**Genetic diversity and the risk for dysglycemia**

**Genetic diversity and the risk for dysglycemia: a study of South Asian and white Caucasian populations.**

**By**

**Zahra Naurish Sohani, B.H.Sc., M.Sc.**

A thesis submitted to the School of Graduate Studies in partial fulfilment of the requirements for the degree Doctor of Philosophy

Department of Clinical Epidemiology and Biostatistics

Health Research Methodology Program

McMaster University © Copyright by Zahra N. Sohani, June 2015

DOCTOR OF PHILOSOPHY (2015) McMaster University

Hamilton, Ontario, Canada

(Department of Clinical Epidemiology and Biostatistics, Health Research Methodology Program)

TITLE: Genetic diversity and the risk for dysglycemia: a study of South Asian and white Caucasian populations.

AUTHOR: Zahra Naurish Sohani, B.H.Sc., M.Sc.

SUPERVISOR: Dr. Sonia Anand

NUMBER OF PAGES: xvi, 153

**Abstract**

**Background:** Type 2 diabetes affects approximately 8% of the world’s population. Individuals of South Asian ancestry tend to develop metabolic abnormalities, leading to diabetes, at lower measures of absolute obesity and approximately 10 years earlier than white Caucasians. Current literature is unclear on the source of this ethnic heterogeneity; the variation in risk cannot be explained by lifestyle factors alone. The overarching aim of this thesis is to explore the role of genetic variants and epigenetic differences to explain the greater risk for type 2 diabetes among South Asians.

**Methods:** We first conducted a systematic review of the literature to ascertain the genetic risk from known single nucleotide polymorphisms (SNPs) among South Asians. We then compared these risk estimates to those from white Caucasians in a cohort of 69,033 individuals. Second, using the EpiDREAM prospective cohort study of individuals at high-risk for diabetes, we assessed the impact of genetic burden for impaired pancreatic beta-cell function alone and together with abdominal obesity on glucose traits. Ethnic heterogeneity in this interaction was also studied. Lastly, using data from two Canadian birth cohorts of South Asian and white Caucasian ancestry, we investigated ethnic differences in the epigenetic architecture for genes known to be implicated birth weight and length, as both are associated with the future risk of adult diabetes.

**Results:** The systematic review identified 15 SNPs robustly associated with type 2 diabetes in both South Asians and white Caucasians. The magnitude of risk and allele frequency of these genetic variants did not differ between the ethnic groups. Additionally, we identified 8 novel polymorphisms implicated in diabetes only among South Asians. Second, using data from the EpiDREAM study, we identified an interaction between cumulative genetic burden of beta-cell impairment, measured using an un-weighted genotype score, and abdominal obesity on glucose traits in South Asians, but not white Caucasians. Third, our investigation of differential DNA methylation between the ethnic groups revealed seven CpG sites for which changes in methylation corresponded to alterations in birth weight among white Caucasians, but not South Asians. An independent agnostic genome-wide search identified methylation levels at three CpG sites that appear to uniquely modulate birth weight in South Asians.

**Conclusions:** Overall, our results indicate that the greater risk for metabolic traits in South Asians likely does not result from common genetic variants shared by both South Asians and white Caucasians. Rather, differences in risk may be additionally influenced by unique risk variants in South Asians. Furthermore, it appears that the risk from a genetic impairment in South Asians may be magnified by abdominal obesity.

**Acknowledgements**

I would first and foremost like to sincerely thank my supervisor, Dr. Sonia Anand, for her mentorship, guidance, and support over the years. She has been an exceptional role model and has given me the confidence to explore my hypotheses and the tools to explore them well. I also thank her for extensively reviewing this thesis.

To my committee members, Dr. Hertzel Gerstein, Dr. David Meyre, and Dr. Guillaume Paré, I extend my gratitude for their direction, patience, and feedback. Additionally, I thank Dr. Gerstein for his guidance and mentorship during my Comprehensive Exams.

The systematic review included in this thesis would not have been possible without the substantial efforts and collaborative work of the co-authors, Wei Deng, Dr. Paré, Dr. Meyre, Dr. Gerstein, and Dr. Anand. Furthermore, I want to express my gratitude to the study teams at the Chanchalani Research Centre and Population Health Research Institute. Dr. Senay Asma, Dr. Philip Joseph, Dipika Desai, Dr. Sebastien Robiou-du-Pont, Natalie Campbell, Randa Stringer, Dr. Michael Zulyniak, and Dr. Russell de Souza, I thank you for coordinating and facilitating use of the study data, teaching me new skills, and patiently answering my many questions.

I am grateful to the Canadian Diabetes Association and Ontario Graduate Scholarship program for providing me with funding to pursue my research. Additionally, I thank the faculty and students in the Department of Clinical Epidemiology & Biostatistics for enriching my learning.

Most of all, I thank my parents, Naseem and Naushad, and siblings, Nida and Nauman, for their continuous and unconditional faith in me. And to Mandark, thank you for inspiring me to pursue my dreams and for your unwavering support in this pursuit.

**Table of Contents**

Chapter 1: Introduction 1

1.1 Overview of knowledge to date 1

1.2 Aims of this thesis 5

1.3 Potential impact 6

Chapter 2: Does genetic heterogeneity account for the divergent risk of type 2 diabetes in South Asian and white Caucasian populations? 8

2.1 Background 9

2.2 Methods 9

2.2.1 Systematic review of studies assessing genetic risk of type 2 diabetes in South Asians 9

2.2.2 Comparison of meta-analysed SNPs in South Asians with white Caucasians 11

2.2.3 Testing novel SNPs discovered from South Asians GWA-studies in white Caucasians 12

2.3 Results 15

2.3.1 Systematic review and meta-analysis 15

2.3.2 Comparison of effect sizes, RAF, and risk alleles between South Asians and white Caucasians 17

2.3.3 Testing novel SNPs discovered from South Asian GWA-studies in white Caucasians 19

2.3.4 Population burden 22

2.4 Discussion 23

2.4.1 SNPs associated with type 2 diabetes from meta-analysis of South Asian studies 23

2.4.2 Comparison of effect sizes, RAF and risk alleles between South Asians and white Caucasians 24

2.4.3 Testing novel SNPs discovered from South Asians GWA-studies in white Caucasians 26

2.4.4 Population burden 27

2.4.5 Strengths and limitations 28

2.4.6 Future directions 29

2.5 Conclusions 29

Chapter 3: Dysglycemia from genetic pancreatic beta-cell failure is worsened by abdominal obesity 31

3.1 Background 31

3.2 Methods 32

3.2.1 Participants 32

3.2.2 SNP selection and genotype score 33

3.2.3 Genotyping and sample quality control 34

3.2.4 Population structure and inbreeding 35

3.2.5 Matching 35

3.2.6 Outcomes 36

3.2.8 Statistical analyses 37

3.3 Results 37

3.3.1 Association of beta-cell genotype score with glucose traits 38

3.3.2 Ethnic heterogeneity in interaction of beta-cell genotype score with abdominal obesity 40

3.3.3 Population structure and inbreeding 41

3.4 Discussion 41

3.4.1 Association of beta-cell genotype score with glucose traits 41

3.4.2 Ethnic heterogeneity in interaction of beta-cell genotype score with abdominal obesity 43

3.4.3 Population stratification and inbreeding 44

3.4.4 Strengths and limitations 45

3.5 Conclusions 46

Chapter 4: Ethnic heterogeneity in DNA Methylation for birth weight and length 47

4.1 Background 47

4.2 Methods 49

4.2.1 Participants 49

4.2.3 Quality control procedures 49

4.2.4 Ethnic heterogeneity in methylation of known birth weight and length genes 51

4.2.5 Regional analysis of significant sites 54

4.2.6 Genome-wide methylation for birth weight in South Asians 54

4.3 Results 55

4.3.1 Ethnic heterogeneity in methylation at known birth weight and length genes 56

4.3.2 Regional analysis of significant sites 65

4.3.3 Genome-wide methylation for birth weight in South Asians 67

4.4 Discussion 69

4.4.1 Ethnic heterogeneity in methylation at known birth weight genes 70

4.4.3 Regional analysis of significant sites 72

4.3.4 Genome-wide methylation for birth weight in South Asians 73

4.4.5 Strengths and limitations 74

4.5 Conclusions 75

Chapter 5: Discussion 76

5.1 Overview of findings 76

5.2 Methodological considerations 79

5.2.1 Internal validity 79

5.2.2 External validity 86

5.3 Other considerations for trans-ethnic genetic studies 87

5.3.1 Population stratification 87

5.3.2 Linkage disequilibrium 88

5.4 Future directions and Conclusions 89

References 90

Appendices 115

**List of Figures**

**Chapter 1:** No figures.

**Chapter 2:**

**Figure 2.1** Overview of the study design 13

**Figure 2.2** Flow diagram of the systematic review of South Asian literature investigating genetic variants predisposing to type 2 diabetes 14

**Figure 2.3** Forest plot of SNPs associated with type 2 diabetes in South Asians from systematic review and white Caucasians from DIAGRAM. Chr: Chromosome; RAF: risk allele frequency; P het: P-value for heterogeneity. SNPs are ordered by level of significance in South Asian meta-analysis. 18

**Figure 2.4** Venn diagram of SNPs common to South Asians and white Caucasians and SNPs unique to both groups. CENTD2 is also known as ARAP1; TMEM195 is also known as AGMO. The green box includes genes with GWAS evidence for association with type 2 diabetes; the pink box includes genes with GWAS evidence in whites and association with type 2 diabetes in this meta-analysis; and yellow with genes identified from a trans-ethnic meta-analysis of South Asians and white Caucasians. 19

**Figure 2.5** Genotype score (with 95% confidence intervals) of SNPs in South Asians and white Caucasians. The genotype score was constructed using effect estimates and RAFs from SNPs common to both groups 22

**Chapter 3:**

**Figure 3.1** Criteria for selection and inclusion of SNPs implicated in beta-cell function 34

**Figure 3.2** Interaction between beta-cell genotype score and abdominal obesity stratified by ethnic group 40

**Chapter 4:**

**Figure 4.1** Plots of Pearson’s r correlations for CpG sites on TCF7L2 and CALCR in white Caucasians and South Asians 66

**Figure 4.2** Manhattan plot showing the distribution of P values of the association between methylation probes and birth weight in South Asians 68

**Figure 4.3** Q-Q plot of P-values acquired from GWA-studies of CpG sites 69

**Chapter 5:**

**Figure 5.1** Power calculations for case-control association studies of unrelated individuals. Curves represent the number of individuals, at a case-control ratio of 1:1, required to achieve 80% power for a disease prevalence of 10% at α = 0.05. 83

**List of Tables**

**Chapter 1:** No Tables.

**Chapter 2:**

**Table 2.1** Comparison of SNPs discovered from South Asian GWA-studies with white Caucasian estimates from the DIAGRAM Consortium 21

**Table 2.2** SNPs associated with type 2 diabetes in published GWA-studies from white Caucasians but not replicated in South Asian meta-analysis 25

**Chapter 3:**

**Table 3.1** Summary characteristics of included participants from the EpiDREAM cohort 39

**Table 3.2** Effect of the genotype score on glucose traits stratified by ethnicity 39

**Chapter 4:**

**Table 4.1** Descriptive characteristics of included participants stratified by ethnicity 56

**Table 4.2** Association between percent methylation at sites showing ethnic variation for birth weight in cord-blood DNA 58

**Table 4.3** Percent methylation at the seven sites showing ethnic heterogeneity in the primary analysis 59

**Table 4.4** Association between methylation level and SNP on birth weight genes 64

**Table 4.5** Regional analysis of significant CpG sites in South Asians and white Caucasians 67

**Table 4.6** Association between percent methylation on CpG sites associated at P<1x10-5 with birth weight in South Asians 68

**Chapter 5:** No Tables.

**List of Appendices**

**Chapter 1:** No appendices.

**Chapter 2:**

**Supplementary Table 2.1** Full search strategy for systematic review 115

**Supplementary Table 2.2** Reported SNPs and their proxies in LD (r2 > 0.8) determined using SNAP 117

**Supplementary Table 2.3** Characteristics of studies included in the systematic review 118

**Supplementary Table 2.4** Risk of bias assessment for studies included in the systematic review 124

**Chapter 3:**

**Supplementary Table 3.1** Risk allele frequencies and HWE-P values for SNPs available in the EpiDREAM cohort 132

**Supplementary Table 3.2** Effect of 11 SNPs included in the genotype score on fasting glucose from GWA-studies 133

**Supplementary Figure 3.1** Histograms of AUC glucose (a) and fasting glucose (b) 135

**Supplementary Figure 3.2** First three principle components for (a) South Asians and (b) white Caucasians 136

**Supplementary Figure 3.3** Histogram of Inbreeding coefficients in (a) South Asians and (b) white Caucasians 138

**Chapter 4:**

**Supplementary Table 4.1** Inclusion and exclusion criteria for the START and CHILD cohorts 140

**Supplementary Table 4.2** Allele frequencies and HWE-P values for SNPs available in the START and CHILD cohorts. 141

**Supplementary Table 4.3** Genes (and SNPs) chosen for investigation in the primary analysis from a literature search 142

**Supplementary Table 4.4** Global methylation differences between START and CHILD cohorts 143

**Supplementary Table 4.5** SNP associations for birth weight genes in South Asians and white Caucasian newborns 144

**Supplementary Table 4.6** SNP associations for birth length genes in South Asian and white Caucasian newborns 145

**Supplementary Table 4.7** Overview of key differences in blood collection protocols of the START and CHILD cohorts 146

**Supplementary Table 4.8** Variation in leukocyte and erythrocyte cell type composition between South Asians and white Caucasians from the CHILD cohort 147

**Supplementary Table 4.9** Association between CpG sites showing ethnic heterogeneity for birth weight in South Asians from the CHILD cohort (n=18) 148

**Supplementary Table 4.10** Effect of processing time on methylation level in the CHILD cohort 149

**Supplementary Table 4.11** Effect of processing time on methylation level in the CHILD cohort 150

**Supplementary Table 4.12** Percent of variance explained by clinical variables and CpG sites on birth weight in white Caucasians 151

**Supplementary Figure 4.1** Histograms of birth weight in South Asians (a) and white Caucasians (b); birth length in South Asians (c) and and fasting glucose (d) 152

**Chapter 5:** No appendices.

**List of Abbreviations**

**2hGlu** Two hour Glucose

**AF** Allele Frequency

**AGEN** Asian Genetic Epidemiology Network Type 2 Diabetes

**AUC** Area Under Curve

**BMI** Body Mass Index

**CGI** CpG Island

**Chr.** Chromosome

**CHILD** Caucasian newborns from the Canadian Healthy Infant Longitudinal Development

**CpG** Cytosine – Guanine dinucleotide

**DIAGRAM** DIAbetes Genetics Replication And Meta-analysis

**DNA** Deoxyribo-Nucleic Acid

**DSAT** Deep Subcutaneous Adipose Tissue

**EpiDREAM** Epi-Diabetes REduction Assessment With Ramipril and Rosiglitazone Medication

**FDR** False Discovery Rate

**FPG** Fasting Plasma Glucose

**GWA** Genome Wide Association

**HC** Hip Circumference

**HOMA** Homeostatic Model of Assessment

**HWE** Hardy-Weinberg Equilibrium

**ICAM-1** Intercellular Adhesion Molecule-1

**IFG** Impaired Fasting Glucose

**IGT** Impaired Glucose Tolerance

**LBW** Low Birth Weight

**LD** Linkage Disequilibrium

**MAGIC** Meta-Analyses of Glucose and Insulin-related traits Consortium

**MRI** Magnetic Resonance Imaging

**NGT** Normal Glucose Tolerance

**ODI** Oral Disposition Index

**OGTT** Oral Glucose Tolerance Test

**OR** Odds Ratio

**RAF** Risk Allele Frequency

**SAT2D** South Asian Type 2 Diabetes

**SD** Standard Deviation

**SNP** Single Nucleotide Polymorphism

**SSAT** Superficial Subcutaneous Adipose Tissue

**START** South Asian Birth Cohort

**SVD** Singular Value Decomposition

**SWAN** Subset-quantile Within Array Normalization

**VAT** Visceral Adipose Tissue

**WaistadjHip** Waist Adjusted for Hip

**WC** Waist Circumference

**Declaration of Academic Achievement**

I am the primary author of the material included in this thesis. I contributed substantially to each chapter by taking a lead role in the conception, development, and analysis of the work discussed herein.

Drs. Anand, Gerstein, Meyre, and Paré supervised me in the completion of this work. Lastly, contributions of co-authors for published work have been included as a note in the relevant chapters.

**Chapter 1**

# Introduction

## 1.1 Overview of knowledge to date

**1.1.1 Pathophysiology of diabetes**

Type 2 diabetes is a heterogeneous disorder that affects approximately 8% of the world’s population1. A triad of metabolic defects characterizes this disease: impairment in insulin secretion, ineffective sensitivity to insulin, and an increase in glucose production by the liver. While the specific contribution of each, and the primary defect needed for the development of type 2 diabetes is the subject of research and debate, it appears that overt diabetes is preceded by failure of the pancreatic beta-cells to compensate for ineffective insulin action2,3. Reduced insulin secretion may be due to genetically mediated beta-cell dysfunction or result from hypothesized damage to the beta-cell from chronic exposure to hyperglycemia and/or free fatty acids. Under normal conditions, presence of insulin in the blood suppresses glucose secretion from the liver. However, with declining insulin levels, this suppression is inhibited, and in turn leads to hepatic glucose production, continuing to worsen dysglycemia in the course of diabetes2–4.

**1.1.2 Genetic basis of the disease**

Although type 2 diabetes is typically discussed as a product of obesity-prone lifestyles, the disease has a genetic basis. Family studies show that as high as 69% of the trait is attributed to genetic factors2. Additionally, Elbein *et al.* examined the heritability of beta-cell function, assessed as insulin response relative to sensitivity, and found heritability of this trait to be 67% in 120 participants who had either impaired glucose tolerance (IGT) or normal glucose tolerance (NGT). This estimate increased to 70% when only participants with IGT were considered5. Lastly, heritability estimates for fasting plasma glucose (FPG) range from 38% to 51% according to studies of twins between 18 and 67 years of age6,7. Genome wide association (GWA) studies investigating genetic origins of type 2 diabetes have found around 100 base pair polymorphisms in the DNA (single nucleotide polymorphisms; SNPs) to date. Similarly, 43 SNPs have been implicated in fasting glucose8–10 and 5 in 2-hour glucose measurements from an oral glucose tolerance test (OGTT)11. A vast majority of these SNPs are associated with differences in pancreatic function through impaired insulin processing or secretion3. Cumulatively, regression models with SNPs from GWA-studies account for only <10% of variation in the traits12. However, Bayesian approaches have found polygenic models with hundreds of SNPs explain approximately 50% of heritability for type 2 diabetes13. Various propositions attempt to account for this ‘missing heritability’, including consideration of epigenetic inheritance and parent of origin effects, among others14,15.

**1.1.3 Risk profile for South Asians**

Observational studies and systematic reviews have emphasized the greater risk of type 2 diabetes in people of South Asian ancestry compared to white Caucasians16–18. White Caucasians are broadly defined as those individuals from the European continent19. Similarly, South Asians are people originating from the Indian subcontinent, falling largely within two linguistic groups – Indo-Aryan and Dravidian20,21. This includes individuals who originate from India, Pakistan, Sri Lanka, Bangladesh, Bhutan, Nepal, and Maldives. Estimates of type 2 diabetes demonstrate a 2-5 fold higher risk in this group compared to white Caucasians17,22. Furthermore, once diagnosed with diabetes, South Asians have a greater predisposition to cardiovascular disease16,23,24 compared to Caucasians. A comparison between South Asians and whites over a follow-up period of 11 years found that the former were twice as likely to die from heart and circulatory disease in those aged 30–64 years at baseline. Lastly, among South Asians in the surviving cohort, there was a 3.8 times greater likelihood of myocardial infarction than white Caucasians25. Various metabolic abnormalities likely lead to this increased risk; a cross-sectional study of 1,711 South Asians and 2,346 Caucasians found a higher prevalence of dyslipidemia, impaired glucose homeostasis, proportion of diagnosed diabetes, and central obesity in South Asians. The age-standardized prevalence of metabolic syndrome was 46.3% in South Asian men and 30.8% in women compared to only 18.8% in white Caucasian men and 9.1% in women26. Studies from urban India report similar age-standardized prevalence of metabolic syndrome. Prasad *et al* found an overall prevalence of 33.5% (24.9 % in males and 42.3% in females)27, though there is some variation between urban and rural regions28,29.

Detailed assessments in South Asians reveal that they possess a greater propensity for insulin resistance and beta-cell dysfunction. A comparison of South Asians and white Caucasians matched on lifestyle factors and body-mass index (BMI) found that while South Asians demonstrated a 30% increase in basal beta-cell responsiveness, this increased function was not sufficient to compensate for a greater degree of insulin resistance, as evidenced by a 60% reduction in disposition index, a measure of beta-cell response to insulin resistance30. Similarly, using a 2-hour oral glucose challenge and 2-hour euglycemic hyperinsulinemic clamps, Raji *et al* show that South Asians exhibit fasting hyperinsulinemia, higher glucose and insulin levels during an OGTT, and reduced glucose disposal rate, despite having similar fasting glucose measurements31. Lastly, a 5-year prospective comparison of fasting glucose, 2-hour post load glucose, fasting insulin, 2-hour post load insulin, homeostatic model assessment of insulin sensitivity (HOMA-%S), and insulin secretion (HOMA-%B) suggest that in addition to lowered sensitivity to the effects of insulin, South Asians also possess reduced beta-cell function. This is demonstrated by steeper age-related trajectories of fasting glucose, higher post-load glucose, fasting insulin, post-load insulin and HOMA-%B, as well as lower HOMA-%S among South Asians compared to whites32.

Anthropometric characteristics also differ in this group. The superficial subcutaneous adipose tissue (SSAT) compartment is the site of primary adipose tissue storage in the body and constitutes the vast majority of adipose tissue in the lower limbs. With increasing levels of adiposity, these stores are used up and fat is stored in the deep subcutaneous (DSAT) and the visceral adipose tissue (VAT) compartments. Whereas the SSAT is relatively metabolically inert, the DSAT and VAT are implicated in transmembrane fatty acid flux and associated with dyslipidemia and dysglycemia33. Comparisons of these regions report that South Asians have greater total abdominal and visceral fat31,34, lower fat-free mass35–38, and preferentially store fat in the deep subcutaneous and visceral depots39, versus white Caucasians, who store fat in the superficial region40. Taken together, these data support the contention that South Asians are comparably more adipose at the same BMI as white Caucasians41.

