Chambers JC, Abbott J, Zhang W, Turro E, Scott WR, Tan ST, et al. The South Asian genome. PLoS One. 2014;9(8):e102645.
67. Moorjani P, Thangaraj K, Patterson N, Lipson M, Loh PR, Govindaraj P, et al. Genetic evidence for recent population mixture in India. Am J Hum Genet. 2013;93(3):422-38.
68. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. Nature. 2009;461(7263):489-94.
69. Li J, Song C, Li C, Liu P, Sun Z, Yang X. Increased risk of cardiovascular disease in women with prior gestational diabetes: A systematic review and meta-analysis. Diabetes Res Clin Pract. 2018;140:324-38.
70. Farahvar S, Walfisch A, Sheiner E. Gestational diabetes risk factors and long-term consequences for both mother and offspring: a literature review. Expert Rev Endocrinol Metab. 2018:1-12.
71. Philipps LH, Santhakumaran S, Gale C, Prior E, Logan KM, Hyde MJ, et al. The diabetic pregnancy and offspring BMI in childhood: a systematic review and meta-analysis. Diabetologia. 2011;54(8):1957-66.
72. Lind JM, Hennessy A, McLean M. Cardiovascular disease in women: the significance of hypertension and gestational diabetes during pregnancy. Curr Opin Cardiol. 2014;29(5):447-53.
73. Anand SS, Samaan Z, Middleton C, Irvine J, Desai D, Schulze KM, et al. A Digital Health Intervention to Lower Cardiovascular Risk: A Randomized Clinical Trial. JAMA Cardiol. 2016;1(5):601-6.
74. Stewart KFJ, Wesselius A, Schreurs MAC, Schols A, Zeegers MP. Behavioural changes, sharing behaviour and psychological responses after receiving direct-to-consumer genetic test results: a systematic review and meta-analysis. J Community Genet. 2018;9(1):1-18.

75. Hollands GJ, French DP, Griffin SJ, Prevost AT, Sutton S, King S, et al. The impact of communicating genetic risks of disease on risk-reducing health behaviour: systematic review with meta-analysis. BMJ. 2016;352:i1102.
76. Ruiz-Palacios M, Ruiz-Alcaraz AJ, Sanchez-Campillo M, Larque E. Role of Insulin in Placental Transport of Nutrients in Gestational Diabetes Mellitus. Ann Nutr Metab. 2017;70(1):16-25.
77. Brunner S, Schmid D, Huttinger K, Much D, Heimberg E, Sedlmeier EM, et al. Maternal insulin resistance, triglycerides and cord blood insulin in relation to post-natal weight trajectories and body composition in the offspring up to 2 years. Diabet Med. 2013;30(12):1500-7.
78. Westgate JA, Lindsay RS, Beattie J, Pattison NS, Gamble G, Mildenhall LF, et al. Hyperinsulinemia in cord blood in mothers with type 2 diabetes and gestational diabetes mellitus in New Zealand. Diabetes Care. 2006;29(6):1345-50.
79. Xie X, Gao H, Wu S, Zhao Y, Du C, Yuan G, et al. Increased Cord Blood Betatrophin Levels in the Offspring of Mothers with Gestational Diabetes. PLoS One. 2016;11(5):e0155646.
80. Guzman-Barcenas J, Hernandez JA, Arias-Martinez J, Baptista-Gonzalez H, Ceballos-Reyes G, Irles C. Estimation of umbilical cord blood leptin and insulin based on anthropometric data by means of artificial neural network approach: identifying key maternal and neonatal factors. BMC Pregnancy Childbirth. 2016;16(1):179.
81. Lawlor DA, West J, Fairley L, Nelson SM, Bhopal RS, Tuffnell D, et al. Pregnancy glycaemia and cord-blood levels of insulin and leptin in Pakistani and white British mother-offspring pairs: findings from a prospective pregnancy cohort. Diabetologia. 2014;57(12):2492-500.
82. Ahmad A, Rukmini MS, Yadav C, Agarwal A, Manjrekar PA, Hegde A. Indices of Glucose Homeostasis in Cord Blood in Term and Preterm Newborns. J Clin Res Pediatr Endocrinol. 2016;8(3):270-5.
83. Huerta-Chagoya A, Vazquez-Cardenas P, Moreno-Macias H, Tapia-Maruri L, Rodriguez-Guillen R, Lopez-Vite E, et al. Genetic determinants for gestational diabetes mellitus and related metabolic traits in Mexican women. PLoS One. 2015;10(5):e0126408.
84. Nielsen T, Sparso T, Grarup N, Jorgensen T, Pisinger C, Witte DR, et al. Type 2 diabetes risk allele near CENTD2 is associated with decreased glucose-stimulated insulin release. Diabetologia. 2011;54(5):1052-6.
85. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47(D1):D1005-D12.
86. Powe CE, Nodzenski M, Talbot O, Allard C, Briggs C, Leya MV, et al. Genetic Determinants of Glycemic Traits and the Risk of Gestational Diabetes Mellitus. Diabetes. 2018;67(12):2703-9.
87. Finer S, Mathews C, Lowe R, Smart M, Hillman S, Foo L, et al. Maternal gestational diabetes is associated with genome-wide DNA methylation variation in placenta and cord blood of exposed offspring. Hum Mol Genet. 2015;24(11):3021-9.