The differences among South Asians and white Caucasians are present in adolescence and at birth42–44. Several cross-sectional and longitudinal studies show that South Asian children have greater %body fat with a higher degree of central adiposity45–48, and a higher prevalence of impaired glucose tolerance45–47,49,50. Similar observations are evident in South Asian youth from studies around the world and the findings seem to be generalizable globally. Peters & Ulijaszek concluded that South Asian children deposit more fat on the trunk and less on upper limb compared to white Caucasians after studying subscapular and tricep skinfold thickness in 2,224 children in England51. Gulliford reported similar results in children of South Asian descent in Trinidad52. Similarly, at birth, South Asians appear to be smaller but with relatively more adipose tissue, measured using skinfold thickness and adipose tissue volume with an MRI53–56. On a global scale, the prevalence of low birth weight (LBW) is highest in the South Asian region, with India accounting for nearly 40% of global LBW infants57. Of particular importance are the observations that the afore-described differences persist in South Asians even after several generations in ‘Western’ countries17,25,42,43,56, advocating for the role of genetic and epigenetic factors as well as gene-environment interactions to explain the observed physiological and epidemiological variation in the risk for metabolic disease and type 2 diabetes.

## 1.2 Aims of this thesis

The overall aim of this thesis is to explore the role of genetic variants and epigenetic differences to explain the greater risk for type 2 diabetes in South Asians. Countless epidemiological studies show that similar to South Asians, many racial / ethnic minorities have a substantially higher risk for cardiovascular and metabolic diseases than do white Caucasians58–60. Studying ethnic variation using white Caucasians as a comparison can help elucidate disease pathogenesis and uncover differences, be they physiological, genetic, environmental or a combination thereof, that mark a greater risk in certain racial groups61.

The thesis is comprised of three parts. In Chapter two, we evaluate genetic susceptibility to type 2 diabetes among South Asians and white Caucasians using existing published literature. There is much speculation in the literature that the higher prevalence of type 2 diabetes results from a greater genetic predisposition among South Asians10,11, either because of higher risk conferred by SNPs or increased frequency of risk alleles. This hypothesis was evaluated first by conducting a meta-analysis of the literature to create a robust list of SNPs predisposing to type 2 diabetes. Next, we compared the risk estimates from this list between South Asians and white Caucasians. We also assessed population burden from these SNPs in both ethnic groups. The last part of the chapter tested eight novel SNPs discovered from South Asian studies in a cohort of 69,033 white Caucasians.

In Chapter three, we explore whether a greater risk for type 2 diabetes in South Asians is influenced by a genetic predisposition to beta-cell dysfunction and if this is magnified in the presence of abdominal adiposity. To do this, we constructed a genotype score of polymorphisms robustly linked to beta-cell dysfunction with evidence from GWA-studies, functional studies, animal models, and/or monogenic diabetes models. We evaluate whether the impact of these SNPs is greater in those with abdominal obesity. We then contrasted findings in South Asians with those from white Caucasians in 5,302 people from the EpiDREAM cohort62.

As differences in body composition between South Asians and whites are present at birth, the variation may in part be a result of differences in the *in-utero* environment. Methylation of genes is the most commonly studied epigenetic mark and results from interactions between the environment and the genetic make-up of an individual. Therefore, Chapter four evaluates genetic and epigenetic differences in birth weight and length between the two ethnic groups using data from two Canadian birth cohorts63.

## 1.3 Potential impact

Findings from this thesis will: (1) further elucidate the etiology and pathogenesis of type 2 diabetes and guide research for diseases with a genetic component, (2) promote an accurate understanding of the differences in risk between South Asians and white Caucasians, and allow for more astute risk stratification in clinical settings, (3) identify genetic and epigenetic targets for the development of novel anti-diabetic drugs, and (4) lastly, allow for recommendations for women during pregnancy based on the epigenetic causes of LBW.

**Chapter 2\***

# Does genetic heterogeneity account for the divergent risk of type 2 diabetes in South Asian and white Caucasian populations?

**\**Nota bene*:** This chapter has been published in: Sohani ZN, Deng WQ, Pare G, Meyre D, Gerstein HC, Anand SS. Does genetic heterogeneity account for the divergent risk of type 2 diabetes in South Asian and white European populations? *Diabetologia*. 2014. doi:10.1007/s00125-014-3354-1.

**Authors’ contributions:** ZNS made substantial contributions to the design of the study, acquisition of data, conducting the analysis, the interpretation of the data and drafting the article, and provided final approval of the version to be published. SSA made substantial contributions to conception and design and analysis and interpretation of data, critically revised the article for important intellectual content and provided final approval of the version to be published. WQD contributed to the acquisition of data, critically reviewed the article for important intellectual content and provided final approval of the version to be published. GP, DM and HCG substantially contributed to conception and design, critically revised the draft for important intellectual content and provided final approval of the version to be published. SSA is the guarantor of this work.

## 2.1 Background

The elevated risk for type 2 diabetes among South Asians has been hypothesized to result from a greater genetic predisposition64,65, either because of larger risk estimates for each single nucleotide polymorphism or increased frequency of risk alleles. This hypothesis has not been systematically evaluated. The purpose of this systematic review is to: (1) establish risk estimates for SNPs predisposing South Asians to type 2 diabetes; (2) compare risk estimates, risk alleles and risk allele frequencies (RAFs) of type 2 diabetes predisposing SNPs between South Asians and white Caucasians; and (3) explore the association of novel SNPs discovered from South Asians in a large cohort of white Caucasians.

## 2.2 Methods

### 2.2.1 Systematic review of studies assessing genetic risk of type 2 diabetes in South Asians

***Search strategy and selection criteria:*** MEDLINE, Embase, the Cumulative Index to Nursing and Allied Health Literature (CINAHL) and the Cochrane registry (from inception to 17 June 2013; MEDLINE and Embase searched using OvidSP) were searched for studies of genetic variants associated with type 2 diabetes in South Asians. South Asians were defined as individuals originating from India, Pakistan, Bangladesh or Sri Lanka. The search strategy was developed in consultation with a research librarian and did not restrict by type of genetic variant, language of study or study design. The full search strategy is presented in Supplementary Table 2.1. Experts were consulted, and reference lists of included articles and relevant excluded articles were searched. Two reviewers (ZNS and WQD) independently assessed each study for eligibility based on four questions: (1) is at least one study population South Asian; (2) is type 2 diabetes the outcome studied; (3) is the exposure a genetic variant; and (4) is this a genetic association study? Disagreements were independently resolved by a third reviewer (SSA). Articles that passed the screening phase were reviewed in depth.

***Full-text review and data extraction:*** Conference abstracts, narrative reviews and other systematic reviews were excluded after full-text review. Primary studies investigating genetic variants other than bi-allelic SNPs (e.g. insertions, deletions, length polymorphisms, haplotypes and complex tri-allelic SNPs) or with fewer than 25 cases were excluded. These limitations were implemented to ensure that estimates from included studies could be pooled and were of sound quality. Additionally, to ensure robustness of our results, SNPs from included studies carried forward to meta-analysis were limited to those that previously reached genome-wide significance (*P*<5x10-8) in any ethnicity. If datasets were published more than once, publication with the largest sample size or the one recommended by the senior author was used for meta-analysis. Two reviewers conducted a full-text assessment and extracted data on participant characteristics (sample size, age, BMI, fasting glucose, waist and hip circumferences, and % men), study design (study methods, region) and results (SNP, risk allele, RAF in cases and controls, and risk estimate with a 95% CI). For GWA-studies in South Asians, SNPs associated with type 2 diabetes at *P*<10−3 were considered for meta-analysis.

***Quality assessment of included studies:***Using recommendations from the STrengthening the REporting of Genetic Association Studies (STREGA) guidelines66, the following three components were considered for quality assessment: testing for Hardy–Weinberg equilibrium (HWE) in controls67, reporting of genotyping call rate or appraisal of genotyping quality by duplication, and sources of case and control ascertainment.

***Statistical analysis:*** Agreement between reviewers was reported with Cohen’s κ statistic. Allelic ORs were calculated using RAFs and sample size for cases and controls and assessed for significance using Fisher’s exact test. The ORs were not adjusted for any covariates. ORs for SNPs in linkage disequilibrium (LD) (*r*2>0.8) obtained from independent cohorts were meta-analysed using random effects model weighted by inverse variance. As data on LD in South Asians was not available using the SNAP software (Broad Institute, Cambridge, MA)68, LD was estimated using 1000 Genome (Pilot 1) Centre d'Etude du Polymorphisme (CEU) (Utah residents with northern and western European ancestry). A two-sided α level of 0.05 was considered significant. For a list of SNPs in LD, see Supplementary Table 2.2. Summary RAFs are presented as averages, weighted by sample size, of all included reports. Heterogeneity was estimated using *I*2 statistic, which indicates the proportion of total variation in estimates attributed to heterogeneity, as well as the *Q* statistic69. A cut-off of 25% for *I*2 was used to represent minimal heterogeneity, 50% to represent moderate and 75% to represent high heterogeneity. Quanto, Los Angeles, CA (version 1.2.4) was used for sample size and power calculations assuming a disease prevalence of 10% and an additive model of inheritance. All other statistical analyses were conducted in R (version 3.0.2).

### 2.2.2 Comparison of meta-analysed SNPs in South Asians with white Caucasians

Estimates for SNPs associated with type 2 diabetes in this meta-analysis were retrieved for white Caucasians from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium. Specific details on the consortium have been previously described70. Briefly, DIAGRAM stage 1 is a publicly available database of 12 GWAS studies with 12,171 cases and 56,862 controls. DIAGRAM authors tested SNPs with minor allele frequency >1% for association with type 2 diabetes under an additive model. The authors combined summary estimates from the 12 GWAS using fixed-effect inverse-variance-weighted meta-analysis.

Effect sizes from white Caucasians in DIAGRAM and South Asians from this meta-analysis were transformed to natural logs and compared using a *Z* test. RAFs in white Caucasians were acquired from published GWAS ([www.genome.gov/gwastudies/](http://www.genome.gov/gwastudies/)). Additionally, a genotype score of SNPs present in both groups was constructed to compare population burden from these SNPs. The genotype score was constructed as below:

Genotype score= Σ [log*e* (ORi) × RAFi]

Variance for the genotype score was estimated as:

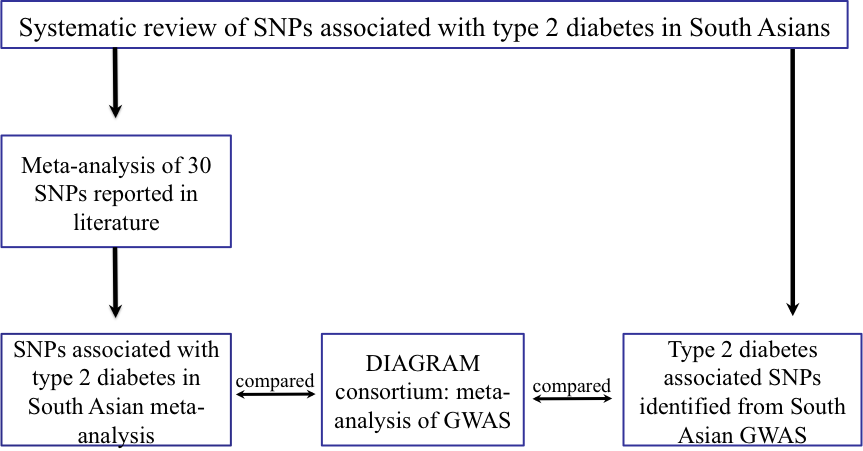
Variance = Σ {RAFi2 × variance[log*e* (ORi)]}

Genotype scores from both ethnicities were compared using a *Z* test.

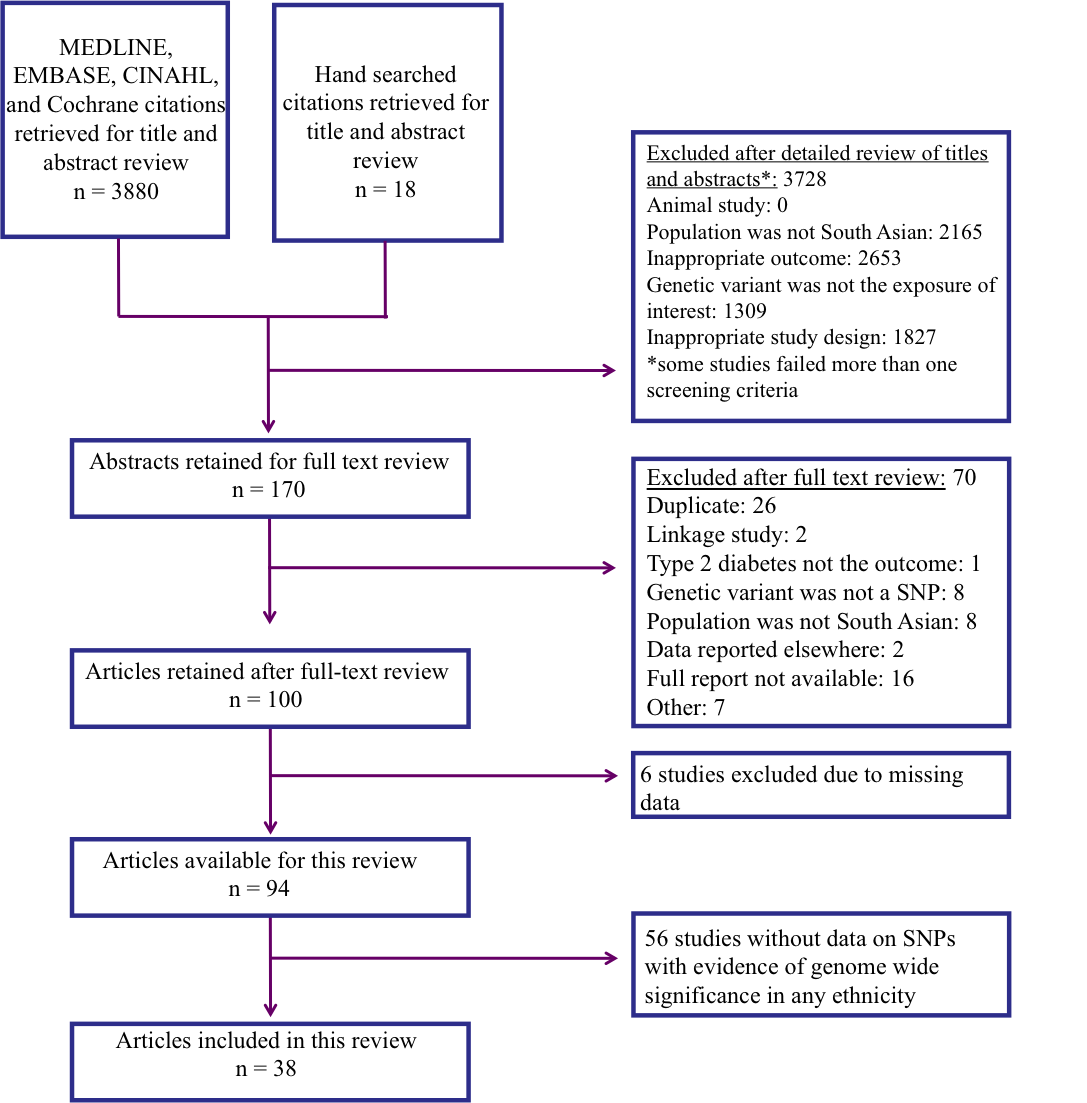
### 2.2.3 Testing novel SNPs discovered from South Asians GWA-studies in white Caucasians

Four GWAS have been conducted in South Asians71–74, which collectively identified nine SNPs (*P*<5×10−8). All studies used a case–control design. Associations with type 2 diabetes for **eight** of the nine SNPs have not been independently discovered from GWAS in any other ethnicity. We compared the association for these eight novel SNPs with type 2 diabetes in white Caucasians from DIAGRAM. An overview of the study design is presented in Figure 2.1.

**Figure 2.1** Overview of the study design



**Figure 2.2** Flow diagram of the systematic review of South Asian literature investigating genetic variants predisposing to type 2 diabetes



## 2.3 Results

### 2.3.1 Systematic review and meta-analysis

***Search yields:*** A total of 3,898 articles were originally screened, of which 170 were carried to full-text review. Ninety-four studies met the inclusion and exclusion criteria, and 38 studies comprising 31 independent cohorts contained data on SNPs with evidence of genome-wide significance in any ethnicity. These 38 studies were included in the systematic review. Figure 2.2 depicts the selection process and lists reasons for exclusion at each stage. There was excellent agreement on study inclusion between reviewers (unweighted κ =0.898).

***Study characteristics:*** All included studies were in English and published in 2004 or later. Most large datasets (>2,500 people) were published after 2008. Eight of the 31 cohorts were from North India, four from South India, five from Pakistan, one from the Eastern region (Orissa), four from Indians residing in Singapore, two from Sri Lanka, one from Indians residing in Trinidad, one from Indians residing in Mauritius and one from all over the South Asian subcontinent; four were unspecified.

***Participant characteristics:*** Studies eligible for analysis included 29,618 cases and 40,329 controls. Controls were described as healthy individuals who were normoglycaemic and/or non-diabetic. Men accounted for 54% of cases and 53% of controls. Age ranged from 34 to 62 years in cases75,76, and from 28 to 62 years in controls75,77. BMI was reported for cases and controls in 79% of studies and ranged from 25.0 to 31.9 kg/m2 in cases78,79 and from 19.4 to 28.1 kg/m2 80,81 in controls. A majority of studies reported fasting glucose: 58% in cases (ranging from 6.41 to 10.60 mmol/l)82,83 and 63% in controls (ranging from 4.60 to 6.89 mmol/l)71,84. Very few datasets reported additional clinical information; 29% reported waist circumference in cases and controls. Only three studies reported hip circumference. Supplementary Table 2.3 provides a summary of all included studies.

***Quality assessment:*** Agreement with HWE in controls, genotyping call rate, and source of controls and cases for included studies are summarised in Supplementary Table 2.4. One study did not report compliance with HWE85. *P*-value thresholds for HWE varied by study (range: *P*=10-2 to 10-4). SNPs in one study deviated from HWE86; rs2237892 and rs2237897 deviated from HWE at *P*=0.002 and *P*=0.001, respectively. The SNPs were excluded from analysis because the HWE threshold in the primary report was fairly conservative. Six (16%) studies did not appraise genotyping quality85,87–91. Furthermore, three studies had a genotyping call rate below 95% for all reported SNPs86,92,93.

***Risk estimates for SNPs associated with type 2 diabetes in South Asians:*** Thirty SNPs with previous evidence of genome-wide significance in other ethnicities were available for meta-analysis. Pooled estimates were significant for 15 of the 30 SNPs (Figure 2.3). Summary ORs, on the whole, suggested between a 1.15- and a 1.35-fold increase in susceptibility for type 2 diabetes per risk allele. Surprisingly, the risk alleles for both *KCNQ1* polymorphisms substantially increased the odds of type 2 diabetes (rs2237892 OR 1.62, 95% CI 1.01-2.59; rs2237897 OR 2.19, 95% CI 1.25-3.82). High degrees of between-study heterogeneity were evident for *CDKAL1* rs7754840 (*I*2=70%, *Q*=19.99, p<0.01) and *TCF7L2* rs7903146 (*I*2=90%, *Q*=99.52, *P* <0.01). In the *TCF7L2* meta-analysis, there was some evidence that larger studies had more conservative estimates, although Chauhan et al92 and Uma Jyothi et al82 were exceptions. The smallest study (n=40 case/control pairs) depicted an association that was directionally inconsistent with the others79. Heterogeneity estimates did not change much without this outlier (*I*2=90%, *Q*=90.31, *P*<0.01). Additional sources of heterogeneity could include consanguinity91,94. One very obvious outlier existed in the meta-analysis for *CDKAL1* (OR 2.22, 95% CI 1.64-2.98)94. Heterogeneity estimates without this outlier diminished substantially (*I*2=0%, *Q*=2.80, *P*=0.73). None of the outliers was excluded from our final analysis.

A majority of genes associated with type 2 diabetes from this meta-analysis (*ADCY5*, *CDKAL1*, *CDKN2A/B*, *HHEX*, *IGF2BP2*, *SLC30A8*, *TCF7L2* and *KCNQ1*) are involved in pancreatic beta cell function, while two are implicated in insulin sensitivity (*PPARG*) and adiposity (*FTO*). This is not surprising as insulin secretion is a more heritable trait than insulin action95. In addition to the above, eight novel SNPs were identified from South-Asian-only GWAS; these are discussed later in this paper.

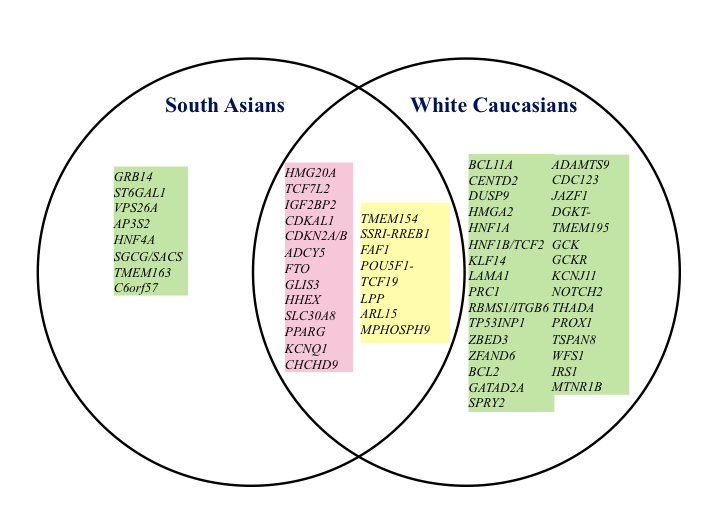
### 2.3.2 Comparison of effect sizes, RAF, and risk alleles between South Asians and white Caucasians

Meta-analysis estimates in South Asians were compared with white Caucasians from DIAGRAM; no significant differences were identified for most SNPs (*P*<0.05) (Figure 2.3), although two SNPs showed heterogeneity (*KCNQ1* rs2237897 and *IGF2BP2* rs4402960). South Asians had a slightly smaller OR for *IGF2BP2* rs4402960 (South Asians: 1.07, 95% CI 1.04−1.09; white Caucasians: 1.13, 95% CI 1.09−1.17). Figure 2.4 presents a Venn diagram of overlapping type 2 diabetes predisposing genes in South Asians and white Caucasians, as well as genes unique to both groups. Variation in RAFs between South Asians and white Caucasians was observed, but no consistent trend was evident; South Asians did not consistently have an RAF greater than white Caucasians (Figure 2.4). While the risk allele for *HHEX* was the same in both ethnicities, the RAF differed; the risk allele is the minor allele in South Asians, but the major/common allele in whites. Risk alleles for six SNPs (*SLC30A8* rs13266634, *ADCY5* rs11708067, *PPARG* rs1801282, *CHCHD9* (also known as also known as *CHCHD2P9*) rs13292136, *KCNQ1* rs2237897 and *CDKN2A*/*B* rs10811661) were major/common alleles in both ethnicities.

**Figure 2.3** Forest plot of SNPs associated with type 2 diabetes in South Asians from systematic review and white Caucasians from DIAGRAM. Chr: Chromosome; RAF: risk allele frequency; P het: P-value for heterogeneity. SNPs are ordered by level of significance in South Asian meta-analysis.



**Figure 2.4** Venn diagram of SNPs common to South Asians and white Caucasians and SNPs unique to both groups. *CENTD2* is also known as *ARAP1*; *TMEM195* is also known as *AGMO.* The green box includes genes with GWAS evidence for association with type 2 diabetes; the pink box includes genes with GWAS evidence in whites and association with type 2 diabetes in this meta-analysis; and yellow with genes identified from a trans-ethnic meta-analysis of South Asians and white Caucasians.



### 2.3.3 Testing novel SNPs discovered from South Asian GWA-studies in white Caucasians

The systematic review included four GWAS in South Asians, which identified nine SNPs associated with type 2 diabetes at genome-wide significance. With the exception of *HMG20A* rs7178572 (South Asian OR: 1.09, 95% CI 1.06-1.12; white Caucasian OR: 1.08, 95% CI 1.05-1.10)71,96, the remaining eight SNPs have not been discovered in any other ethnic group. Five of the eight (*GRB14* rs3923113, *VPS26A* rs1082295, *HNF4α* [also known as *HNF4A*] rs4812829, *ST6GAL1* rs16861329, and *AP3S2* rs2028299) were independently replicated in South Asians72. We tested the eight novel SNPs for association with type 2 diabetes in DIAGRAM. *SGCG* rs9552911 was monomorphic in white Caucasians. The remaining SNPs were directionally consistent with South Asian estimates and three (*AP3S2* rs2028299, *GRB14* rs3923113 and *HNF4α* rs4812829) were nominally (*P*<0.05) associated with type 2 diabetes (Table 2.1) in DIAGRAM. The other four were not associated with type 2 diabetes in white Caucasians. It should be noted that after meta-analysis of the novel SNPs in South Asians, *C6orf57* rs1048886 and *GRB14* rs3923113 were slightly above the genome-wide threshold. Lastly, no consistent trend in RAFs was observed. The risk allele for *TMEM163* rs6723108 is the same in both ethnicities, but the RAF differs. Specifically, the risk allele is the minor allele in white Caucasians, but the major/common allele in South Asians.

**Table 2.1** Comparison of SNPs discovered from South Asian GWA-studies with white Caucasian estimates from the DIAGRAM Consortium

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Chr.** | **Gene** | **SNP** | **Risk allele** |  |  | | | | |  |  | | |  | ***P*heterogeneity** |
|  | **South Asian** | | | | |  | **White Caucasian** | |  |  |
|  | **Reference** | **Sample size (case/control)** | **RAF** | **OR (95% CI)** | **Meta-analysed OR (95% CI)a** |  | **RAFb** | **DIAGRAM OR (95% CI)** | **Sample size (case/control)** |  |
| 13 | *SGCG* | rs9552911 | G |  | Saxena et al, 2013 73 | 2620/4284 | 0.08 | 1.49 (1.30-1.72) | 1.49 (1.30-1.72)  (p=4.25×10–8) |  | – | – | – |  | – |
| 15 | *AP3S2* | rs2028299 | C |  | Kooner et al, 2011 71 | 18731/39856 | 0.31 | 1.10 (1.07-1.13) | 1.10 (1.07-1.13)  (p=2.49×10–11) |  | 0.26 | 1.04 (1.00-1.09)  (p=0.03) | 9580/53810 |  | 0.03 |
|  | Tabassum et al, 2013 72 | 1256/1209 | 0.30 | 1.10 (0.96-1.25) |  |  |
| 3 | *ST6GAL1* | rs16861329 | C |  | Kooner et, 2011 71 | 18731/39856 | 0.75 | 1.09 (1.06-1.12) | 1.09 (1.06-1.12)  (p=1.94×10–9) |  | 0.87 | 1.03 (0.97-1.09)  (p=0.39) | 6201/48359 |  | 0.09 |
|  | Tabassum et al, 2013 72 | 1256/1209 | 0.75 | 1.14 (0.99-1.31) |  |  |
| 6 | *C6orf57* | rs1048886 | G |  | Sim et al, 2011 74 | 977/1169 | 0.18 | 1.54 (1.32-1.80) | 1.54 (1.32-1.80)  (p=8.32×10–8) |  | 0.17 | 1.01 (0.97-1.06)  (p=0.56) | 12171/56862 |  | <0.01 |
| 2 | *GRB14* | rs3923113 | A |  | Kooner et al, 2011 71 | 18731/39856 | 0.74 | 1.09 (1.06-1.13) | 1.09 (1.06-1.13)  (p=2.06×10–7) |  | 0.61 | 1.04 (1.00-1.08)  (p=0.03) | 9580/53810 |  | 0.07 |
|  | Tabassum et al, 2013 72 | 1256/1209 | 0.76 | 1.15 (0.99-1.33) |  |  |
| 15 | *HMG20A* | rs7178572 | G |  | Kooner et al, 2011 71 | 18731/39856 | 0.52 | 1.09 (1.06-1.12) | 1.09 (1.06-1.12)  (p=1.94×10–9) |  | 0.68 | 1.08 (1.05-1.10)c  (p=2.17×10–8) | 22669/58119 |  | 0.62 |
|  | Tabassum et al, 2013 72 | 1256/1209 | 0.51 | 1.15 (1.02-1.30) |  |  |
| 20 | *HNF4α* | rs4812829 | A |  | Kooner et al, 2011 71 | 18731/39856 | 0.29 | 1.09 (1.06-1.12) | 1.13 (1.02-1.26)  (p=0.02) |  | 0.16 | 1.07 (1.02-1.12)  (p=0.01) | 9580/53810 |  | 0.35 |
|  | Tabassum et al, 2013 72 | 1256/1209 | 0.28 | 1.15 (1.02-1.30) |  |  |
| 10 | *VPS26A* | rs1802295 | T |  | Kooner et al, 2011 71 | 18731/39856 | 0.26 | 1.08 (1.05-1.12) | 1.08 (1.06-1.10)  (p=3.02×10–15) |  | 0.27 | 1.02 (0.98-1.06)  (p=0.28) | 12171/56862 |  | 0.01 |
|  | Tabassum et al, 2013 72 | 1256/1209 | 0.27 | 1.09 (0.95-1.25) |  |  |
| 2 | *TMEM163*d | rs6723108 | T |  | Tabassum et al, 2013 72 | 1256/1209 | 0.87 | 1.31 (1.20-1.44) | 1.31 (1.20-1.44)  (p=1.26×10–8) |  | 0.62 | 1.01 (0.97-1.04)  (p=0.71) | 9580/53810 |  | <0.01 |

aORs are from meta-analysis of studies, where applicable

bRAF from 1,000 genomes (EUR)

cEstimate obtained from stage 2 (sample size is an approximate); found to be significantly associated with type 2 diabetes in white Caucasians

dOR adjusted for age and sex

### 2.3.4 Population burden

All sixteen SNPs, which were significantly associated with type 2 diabetes in South Asians, were included in the genotype score (*ADCY5*, *CDKAL1* rs7754840andrs7756992, *CDKN2A/2B*, *FTO*, *GLIS3*, *HHEX*, *PPARG*, *SLC30A8*, *TCF7L2* rs7903146 and rs12255372, *CHCHD9*, *KCNQ1* rs2237892 and rs2237897, *IGF2BP2* and *HMG20A* SNPs from Table 2.1 and Figure 2.3). RAFs for both ethnicities were weighed by White Caucasian summary ORs from DIAGRAM because our primary analysis showed that individual risk estimates did not vary, and due to a larger sample, we expect the White Caucasian estimates to be more precise. Using this approach, no significant difference in the population burden was observed (*P*=0.85) (Figure 2.5).

**Figure 2.5** Genotype score (with 95% confidence intervals) of SNPs in South Asians and white Caucasians. The genotype score was constructed using effect estimates and RAFs from SNPs common to both groups



## 2.4 Discussion

Twenty-four SNPs were associated with type 2 diabetes in South Asians, eight of which were novel, discovered from GWAS. There were some variations in RAFs but the effect sizes for common SNPs did not differ between the ethnic groups. Interestingly, only three of the novel SNPs discovered from the South Asian GWAS were nominally associated with type 2 diabetes in white Caucasians. Overall, the population burden from type 2 diabetes SNPs estimated using a genotype score appears to be comparable in both ethnicities.

### 2.4.1 SNPs associated with type 2 diabetes from meta-analysis of South Asian studies

Meta-analysis of risk alleles for most SNPs increased the odds of type 2 diabetes by 15-35% among South Asians. Notable exceptions were the *KCNQ1* SNPs, which showed prominent ORs for type 2 diabetes. *KCNQ1* SNPs have been shown as a more significant contributor to type 2 diabetes than other loci in other Asian populations97 and because both SNPs (in weak LD) with larger ORs are located on the same gene, it appears that *KCNQ1* may truly have a stronger signal in South Asians than in whites. However, the exceptionally large effect size for *KCNQ1* rs2237897 is likely to be inaccurate, given the wide CIs and inconsistency with GWAS estimates from Caucasians and East Asians (OR 1.33, 95% CI 1.24-1.41)98. Moreover, the *KCNQ1* rs2237897 association was reported in only two studies among South Asians and thus could be a product of the winner’s curse99.

Fifteen GWAS signals for type 2 diabetes in white Caucasians, East Asians, and Singaporean Malay100–103 were not associated with type 2 diabetes in our South Asian meta-analysis. This may reflect the low power to detect similar effect sizes for *ADAMTS9* rs4607103, *CDC123* rs12779790, *JAZF1* rs864746, *KCNQ1* rs231362, *DGKT* rs2191349, *GCK* rs1799884, *GCKR* rs780094, *MTNR1B* rs10830963, *PROX1* rs340874, and *TSPAN8* rs7961581 (Table 2.2) and either a true lack of association for *KCNJ11* rs5219, *KCNQ1* rs2237895, *NOTCH2* rs10923931, *THADA* rs7578597 and *WFS1* rs10010131 or significant between-study heterogeneity as we were adequately powered for the last five SNPs (Table 2.2). *KCNJ11* rs5219 in particular was close to significance (OR 1.19. 95% CI 0.98-1.45) and demonstrated a high degree of heterogeneity (*I*2=81%, *P*<0.01).

### 2.4.2 Comparison of effect sizes, RAF and risk alleles between South Asians and white Caucasians

In general when comparing effect sizes of SNPs associated with type 2 diabetes from our meta-analysis with DIAGRAM estimates in white Caucasians, the risk from SNPs predisposing to type 2 diabetes did not differ substantially between the groups. However, the point estimates were more precise and CIs tighter among the white Caucasians because of the larger sample size. Our observation of no significant difference in risk estimates is not surprising because SNPs evaluated in South Asians are selected for homogeneity as they were first discovered in white Caucasians and then replicated in South Asians. If there is a difference in genetic risk between the ethnic groups, it probably does not result from polymorphisms common to both groups. The results of our paper are supported by those recently published by DIAGRAM, Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium, and South Asian Type 2 Diabetes (SAT2D) Consortium104. Their study also found effect estimates for common SNPs predisposing to type 2 diabetes to be homogenous among South Asians and white Caucasians. In addition to the recently published study, we show that the genotype score, which measures population burden in both groups, does not differ. Our conclusion that the genetic risk for type 2 diabetes probably does not differ between the two ethnicities is greatly strengthened by this recent publication.

**Table 2.2** SNPs associated with type 2 diabetes in published GWA-studies from white Caucasians but not replicated in South Asian meta-analysis

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Chr.** | **Gene** | **SNP** | **GWAS estimate** | |  | **South Asian meta-analysis** | |  | **Directionc** | **Cases in current meta-analysis** | **Power in this meta-analysis** | **Case–control pair needed for 80% power** |
| **Reference** | **OR (95% CI)a** |  | **OR (95% CI)b** | **RAF** |  |
| 3 | *ADAMTS9* | rs4607103 | Zeggini et al, 2008 101 | 1.09 (1.06-1.12) |  | 1.01 (0.90-1.14) | 0.50 |  | + | 2307 | 0.54 | 4231 |
| 10 | *CDC123* | rs12779790 | Zeggini et al, 2008 101 | 1.11 (1.07-1.14) |  | 1.11 (0.97-1.27) | 0.14 |  | + | 4049 | 0.65 | 6135 |
| 7 | *JAZF1* | rs864745 | Zeggini et al, 2008 101 | 1.10 (1.07-1.13) |  | 1.08 (0.96-1.22) | 0.71 |  | + | 4055 | 0.78 | 4285 |
| 11 | *KCNQ1* | rs2237895 | Tsai et al, 2010 100 | 1.29 (1.19-1.40) |  | 1.12 (0.96-1.31) | 0.42 |  | + | 1424 | 0.99 | 496 |
| 7 | *DGKT* | rs2191349 | Dupuis et al, 2010 10 | 1.06 (1.04-1.08) |  | 1.04 (0.85-1.27) | 0.61 |  | + | 1632 | 0.21 | 9771 |
| 7 | *GCK* | rs1799884 | Dupuis et al, 2010 10 | 1.07 (1.05-1.10) |  | 0.92 (0.73-1.15) | 0.14 |  | + | 1670 | 0.16 | 13,971 |
| 7 | *GCKR* | rs780094 | Dupuis et al, 2010 10 | 1.06 (1.04-1.08) |  | 1.07 (0.96-1.19) | 0.76 |  | + | 4055 | 0.35 | 12,833 |
| 11 | *KCNJ11* | rs5219 | Scott et al., 2007 105 | 1.14 (1.10-1.19) |  | 1.19 (0.98-1.45) | 0.37 |  | + | 5643 | 1.00 | 1939 |
| 7 | *PROX1* | rs340874 | Dupuis et al, 2010 10 | 1.07 (1.05-1.09) |  | 0.96 (0.83-1.11) | 0.58 |  | + | 1626 | 0.27 | 7073 |
| 11 | *MTNR1B* | rs10830963 | Dupuis et al, 2010 10 | 1.09 (1.06-1.12) |  | 1.01 (0.91-1.11) | 0.40 |  | + | 3475 | 0.70 | 4377 |
| 2 | *THADA* | rs7578597 | Zeggini et al, 2008 101 | 1.15 (1.10-1.20) |  | 1.05 (0.92-1.20) | 0.86 |  | + | 4090 | 0.86 | 3313 |
| 11 | *KCNQ1* | rs231362 | Voight et al, 2010 102 | 1.08 (1.06-1.10) |  | 1.15 (0.97-1.36) | 0.75 |  | + | 3052 | 0.45 | 6847 |
| 1 | *NOTCH2* | rs10923931 | Zeggini et al, 2008 101 | 1.13 (1.08-1.17) |  | 1.06 (0.94-1.19) | 0.21 |  | + | 4035 | 0.89 | 3581 |
| 12 | *TSPAN8* | rs7961581 | Zeggini et al, 2008 101 | 1.09 (1.06-1.12) |  | 1.05 (0.95-1.16) | 0.33 |  | + | 3977 | 0.73 | 4865 |
| 4 | *WFS1* | rs10010131d | Voight et al, 2010 102 | 1.13 (1.08-1.18) |  | 1.05 (0.96-1.15) | 0.72 |  | + | 5022 | 0.97 | 2472 |

aORs are from the referenced GWAS

bORs are from the South Asian meta-analysis

c+ is the same direction of effect

dGWAS significant SNP rs1801214 is in LD (*r*2=1) with reported SNP

Figure 2.4 depicts overlap between genes associated with type 2 diabetes in South Asians and white Caucasians. Considerably more signals have been identified in whites due to the greater number of GWAS. Because a majority of signals were replicated in our meta-analysis, it is unlikely that they are unique to white Caucasians. Rather, larger GWAS with greater than 36,000 case/control pairs are required to detect ORs as low as 1.05 with an RAF of 10%. The four GWAS in South Asians represent about 25,704 cases and 43,688 controls, therefore more GWAS in South Asians alone will detect small effects and discover SNPs unique to this group, and ultimately facilitate further elucidation of the genetic basis of type 2 diabetes in this group.

We did not find a trend in RAFs; specifically, RAFs were not consistently higher in South Asians. Risk alleles from six SNPs in this analysis were the major/common alleles in both ethnicities. Negative selection usually prevents risk alleles from becoming common unless they are advantageous106. Risk alleles for genes encouraging fat storage may have been advantageous in ancient environments with unpredictable food supply and high level of activity, but now could predispose to type 2 diabetes106. To this end, literature examining evidence of selection for type 2 diabetes variants has shown inconsistent results107–109.

### 2.4.3 Testing novel SNPs discovered from South Asians GWA-studies in white Caucasians

Finally, we tested novel SNPs derived from South Asians in white Caucasians; only three were associated with type 2 diabetes. Interestingly, one of the novel SNPs, rs1048886, is a functional missense mutation resulting in a change from glutamine to arginine in the UPF0369 protein involved in immune response110. Additionally, the G risk allele for *SGCG* rs9552911 is not present in white Caucasians. The putative existence of polymorphisms existing in South Asians that are not present in Caucasians is supported by presence of assortative mating resulting from prolonged geographical and cultural isolation of this region from European whites21 as well as founder effects20. Non-association of the remaining seven dimorphic SNPs with type 2 diabetes in white Caucasians is quite puzzling as the risk alleles are present in this group. DIAGRAM’s large sample ensures reasonable power; for the lowest RAF of all eight SNPs, DIAGRAM sample had greater than 80% power to detect an OR of at least 1.07, and therefore non-association may be due to substantially different LD structures or lower effect sizes in white Caucasians because of ethnic specific gene–environment or gene–gene interactions. We compared the LD structure for the seven SNPs using HapMap data, and while not substantial, *r*2 values with neighbouring SNPs have some differences (for example, *r*2 of rs1048886 with neighbouring rs9455158 is 0.75 in white Caucasians and 0.87 in South Asians; *r*2 of rs3923113 with neighbouring rs13432797 is 0.93 in white Caucasians and 0.64 in South Asians). We also compared D’ values since *r*2 tends to depend heavily on allele frequencies111. Fine mapping analyses will determine whether the same functional variants are responsible for increased risk in both groups and if the effect is comparable, or if the seven SNPs tag unique functional variants in South Asians.

### 2.4.4 Population burden

No difference in genotype score between South Asians and white Caucasians was observed, which is consistent with our finding that no trend exists for variation in RAFs. It should be noted that our conclusion is based on an assumption of homogeneity in effect sizes between the groups, supported by our primary analysis. Our conclusion is not consistent with some literature, which shows a greater population burden in South Asians based on unweighted genotype scores 62,11233,72ref)e exception of Table 1 at baseline.sing data from the DREAM randomized control trial. Areduced ls was claculated cise . However, while the cumulative genotype score for South Asians was statistically higher than white Caucasians in the referenced study, the magnitude of the difference is small (0.99 points on a scale that varied from 0 to 32)62. Moreover, unweighted scores consider the contribution from SNPs with larger effects to be the same as that from SNPs with small effects, and appear to discriminate less effectively between disease states113–116. It should be noted, though, that the genotype score calculation in the referenced paper is much more direct than ours as it is based on primary data.

### 2.4.5 Strengths and limitations

This is the first meta-analysis that compares genetic risk of type 2 diabetes in South Asians and white Caucasians. Potential limitations of our study include use of unadjusted allelic OR, which precluded us from accounting for SNP–type 2 diabetes associations altered by adiposity. Use of unadjusted allelic OR depended largely on the unavailability of appropriately adjusted data. However, we informally compared adjusted estimates reported in the study with our unadjusted allelic OR and found the adjustment to make little difference; for example, the OR for *GCKR* rs780094 in the primary report112 was 0.87 (0.72-1.05) after adjustment for age and sex. In our study, the OR is 0.86 (0.65-1.13). Furthermore, by using allelic OR, we assume an additive model of inheritance, which may not apply to some loci. Second, publication and time-lag biases may exist whereby significant results are published more often than negative studies. To minimise these biases, we conducted a thorough search of literature and consulted experts to identify as many eligible studies as possible. We were unable to produce funnel plots to formally assess publication bias because fewer than ten studies were used to meta-analyse a majority of SNPs. Third, some SNPs in this meta-analysis were only investigated in two cohorts and their summary estimates should be interpreted with caution. Fourth, sample sizes varied with each SNP in South Asians but were consistent for all SNPs in white Caucasians, which may create artificial differences between the groups as a result of differences in precision. Lastly, meta-analyses of candidate gene studies face the same limitations as individual such studies, including improper assessment of population structure and inability to apply stringent multiple testing control to circumvent false-discoveries, among others117–119. Our conclusions regarding the genetic heterogeneity between South Asians and white Caucasians are applicable only to genetic risk from bi-allelic common SNPs as our investigation was limited to this type of variant.

### 2.4.6 Future directions

The question of why South Asians have a greater risk for type 2 diabetes remains unanswered from a genetic perspective as risk from common SNPs predisposing to type 2 diabetes does not differ. Our results are particularly important as they emphasise the need for future research to explore ethnic-specific heritable epigenetic changes, epistasis and gene–environment interactions, homozygocity mapping in consanguineous pedigrees, as well as exome-sequencing to answer this question14,15, rather than focusing on exploring differences in risk estimates and RAFs. Additionally, differences in genetic risk may result from low-frequency causal variants with large effects (rare variants)120,121. In fact, there has been recent evidence supporting the existence of rare ethnic specific exonic variants associated with type 2 diabetes122–124 The existence of such rare variants can be investigated through candidate gene, whole-genome or whole-exome sequencing. In fact, discovery of rare variants for Crohn’s disease greatly supports this paradigm125. Large-scale exon resequencing around an *MTNR1B* SNP modestly associated with type 2 diabetes (OR between 1.10 and 1.15) identified four rare variants that led to loss of melatonin binding and signalling capacity (OR 5.67, 95% CI 2.17-14.82)126. Using a family-based approach with multiple affected members or studying founder populations can facilitate the identification of rare variants in sequencing studies. Last, systematic reviews and primary studies comparing dietary intake and physical activity patterns between South Asians and white Caucasians have noted considerable variation127,128, which may also contribute to the differences observed between the two ethnic groups.

## 2.5 Conclusions

Similar effect sizes for SNPs predisposing to type 2 diabetes are apparent among South Asians and white Caucasians, but there is variation in RAFs. Additionally, some novel SNPs are present in South Asians. Given the current literature, there is no strong evidence to indicate that currently known genetic variants explain the higher risk of type 2 diabetes in South Asians compared with white Caucasians.

**Chapter 3**

# Dysglycemia from genetic pancreatic beta-cell failure is worsened by abdominal obesity

## 3.1 Background

In healthy individuals, glucose tolerance is maintained by insulin secreted from pancreatic beta-cells in response to physiological demands. Blood glucose levels increase when beta-cells fail to respond to these demands or as a result of ineffective action of the secreted insulin2,129,130. The resulting elevation in glucose and eventually lipids in turn compromise the beta-cells’ ability to compensate by inhibiting glucose-induced insulin secretion, reducing insulin gene expression, and/or inducing apoptosis of islet cells, a phenomenon known as ‘glucolipotoxicity’131.