88. Vaag A, Brons C, Gillberg L, Hansen NS, Hjort L, Arora GP, et al. Genetic, nongenetic and epigenetic risk determinants in developmental programming of type 2 diabetes. Acta Obstet Gynecol Scand. 2014;93(11):1099-108.
89. Dunger DB, Petry CJ, Ong KK. Genetics of size at birth. Diabetes Care. 2007;30 Suppl 2:S150-5.
90. Anand SS, Yusuf S, Vuksan V, Devanesen S, Teo KK, Montague PA, et al. Differences in risk factors, atherosclerosis and cardiovascular disease between ethnic groups in Canada: the study of health assessment and risk in ethnic groups (SHARE). Indian Heart J. 2000;52(7 Suppl):S35-43.
91. International Schizophrenia C, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460(7256):748-52.
92. Team RC. R: A language and environment for statistical compting. Vienna, Austria: R Foundation for Statistical Computing; 2017.
93. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.
94. Purcell SC, C. PLINK. 1.9 ed.
95. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010;42(2):105-16.
96. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 2005;21(16):3439-40.
97. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184-91.
98. Warrington NM, Freathy RM, Neale MC, Evans DM. Using structural equation modelling to jointly estimate maternal and fetal effects on birthweight in the UK Biobank. Int J Epidemiol. 2018;47(4):1229-41.
99. Stordal K, Marild K, Tapia G, Haugen M, Cohen AS, Lie BA, et al. Fetal and Maternal Genetic Variants Influencing Neonatal Vitamin D Status. J Clin Endocrinol Metab. 2017;102(11):4072-9.
100. Tyrrell J, Richmond RC, Palmer TM, Feenstra B, Rangarajan J, Metrustry S, et al. Genetic Evidence for Causal Relationships Between Maternal Obesity-Related Traits and Birth Weight. JAMA. 2016;315(11):1129-40.
101. Li A, Teo KK, Morrison KM, McDonald SD, Atkinson SA, Anand SS, et al. A genetic link between prepregnancy body mass index, postpartum weight retention, and offspring weight in early childhood. Obesity (Silver Spring). 2017;25(1):236-43.
102. Li M, Erickson SW, Hobbs CA, Li J, Tang X, Nick TG, et al. Detecting maternal-fetal genotype interactions associated with conotruncal heart defects: a haplotype-based analysis with penalized logistic regression. Genet Epidemiol. 2014;38(3):198-208.