Among South Asians, longitudinal studies of glycemic traits show steeper age-related increases in fasting glucose with lower age-related increases in HOMA2-%B32, suggesting a comparatively lower beta-cell output than white Caucasians. Yet, the evidence for a greater genetic propensity for beta-cell impairment in this group is inconsistent104,132. Conversely, presence of diminished sensitivity to insulin, after accounting for differences in BMI, among South Asians is well characterized31,37,133. This variation in insulin response may result from South Asians’ tendency to store fat in ectopic regions and the liver39,134,135. In this context, we hypothesize that the higher risk of dysglycemia observed in South Asians results from a compounded effect of genetic beta-cell impairment in the presence of excess abdominal obesity39,136.

This hypothesis was tested using data from an international prospective cohort study62. A genotype score of SNPs involved in beta-cell impairment was created and used to investigate: (1) ethnic differences in the effect of beta-cell impairment on glycemic traits, and (2) whether this genetic risk for dysglycemia is compounded by the presence of abdominal obesity.

## 3.2 Methods

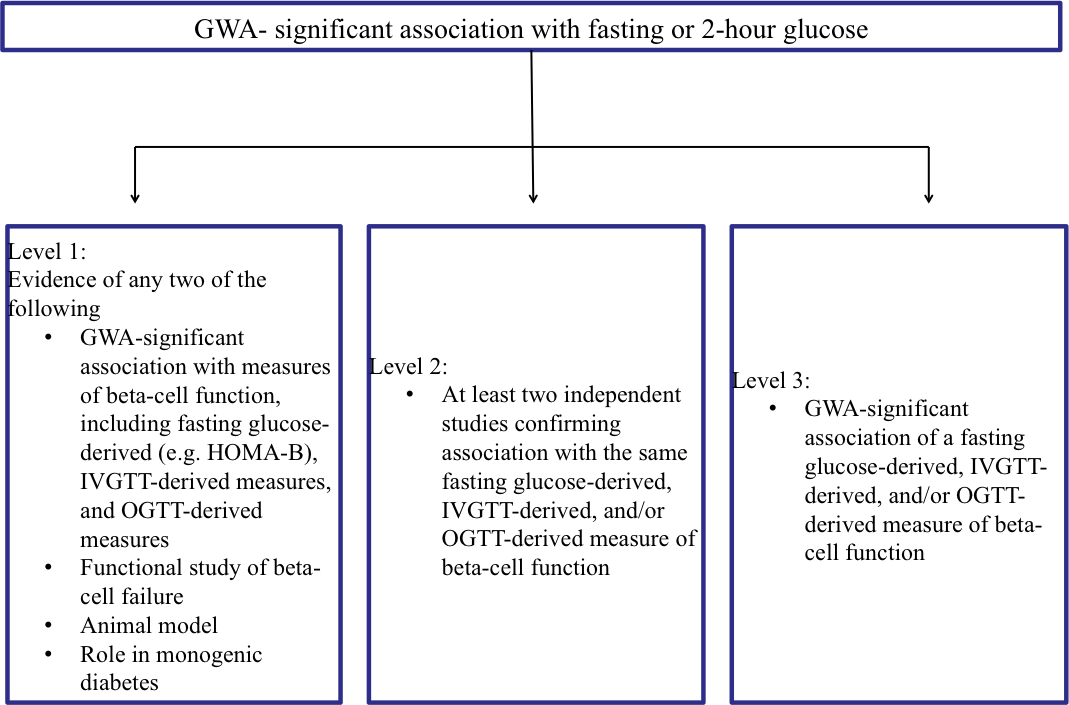
### 3.2.1 Participants

Details of the EpiDREAM cohort and primary results have been reported previously62. Briefly, 25,063 people from 191 centres in 21 countries were screened and 5,269 were included in the DREAM randomized control trial, which assessed the efficacy of rosiglitazone in reducing progression to diabetes from impaired fasting glucose (IFG) and / or IGT137. Individuals ineligible for the DREAM trial or those who declined to participate were enrolled in the EpiDREAM cohort. Inclusion in the cohort was determined using family history, ethnicity, and abdominal obesity. Participants were followed for a median of 3.3 years. DNA was obtained at baseline. Buffy coat DNA extractions and genotyping was successful for 19,197 subjects; 711 of which were excluded because of genotyping failure (defined as <97% of genotypes called), leaving 18,486 eligible for genetic study. For the present analysis, only baseline biochemical and anthropometric measures of 2,764 participants of South Asian descent and 9,408 white Caucasians of European origin were used.

### 3.2.2 SNP selection and genotype score

Evidence from published literature was used to identify genes robustly implicated in beta-cell impairment. Specifically, GWA-studies, functional, and animal studies were used to identify genes associated with beta-cell function. Details of the evidence considered are presented in Figure 3.1. SNPs on genes that had conflicting evidence for beta cell impairment, such as genes shown to be associated with both beta-cell impairment and insulin sensitivity were excluded. Based on these criteria, 29 SNPs on 25 genes were identified. Data for 11 of these SNPs were available in the EpiDREAM cohort in both South Asians and Caucasians, and were used to create a beta-cell impairment genotype score. For each SNP, the glucose-elevating allele was used as the risk allele. An unweighted genotype score was created in accordance with recommendations by Janssens *et al*138 and since weights from large GWA-studies are not readily available for South Asians. Genotypes were coded as 0, 1, and 2, designating copies of the risk allele and the score was calculated for each individual by allele counting. The score could range from 0 to 22. A higher genotype score designated greater beta-cell impairment.

**Figure 3.1** Criteria for selection and inclusion of SNPs implicated in beta-cell function



### 3.2.3 Genotyping and sample quality control

Genotyping was performed at McGill University and the Genome Quebec Innovation Centre using the 50 K Illumina CVD bead chip microarray ITMAT Broad Care (IBC) array139. This array includes 49,234 genetic markers enriched in SNPs of candidate genes and pathways for cardiovascular, inflammatory and metabolic phenotypes. Design of this array was led by Institute of Translational Medicine and Therapeutics (ITMAT), the Broad Institute, and the National Heart Lung and Blood Institute (NHLBI). Further details of the design and conceptualization are presented by Keating *et al*139. Standard assessments were conducted to ensure quality of genotyping. All SNPs included in our analysis (described above) were in agreement with Hardy Weinberg Equilibrium (*P<*1.00x10-3), had a minor allele frequency ≥ 5%, and displayed call rates greater than 95% (Supplementary Table 3.1). Individuals with more than 5% missing genotypes were excluded. Missing values for the remaining individuals in the genotype score were imputed using the arithmetic average of the coded genotypes. After quality control, data from 2,651 South Asians and 9,408 Caucasians were available for analysis.

### 3.2.4 Population structure and inbreeding

Presence of population stratification and inbreeding was assessed using array-wide genotyping data for South Asians and white Caucasians from the 50K IBC array. Principle components analysis (PCA) with 20 axes of variation was conducted to assess population stratification within South Asian and white Caucasian groups140. Secondly, coefficients of inbreeding were calculated based on observed and expected number of homozygous genotypes. The coefficient of inbreeding (F) measures the probability that two genes at any locus are identical by descent from the common ancestor(s)141 – the degree to which two alleles are more likely to be homozygous in an individual because the parents are related. As F is a relative measure, it estimates the increase in homozygosity from the base population. In the present study, F was estimated using a sub-set of SNPs that were pruned to be in approximate linkage equilibrium (18,424 SNPs for South Asians and 21,849 SNPs for white Caucasians). Analyses were conducted using PLINK (version 1.9)142 and Genome-wide Complex Trait Analysis (http://cnsgenomics.com/software/gcta/) version 1.02143.

### 3.2.5 Matching

The baseline characteristics of the eligible South Asians (n=2,651) and white Caucasians (n=9,408) from the EpiDREAM cohort varied substantially in age and male to female ratio. Specifically, the South Asian cohort was 10 years younger (Mean age South Asians: 44.97 (SD=9.38); Mean age white Caucasians: 54.98 (SD=10.82)) and had a lower proportion of females (48.4% compared to 60.8%, respectively). To avoid bias from differences in these baseline characteristics, the ethnic groups were matched 1:1 by age (±5 years) and sex144. As a result, 2,651 South Asians were matched to 2,651 white Caucasians and included in this analysis.

### 3.2.6 Outcomes

***Main outcomes*:** The association with beta-cell genotype score was tested with two glycemic traits: i) the area under the curve (AUC), which uses the FPG and 2 hour post 75 gram OGTT load glucose values (2hrGlu), and is calculated as 145: AUC = (2-0)\*{[FPG + 2hGlu]/2}, and ii) FPG level, measured after an overnight fast.

***Beta-cell impairment:*** While the primary purpose of this paper was to assess the effect of the beta-cell genotype score on glycemic traits, an oral disposition index (ODI) was also used to quantify the effect of the genotype score on beta-cell function. This analysis was performed in a subset of 701 individuals of both South Asian and white Caucasian ethnicities in whom the ODI was available146. This subset of individuals was selected for a DREAM sub-study assessing the effect of rosiglitazone and ramipril on beta-cell function. The participants in this subset were from Canadian study centers and had OGTTs at baseline, after 2 years, and at the end of the study, with blood samples drawn fasting as well as 30 and 120 min after the glucose challenge146. The ODI uses fasting and post glucose load glucose and insulin values at baseline and 30 minutes, and represents the relationship between insulin sensitivity and action. The ODI has been shown to correlate with pancreatic beta-cell impairment; a low ODI indicates poor beta-cell function147. The ODI is calculated as {ΔI0-30/ΔG0-30} x {1/fasting insulin}, where ΔI0-30/ΔG0-30 is the ratio of the change in insulin to the change in glucose from 0 to 30 min.

**3.2.7 Abdominal obesity**

The primary measure of abdominal obesity utilized in this analysis was the waist circumference adjusted for hip circumference (WaistadjHip) since this has been shown to most robustly predict type 2 diabetes and cardiovascular disease in the EpiDREAM cohort and other studies148–150. WaistadjHip was ascertained using residuals from a regression of waist circumference on hip circumference.

### 3.2.8 Statistical analyses

Analyses were performed using PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/) (version 1.9)142 and R (version 3.0.2). The total genotype score was compared between the ethnic groups using an independent sample’s t-test. The association between the genotype score and glucose traits was evaluated using linear regression, adjusted for age, sex, ethnicity, and BMI. An ethnicity x genotype score interaction term was included to assess the presence of ethnic heterogeneity in the effect of the genotype score on glucose traits. To investigate if beta-cell impairment varies by levels of abdominal obesity in the ethnic groups, a three-way interaction term between the beta-cell genotype score, measure of abdominal obesity, and ethnicity was included in the regression analyses. The analysis was then stratified by ethnicity to quantify the differences in the interaction of genotype score with abdominal obesity comparing South Asians and white Caucasians. Lastly, a linear regression analysis was performed between the genotype score and ODI, adjusted for age, sex, and ethnicity, to estimate the effect of the genotype score on pancreatic beta-cell dysfunction. A *P* value of ≤ 0.05 was considered statistically significant.

## 3.3 Results

As shown in Table 3.1, after matching South Asians and White Caucasians for age and sex, difference in some demographic and physical characteristics were identified. Briefly, South Asians had comparable abdominal obesity (WHR) at a lower BMI as white Caucasians. The proportion of individuals with type 2 diabetes was greater in South Asians, and there was less IFG. No difference in the proportion of individuals with IGT between the two groups. Histograms of AUC glucose and fasting glucose are provided in Supplementary Figure 3.1.

### 3.3.1 Association of beta-cell genotype score with glucose traits

The average genotype score was lower in South Asians (mean=11.65, SD=2.02) compared to white Caucasians (mean=12.22 (SD=2.13) (*P* for mean difference < 0.001). However, the median genotype score was the same (median = 12.00), indicating that a comparable burden of these 11 beta-cell impairment SNPs likely exists among the ethnic groups.

Although the absolute values of the beta-cell genotype score did not substantially differ between South Asians and white Caucasians, its effect on AUC glucose and fasting glucose significantly varied. The per-allele increase in AUC glucose was 0.23 mmol/L in South Asians, compared to only 0.08 mmol/L in white Caucasians (*P* for ethnic heterogeneity = 1.79x10-2). Similarly, the per-allele increase in fasting glucose was 0.07 mmol/L in South Asians compared to 0.02 mmol/L in white Caucasians (*P* ethnic heterogeneity = 6.29x10-3) (Table 3.2). Furthermore, the genotype score was associated with score predicted a decrease in beta-cell function, measured using an ODI; ODI decreased by 0.01 pM-1 for each risk allele (*P*=4.80x10-2).

**Table 3.1** Summary characteristics of included participants from the EpiDREAM cohort

|  |  |  |  |
| --- | --- | --- | --- |
|  | **South Asians (n=2,651)**  **Mean (SD) / %** | **White Caucasians (n=2,651)**  **Mean (SD) / %** | **P ethnic heterogeneity** |
| Age (years) | 44.97 (9.38) | 45.49 (8.96) | 3.72x10-2 |
| % women | 51.57 | 47.19 | 1.44x10-3 |
| Waist adjusted for hip | -0.32 | 0.44 | 1.61x10-2 |
| WHR | 0.89 (0.09) | 0.89 (0.09) | 3.15x10-1 |
| BMI (in kg/m2) | 26.46 (4.35) | 30.71 (6.05) | 2.82x10-174 |
| Fasting glucose (mmol/l) | 5.39 (1.87) | 5.56 (1.17) | 7.10x10-5 |
| AUC glucose (mmol/l) | 13.31 (5.82) | 12.68 (3.80) | 3.14x10-6 |
| % IGT | 20.60 | 22.18 | 1.60x10-1 |
| % IFG | 6.37 | 13.47 | 4.50x10-18 |
| % Baseline type 2 diabetes | 16.03 | 11.66 | 3.91x10-6 |

**Table 3.2** Effect of the genotype score on glucose traits stratified by ethnicity

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Trait** |  | **South Asians** | **White Caucasians** | **P ethnic heterogeneity** |
| AUC glucose | β-coefficient (SE)\* | 0.23 (0.05) | 0.08 (0.03) | 1.79x10-2 |
| *P* | *2.38x10-5* | *2.32x10-2* |
| Fasting glucose | β-coefficient (SE)\* | 0.07 (0.02) | 0.02 (0.01) | 6.29x10-3 |
| *P* | *6.17x10-5* | *8.08x10-2* |

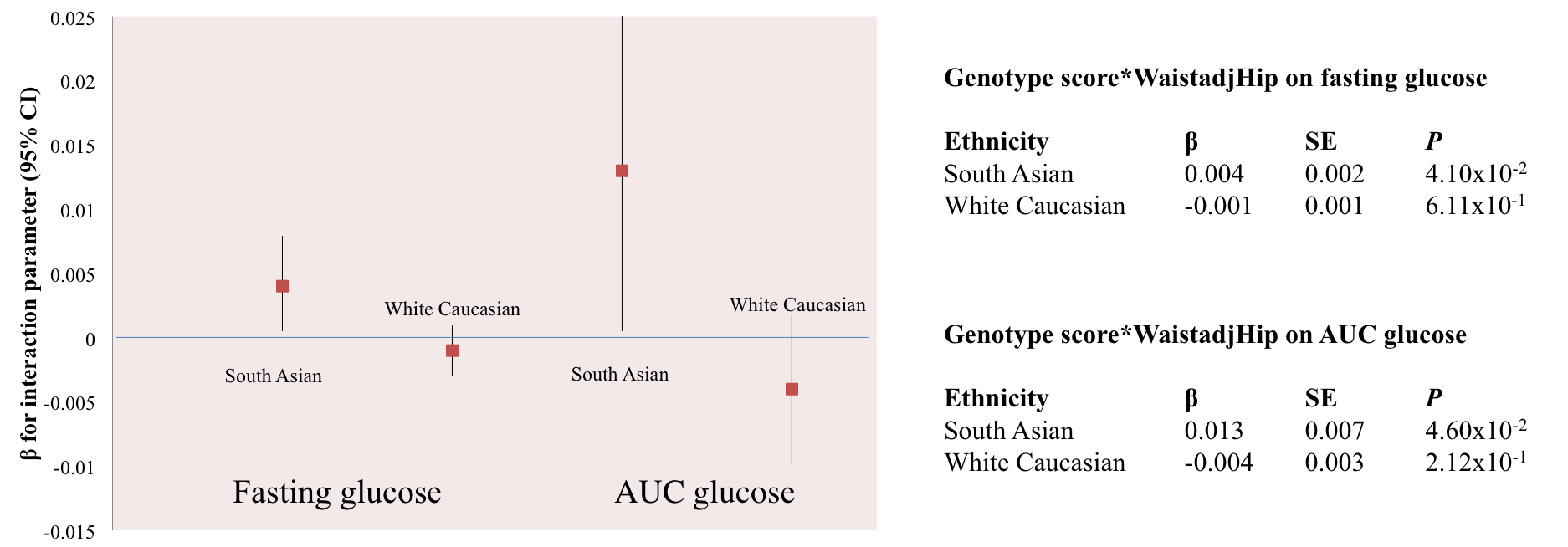
\* β-coefficients are adjusted for age, sex, and BMI

### 3.3.2 Ethnic heterogeneity in interaction of beta-cell genotype score with abdominal obesity

To assess the hypothesis that ethnic difference in the effect of the genotype score on glucose traits may be influenced by abdominal adiposity, a three-way interaction of abdominal obesity measured by WaistadjHip, beta-cell genotype score, and ethnicity was tested. The three-way interaction was significant for both AUC (*P*=1.51x10-2) and fasting glucose (*P*=2.55x10-2).

When stratified by ethnicity, the interaction between genotype score and WaistadjHip was significant in South Asians, but not in white Caucasians (Figure 3.2), indicating that the genetic predisposition to beta-cell impairment is magnified by increasing abdominal obesity in South Asians. The same trend was observed for the interaction between waist circumference and the genotype score (data not shown).

**Figure 3.2** Interaction between beta-cell genotype score and abdominal obesity stratified by ethnic group



### 3.3.3 Population structure and inbreeding

Analysis of population stratification using PCA revealed no evidence of clustering within white Caucasians or South Asians (Supplementary Figure 3.2). An analysis of inbreeding was also performed in both ethnic groups. Coefficients of inbreeding (F) were calculated to estimate the degree of pedigree collapse within an individual’s genealogy. On average, F was 1.81% for South Asians and 0.57% for white Caucasians (Supplementary Figure 3.3), suggesting a greater degree of inbreeding among South Asians.

## 3.4 Discussion

In an at risk population for dysglycemia, the genetic load of beta-cell impairment between South Asian and white Caucasians was similar, but the genotype score was more strongly associated with glucose traits among South Asians than white Caucasians. Additionally the per-allele increase in dysglycemia was magnified in South Asians with greater abdominal obesity, which was not observed in white Caucasians.

### 3.4.1 Association of beta-cell genotype score with glucose traits

A genotype score of 11 beta-cell impairment SNPs was similar between the two ethnicities (an absolute difference of 0.57 on a scale of 0 to 22), suggesting that a comparable genetic load for impaired pancreatic function exists. However, the association of the per-allele effect size on glucose traits was greater in South Asians. For example, the per-allele increase for AUC glucose in South Asians was 0.23 mmol/L compared to 0.08 mmol/L in white Caucasians. Such a heterogeneous effect in the genetic association of glucose traits between South Asians and white Caucasians has not been reported previously81,104,132,151–153. Four of the published studies81,151–153 tested SNPs derived from white Caucasian populations on glucose and type 2 diabetes in various South Asian groups, including Pakistani, Sri Lankan, and South Indians. These SNP effects were then compared with literature-based estimates from white Caucasians. In a study from the UKADS/DGP cohorts, the South Asian estimate of glucose SNPs in 1,163 individuals were found to be similar to white Caucasians from the MAGIC consortium (n= 76,558)151. Similarly, a study of type 2 diabetes genes compared 3,262 South Asians, again from the UKADS/DGP cohorts, to literature estimates from white Caucasian GWA-studies81, and reported homogeneity. As these were indirect comparisons with literature estimates, differences in ascertainment carry the potential to bias comparisons. However, an additional study104 performed a direct comparison of type 2 diabetes SNPs among white Caucasians and South Asians and again found similar effect sizes. There are a few key differences between our study and those previously conducted, which may explain why our results differ. Firstly, the EpiDREAM cohort included white Caucasians at a high-risk for developing diabetes, and thus differs from population-based cohorts included in GWA-studies and consortiums. Though some of these cohorts also investigate individuals at a high-risk for dysglycemia, a majority sample was from the general population or focused on other disease pathologies154. Second, the South Asians in our study are younger in age but with comparable BMI and baseline FPG than those from UKADS/DGP cohorts, and as such it is possible that the effect of a genetic load on glucose traits in this group varies from that reported in previous studies. Overall, among the high-risk individuals from the EpiDREAM cohort, genetics may play a greater role in South Asians than in white Caucasians, explaining why a greater effect on glucose traits is seen for this ethnic group. Third, unlike previous studies, we undertook a direct comparison of a genotype score, representing cumulative genetic burden, between South Asians and Caucasians.

The genotype score is comprised of 11 SNPs that vary in their effect on glucose traits. The contribution of each SNP, as identified by published GWA-studies, is presented in Supplementary Table 3.2. Based on these estimates, it appears that among the 11 SNPs included, the two *MTNR1B* SNPs rs10830963 and rs2166706 and *GCK* rs4607517 has the biggest impact on fasting glucose. The genotypic effects of each SNP on glucose traits in both ethnicities are presented in Supplementary Table 3.3. However, it should be noted that the present analysis was underpowered; specifically, to detect an effect equivalent to the literature for *GCK* rs4607517, we had 74% power with a Bonferroni corrected type 1 error rate of 5%.

### 3.4.2 Ethnic heterogeneity in interaction of beta-cell genotype score with abdominal obesity

Given the strong relationship between abdominal obesity and dysglycemia, as well as its high prevalence among South Asians, we assessed if the impact of the beta-cell genotype score on glucose traits was accentuated by the presence of abdominal obesity, and whether this varied by ethnicity. Since a three-way interaction between ethnicity, abdominal obesity, and the genotype score was significant for both glucose traits, the interaction of abdominal obesity with genotype score was investigated separately for South Asians and white Caucasians. The findings showed a positive interaction of abdominal obesity (measured by WaistadjHip) with beta-cell genotype score on AUC glucose and fasting glucose in South Asians, but not in white Caucasians. This suggests that South Asians experience a greater impact on glucose levels of the genetic load of beta cell impairment as they become more abdominally obese.