103. Freathy RM, Weedon MN, Bennett A, Hypponen E, Relton CL, Knight B, et al. Type 2 diabetes TCF7L2 risk genotypes alter birth weight: a study of 24,053 individuals. Am J Hum Genet. 2007;80(6):1150-61.
104. Ahmad A, Mysore Srikantiah R, Yadav C, Agarwal A, Ajay Manjrekar P, Hegde A. Cord Blood Insulin Levels: It's Correlation with Gender, Birth Weight and Placental Weight in Term Newborns. Indian J Clin Biochem. 2016;31(4):458-62.
105. Sahasrabuddhe A, Pitale S, Raje D, Sagdeo MM. Cord blood levels of insulin and glucose in full-term pregnancies. J Assoc Physicians India. 2013;61(6):378-82.
106. Beaumont RN, Warrington NM, Cavadino A, Tyrrell J, Nodzenski M, Horikoshi M, et al. 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. Hum Mol Genet. 2018;27(4):742-56.
107. Ong KK, Dunger DB. Birth weight, infant growth and insulin resistance. Eur J Endocrinol. 2004;151 Suppl 3:U131-9.
108. Hales CN, Barker DJ. Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. 1992. Int J Epidemiol. 2013;42(5):1215-22.
109. Yajnik C. Interactions of perturbations in intrauterine growth and growth during childhood on the risk of adult-onset disease. Proc Nutr Soc. 2000;59(2):257-65.
110. Chasioti D, Yan J, Nho K, Saykin AJ. Progress in Polygenic Composite Scores in Alzheimer's and Other Complex Diseases. Trends Genet. 2019.
111. Nakamura S, Narimatsu H, Sato H, Sho R, Otani K, Kawasaki R, et al. Gene-environment interactions in obesity: implication for future applications in preventive medicine. J Hum Genet. 2016;61(4):317-22.
112. Schoeler T, Choi SW, Dudbridge F, Baldwin J, Duncan L, Cecil CM, et al. Multi-Polygenic Score Approach to Identifying Individual Vulnerabilities Associated With the Risk of Exposure to Bullying. JAMA Psychiatry. 2019.
113. Natarajan P, Young R, Stitziel NO, Padmanabhan S, Baber U, Mehran R, et al. Polygenic Risk Score Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From Statin Therapy in the Primary Prevention Setting. Circulation. 2017;135(22):2091-101.
114. Krapohl E, Patel H, Newhouse S, Curtis CJ, von Stumm S, Dale PS, et al. Multi-polygenic score approach to trait prediction. Mol Psychiatry. 2018;23(5):1368-74.
115. Medina-Gomez C, Felix JF, Estrada K, Peters MJ, Herrera L, Kruithof CJ, et al. Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. Eur J Epidemiol. 2015;30(4):317-30.
116. De La Vega FM, Bustamante CD. Polygenic risk scores: a biased prediction? Genome Med. 2018;10(1):100.
117. Shing Wan Choi, Mak TSH, O'Reilly PF. A guide to performing Polygenic Risk Score analyses. bioRxiv. 2018.
118. Lin R, Ju H, Yuan Z, Zhang C, Zeng L, Sun Y, et al. Effects of maternal and fetal LEP common variants on maternal glycemic traits in pregnancy. Sci Rep. 2017;7(1):17710.
119. Kathiresan S, Newton-Cheh C, Gerszten RE. On the interpretation of genetic association studies. Eur Heart J. 2004;25(16):1378-81.

120. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013;9(3):e1003348.
121. Haljas K, Amare AT, Alizadeh BZ, Hsu YH, Mosley T, Newman A, et al. Bivariate Genome-Wide Association Study of Depressive Symptoms With Type 2 Diabetes and Quantitative Glycemic Traits. Psychosom Med. 2018;80(3):242-51.
1

TABLES AND FIGURES

Table 5: Characteristics of START study participants included in the analysis

	No GDM (n= 531)	GDM (n=301)	P value *
Maternal Age, yr (SD)	30 (4)	31 (4)	<0.001
Maternal Height, cm (SD)	162.7 (6.2)	161.5 (6.2)	0.007
Pre-pregnancy weight, kg (SD)	61.4 (11.7)	64.9 (12.2)	<0.001
Pre-pregnancy BMI, kg/m^2 (SD)	23.2 (4.2)	24.9 (4.6)	<0.001
Low diet quality, n (%)	123 (23.5%)	96 (32.2%)	0.006
Family history of diabetes, n (%)	180 (34.0%)	154 (51.2%)	<0.001
Maternal Polygenic Risk Score (SD)	-0.2 (1.0)	0.3 (1.0)	<0.001
Maternal Polygenic Risk Score, n (%)			
Tertile 1	205 (38.6%)	72 (23.9%)	<0.001
Tertile 2	180 (33.9%)	98 (32.6%)	
Tertile 3	146 (27.5%)	131 (43.5%)	
Tertile 2+3 (vs Tertile 1)	326 (61.4%)	229 (76.1%)	<0.001