A possible explanation for this observation may be that South Asians experience greater functional beta-cell impairment from glucolipotoxicity131,155,156. Comparison of anthropometric profiles, MRI measurement of abdominal fat, and MR spectroscopy of the liver indicate that when matched by BMI, South Asians have more deep superficial subcutaneous and visceral adipose tissue, and more ectopic fat deposition in the liver39,134,135 than do white Caucasians. These differences are associated with greater dyslipidemia, lower adiponectin, and lowered insulin sensitivity in South Asians39,157. *In vivo* and *in vitro* studies have demonstrated that excess fatty acids in the presence of hyperglycemia causes glucolipotoxicity by impairing gene expression158, inducing beta-cell death156, and reducing insulin secretion from existing islet cells159. Because white Caucasians can expand fat stores in superficial and deep subcutaneous regions, they may be less susceptible to glucolipotoxicity155 than South Asians. Congruent with this, a mechanistic study comparing human urinary 2,3-dinor-8- iso-prostaglandin-F1α and plasma Intercellular Adhesion Molecule 1 (ICAM-1), measures of oxidative stress, in South Asians and Caucasians found that the former are more susceptible to the detrimental effects of oxidative stress on pancreatic beta-cells, a contributing factor to glucolipotoxicity, from hyperglycaemia at lower glucose thresholds155.

Moreover, using whole genome sequencing Chambers *et al* recentlydemonstrated divergence in risk allele frequency in South Asians for genes involved in metabolism and energy storage compared to white Caucasians. This whole genome analysis provides support for variation in the metabolic process among ethnic groups108 and also supports the possibility of novel, yet unexplored genetic variants that contribute to the divergent risk for hyperglycemia and the metabolic syndrome. While not exhaustive, some evidence supports presence of exclusive genetic correlates for components of metabolic syndrome in South Asians160–162. Future investigation of novel SNPs and insertion/deletion polymorphisms involved in storage and metabolism may help explain the divergent metabolic phenotype of South Asians, including variation in deposition of fat as well as adipocyte cell size and type, and may further explain the interaction between abdominal adiposity and beta cell function observed in this high risk ethnic group.

### 3.4.3 Population stratification and inbreeding

Population stratification was assessed using PCA as outlined by Price and colleagues140. Our results suggest that within the South Asian and white Caucasian groups from the EpiDREAM cohort, there is no evidence of population substructure. It should be noted that the 50K IBC array is designed with admixture and ancestry informed markers for only European and American Caucasians as well as African American groups163, as such it may not be equipped to appropriately address the question of substructure among South Asians. However, as the South Asians in this cohort were recruited mainly from one geographical region in South India, there likely is no substructure in this group. Conversely, the white Caucasians from EpiDREAM were recruited from several countries in Europe, North America, and Australia. While some previous literature investigating white Caucasian groups has shown evidence of a North-South gradient164, no substructure was found in our cohort. Secondly, we investigated inbreeding in both ethnicities. Our findings show no evidence of substantial inbreeding in either group, but indicate that South Asians were more inbred than Caucasians. This is congruent with previous research20,21. The coefficient of inbreeding was calculated using pruned SNPs in linkage equilibrium and therefore approximately 20,000 SNPs were used in this estimation. As the procedure performs optimally in larger number of SNPs, preferably greater than 50,000, the estimate for the coefficient of inbreeding may be biased142,165.

### 3.4.4 Strengths and limitations

This study has several strengths. First, the sample was derived from a large multi-ethnic study and contained detailed assessment of dysglycemia and abdominal obesity. Additionally, evidence of genes associated with beta-cell impairment was systematically gathered in order to create a robust genotype score. Some limitations exist; first, as some genes implicated in pancreatic dysfunction were unavailable on the Illumina 50K IBC array, the genotype score was not exhaustive. Secondly, since EpiDREAM is a cohort of high-risk individuals, as determined by family history and abdominal obesity, the estimates ascertained from this analysis may not reflect the general population. Nevertheless, this should not affect our inter-ethnic comparisons as all individuals were recruited using the same criteria, and hence the results are internally valid. Lastly, our study used glucose testing to establish dysglycemia. Many institutions, including American Diabetes Association, the Endocrine Society, and the American Association of Clinical Endocrinologists/American College of Endocrinology, have endorsed the use of HbA1c in screening for and monitoring dysglycemia166–168. These recommendations were made based on overview of factors, such as inconvenience in acquiring fasting and OGTT measures, sample instability (i.e. glucose concentrations decrease in the test tube by 5–7% per hour due to glycolysis169), considerable intra-individual variation as a result of acute stressors (such as not finding parking prior to a visit), and nature of collection (including site of collection)170. Use of HbA1c would overcome many of these limitations that can bias results and as such future investigations should seek to incorporate HbA1c as a measure of dysglycemia. However, researchers and reviewers should be cognizant of reported racial and ethnic differences in A1c concentrations171,172 when incorporating this measure in studies similar to ours.

## 3.5 Conclusions

A beta-cell impairment genotype score is associated with dysglycemia in South Asians and white Caucasians, although the per-allele effect appears to be greater in South Asians. The impact of genetic beta-cell impairment is magnified in South Asians with abdominal obesity.

**Chapter 4**

# Ethnic heterogeneity in DNA Methylation for birth weight and length

## 4.1 Background

DNA methylation is the reversible and heritable attachment of a methyl group to a nucleotide, commonly at the 5’ carbon of cytosine in dinucleotide consisting of cytosine and guanine (CpG). CpG sites are distributed throughout the genome; because methylcytosine spontaneously deaminates to thymine, there is a substantial underrepresentation of CpG sites in the genome, except in densely populated repeat elements / clusters and CpG islands (CGI)173–175. Considerable efforts have been made to identify these clusters; based on a relaxed correlation threshold 307,000 clusters were identified and 24,163 based on strict thresholds176. Most sites within clusters show a similar level of methylation and together may be involved in transcriptional regulation of genes and miRNAs, control of alternative promoter usage, and alternative splicing. Methylation in this manner is considered a mark of long-term inactivation of the associated gene173,174 and ultimately can impact how specific phenotypes or disease states develop.

Considerable variation in body composition35–38 and adipose tissue distribution / storage40 exists between South Asians and white Caucasians, both at birth53–56 and as adults. These differences contribute to cardiometabolic traits such as dyslipidemia and dysglycemia33. As differences in body composition are present at birth, the variation is likely a result of altered genetic predisposition and/or the *in-utero* environment, or their interactions. Since both genetic and environmental factors can independently and together alter methylation of genes, having downstream effects on the expression of genes, it is an intriguing field of study to explain the ethnic heterogeneity between South Asians and white Caucasians.

Despite a strong rationale to explain ethnic heterogeneity, very few studies have explored whether DNA methylation is associated with differences in birth weight and length between ethnic groups177,178. Furthermore, of the work published, both GWA-179–181 and methylation studies182,183 are predominantly performed in white Caucasians. No genome-wide search of SNPs or CpG sites for birth weight has been undertaken in South Asians alone or comparing South Asians and white Caucasians. Since methylation can be important in identifying the phenotypic differences, we investigated whether ethnic heterogeneity between South Asians and white Caucasians exists in the effect of methylation on birth weight and length.

The purpose of this chapter was to: (1) explore ethnic heterogeneity in methylation at known birth weight and length genes; (2) conduct a regional analysis to assess the relationship of the entire CpG cluster with the birth weight / length in genes that were significant from our primary analysis; and (3) perform a genome-wide search of CpG sites related to birth weight in South Asians.

## 4.2 Methods

### 4.2.1 Participants

Data for 250 unrelated South Asian newborns from the South Asian Birth Cohort (START) and 274 unrelated white Caucasian newborns from the Canadian Healthy Infant Longitudinal Development (CHILD) study were acquired. Individuals were considered South Asian if both parents originated from the Indian subcontinent. Newborns were white Caucasian if both parents were whites of European origin. Details of these cohorts have been reported previously63,184 and the specific inclusion and exclusion criteria are presented in Supplementary Table 4.1. Briefly, the START and CHILD cohorts recruited pregnant mothers above the age of 18 during the first or second trimester; the pregnant women, along with their newborns, are prospectively followed for 3 to 5 years. Anthropometric measurements and cord blood are collected for the mother-baby pairs at birth and at each subsequent visit. Informed consent was obtained from all participants. Ethics boards at Hamilton Health Sciences, Trillium Health Partners, and the William Osler Health System approved the START study and ethics boards at St. Joseph’s Healthcare, University of Alberta, Covenant Health, Alberta Health Services, BC Children’s hospital, University of Manitoba, Health Science Centre Winnipeg, St. Boniface General Hospital, Regional Health Authority – Central Manitoba Inc., Manitoba Health Information Privacy Committee, University of British Columbia, Children’s and Women’s Hospital, St. Paul’s Hospital – Providence Health Care, Simon Fraser University, and Vancouver Coastal Health Research Institute approved the CHILD study.

### 4.2.3 Quality control procedures

The following quality control procedures were undertaken prior to conducting the main analysis. Methods describing our main analysis follow (Sections 4.2.4 – 4.2.6).

***Methylation probe site quality control****:* DNA was extracted from cord blood. Data on methylation of DNA were available for 274 CHILD and 250 START participants. Quality control assessments were conducted in R using the ChAMP and minfi bioconductor packages.Raw intensity signals (\*.idat files) were imported from the Infinium Illumina HumanMethylation450 BeadChip. Singular Value Decomposition (SVD) analysis was performed to identify technical variation between the samples (i.e., array, slide, or batch effect). The analysis demonstrated that the primary sources of variation within and between samples were not technical. The following exclusions were made for CpG probe sites: probes on the X- or Y-chromosomes and any probe with a detection p-value of ≤ 0.05185. Based on these criteria, 8,634 (1.82%) and 12,598 (2.66%) of 474,209 probes were removed from the START and CHILD cohorts, respectively, leaving 465,575 probes in START and 461,611 probes in CHILD available for analysis. The following exclusions were made at the participant-level: individuals with discordant methylation-identified and reported phenotypic sex and those with ≥ 0.01 missing probes. Nine samples from START and 16 from CHILD demonstrated discordant sex, and 1 sample from START and 9 samples from CHILD failed probe detection. These samples were removed from the analysis. After quality control, 234 South Asian and 250 white Caucasian newborns were available for methylation analysis. Finally, to correct for differences in distribution and intensity between the Infinium type-I and type-II methylation probes used by the HumanMethylation450k platform, Subset-quantile Within Array Normalization (SWAN) was performed.

***Genotyping quality control****:* Genetic data from newborns were available for 244 START participants and 204 CHILD participants. Genotyping was performed using the Illumina Human Core Exome 12 (version 1.1). Standard assessments were conducted to ensure quality of genotyping. SNPs chosen for the main analysis (see Section 4.2.4 below) were in agreement with Hardy Weinberg Equilibrium (*P*>1.00x10-3), had a minor allele frequency ≥ 5%, and displayed call rates greater than 95% (Supplementary Table 4.2). Presence of inbreeding was assessed using SNPs in approximate linkage equilibrium, automatically selected from genome-wide data (542,585 markers) using LD-based pruning in PLINK142. 175,217 pruned markers were used to calculate coefficients of inbreeding (F) based on observed and expected number of homozygous genotypes. An F-value of >0.2 was used to exclude individuals. No exclusions were made based on this procedure.

### 4.2.4 Ethnic heterogeneity in methylation of known birth weight and length genes

The primary objective was to explore whether ethnic heterogeneity exists in methylation on known birth weight and length genes. This section describes methods used to accomplish this.

***Identification of known birth weight and length genes****:* We searched the literature for candidate genes reported in GWA-studies or at a genome-wide significance level (*P*<5x10-8). Based on this criterion, we identified 12 independent SNPs for birth weight and 18 for birth length (Supplementary Table 4.3). Additionally, though SNPs on *IGF1* and *IGF2* have not been reported in GWA-studies, evidence from previous methylation studies strongly suggests its role in fetal growth186–188. As a result, these two genes were included. Overall, 15 candidate genes for birth weight and 20 genes for birth length were investigated.

***Selection of CpG probe sites within candidate genes***: The Illumina 450k database in R was used to identify CpG probe sites that resided within 100 kilobase pairs of the candidate SNPs. Since no index SNP was available for *IGF1* and *IGF2*, all CpG sites on these genes were included. CpG probe sites with missing data for either study (n=25) and probes previously reported as being sensitive to SNPs (n=51) or cross-reactivity (n=3)189 were excluded. None of the 7 probes available on *ACTBL2* resided within 100 kilobase pairs of the SNP and the only probe available on *DCST2* was polymorphic. Furthermore, no CpG probe sites on *IGF2* were available. As a result, probe sites on these three genes were not tested. Overall, 216 probe sites on birth weight genes and 308 probe sites on birth length genes were available for analysis.

***Statistical analysis of selected CpG sites:*** All analyses were performed using R (version 3.0.2) and PLINK (<http://pngu.mgh.harvard.edu/~purcell/plink/>) (version 1.9)142. The main outcomes of birth weight and length were normally distributed (Supplementary Figure 4.1). Newborn and mother characteristics were compared between the ethnic groups using an independent sample’s t-test. Linear regression analyses were conducted to test the association between percent methylation at CpG sites identified above with birth weight / length175. Due to marked difference in global methylation observed between the two cohorts (Supplementary Table 4.4), all analyses were conducted separately for South Asians and white Caucasians. Heterogeneity in the effect of methylation on birth weight / length between ethnicities was assessed using Cochran’s *Q* statistic based on a fixed effects meta-analysis. To correct for multiple testing, we used the false-discovery rate (FDR) procedure to estimate q values; associations with a q value of 0.05 or less were considered statistically significant185.

For seven sites showing significant ethnic heterogeneity on birth weight from the primary analysis, an investigation of genetic confounding was carried out to assess whether observed associations between birth weight / length and methylation levels at CpG sites were a result of adjacent SNPs being correlated with both methylation and birth weight / length190. An additive model of inheritance for the SNPs was assumed. We report both the unadjusted association between CpG sites and birth weight / length, and an adjusted association for gestational age, newborn sex, and genotype.

Additionally, we assessed whether selected SNPs (or proxies, Supplementary Table 4.5 and 4.6) were associated with birth weight / length in our sample and if ethnic heterogeneity in this association was present. To accomplish this, we tested the association of a genotype score with the respective outcome. Genotype data was available for nine of the 12 chosen birth weight SNPs and nine of the 18 birth length SNPs from both cohorts. A genotype score was also created for both traits. For each SNP, the weight/length-reducing allele was used as the risk allele. Genotypes were coded as 0, 1, and 2, designating copies of the risk allele. The un-weighted genotype score was calculated for each individual by allele counting.

***Exploring sources of heterogeneity:*** The primary analysis revealed ethnic heterogeneity in the effect of methylation on birth weight. To preclude the possibility that this heterogeneity stemmed from technical or other biological variation, we conducted the following investigations. The CHILD offers a unique opportunity to investigate possible sources of heterogeneity as a small number of individuals in this cohort were of South Asian descent. As these individuals were recruited and their samples processed in the same manner as the white Caucasians included in our analysis, we were able to investigate whether observed ethnic differences were true or result from variation between the cohorts. Firstly, as leukocytes and erythrocytes in blood can vary in methylation levels, we investigated whether differences in cell type existed between South Asians and white Caucasians that can manifest as differences in the effect of methylation. Additionally, there were some differences in the blood collection protocols for the START and CHILD cohorts, which may have impacted the relationship between methylation and the outcomes tested. The main protocol variations are highlighted in Supplementary Table 4.7. To this end, we (1) investigated the relationship between methylation and birth weight for both South Asians and white Caucasians in the CHILD cohort as data on a small number of South Asians were available in this cohort; and (2) assessed the effect of processing time on methylation levels in the CHILD cohort, as this was among the key protocol differences between the START and CHILD cohorts.

### 4.2.5 Regional analysis of significant sites

CpG sites can be highly correlated with neighboring sites and may occur within clusters that are co-regulated174. Therefore, for each CpG site that was found to be significant in our primary analysis (Section 4.2.4), we conducted an additional regional analysis whereby the presence of a CpG cluster and its effect on the phenotype was investigated.

***Identification of correlated sites:*** The Illumina 450k database was used to identify known CGI. For sites for where CGI have not been previously identified or existence of a cluster is not known, a correlation with neighboring sites with Pearson’s *r* ≥0.5 was used to establish a CpG cluster173.

***Statistical analysis:***Within each cluster, the percent methylation for all CpG sites was averaged; the average percent methylation was then used as a predictor for birth weight. The analysis was adjusted for gestational age, sex, and genotype. For each cluster, analyses were conducted separately in South Asians and white Caucasians, and heterogeneity in the effect sizes between ethnic groups was assessed.

### 4.2.6 Genome-wide methylation for birth weight in South Asians

Since the primary analysis consistently found CpG sites were associated with birth weight in white Caucasians, but not South Asians, we conducted an agnostic search in the START cohort to investigate whether novel CpG sites were associated with birth weight in South Asians.

***Selection of CpG probe sites:***After quality control, 465,575 CpG probe sites were available from the Ilumina Methylation BeadChip in the START cohort, as described earlier (Section 4.2.3). We further excluded CpG sites known to be polymorphic and/or cross-reactive (n=91,698), leaving 373,877 CpG sites for analysis.

***Statistical analysis:***Linear regression was performed between percent methylation at CpG probe sites and birth weight. The genome-wide analysis was adjusted for gestational age and sex. To correct for multiple comparisons, a genome-wide significance level of 1.3x10-7 was used, determined according to the Bonferroni procedure. Similar to previous genome-wide analyses, a second ‘suggestive’ threshold (*P*<1x10-5) was set191.

## 4.3 Results

As shown in Table 4.1, some differences in baseline physical characteristics between South Asians and white Caucasians were evident. Briefly, South Asian newborns were lower birth weight, had a greater birth length, and a lower gestational age than white Caucasians. Ethnic differences in birth weight and length persisted after adjustment for gestational age and sex (*P* for birth weight = 8.61x10-9; *P* for birth length = 1.59x10-4 after adjustment). Additionally, South Asian mothers were younger, had a lower pre-pregnancy BMI, were less likely to be smokers, but had a higher prevalence of gestational diabetes and hypertension during pregnancy.

**Table 4.1** Descriptive characteristics of included participants stratified by ethnicity

|  |  |  |  |
| --- | --- | --- | --- |
|  | **South Asian**  **(n=234)** | **White Caucasian**  **(n=250)** | **P** |
| **Mean (SD) / %** | **Mean (SD) / %** |
| **Newborns** |  |  |  |
| Birth weight (in kg) | 3.21 (0.51) | 3.48 (0.48) | 7.73x10-9 |
| Birth length (in cm) | 52.22 (2.89) | 51.41 (2.72) | 3.68x10-3 |
| Gestational age (weeks) | 38.96 (1.48) | 39.17 (1.35) | 9.58x10-2 |
| % Female | 50.0% | 44.0% | 1.93x10-1 |
| **Mothers** |  |  |  |
| Age (years) | 29.96 (4.05) | 32.53 (4.54) | 1.30x10-10 |
| Gestational weight gain (in kg)\* | 13.28 (5.97) | 15.14 (5.93) | 3.35x10-3 |
| Pre-pregnancy BMI (in kg/m2)\*\* | 23.87 (4.66) | 24.14 (6.68) | 6.73x10-1 |
| % Gestational diabetes | 17.5%† | 5.20%‡ | 1.31x10-3 |
| % Smokers | 0% | 2.83% | 7.89x10-3 |
| % Hypertension during pregnancy | 4.32% | 1.65% | 8.86x10-2 |

\* Available in n=146 White Caucasians

\*\* Available in n=141 White Caucasians

† Established using self-report and diagnosis using an OGTT

‡ Established using self-report (OGTT data was not available from the CHILD cohort)

### 4.3.1 Ethnic heterogeneity in methylation at known birth weight and length genes

**Birth weight**

We first conducted an unadjusted analysis of methylation at CpG sites (specified in the methods, Section 4.2.4) and birth weight. Thirty-seven out of the 216 CpG sites that were investigated showed nominal evidence of ethnic heterogeneity (*P heterogeneity* <0.05) in the association between methylation and birth weight. After adjusting for multiple testing using the FDR procedure, seven of these CpG sites still showed significant ethnic heterogeneity (Table 4.2). These included CpG sites in *TCF7L2* on chromosome 10, *CALCR* on chromosome 7, *CDKAL1* on chromosome 6, *IGF1* on chromosome 12, and *HMGA2* on chromosome 12. Among white Caucasians, six of these seven sites (*TCF7L2* cg09022607, *CALCR* cg23061150, *IGF1* cg01305421, *HMGA2* cg24776736, *TCF7L2* cg11748187, and *CDKAL1* cg06512263) were associated with birth weight after an FDR correction, while none were significantly associated with birth weight in South Asians.

While increasing methylation demonstrated a positive relationship with birth weight among South Asians, it led to a decrease in birth weight in white Caucasians, indicating the presence of a quantitative interaction. Methylation at *HMGA2* cg24892571 showed the largest change in birth weight among South Asians – a 10 percent increase in methylation led to a 0.6 kg higher birth weight (Table 4.2). In white Caucasians, methylation of cg09022607 on *TCF7L2* had the greatest impact on birth weight; specifically, a decrease in birth weight of 0.4 kg for every 10 percent increase in methylation. Average methylation at all seven sites is summarized in Table 4.3. Notably, white Caucasians were consistently hypo-methylated at all sites compared to South Asians.

**Table 4.2** Association between percent methylation at sites showing ethnic variation for birth weight in cord-blood DNA

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Gene** | **CpG site** | **Base pair position** | **Model** | **South Asian** | | **White Caucasians** | | **P heterogeneity** |
|  | **β-coefficient**† | ***P*** | **β-coefficient**† | ***P*** |
| *TCF7L2* | cg09022607 | 114712695 | Unadjusted | 0.11 | 1.70x10-1 | -0.37 | 1.51x10-5 | 2.72x10-5 |
|  | Adjusted | 0.05 | 4.03x10-1 | -0.21 | 3.46x10-2 | 2.46x10-2 |
| *CALCR* | cg23061150 | 93181353 | Unadjusted | 0.26 | 7.71x10-3 | -0.26 | 9.70x10-4 | 2.82x10-5 |
|  | Adjusted | 0.18 | 2.79x10-2 | -0.23 | 7.41x10-3 | 4.78x10-4 |
| *HMGA2* | cg24892571 | 66284828 | Unadjusted | 0.57 | 9.55x10-4 | -0.20 | 7.87x10-2 | 1.64x10-4 |
|  | Adjusted | 0.45 | 1.56x10-3 | -0.12 | 3.22x10-1 | 2.13x10-3 |
| *HMGA2* | cg24776736 | 66356357 | Unadjusted | 0.08 | 2.40x10-1 | -0.22 | 1.97x10-5 | 3.43x10-4 |
|  | Adjusted | 0.04 | 4.28x10-1 | -0.13 | 2.95x10-2 | 3.34x10-2 |
| *TCF7L2* | cg11748187 | 114713108 | Unadjusted | 0.09 | 1.09x10-1 | -0.16 | 1.36x10-3 | 6.80x10-4 |
|  | Adjusted | 0.03 | 4.47x10-1 | -0.10 | 6.18x10-2 | 5.12x10-2 |
| *IGF1* | cg01305421 | 102874286 | Unadjusted | 0.08 | 2.10x10-1 | -0.23 | 1.68x10-4 | 3.40x10-4 |
|  |  |  | Adjusted\* | 0.01 | 7.63x10-1 | -0.15 | 6.64x10-4 | 2.39x10-2 |
| *CDKAL1* | cg06512263 | 20709867 | Unadjusted | 0.05 | 3.06x10-1 | -0.17 | 6.40x10-4 | 1.31x10-3 |
|  | Adjusted | 0.03 | 4.74x10-1 | -0.06 | 2.92x10-1 | 2.02x10-1 |

\* *IGF1* cg01305421 was only adjusted for gestational age and sex

† β-coefficients are reported as change in birth weight (in kg) for a 10% increase in methylation at the respective CpG probe site

**Table 4.3** Percent methylation at the seven sites showing ethnic heterogeneity in the primary analysis

|  |  |  |  |
| --- | --- | --- | --- |
| **Gene** | **CpG site** | **Average percent methylation (interquartile range)** | |
| **South Asians** | **White Caucasians** |
| *TCF7L2* | cg09022607 | 33% (30% - 36%) | 29% (27% - 31%) |
| *TCF7L2* | cg11748187 | 44% (40% - 48%) | 37% (33% - 41%) |
| *CALCR* | cg23061150 | 86% (84% - 88%) | 82% (79% - 84%) |
| *HMGA2* | cg24892571 | 90% (89% - 91%) | 86% (84% - 88%) |
| *HMGA2* | cg24776736 | 66% (62% - 70%) | 59% (54% - 63%) |
| *CDKAL1* | cg06512263 | 34% (29% - 39%) | 27% (23% - 31%) |
| *IGF1* | cg01305421 | 27% (22% - 31%) | 19% (16% - 22%) |

Of the seven CpG sites showing ethnic heterogeneity, the CpG sites on *TCF7L2* and *HMGA2* were correlated with other sites on those genes. The two CpG sites on *TCF7L2* are neighbouring sites in the body of the gene. Methylation levels at cg11748187 and cg09022607 were highly correlated with each other (Pearson’s *r*=0.767 and 0.782 in the CHILD and START cohorts, respectively). Similarly, on *HMGA2*, cg24892571 and cg24776736 were weakly correlated neighboring probes (Pearson’s *r*=0.386 in the CHILD cohort and 0.420 in the START cohort).