Presented data are means (Standard Deviation) unless otherwise indicated. There is missing data for some variables. * P-Values are calculated from Chi-squared test for categorical variables and independent t-test for continuous variables

Table 6: Association results between GDM risk factors and GDM in mothers from the START cohort

Risk Factor	OR (95%CI) *	P-value*
PRS (Tertile 2+3 *vs* 1)[b]	1.81 (1.30 - 2.53)	<0.001
Maternal Age	1.08 (1.04 - 1.12)	<0.001
Pre-pregnancy weight, kg	1.03 (1.01 - 1.04)	<0.001
Maternal Height, cm	0.96 (0.94 - 0.99)	0.003
Diet quality (low *vs* medium + high)	1.44 (1.03 - 2.01)	0.032
Family history of T2D	1.62 (1.20 - 2.20)	0.002

Abbreviations: CI, confidence interval; OR, odds ratio

*from multivariable logistic regression model.

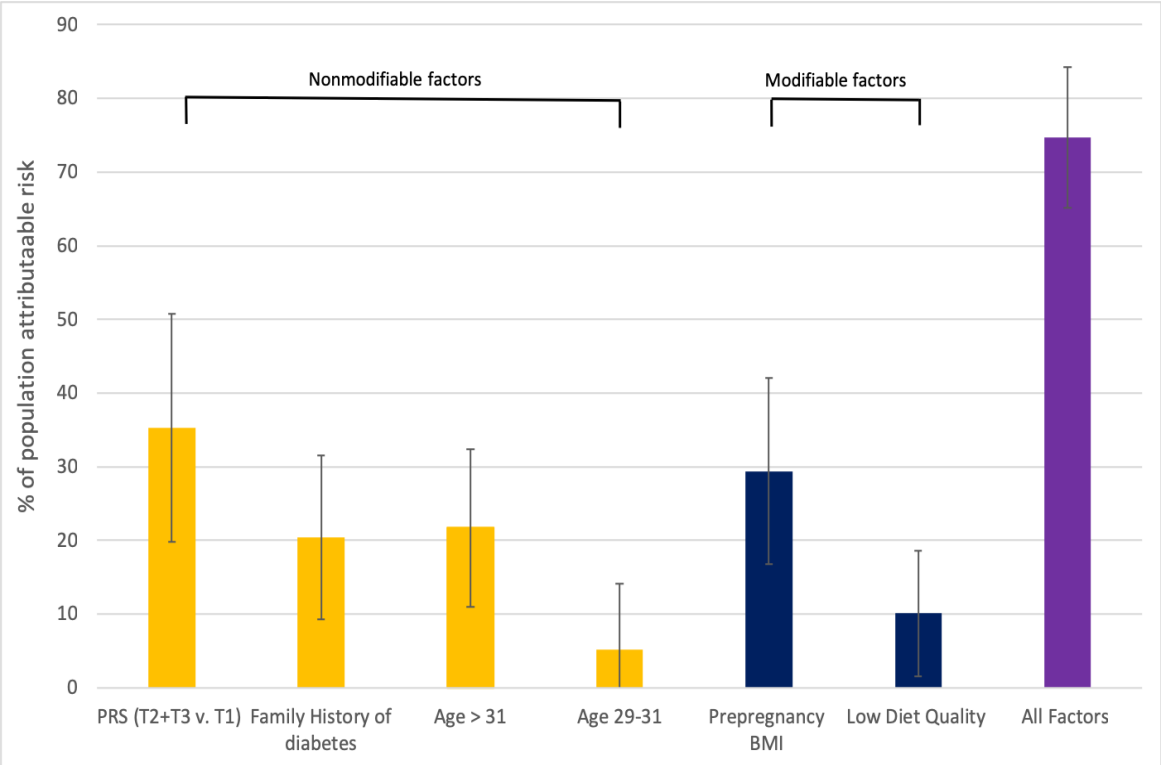
b: Maternal categorical PRS is further organized as a dichotomous variable with Tertile 2 and 3 being compared to Tertile 1 to improve interpretability and since these categories were the most significantly different when compared to each other.