***Confounding and adjustment for covariates***

**Genetic confounding:** Genetic confounding can result if genotype variation at SNPs is correlated with both methylation levels at nearby CpG sites and the phenotype. Conversely, the SNP can also be causal in the association with the phenotype192. To investigate presence of genetic confounding, we tested whether SNPs on *TCF7L2*, *CALCR*, *HMGA2*, and *CDKAL1* were associated with percent methylation at the related CpG site and birth weight. We then examined whether any ethnic heterogeneity in these associations existed (Table 4.4). It should be noted that genetic confounding at *IGF1* cg01305421 was not tested, as a SNP associated with birth weight on this gene has not been reported from GWA-studies.

Of the SNPs tested, rs4132670 and rs10784502 appeared to be associated with methylation levels at nearby CpG sites. Variation in methylation of cg11748187 by genotype of rs4132670 on *TCF7L2* was significant in white Caucasians, but not South Asians (*P heterogeneity*=2.99x10-2). Genetic variation on rs4132670 explained 4.04% of the variation in methylation at cg11748187 in white Caucasians, and as expected, explained none of the variation in South Asians. Additionally, methylation at cg24776736 on *HMGA2* was significantly associated with rs10784502 in both South Asians and white Caucasians. Though this association explained 4.44% of the variation in methylation at cg24776736 in Caucasians and only 1.97% in South Asians, there was no statistically significant ethnic heterogeneity. Given the association of these two SNPs with methylation, we next needed to test whether the SNPs were associated with birth weight to uncover genetic confounding. Similar to the relationship of the SNP with methylation, we found that *TCF7L2* rs4132670 was associated with birth weight in Caucasians, suggesting a possibility for genetic confounding at this site. The SNP was not associated with birth weight in South Asians. *HMGA2* rs10784502, on the other hand, was not associated with birth weight in either ethnicity. An important consideration is that the power of these analyses was quite low (Supplementary Table 4.5).

**Confounding from gestational age and sex**: No substantial differences were evident for percent methylation at any of the seven sites between males and females or by gestational age in South Asians. Conversely, in female Caucasians, percent methylation was significantly lower compared to males at 4 sites (*IGF1* cg01305421, *CDKAL1* cg06512263 and both cg09022607 and cg11748187 on *TCF7L2*). Furthermore, 5 sites also varied by gestational age in white Caucasians; with the percent methylation at *IGF-1* cg01305421, *HMGA2* cg24776736, *TCF7L2* cg11748187 and cg09022607, and *CDKAL1* cg06512263 lower in newborns with a gestational age ≥ 39 weeks.

**Adjustment for confounders:** To account for possible confounding by gestational age, sex, and genotype, the models for each CpG–birth weight pair were recalculated after adjustment in both ethnic groups. After adjustment, significant heterogeneity between ethnic groups persisted at five of the seven sites (*IGF1* cg01305421, *TCF7L2* cg09022607, *CALCR* cg23061150, *HMGA2* cg24776736 and cg24892571). After adjustment for gestational age, sex, and genotype, percent methylation at *TCF7L2* cg11748187, *CDKAL1* cg06512263, and *HMGA2* cg24892571 was no longer significantly associated with birth weight in white Caucasians. Because *HMGA2* cg24892571 remained nominally significant among South Asians, ethnic heterogeneity at this site persisted. For the other two sites, it appears that a combination of gestational age, sex, and genotype explains variation in birth weight and as such there is no ethnic heterogeneity in the effect of methylation after adjustment for these confounders, i.e. methylation status is no longer associated with outcome in either ethnic group.

***Exploring sources of heterogeneity***

**Variation in cell-type composition:** Differences in the composition of leukocytes and erythrocytes by ethnicity was investigated in the CHILD cohort as this data was available for both ethnic groups. Results are presented in Supplementary Table 4.8 and suggest no statistically significant differences between South Asians and white Caucasians.

**Methylation-birth weight relationship in South Asians of the CHILD cohort:** To investigate whether the discordance in direction of effect for methylation on birth weight between South Asians and white Caucasians was a result of between-study differences, we assessed the seven sites in a sample of 18 South Asian newborns available from the CHILD cohort. Our analyses revealed that at five out of the seven sites, an increase in methylation led to a decrease in birth weight, consistent with white Caucasians from CHILD (Supplementary Table 4.9). Though, it should be noted, this was not evident at all CpG sites and was based only on a sample of 18 individuals.

**Effect of processing time on methylation:** Among the most notable protocol differences between the START and CHILD cohorts is the time elapsed between collection and processing of the samples. In the START cohort, this time was less than or equal to 2 hours, while in the CHILD cohort, the time stipulation was 24 hours, with an average of 19.8 hours. To this end, we first investigated whether processing time affected methylation levels. A statistically significant relationship between methylation and processing time was evident for only one CpG site among the seven tested; methylation at *TCF7L2* cg11748187 was associated with processing time with a *P*-value of 4.86x10-5. Methylation at the remaining six sites did not show any relationship with processing time (Supplementary Table 4.10). Secondly, we stratified the CHILD samples by processing time (greater than 2 hours elapsed versus less than or equal to 2 hours) and compared the relationship between methylation at the seven CpG sites and birth weight in the sub-groups. The results are presented in Supplementary Table 4.11. While not statistically significant, methylation at three of the seven sites in the less than 2-hour subgroup showed a positive relationship with birth weight, congruent with results from the START cohort. This lends some support to the speculation that differences in processing time may be responsible for the observed discordance in the direction of effect between South Asians and white Caucasians.

**Birth length**

No CpG sites showed evidence of ethnic heterogeneity after correction for multiple testing using a FDR q value of 0.05. As no sites were significant in our primary analysis, we did not carry out any further analyses (i.e. regional or genome-wide analyses) for this phenotype. Additionally, our genetic analysis revealed that no individual SNPs or the genotype score of birth length SNPs was associated with the trait in either ethnic group (Supplementary Table 4.6). The power for these analyses, however, was very low.

**Table 4.4** Association between methylation level and SNP on birth weight genes

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Gene** | **SNP** ‡ | **Risk allele** | **CpG site** | **% Methylation** | | | **β-coefficient** † | **P** | **P heterogeneity** |
| ***TCF7L2*** | rs4132670 | G | **cg09022607** | **GG** | **GA** | **AA** |  |  |  |
| South Asians | 33.05 | 33.11 | 32.79 | 0.05 | 9.21 x10-1 | 4.74 x10-1 |
| White Caucasians | 28.96 | 29.51 | 29.60 | -0.38 | 3.23 x10-1 |
| **cg11748187** | **GG** | **GA** | **AA** |  |  |  |
| South Asians | 44.01 | 44.20 | 43.07 | -0.19 | 7.73 x10-1 | 2.99 x10-2 \* |
| White Caucasians | 36.37 | 38.42 | 39.84 | 1.83 | 5.92 x10-3 \* |
| ***CALCR*** | rs6968642 | G | **cg23061150** | **GG** | **GA** | **AA** |  |  |  |
| South Asians | 86.19 | 85.96 | 85.52 | 0.31 | 3.29 x10-1 | 1.24 x10-1 |
| White Caucasians | 81.32 | 81.93 | 82.19 | -0.45 | 2.36 x10-1 |
| ***HMGA2*** | rs10784502 | A | **cg24892571** | **GG** | **GA** | **AA** |  |  |  |
| South Asians | 89.52 | 89.67 | 89.80 | 0.14 | 4.97 x10-1 | 6.00 x10-1 |
| White Caucasians | 85.87 | 85.86 | 85.27 | -0.32 | 2.61 x10-1 |
| **cg24776736** | **GG** | **GA** | **AA** |  |  |  |
| South Asians | 62.46 | 65.97 | 66.15 | 1.11 | 3.25 x10-2 \* | 3.86 x10-1 |
| White Caucasians | 56.61 | 60.09 | 60.44 | 1.81 | 3.79 x10-3 \* |
| ***CDKAL1*** | rs9368222 | A | **cg06512263** | **CC** | **CA** | **AA** |  |  |  |
| South Asians | 34.58 | 33.54 | 34.86 | -0.43 | 5.72 x10-1 | 7.33 x10-1 |
| White Caucasians | 28.11 | 27.45 | 26.32 | -0.79 | * 1. x10-1 |

‡ Proxy SNPs are presented here; Lead SNPs are available in Supplementary Table 4.2

† β-coefficients represent percent change in methylation for each risk allele

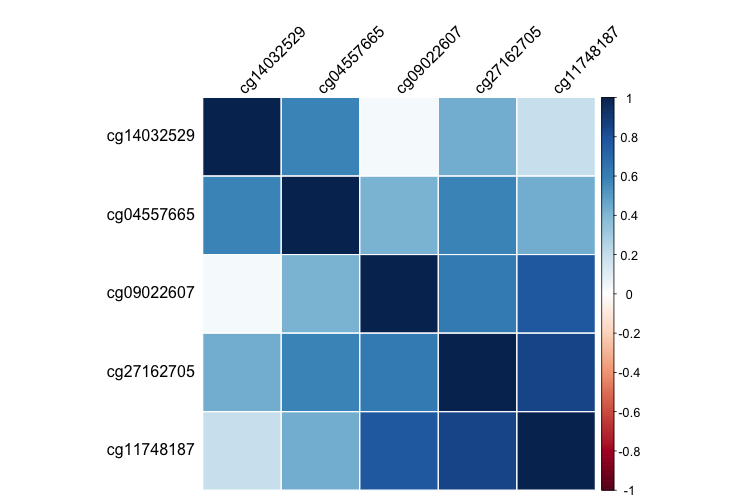
\* *P* < 0.05

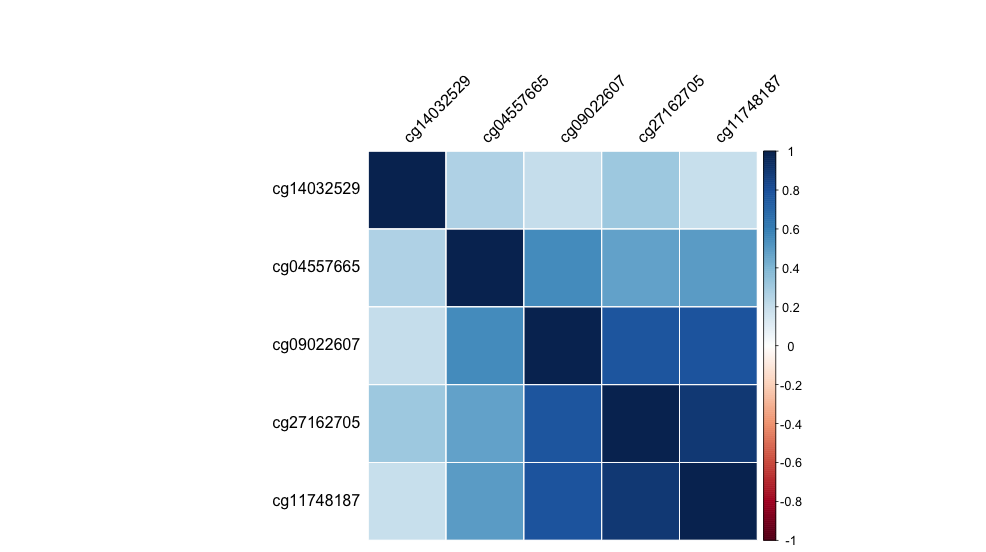
### 4.3.2 Regional analysis of significant sites

As neighboring CpG sites can be correlated and co-regulated, we conducted a regional analysis whereby the effect of CpG clusters (neighboring sites with Pearson’s *r* ≥0.5) on birth weight was investigated in both ethnic groups. Only CpG sites on *TCF7L2* and *CALCR* showed significant correlations with neighboring sites. CpG sites near *IGF1* cg01305421, *HMGA2* cg24776736 and cg24892571 as well as *CDKAL1* cg06512263 were not greatly correlated and therefore regional analyses on these sites were not performed.

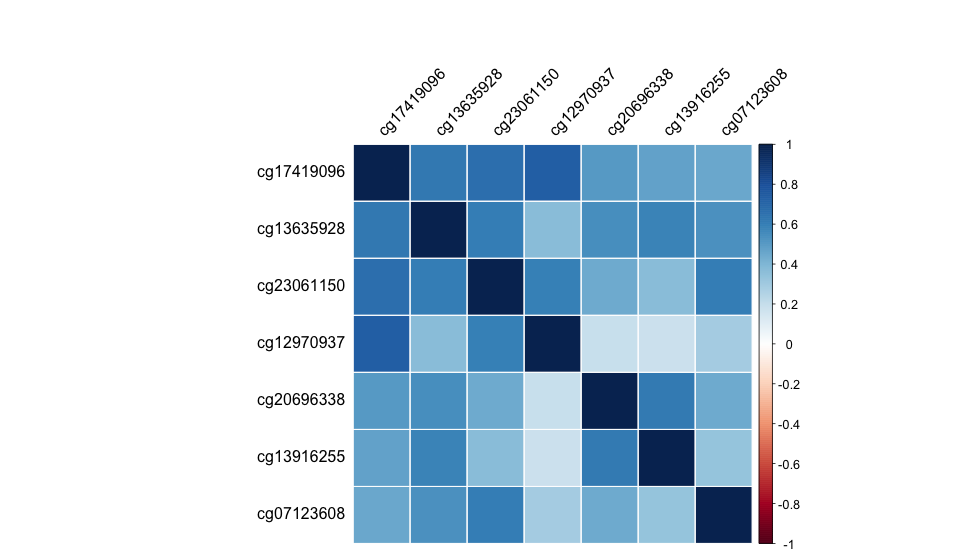
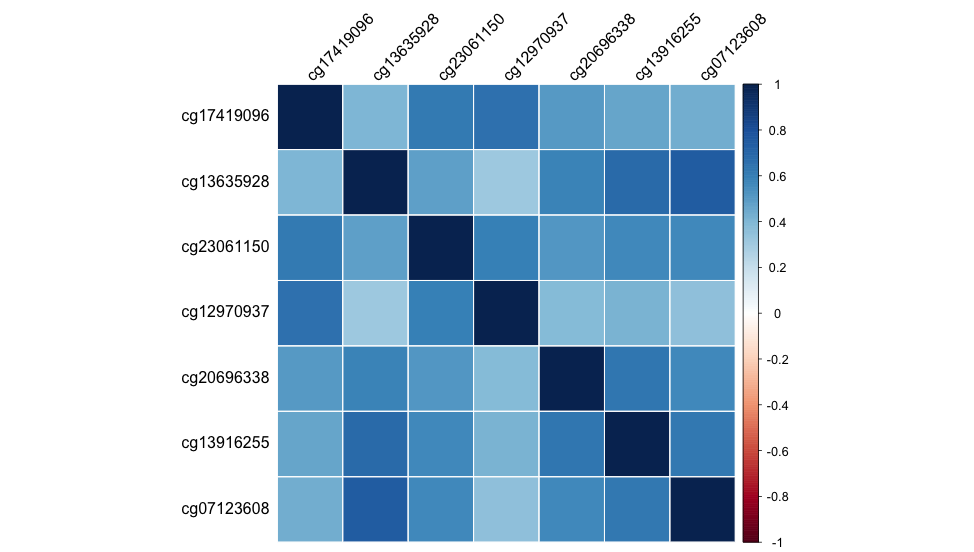
The two probe sites on *TCF7L2* are located on the south shore of a known CGI that spans from bp114712115 and bp114712544 on Chromosome 10. Consequently, percent methylation at sites encompassed in this region was averaged. These included cg04557665, cg09022607, cg27162705, and cg11748187. A 10 percent increase in the average methylation at these 4 sites led to a 0.1 kg increase in birth weight for South Asians, and 0.2 kg decrease for white Caucasians. Four CpG probe sites (cg12970937, cg20696338, cg13916255, and cg07123608) neighboring cg23061150 on *CALCR* had a Pearson’s *r* *≥* 0.5 (Figure 4.1). However, the effect of the average methylation across these four sites was not statistically associated with birth weight in either ethnicity (Table 4.5).

**Figure 4.1** Plots of Pearson’s *r* correlations for CpG sites on *TCF7L2* and *CALCR* in white Caucasians and South Asians

**(a)** CpG sites on *TCF7L2* in white Caucasians (left) and South Asians (right)



**(b)** CpG sites on *CALCR* in white Caucasians (left) and South Asians (right)



**Table 4.5** Regional analysis of significant CpG sites in South Asians and white Caucasians

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Gene** | **South Asians** | | **White Caucasians** | | **P heterogeneity** |
| **β-coefficient \*** | ***P*** | **β-coefficient \*** | ***P*** |
| *TCF7L2* ‡ | 0.06 | 3.68 x10-1 | -0.21 | 2.48x10-2 | 1.92x10-2 |
| *CALCR* † | 0.14 | 8.18x10-2 | -0.22 | 5.31x10-2 | 9.06x10-3 |

\* β-coefficients are reported as change in birth weight (in kg) for a 10% increase in methylation at the respective CpG probe site

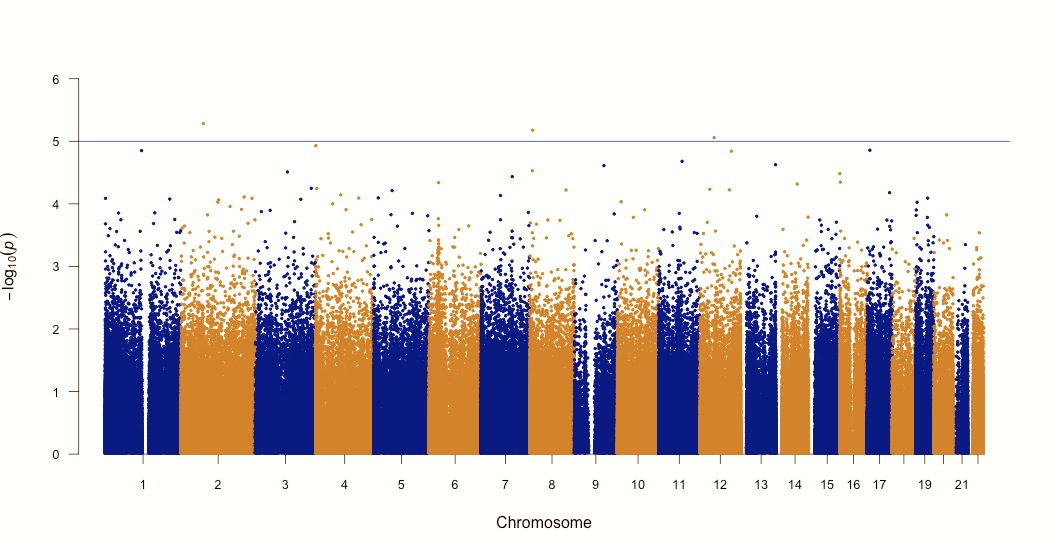
‡ Methylation was averaged over the following sites: cg04557665, cg09022607, cg27162705, and cg11748187

† Methylation was averaged over the following sites: cg12970937, cg20696338, cg13916255, cg07123608, and cg23061150

### 4.3.3 Genome-wide methylation for birth weight in South Asians

As the primary analysis of ethnic heterogeneity revealed that methylation of CpG sites on known birth weight genes were not associated with the phenotype in South Asians, we conducted a genome-wide search in this group. Associations between methylation at CpG probe sites with birth weight were adjusted for gestational age and sex. After Bonferroni correction, none of the sites were significant at the genome-wide threshold of *P<*1.3x10-7, however three sites showed suggestive significance (*P*<1x10-5) (Figure 4.2). These included probe site cg07605236 on *SFXN5* (*P*=5.20x10-6), cg20408693 on *SLC38A2* (*P*=8.76x10-6), and cg06032483 on a gene on chromosome 8 (*P*=6.65x10-6). Average methylation values for cg06032483, cg07605236, and cg20408693 were 90.0%, 91.9%, 6.5% in South Asians, respectively. A Q-Q plot of the *P*-values from the GWA-study of methylation sites was generated (Figure 4.3). We then investigated the association of these sites in white Caucasians; none were associated with birth weight in this group (Table 4.6).