Table 7: Population attributable risk of GDM risk factors in mothers from the START cohort

Risk Factor	OR (95% CI) from multivariable logistic regression model	Prevalence %	Population attributable risk % (95% CI)
PRS (Tertile 2+3 *vs.* 1)	1.86 (1.34 - 2.59)	66.6	35.3 (19.8 - 50.7)
Age 32-43 yr *vs.* < 29 yr	1.89 (1.32 - 2.70)	36.6	21.7 (11.0 - 32.4)
Age 29-31 yr *vs.* <29 yr	1.25 (0.84 - 1.84)	27.7	5.2 (-3.7 - 14.2)
Body mass index >23 *vs.* $\leq$ 23	1.85 (1.36 - 2.51)	51.8	29.4 (16.8 - 42.0)
Low diet quality	1.46 (1.04 - 2.03)	26.7	10.1 (1.5 - 18.6)
Family history of T2D	1.66 (1.22 - 2.25)	40.4	20.4 (9.2 - 31.5)
Total PAR			74.7 (65.1 - 84.3)

Abbreviations: CI, confidence interval; PRS: Polygenic risk score, OR, odds ratio

Figure 2: Partial population attributable risk for individual risk factors associated with GDM among South Asian women.



Note: Error bars represent 95% confidence intervals

Table 8: Association results between insulin-based PRS and cord blood insulin in participants from the START cohort

	Continuous β [95% CI, p-value]	**Categorical β*[b] [95%CI]**		**P-value***
Maternal PRS **(n=523)**	-0.007 [-0.36 to 0.023; p=0.66]	T2 vs. 1	-0.003 [-0.065 to 0.060]	0.93
		T3 vs. 1	-0.014 [-0.078 to 0.049]	0.66
		T2+T3 vs. 1	-0.016 [-0.078 to 0.045]	0.60
Newborn PRS **(n=514)**	0.021 [-0.008 to 0.050; p=0.12]	T2 vs. 1	0.026 [-0.037 to 0.090]	0.41
		T3 vs. 1	0.018 [-0.045 to 0.080]	0.58
		T2+T3 vs. 1	0.043 [-0.019 to 0.11)	0.18

Abbreviations: CI, confidence interval; β, effect estimate

*from univariate logistic regression model.

b: Maternal and Newborn categorical PRS are further organized as a dichotomous variable with Tertile 2 and 3 being compared to Tertile 1 to improve interpretability and since these categories were the most significantly different when compared to each other.

Table 9: Association results between insulin-based newborn PRS and cord blood insulin, adjusted for maternal PRS in participants from the START cohort

Risk Factor	**β [95%CI] ***	**P-value***
Newborn PRS[b]	0.021 [-0.011 – 0.052]	0.19
Newborn age, days	0.04 [-0.007 – 0.14]	0.48
Gestational age, wks	-0.066 [(-0.090) – (-0.42)]	<0.001
Newborn weight, g	<0.001 [<0.001 – <0.001]	<0.001
Newborn length, cm	-0.003 [-0.017 – 0.011]	0.66
Sex, M/F	0.13 [0.072 – 0.19]	<0.001
Maternal PRS[b]	-0.008 [-0.040 – 0.023]	0.61

Abbreviations: CI, confidence interval; β, effect estimate

*from multivariable linear regression model.

b: Maternal and newborn continuous PRS were used in this model to maintain adequate power

Table 10: Association results between insulin-based maternal PRS and cord blood insulin, adjusted for newborn PRS in participants from the START cohort

Risk Factor	**β [95%CI] ***	**P-value***
Maternal PRS[b]	-0.17 [-0.051 – 0.17]	0.33
Maternal age, years	0.001 [-0.008 – 0.010]	0.82
Pre-pregnancy BMI, kg/m^2	0.010 [0.002 – 0.017]	0.016
Parity	0.022 [-0.021 – 0.066]	0.32
Gestational weight gain, kg	0.004 [-0.001 – 0.008]	0.018
Maternal smoking history, Yes/No	-0.31 [-0.64 – 0.021]	0.067
GDM during pregnancy	0.087 [0.021 – 0.15]	0.010
Newborn PRS[b]	0.035 [0.002 – 0.069]	0.039