**Figure 4.2** Manhattan plot showing the distribution of *P* values of the association between methylation probes and birth weight in South Asians

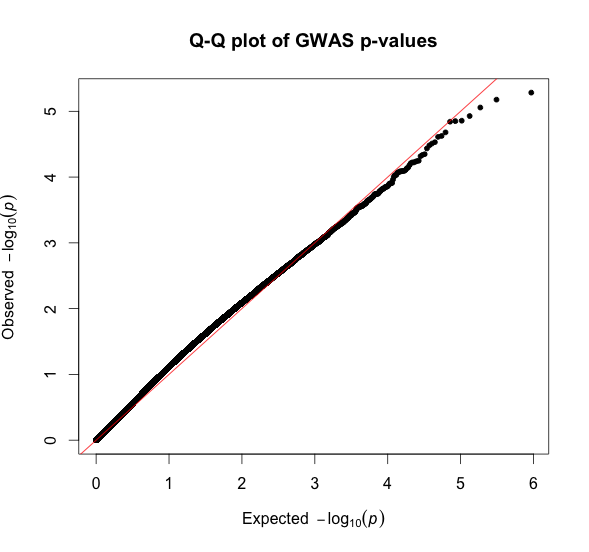


**Table 4.6** Association between percent methylation on CpG sites associated at *P*<1x10-5 with birth weight in South Asians

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Gene** | **CpG site** | **South Asians** | | **White Caucasians** | |
|  |  | **β-coefficient** † | ***P*** | **β-coefficient** † | ***P*** |
| *SFXN5* | cg07605236 | 0.91 | 5.20x10-6 | -0.04 | 7.54x10-1 |
| *NA* | cg06032483 | 0.46 | 6.65x10-6 | -0.04 | 6.18x10-1 |
| *SLC38A2* | cg20408693 | 1.34 | 8.76x10-6 | -0.30 | 2.44x10-1 |

† β-coefficients are reported as change in birth weight (in kg) for a 10% increase in methylation at the respective CpG probe site

**Figure 4.3** Q-Q plot of *P*-values acquired from GWA-studies of CpG sites



## 4.4 Discussion

Our investigation demonstrates significant ethnic heterogeneity in methylation levels of birth weight genes comparing South Asians and white Caucasians. We report seven CpG probe sites on *TCF7L2*, *CALCR*, *IGF1*, *HMGA2*, and *CDKAL1* at which percent methylation is robustly associated with birth weight in white Caucasians, but not in South Asians. We also report three sites associated with birth weight in South Asians at a genome-wide *P* < 1x10-5. These three sites were not associated with birth weight in white Caucasians. None of the sites tested were robustly associated with birth length in either cohort and did not show ethnic heterogeneity. Although some previous studies have assessed methylation profiles for birth weight193–198 and ethnic variation in these profiles177, we are the first to undertake a large-scale comparison of differentially methylated regions for birth weight between South Asians infants and Caucasians.

### 4.4.1 Ethnic heterogeneity in methylation at known birth weight genes

Birth weight is a polygenic trait, influenced by multiple factors including genetic, constitutional, and environmental (e.g. *in-utero*) exposures. These include gestational age, infant sex, parental height, maternal risk factors such as pre-pregnancy weight, nutrition, hypertension, gestational diabetes and exposure to smoking during pregnancy, among others. In our analysis, percent methylation at all seven sites (from our primary analysis) explained 3.2% of the variation in birth weight in white Caucasians after considering a multivariate model including infant’s gestational age, sex, mothers pre-pregnancy BMI, smoking exposure, gestational diabetes and hypertension in pregnancy (Supplementary Table 4.12).

Among the sites showing significant heterogeneity by ethnicity were probe sites on two genes associated with diabetes among other phenotypes, *TCF7L2* and *CDKAL1*. Methylation at these CpG sites influence birth weight in white Caucasians, but this was not observed in South Asians. Association for *TCF7L2* persisted after adjustment for gestational age, sex, and genotype at near-by SNP. *TCF7L2* is a transcription factor involved in the Wnt signaling pathway. The rs7903146 SNP in the *TCF7L2* gene is the hallmark type 2 diabetes variant. Functional studies and animal models suggest that *TCF7L2* is involved in distribution of voltage-gated Ca2+ channels199, which in turn affect exocytosis of insulin from pancreatic beta-cells200. Similarly, while the specific function of *CDKAL1*, a member of a methylthiotransferase family, is not known, it too is implicated in insulin secretion201–203. Both these genes may affect birth weight via *in-utero* reduction of insulin secretion and consequent impaired insulin-mediated fetus growth, as suggested by the fetal insulin hypothesis204,205.

Furthermore, methylation of probe sites on *IGF1*, *HMGA2* and *CALCR* were also associated with birth weight in Caucasians, but not South Asians. *HMGA2* encodes proteins that belong to the non-histone chromosomal high-mobility group. Functional, knockout, and amplification studies suggest a role of these proteins in adipogenesis and mesenchymal differentiation. Subsequently, at a genome-wide level, this gene is robustly associated with various dimensions of growth, including height, infant head circumference, and tooth development. Studies of *HMGA2* elucidating its role in adipocyte206 and muscle development and proliferation207 makes this gene an obvious candidate for implication in birth weight. *CALCR* codes a G-coupled receptor that binds calcitonin and is important in maintaining calcium homeostasis. Although its specific role in birth weight is not clear, it is known to interact with *LRP1*208, involved in energy homeostasis209,210. Lastly, *IGF1* encodes a protein, similar to insulin in structure and function, which plays an integral role in mediating growth and development. In adults, SNPs on *IGF1* are associated with height at a genome-wide level211. Thus, involvement of a CpG site on this gene in birth weight is unsurprising. Notably, SNPs on *IGF1* have also been associated with fasting glucose and insulin in adults9,212.

The fact that there is ethnic variation in the methylation of these seven CpG sites may be suggestive of the multitude of pathways for modulating weight and their relative importance in different ethnic groups. It is possible that while pathways involving *TCF7L2, CDKAL1, CALCR, IGF1,* and *HMGA2* are important in governing birth weight for white Caucasians, they may be less marked in South Asians as a result of differential selection / *in-utero* environmental pressures61,213. This is supported by a lack of association for the above-described CpG sites in South Asians, but a very robust one in white Caucasians. In fact, in addition to methylation levels, the SNPs themselves appear to be more strongly associated with birth weight in white Caucasians. When we tested the association of a genotype score of the 9 birth weight SNPs available in both South Asian and white Caucasians, we found that the score significantly predicted a decrease in birth weight for each risk allele among Caucasians (β-coefficient=-0.042, *P*=1.36x10-2), but was not associated with birth weight in South Asians (β-coefficient= -0.021, *P*=2.05x10-1) (Supplementary Table 4.5).

Our methylation analysis revealed that in addition to a variation in strength of association by ethnicity for the seven sites, the direction of association also varied. Specifically, an increase in methylation led to an increase in birth weight for South Asian newborns, but a decrease for white Caucasians. Though previous literature on ethnic variation in methylation is sparse, one study shows a similar observation. King *et al*. have also reported a variable direction of effect of CpG sites on *IGF2*, *H19*, *MEG3*, and *NNAT* with parameters of gestational growth depending on parental ethnicity178. But, as a parental-effect is also considered, it is difficult to disentangle the role ethnicity plays in the direction of effect. This divergence in the direction of effect by ethnicity may explain the observed ethnic heterogeneity in birth weight, however studies of gene expression are needed to assess whether changes in methylation at these sites translate to changes in the expression of these genes and whether ethnic heterogeneity is present in the level of gene expression.

### 4.4.3 Regional analysis of significant sites

Since CpG sites often occur within clusters and CGIs which together control transcription and gene expression173,174, we investigated the presence of clusters around the seven sites that were significant in our primary analysis. If there was evidence of clusters, we estimated the effect of methylation at the entire element on birth weight. CpG sites on *TCF7L2* and *CALCR* from our primary analysis appeared to be greatly correlated with neighboring sites. Average methylation over the entire element at *TCF7L2* revealed a strengthened effect of methylation on birth weight, adjusted for gestational age, sex, and genotype, in white Caucasians. Conversely, averaging methylation at neighboring sites for cg23061150 on *CALCR* diminished the effect of methylation on birth weight. While correlated, neighboring sites for cg23061150 are likely not part of a CpG cluster as they span a region of 17,108 base pairs. Previous work in mammalian promoters has suggested that these sequences span between 300 and 3,000 base pairs214. Even the closest CpG probe site to cg23061150 constituted a distance of 6,772 base pairs. Therefore, it is either possible that the probe sites available for testing do not cover the region within 3,000 base pairs of cg23061150 or that it is not constituted within a CpG cluster.

### 4.3.4 Genome-wide methylation for birth weight in South Asians

It is important to note that GWA-studies identifying the SNPs/genes investigated in our analysis were mostly performed in white Caucasians. The fact that methylation at these genes was not robustly associated with birth weight in South Asians underscores the possibility that birth weight may be influenced by unique genes in this group. To address this current limitation of knowledge, we conducted a genome-wide search of CpG sites associated with birth weight in South Asians and found three sites possibly unique to this group. Importantly, these sites were not associated with the outcome in white Caucasians. As these sites show suggestive significance (P<10-5), we recommend further testing in a well-powered sample and independent replication if the sites reach genome-wide significance. The Q-Q plot provides evidence that there was likely no inflation of type 1 error215 (Figure 4.2).

The importance of ethnic specific evaluation was recently highlighted by a whole genome sequencing experiment of 168 South Asians and whole-exome sequencing of 147 South Asians, which identified ethnic stratification in genes involved in energy conservation, including *APOH*, *IGF1*, *IGF2*, *LYN*, *LDLR*, *NEUROD1*, *PNPLA2* and *VLDLR*108. As with metabolic traits, this discovery underscores the possibility that an entirely different set of genes may be important in governing birth weight for South Asians compared to those that have been identified as important in white Caucasians. This contention is supported by our findings that methylation of known birth weight genes were only associated with the outcome in white Caucasians and furthermore a genotype score of SNPs on these genes was also associated with birth only in this ethnic group. Large studies of CpG sites and SNPs associated with birth weight in South Asians may uncover genes representing pathways of importance in this group.

### 4.4.5 Strengths and limitations

This study has several key strengths. First of all, as recruitment for the START and CHILD cohorts was undertaken at various cities around Canada, our results are greatly generalizable. Second, as we conducted a stringent control for multiple testing, our results are likely robust and avoid false-positives discoveries. There were some limitations. Firstly, as our primary analysis included 234 South Asian and 250 Caucasian newborns, we were appropriately powered (>80%) to detect a change in birth weight as low as 0.05 kg for a 10% increase in methylation. It is thus possible that effect sizes lower than this were missed. Furthermore, if effect sizes for CpG sites showing heterogeneity were lower than 0.05 in South Asians, we were likely unable to detect them. Additionally, we were substantially underpowered to assess the effect of the tested SNPs with birth weight in our sample. We therefore explored this association with a genotype score. Thirdly, as blood collection protocol differed for the START and CHILD cohorts and the methylation data were processed separately, there exists a possibility of technical differences between studies, which may manifest as ethnic heterogeneity. Our analyses to explore sources of heterogeneity lend support to the possibility of this technical variation. However, these *post-hoc* analyses were comprised of small subgroups (i.e. 18 South Asians from CHILD cohort and 14 white Caucasians with a time of less than 2 hours between time of collection and processing) and thus should be interpreted with caution.

## 4.5 Conclusions

The results from our study support presence of ethnic heterogeneity at seven CpG sites on *TCF7L2, IGF1, CALCR, HMGA2,* and *CDKAL1*. Probe sites on these genes were robustly linked to birth weight in white Caucasians, but not in newborns of South Asian origin. Conversely, a genome-wide search of CpG sites associated with birth weight in South Asians revealed three novel suggestive signals, which were not associated with the outcome in Caucasians. Our results suggest that a unique sub-set of genes may be more important in modulating birth weight among South Asians.

**Chapter 5**

# Discussion

## 5.1 Overview of findings

Numerous studies have investigated reasons for variation in type 2 diabetes prevalence and related risk factors among people of South Asian and white Caucasian origin. Previous work from various geographic regions and age groups in South Asians has shown that this group is at a substantially higher risk for developing diabetes17,18,33,47,127. Literature to date supports differences in insulin sensitivity, compensatory beta-cell function, body fat distribution, and lipid profiles as underlying this increased risk26,53,157. However, it is unclear whether this risk is predominantly due to genetic, epigenetic, environmental / behavioral factors, or a combination of all108,127,128. The specific aim of this thesis was to explore the role of genetic variants and epigenetic changes to explain the greater risk for type 2 diabetes and associated traits in South Asians.

Findings from the meta-analysis of existing data on SNPs associated with type 2 diabetes (Chapter 2) show that no substantial differences in genetic risk likely exist for cosmopolitan SNPs present in both ethnicities. When population burden from all known common variants was examined using a genotype score, again no differences were found132. Our conclusion was supported by additional advancements in literature104. This was not entirely surprising as the SNPs tested were identified from white Caucasian populations and then investigated in South Asians, increasing the likelihood of homogeneity. Therefore, if there is a difference in genetic risk between South Asians and Caucasians, it probably does not result from polymorphisms common to both groups. Consequently, we tested eight SNPs identified from South Asians71–74 in Caucasians. The results revealed that one SNP, *SGCG* rs9552911, was monomorphic in white Caucasians. The remaining SNPs, while polymorphic, were not robustly associated with type 2 diabetes in Caucasians, suggesting either existence of substantially different LD structures such that these seven SNPs tag lower frequency causal variants unique to South Asians, or lower effect sizes in white Caucasians due to ethnic specific gene–environment and/or gene–gene interactions.

Although the meta-analysis in Chapter 2 presented important findings, the question of why South Asians are at a higher risk was not fully answered since the risk from individual common SNPs did not vary between the ethnic groups. Subsequently, Chapter 3 explored if variation in risk arises from an interaction of the genetic risk with abdominal obesity, a notable clinical risk factor among South Asians40. In agreement with the meta-analysis from Chapter 2, this primary analysis in an international cohort revealed that absolute genetic risk, measured as an un-weighted genotype score of SNPs involved in pancreatic beta-cell function did not vary among South Asians and Caucasians. The findings suggested that glycemic control deteriorated more in South Asians compared to white Caucasians even though the genetic risk was similar. In fact, an interaction existed between abdominal obesity and this genotype score in South Asians, such that those individuals with a genetically weak pancreas showed worse glucose control with increasing abdominal obesity. This interaction was not present in Caucasians suggesting that it is not innate differences in beta-cell function by itself, but a combined impact of reduced insulin secretion, from genetic impairments, and the effects of excess fat, from abdominal obesity, that leads to the ethnic variation in the prevalence of dysglycemia.

The conclusions from Chapter 2 and 3 may appear at odds as the genotype score in Chapter 2 showed no ethnic heterogeneity, but the score tested in Chapter 3 had a bigger impact on glucose traits in South Asians compared to white Caucasians. These seemingly incongruent findings can be reconciled with the following considerations: (1) the outcomes tested were different between the studies; while Chapter 2 assessed type 2 diabetes, Chapter 3 explored the impact of a genotype score on fasting and AUC glucose. Previous research has shown a different subset of genetic variants at work for glucose traits versus type 2 diabetes, with only a partial overlap10. (2) the SNPs tested did not overlap; in Chapter 3, we specifically examined SNPs with strong evidence for beta-cell function, while Chapter 2 was a meta-analysis of any reported SNP implicated in type 2 diabetes, regardless of the biological role. In fact, only 2 SNPs, *TCF7L2* rs7903146 and *SLC30A8* rs13266634, were common to both analyses. Lastly, (3) the populations tested varied substantially. A majority of the studies included in the meta-analysis from Chapter 2 recruited cases and controls from the general population (Supplementary Table 2.4), while some recruited cases from hospitals and diabetes clinics. Conversely, the EpiDREAM study solely included participants at high risk for developing diabetes, measured using family history and abdominal obesity, among other markers of metabolic health. Many of the participants had impaired fasting glucose or were impaired glucose intolerant. Consequently, it is possible that in the context of higher risk, where other metabolic abnormalities are already present, a genetic predisposition has a greater impact in South Asians than in the general population.

Since differences in body fat composition among South Asians and Caucasians are present at birth and because low birth weight is strongly linked to type 2 diabetes216, Chapter 4 investigated the role of DNA methylation, an epigenetic mark that incorporates the genetic architecture and *in-utero* environment, in explaining ethnic variation of birth weight. Our findings suggest that genes involved in modifying birth weight in white Caucasians, and consequently the methylation levels at CpG sites on these genes, are less important in South Asians. To this end, our genome-wide search revealed three suggestive methylation sites important in modifying birth weight in South Asians, which were not associated in white Caucasian newborns. As these are suggestive signals, replication is warranted. Overall, our findings indicate that increased risk in metabolic traits among South Asians maybe due to a unique set of genes in this group.

## 5.2 Methodological considerations

Interpretation of research findings is contingent on the conduct of methodological sound studies. Herein, the methodological context for the studies used in this thesis is discussed. In addition to challenges in classical epidemiological designs, such as selection bias and confounding, studies of genetic associations commonly face additional risk of bias, including appropriate sample size and power as well as quality of genotyping217. Details on genotyping quality have been discussed in the methods section of each chapter and will not be further elaborated here. Rather, the following discussion will focus on issues around sample selection, power, and confounding for the data used in this thesis. Subsequently, the next section will discuss the extent to which the findings from this thesis are generalizable and therefore externally valid.

### 5.2.1 Internal validity

**5.2.1.1 Selection bias**

Selection bias results from systematic differences between participants included in the study and those not included; presumably, these differences can bias risk estimates reflecting the relationship between the exposure and outcome218. Case-control and cross-sectional design is the mainstay of genetic association studies. In these types of studies, the chosen control group represents the population without the outcome of interest. Incorrect selection of the comparison group is a likely source of bias217. For example, if very sick cases are compared with healthy population controls, the effect size to estimate the impact of genetic variant on the disease may be inflated, and thus the most likely source of bias can come from choice of comparison groups.

***Meta-analysis:*** Selection bias in the context of a systematic review was examined by considering how representative the included studies were of the source population. All control and a majority of cases used in the meta-analysis were derived from general clinics through community and health promotion fairs. Other cases were recruited from diabetes / endocrinology clinics as well as general clinics in hospitals. Although this lowers the possibility of systematic bias, it cannot be absolutely ruled out since some cases were taken from high-risk diabetes clinics and hospitals, possibly representing very sick individuals with comorbidities. Furthermore, none of the studies compared characteristics of included participants with those refusing to participate making it difficult to holistically examine possible bias.

***The EpiDREAM study:*** As an international prospective design, the possibility of selection bias in the EpiDREAM study is minimized since the sample is representative of the overall population as individuals from many socio-economic, regional, and cultural backgrounds are included. Additionally, the comparison groups were taken from the same source. The study includes a multi-ethnic sample of 24,595 individuals from 191 sites in 21 countries62,149. However, it should be noted that the study aimed to include only those individuals at a high-risk for developing dysglycemia, established using family history, ethnicity, gestational diabetes, and abdominal obesity. As a result, data used to establish the genetic effect of variants on type 2 diabetes and other glycemic measures does not reflect risk in the general population. Since this is a high-risk group, the effect of genes is likely over-represented. Nonetheless, as ours was an ethnic comparison for which both South Asians and Caucasians were recruited using the same study protocol, use of a high-risk group does not bias our results on ethnic variation. Furthermore, to circumvent bias from ascertainment, we matched South Asians and Caucasians on age and sex. Since the match was imperfect, the possibility of bias from selection and ascertainment cannot be completely obviated.

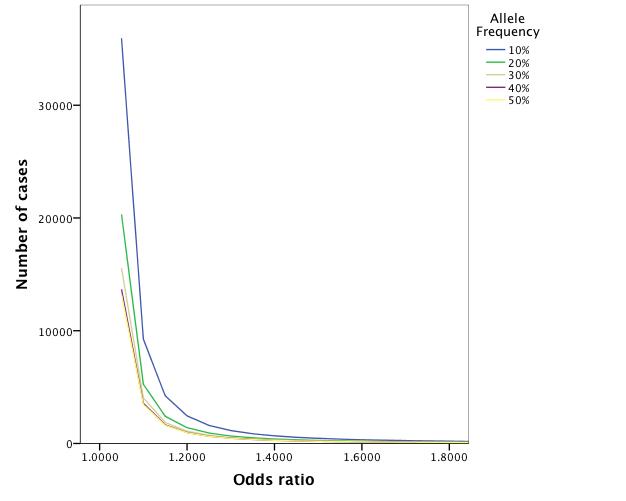
***START and CHILD birth cohorts:*** START is a birth cohort of South Asian mother-baby pairs living in Canada. Mothers in the first trimester of their pregnancy were recruited from three hospitals in the Peel region, where South Asians comprise 18% of the population63. Similarly, CHILD is a population-based cohort of mother-baby dyads across four major Canadian cities184. Though similar in design, small differences in inclusion and exclusion criteria do exist. Furthermore, both studies recruited from hospital settings and as a result missed pregnant mothers using birthing centers and other alternative settings. There may be systematic differences between these groups, including infant outcomes and socioeconomic status, among others219. Additionally, as the study recruitment and follow-up largely took place in urban / metropolitan regions, rural families were inherently excluded or under-represented220. Previous research on our main outcome, birth weight, suggests that possible bias against inclusion from rural regions likely led to an under-representation of children born with extremely low birth weight. Conversely, planned non-hospital births have been associated with healthier babies, usually with higher birth weights. Overall, the mean birth weights for both South Asian and European groups reflected published Canadian averages221, and though the possibility of selection bias cannot be entirely precluded, there is likely little effect of bias from ascertainment in these cohorts.

**5.2.1.2 Sample size and power**

Appropriately powered samples are a key determinant of quality in genetic association studies, as many common variants can have a low relative-risk and rare variants have lower frequencies. Inappropriately powered studies are either unable to detect the effect of a genetic variant on an outcome or report effect sizes for genetic variants that are overestimates217.

***Meta-analysis:*** Our meta-analysis found 15 SNPs associated with type 2 diabetes in South Asians that also showed an effect in Caucasians. Data were available on an additional 9 SNPs that are associated with the outcome in Caucasians, but were not replicated in South Asians. Our main analysis, with a maximum of 25,984 and a minimum of 1,024 cases, was appropriately powered to detect an effect size of 1.05 at an allele frequency as low as 15% and 1.20 at an allele frequency of 30% (Figure 5.1). For the SNPs not associated with type 2 diabetes in South Asians, it is likely that we were not appropriately powered to detect a significant association.

**Figure 5.1** Power calculations for case-control association studies of unrelated individuals. Curves represent the number of individuals, at a case-control ratio of 1:1, required to achieve 80% power for a disease prevalence of 10% at α = 0.05.

****

***The EpiDREAM study***: To assess whether an ethnic interaction existed in the effect of the beta-cell genotype score on glucose traits, we had >90% power (at α=0.025 to account for inclusion of a 2-way interaction term) in our sample of 5,302 individuals to detect an interaction parameter as low as 0.005. To test a three-way interaction (ethnicity\*genotype score\*WaistadjHip), our power decreased to 85% to detect an interaction parameter as low as 0.005 using an α=0.01. Within our sample, we report a two-way interaction parameter of 0.05 on fasting glucose and 0.15 on AUC glucose. Similarly, we found a three-way interaction parameter of 0.005 on fasting glucose and 0.01 on AUC glucose.

***START and CHILD birth cohorts:*** Our analysis investigating the effect of CpG probe sites on birth weight (adjusted for gestational age and sex), separately for white Caucasians and South Asian newborns, had a power of 80% to detect a nominal association for a β-coefficient of at least 0.05 for a 10% increase in methylation. All detected effect sizes, however, exceeded this minimum. We applied a false discovery rate q-value of 0.05 to avoid bias from multiple testing. For our genome-wide analysis, at a Bonferroni corrected α threshold of 1.3x10-7, we had 80% power to detect a β-coefficient of 0.2. Hence, at this genome-wide level, we were underpowered.