Abbreviations: CI, confidence interval; β, effect estimate

*from multivariable linear regression model.

b: Maternal and Newborn continuous PRS were used in this model to maintain adequate power

Table 11: Association results between insulin-based PRSs and cord blood insulin, adjusted for maternal and newborn factors in participants from the START cohort

Risk Factor	**β [95%CI] ***	**P-value***
Maternal PRS[b]	-0.008 [-0.040 – 0.024]	0.62
Newborn PRS[b]	0.022 [-0.010 – 0.055]	0.17
Maternal age, years	0.001 [-0.007 – 0.010]	0.80
Pre-pregnancy BMI, kg/m^2	0.005 [0.002 – 0.013]	0.18
Parity	0.025 [-0.017 – 0.066]	0.24
Gestational weight gain, kg	0.004 [<0.001 – 0.007]	0.038
Maternal smoking history, Yes/No	-0.18 [-0.50 – 0.14]	0.26
GDM during pregnancy	0.078 [0.014 – 0.14]	0.016
Newborn age, days	0.003 [-0.008 – 0.015]	0.56
Gestational age, wks	-0.057 [(-0.082) – (-0.032)]	<0.001
Newborn weight, g	<0.001 [<0.001 – <0.001]	<0.001
Newborn length, cm	-0.002 [-0.016 – 0.012]	0.79
Sex, M/F	0.13 [0.069 – 0.19]	<0.001

Abbreviations: CI, confidence interval; β, effect estimate

*from multivariable linear regression model.

b: Maternal and Newborn continuous PRS were used in this model to maintain adequate power

Table 12: Influential maternal characteristics by cord blood insulin quartiles

	N	Q1	Q2	Q3	Q4	P-value for trend
N	638	160	166	152	160	
Maternal Phenotypes	.					
Newborn PRS	638	-0.086 (1.05)	-0.025 (1.00)	0.022 (0.96)	0.042 (1.02)	0.29
Pre-pregnancy BMI, kg/m^2	638	23.70 (4.4)	22.70 (3.50)	24.50 (4.90)	24.30 (4.60)	0.022
Gestational weight gain, kg	638	12.85 (5.72)	14.82 (8.58)	14.14 (9.93)	15.26 (6.73)	0.008
GDM during pregnancy (yes)	231	40 (25%)	59 (35.5%)	59 (39.1%)	73 (46.5%)	<0.001

[a] Continuous newborn PRS was used in this model to maintain adequate power

Table 13: Association results between insulin-based maternal PRS and cord blood insulin, adjusted for newborn PRS in non-GDM mothers only from the START cohort (n=287)

Risk Factor	**β [95%CI] ***	**P-value***
Maternal PRS[b]	-0.011 [-0.052 – 0.31]	0.61
Maternal age, years	0.006 [-0.006 – 0.017]	0.33
Pre-pregnancy BMI, kg/m^2	0.008 [-0.002 – 0.018]	0.11
Parity	0.007 [-0.048 – 0.061]	0.80
Gestational weight gain, kg	0.003 [-0.001 – 0.007]	0.13
Maternal smoking history, Yes/No	-0.10 [-0.49 – 0.28]	0.61
Newborn PRS[b]	0.042 [<0.001 – 0.084]	0.051

Abbreviations: CI, confidence interval; β, effect estimate

*from multivariable linear regression model.

b: Maternal and newborn continuous PRS were used in this model to maintain adequate power

Table 14: Association results between insulin-based maternal PRS and cord blood insulin, adjusted for newborn PRS in GDM mothers only from the START cohort (n=161)

Risk Factor	**β [95%CI] ***	**P-value***
Maternal PRS[b]	-0.024 [-0.082 – 0.033]	0.41
Maternal age, years	-0.004 [-0.019 – 0.011]	0.61
Pre-pregnancy BMI, kg/m^2	0.012 [-0.001 – 0.025]	0.062
Parity	0.054 [-0.020 – 0.13]	0.15
Gestational weight gain, kg	0.010 [0.001 – 0.020]	0.028
Maternal smoking history, Yes/No	-0.84 [-1.51 – (-0.17]	0.014
Newborn PRS[b]	0.036 [-0.020 – 0.092]	0.21

Abbreviations: CI, confidence interval; β, effect estimate

*from multivariable linear regression model.

b: Maternal and newborn continuous PRS were used in this model to maintain adequate power

Table 15: Association results between insulin-based PRS and cord blood glucose/insulin ratio in participants from the START cohort

	Continuous β [95% CI, p-value]	**Categorical β*[b] [95%CI]**		**P-value***
Maternal PRS **(n=522)**	0.009 [-0.022 to 0.039; p=0.57]	T2 vs. 1	-0.004 [-0.068 to 0.060]	0.90
		T3 vs. 1	0.019 [-0.045 to 0.084]	0.56
		T2+T3 vs. 1	0.014 [-0.049 to 0.077]	0.66
Newborn PRS **(n=513)**	-0.018 [-0.048 to 0.013; p=0.25]	T2 vs. 1	-0.004 [-0.061 to 0.069]	0.90
		T3 vs. 1	-0.032 [-0.097 to 0.032]	0.32
		T2+T3 vs. 1	-0.028 [-0.092 to 0.036)	0.39

Abbreviations: CI, confidence interval; β, effect estimate

*from univariate linear regression model.

b: Maternal and Newborn categorical PRS are further organized as a dichotomous variable with Tertile 2 and 3 being compared to Tertile 1 to improve interpretability and since these categories were the most significantly different when compared to each other.

Table 16: Association results between insulin-based newborn PRS and cord blood glucose/insulin ratio, adjusted for maternal PRS in participants from the START cohort

Risk Factor	**β [95%CI] ***	**P-value***
Newborn PRS[b]	-0.018 [-0.050 – 0.014]	0.27
Newborn age, days	0.002 [-0.009) – (0.013)]	0.72
Gestational age, wks	0.077 [0.053 – 0.10]	<0.001
Newborn weight, g	<0.001 [<0.001 – <0.001]	<0.001
Newborn length, cm	0.002 [-0.012 – 0.016]	0.76
Sex, M/F	-0.15 [-0.21 – (-0.087)]	<0.001
Maternal PRS[b]	0.011 [-0.022 – 0.043]	0.52

Abbreviations: CI, confidence interval; β, effect estimate

*from multivariable linear regression model.

b: Maternal and newborn continuous PRS were used in this model to maintain adequate power

Table 17: Association results between maternal based newborn PRS and cord blood glucose/insulin ratio, adjusted for newborn PRS in participants from the START cohort

Risk Factor	**β [95%CI] ***	**P-value***
Maternal PRS[b]	0.020 [-0.015 – 0.054]	0.26
Maternal age, years	-0.002 [-0.011 – 0.007]	0.63
Pre-pregnancy BMI, kg/m^2	-0.010 [-0.0018 – (-0.002)]	0.012
Parity	-0.037 [-0.082 – 0.008]	0.11
Gestational weight gain, kg	-0.004 [-0.007 – <0.001]	0.049
Maternal smoking history, Yes/No	0.26 [-0.067 – 0.61]	0.13
GDM during pregnancy	-0.082 [-0.15 – (-0.014)]	0.018
Newborn PRS[b]	-0.030 [-0.064 – 0.004]	0.085

Abbreviations: CI, confidence interval; β, effect estimate

*from multivariable linear regression model.

b: Maternal and newborn continuous PRS were used in this model to maintain adequate power

Appendix 1:

1. **Power analysis for a polygenic risk score:**

We used the R code computing the formulae provided by Dudbridge(120) to calculate the power of the insulin-based polygenic risk scores (PRSs). There were 1128017 independent markers in the PRSs and N=638 START participants with complete cord blood insulin measurements. We used a heritability estimate of 0.07 [0.05, 0.10] as the proportion of variance explained by genetic effects of fasting insulin(121) in our GWAS sample for the analysis and left the other parameters set to default. We found that the power of the chi-squared test of association between the PRSs and cord blood insulin is only 50%. Thus, there is less than adequate statistical power to test the association (between an insulin-based PRS and cord blood insulin) being tested in this study.

The paper by Dudbridge, however, does not support a calculation for the sample size required to achieve sufficient power (80%) when testing the association between a PRS and a trait. Thus, we are unable to calculate that for our study.