**5.2.1.3 Confounding**

Bias from confounding arises when an extraneous variable, correlated with the exposure and outcome, is not considered in the analysis of the exposure-outcome relationship218. Inclusion of this confounder in a statistical model can affect the magnitude, strength, and direction of the exposure-outcome relationship218. Confounding bias can be avoided by appropriate identification and adjustment of potential confounders in the study design and statistical analysis. On the other hand, it should be noted that adjusting for confounders that share heritability with the outcome can bias results of the association test222. Therefore, covariates and confounders should be chosen carefully.

***Meta-analysis:*** Effects for SNPs associated with type 2 diabetes were pooled under an additive model of inheritance. Because of the unavailability of consistently adjusted data, since the included studies adjusted for different combinations of potential confounders, we pooled unadjusted allelic odds ratios. Although age and sex are often adjusted for in genetic case-control studies, they are likely associated with the outcome, i.e. type 2 diabetes, but are unlikely to be related to the genotype. As a result, these variables are not confounders and adjustment can often come at the cost of loss in power223. A possible confounder to consider, however, is BMI as some type 2 diabetes variants can also be associated with BMI; a classic example of this is *FTO*224. Since our analysis did not adjust for BMI, confounding bias cannot be entirely avoided. Particularly, if BMI is more strongly associated with a genetic variant in one ethnic group compared to another, the ethnic comparison may be biased. Often, the most important confounder in genetic studies is population stratification225. As the meta-analysis was restricted to only South Asians, this confounder was less concerning. Although, it is noteworthy that sub-structure within the South Asian group was not addressed.

***The EpiDREAM study:*** The analysis using data from this study examined the variation in association of a beta-cell genotype score and glucose traits for South Asians and Caucasians (using a 2-way interaction term) and a variation in the relationship of adiposity with the genotype score by ethnicity on the glucose traits (using a 3-way interaction term). The South Asian and white Caucasian cohorts were matched on age and sex to avoid ascertainment bias. Furthermore, all analyses were adjusted for age, sex, ethnicity, and BMI to address any possibility of residual confounding for age and sex since this was an imperfect match and to avoid confounding from BMI and population stratification. Our analyses showed a strong and statistically significant effect of the genotype score independent of these other factors.

***START and CHILD birth cohorts:*** As DNA methylation analyses are a relatively new field of study, commonly accepted norms around confounding variables have not yet been established. To select confounders, we investigated the association of potential confounders with the outcome (i.e. birth weight and length) and predictor (i.e. methylation at CpG sites). While sex and gestational age were consistently associated with birth weight / length, there was inconsistent evidence of association with methylation levels, i.e. some probe sites showed a significant association while others did not. Moreover, we found genotypes to also be inconsistently associated with methylation. Since the outcome was a continuous variable, we did not expect to lose power from adjusting for multiple covariates; in fact, including a non-confounding covariate can be beneficial as it explains some of the variability in the outcome, hence reducing noise and increasing power to detect novel associations223. Our analysis included the aforementioned covariates and thus reduced the potential for bias from confounding.

### 5.2.2 External validity

External validity pertains to the process of generalizing the findings from the sample studied to the population at-large. Within the meta-analysis, to ensure external validity, a representative sample of the literature was acquired using a thorough search strategy, with input from a McMaster University librarian, and *via* search of a variety of databases. As previously described, a majority of studies were derived from the general population, with the exception of some for which cases were recruited from diabetes clinics / hospitals. This type of study design is more feasible than a community design where the study team must visit individual households or community dwellings. Though feasible, the risk estimates established for the effect of genetic variants on diabetes from such studies may not be applicable to less sick, more general populations. It should be noted, however, that no substantial between-study heterogeneity was uncovered from inclusion of such studies, and thus the estimates are likely representative of a general population. Furthermore, with the exception of the Sikh Diabetes Study94 and an Aggarwal population study91, none of the cases or controls included endogamous samples. An important consideration to external validity was the sub-structure of the included samples; a minority of the included datasets (6/31) were of Dravidian South Asians. There is some evidence of variation in risk allele frequencies and risk for complex diseases among Indo-Aryan and Dravidian groups, which may restrict the applicability of our findings to the latter group. A more thorough discussion of this population sub-structure is undertaken in the next section.

In the EpiDREAM cohort, South Asians were largely sampled from South India (88%), including Hyderabad, Bangalore, and Chennai, representing a Dravidian population. Therefore, generalizability to Indo-Aryan groups may be limited20,21. White Caucasians, on the other hand, were sampled from Canada, United States, Europe, and Australia, making the results very generalizable.

Lastly, the two birth cohorts used for the methylation analysis were drawn from Canadian cities and therefore the findings are likely to be generalizable to South Asian and white populations with similar social and economic groups.

## 5.3 Other considerations for trans-ethnic genetic studies

### 5.3.1 Population stratification

Population stratification refers to systematic differences between sub-populations that arise from ancestry226. These differences include frequency of risk alleles and variation in methylation irrespective of disease status227. Variation in allele frequencies attributable to diversity in population and unrelated to outcome status can lead to the detection of spurious associations. Notably, population stratification is the most commonly cited reason for non-replication226. Association studies, in an attempt to avoid bias from population stratification, have used methods such as genomic control to correct for inflated test statistics and adjustment with principle components. The issue of population stratification also extends to methylation analyses, which show evidence of variation by ethnic / racial groups228,229. These differences could arise from epigenetic inheritance230, differences in environmental factors by population, and allele-specific DNA methylation, which is affected by variation in allele frequencies190,231,232. The literature supports a North-South gradient amongst South Asians, reflecting a historic admixture between Indo-Aryan and Dravidian people; Indo-Aryans tend to be more closely related to white Caucasians20,21,108. Additionally, as South Asian populations are endogamous, genetic stratification is marked and alleles of rarer variants may be observed with a higher frequency in this group compared to groups with greater gene flow233, such as many white Caucasian populations. These make considerations of population stratification particularly important in this group. Subgroups within white Caucasians also exist19, though there is less evidence of genetic stratification as a result of consanguineous and endogamous unions. We did not find evidence of population substructure within the ethnic groups in our cohort and therefore the analyses were not adjusted for population stratification.

### 5.3.2 Linkage disequilibrium

The non-random association of alleles at two or more loci, formally known as linkage disequilibrium, is an important concept when studying ethnically diverse populations. Linkage equilibrium implies statistical independence, such that frequency of an AB haplotype is the product of allele frequencies *pA* and *pB*234.Population genetic forces, such as selection, polymorphism, genetic drift, and gene flow will create LD. Among humans, haplotype blocks vary; they tend to be shortest in African populations and larger in Caucasians235. Deviations from equilibrium can be partitioned into average disequilibrium within sub-populations and differences in allele frequencies between sub-populations. The latter is commonly studied using Wright’s fixation index (FST)234. A recent study of the South Asian genome revealed that while a majority of genetic variants is universally shared, substantial heterogeneity in allele frequency exists between populations. Notably, the mean FST between South Asians and white Caucasians showed the greatest stratification (FST = 0.010) compared to other ethnic groups. Loci showing the greatest heterogeneity were enriched for genes involved in metabolism. Additionally, the authors report discovery of SNPs unique to South Asians, again mostly involved in lipid and carbohydrate metabolism. This discovery is congruent with our report of associations of 7 SNPs on *AP3S2*, *ST6GAL1*, *C6orf57*, *GRB14*, *HNF4α*, *VPS26A*, and *TMEM163d* with type 2 diabetes in South Asians, but not Caucasians, a pattern that has been shown for other traits236. Fine mapping analyses of these SNPs may reveal causal variants tagged by the above unique to South Asians as a result of varying LD. Furthermore, future such investigations may reveal more genetic variants contributing to cardio-metabolic phenotypes that together can explain the divergent risk in this group.

## 5.4 Future directions and Conclusions

Overall, our results indicate that the greater risk for metabolic traits in South Asians likely does not result from shared genetic variants, but those unique to South Asians. Evidence for this comes from our meta-analysis (Chapter 2) and genome-wide study of DNA methylation (Chapter 4). Furthermore, there is evidence for differences in interaction of an underlying genetic risk with other physiological traits, notably fat storage, which also contributes to this ethnic variation (Chapter 3). Future studies should: 1) conduct large-scale genome-wide investigations of SNPs and methylation loci implicated in metabolic traits, 2) focus on fine-mapping already identified variants showing heterogeneity in effect between South Asians and white Caucasians, 3) homozygocity mapping in consanguineous South Asian families20,21, and 4) study differences in gene-gene and gene-environment interactions which may explain differential risk from some genetic variants.

# References

1. World Health Organization. *Global status report on noncommunicable diseases 2014*. Geneva, Switzerland; 2012.

2. Stumvoll M, Goldstein BJ, van Haeften TW. Type 2 diabetes: principles of pathogenesis and therapy. *Lancet*. 365(9467):1333–46. doi:10.1016/S0140-6736(05)61032-X.

3. Kahn SE. The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of Type 2 diabetes. *Diabetologia*. 2003;46(1):3–19. doi:10.1007/s00125-002-1009-0.

4. Weyer C, Tataranni PA, Bogardus C, Pratley RE. Insulin resistance and insulin secretory dysfunction are independent predictors of worsening of glucose tolerance during each stage of type 2 diabetes development. *Diabetes Care*. 2001;24(1):89–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11194248. Accessed March 13, 2014.

5. Elbein SC, Hasstedt SJ, Wegner K, Kahn SE. Heritability of pancreatic beta-cell function among nondiabetic members of Caucasian familial type 2 diabetic kindreds. *J Clin Endocrinol Metab*. 1999;84(4):1398–403. doi:10.1210/jcem.84.4.5604.

6. Schousboe K, Visscher PM, Henriksen JE, Hopper JL, Sørensen TIA, Kyvik KO. Twin study of genetic and environmental influences on glucose tolerance and indices of insulin sensitivity and secretion. *Diabetologia*. 2003;46(9):1276–83. doi:10.1007/s00125-003-1165-x.

7. Katoh S, Lehtovirta M, Kaprio J, et al. Genetic and environmental effects on fasting and postchallenge plasma glucose and serum insulin values in Finnish twins. *J Clin Endocrinol Metab*. 2005;90(5):2642–7. doi:10.1210/jc.2004-2471.

8. Scott RA, Lagou V, Welch RP, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet*. 2012;44(9):991–1005. doi:10.1038/ng.2385.

9. Manning AK, Hivert M-F, Scott RA, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet*. 2012;44(6):659–69. doi:10.1038/ng.2274.

10. Dupuis J, Langenberg C, Prokopenko I, 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. doi:10.1038/ng.520.

11. Saxena R, Hivert M-F, Langenberg C, et al. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet*. 2010;42(2):142–8. doi:10.1038/ng.521.

12. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet*. 2012;90(1):7–24. doi:10.1016/j.ajhg.2011.11.029.

13. Stahl EA, Wegmann D, Trynka G, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet*. 2012;44(5):483–9. doi:10.1038/ng.2232.

14. Maher B. Personal genomes: The case of the missing heritability. *Nature*. 2008;456(7218):18–21. doi:10.1038/456018a.

15. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747–53. doi:10.1038/nature08494.

16. Anand SS, Yusuf S, Vuksan V, 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). *Lancet*. 2000;356(9226):279–284. doi:10.1016/S0140-6736(00)02502-2.

17. Rana A, de Souza RJ, Kandasamy S, Lear SA, Anand SS. Cardiovascular risk among South Asians living in Canada: a systematic review and meta-analysis. *C open*. 2014;2(3):E183–91. doi:10.9778/cmajo.20130064.

18. Valera B, Sohani ZN, Rana A, Poirier P, Anand SS. The ethno-epidemiology of obesity. *Can J Cardiol*. 2014. doi:10.1016/j.cjca.2014.10.005.

19. Tian C, Kosoy R, Nassir R, et al. European population genetic substructure: further definition of ancestry informative markers for distinguishing among diverse European ethnic groups. *Mol Med*. 15(11-12):371–83. doi:10.2119/molmed.2009.00094.

20. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. *Nature*. 2009;461(7263):489–94. doi:10.1038/nature08365.

21. Moorjani P, Thangaraj K, Patterson N, et al. Genetic evidence for recent population mixture in India. *Am J Hum Genet*. 2013;93(3):422–38. doi:10.1016/j.ajhg.2013.07.006.

22. Barnett AH, Dixon AN, Bellary S, et al. Type 2 diabetes and cardiovascular risk in the UK south Asian community. *Diabetologia*. 2006;49(10):2234–46. doi:10.1007/s00125-006-0325-1.

23. Yusuf S, Reddy S, Ounpuu S, Anand S. Global Burden of Cardiovascular Diseases: Part I: General Considerations, the Epidemiologic Transition, Risk Factors, and Impact of Urbanization. *Circulation*. 2001;104(22):2746–2753. doi:10.1161/hc4601.099487.

24. Yusuf S, Reddy S, Ounpuu S, Anand S. Global Burden of Cardiovascular Diseases: Part II: Variations in Cardiovascular Disease by Specific Ethnic Groups and Geographic Regions and Prevention Strategies. *Circulation*. 2001;104(23):2855–2864. doi:10.1161/hc4701.099488.

25. Mather HM, Chaturvedi N, Fuller JH. Mortality and morbidity from diabetes in South Asians and Europeans: 11-year follow-up of the Southall Diabetes Survey, London, UK. *Diabet Med*. 1998;15(1):53–9. doi:10.1002/(SICI)1096-9136(199801)15:1<53::AID-DIA521>3.0.CO;2-V.

26. Tillin T, Forouhi N, Johnston DG, McKeigue PM, Chaturvedi N, Godsland IF. Metabolic syndrome and coronary heart disease in South Asians, African-Caribbeans and white Europeans: a UK population-based cross-sectional study. *Diabetologia*. 2005;48(4):649–56. doi:10.1007/s00125-005-1689-3.

27. Prasad DS, Kabir Z, Dash AK, Das BC. Prevalence and risk factors for metabolic syndrome in Asian Indians: A community study from urban Eastern India. *J Cardiovasc Dis Res*. 2012;3(3):204–11. doi:10.4103/0975-3583.98895.

28. Mohan V, Shanthirani S, Deepa R, Premalatha G, Sastry NG, Saroja R. Intra-urban differences in the prevalence of the metabolic syndrome in southern India - the Chennai Urban Population Study (CUPS No. 4). *Diabet Med*. 2001;18(4):280–287. doi:10.1046/j.1464-5491.2001.00421.x.

29. Deepa M, Farooq S, Datta M, Deepa R, Mohan V. Prevalence of metabolic syndrome using WHO, ATPIII and IDF definitions in Asian Indians: the Chennai Urban Rural Epidemiology Study (CURES-34). *Diabetes Metab Res Rev*. 2007;23(2):127–34. doi:10.1002/dmrr.658.

30. Petersen KF, Dufour S, Feng J, et al. Increased prevalence of insulin resistance and nonalcoholic fatty liver disease in Asian-Indian men. *Proc Natl Acad Sci U S A*. 2006;103(48):18273–7. doi:10.1073/pnas.0608537103.

31. Raji A, Seely EW, Arky RA, Simonson DC. Body fat distribution and insulin resistance in healthy Asian Indians and Caucasians. *J Clin Endocrinol Metab*. 2001;86(11):5366–71. doi:10.1210/jcem.86.11.7992.

32. Ikehara S, Tabák AG, Akbaraly TN, et al. Age trajectories of glycaemic traits in non-diabetic South Asian and white individuals: the Whitehall II cohort study. *Diabetologia*. 2015;58(3):534–42. doi:10.1007/s00125-014-3448-9.

33. Sniderman AD, Bhopal R, Prabhakaran D, Sarrafzadegan N, Tchernof A. Why might South Asians be so susceptible to central obesity and its atherogenic consequences? The adipose tissue overflow hypothesis. *Int J Epidemiol*. 2007;36(1):220–5. doi:10.1093/ije/dyl245.

34. Gujral UP, Pradeepa R, Weber MB, Narayan KMV, Mohan V. Type 2 diabetes in South Asians: similarities and differences with white Caucasian and other populations. *Ann N Y Acad Sci*. 2013;1281:51–63. doi:10.1111/j.1749-6632.2012.06838.x.

35. Yajnik CS, Yudkin JS. The Y-Y paradox. *Lancet*. 2004;363(9403):163. doi:10.1016/S0140-6736(03)15269-5.

36. Banerji MA, Faridi N, Atluri R, Chaiken RL, Lebovitz HE. Body Composition, Visceral Fat, Leptin, and Insulin Resistance in Asian Indian Men1. *J Clin Endocrinol Metab*. 2013. Available at: http://press.endocrine.org/doi/abs/10.1210/jcem.84.1.5371. Accessed January 26, 2015.

37. Chandalia M, Lin P, Seenivasan T, et al. Insulin Resistance and Body Fat Distribution in South Asian Men Compared to Caucasian Men. Maedler K, ed. *PLoS One*. 2007;2(8):e812. doi:10.1371/journal.pone.0000812.

38. Deurenberg P, Deurenberg-Yap M, Guricci S. Asians are different from Caucasians and from each other in their body mass index/body fat per cent relationship. *Obes Rev*. 2002;3(3):141–146. doi:10.1046/j.1467-789X.2002.00065.x.

39. Anand SS, Tarnopolsky MA, Rashid S, et al. Adipocyte hypertrophy, fatty liver and metabolic risk factors in South Asians: the Molecular Study of Health and Risk in Ethnic Groups (mol-SHARE). Stanojevic S, ed. *PLoS One*. 2011;6(7):e22112. doi:10.1371/journal.pone.0022112.

40. Kohli S, Sniderman AD, Tchernof A, Lear SA. Ethnic-specific differences in abdominal subcutaneous adipose tissue compartments. *Obesity (Silver Spring)*. 2010;18(11):2177–83. doi:10.1038/oby.2010.94.

41. Balakrishnan P, Grundy SM, Islam A, Dunn F, Vega GL. Influence of upper and lower body adipose tissue on insulin sensitivity in South Asian men. *J Investig Med*. 2012;60(7):999–1004. doi:10.231/JIM.0b013e3182650a09.

42. Leon DA, Moser KA. Low birth weight persists in South Asian babies born in England and Wales regardless of maternal country of birth. Slow pace of acculturation, physiological constraint or both? Analysis of routine data. *J Epidemiol Community Health*. 2012;66(6):544–51. doi:10.1136/jech.2010.112516.

43. Margetts BM, Mohd Yusof S, Al Dallal Z, Jackson AA. Persistence of lower birth weight in second generation South Asian babies born in the United Kingdom. *J Epidemiol Community Health*. 2002;56(9):684–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1732245&tool=pmcentrez&rendertype=abstract. Accessed May 20, 2014.

44. Janssen PA, Thiessen P, Klein MC, Whitfield MF, Macnab YC, Cullis-Kuhl SC. Standards for the measurement of birth weight, length and head circumference at term in neonates of European, Chinese and South Asian ancestry. *Open Med*. 2007;1(2):e74–88. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2802014&tool=pmcentrez&rendertype=abstract. Accessed April 20, 2015.

45. Whincup PH, Gilg JA, Owen CG, Odoki K, Alberti KGMM, Cook DG. British South Asians aged 13-16 years have higher fasting glucose and insulin levels than Europeans. *Diabet Med*. 2005;22(9):1275–7. doi:10.1111/j.1464-5491.2005.01587.x.

46. Rosenbaum M, Fennoy I, Accacha S, et al. Racial/ethnic differences in clinical and biochemical type 2 diabetes mellitus risk factors in children. *Obesity (Silver Spring)*. 2013;21(10):2081–90. doi:10.1002/oby.20483.

47. Ranjani H, Sonya J, Anjana RM, Mohan V. Prevalence of glucose intolerance among children and adolescents in urban South India (ORANGE-2). *Diabetes Technol Ther*. 2013;15(1):13–9. doi:10.1089/dia.2012.0236.

48. Shaw NJ, Crabtree NJ, Kibirige MS, Fordham JN. Ethnic and gender differences in body fat in British schoolchildren as measured by DXA. *Arch Dis Child*. 2007;92(10):872–5. doi:10.1136/adc.2007.117911.

49. Sinha R, Fisch G, Teague B, et al. Prevalence of impaired glucose tolerance among children and adolescents with marked obesity. *N Engl J Med*. 2002;346(11):802–10. doi:10.1056/NEJMoa012578.

50. Wiegand S, Maikowski U, Blankenstein O, Biebermann H, Tarnow P, Gruters A. Type 2 diabetes and impaired glucose tolerance in European children and adolescents with obesity -- a problem that is no longer restricted to minority groups. *Eur J Endocrinol*. 2004;151(2):199–206. doi:10.1530/eje.0.1510199.

51. Peters J, Ulijaszek SJ. Population and sex differences in arm circumference and skinfold thicknesses among Indo-Pakistani children living in the East Midlands of Britain. *Ann Hum Biol*. 19(1):17–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1734819. Accessed April 27, 2015.

52. Gulliford MC, Mahabir D, Rocke B, Chinn S, Rona R. Overweight, obesity and skinfold thicknesses of children of African or Indian descent in Trinidad and Tobago. *Int J Epidemiol*. 2001;30(5):989–98. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11689509. Accessed April 27, 2015.

53. Stanfield KM, Wells JC, Fewtrell MS, Frost C, Leon DA. Differences in body composition between infants of South Asian and European ancestry: the London Mother and Baby Study. *Int J Epidemiol*. 2012;41(5):1409–18. doi:10.1093/ije/dys139.

54. Modi N, Thomas EL, Uthaya SN, Umranikar S, Bell JD, Yajnik C. Whole Body Magnetic Resonance Imaging of Healthy Newborn Infants Demonstrates Increased Central Adiposity in Asian Indians. *Pediatr Res*. 2009;65(5):584–587. doi:10.1203/PDR.0b013e31819d98be.

55. Yajnik CS, Lubree HG, Rege SS, et al. Adiposity and Hyperinsulinemia in Indians Are Present at Birth. *J Clin Endocrinol Metab*. 2002;87(12):5575–5580. doi:10.1210/jc.2002-020434.

56. Nightingale CM, Rudnicka AR, Owen CG, Cook DG, Whincup PH. Patterns of body size and adiposity among UK children of South Asian, black African-Caribbean and white European origin: Child Heart And health Study in England (CHASE Study). *Int J Epidemiol*. 2011;40(1):33–44. doi:10.1093/ije/dyq180.

57. Ramachandran P. Nutrition and child survival in India. *Indian J Pediatr*. 2010;77(3):301–5. doi:10.1007/s12098-010-0038-9.

58. Brancati FL, Whelton PK, Kuller LH, Klag MJ. Diabetes mellitus, race, and socioeconomic status. A population-based study. *Ann Epidemiol*. 1996;6(1):67–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8680628. Accessed May 16, 2015.

59. Young BA, Maynard C, Boyko EJ. Racial Differences in Diabetic Nephropathy, Cardiovascular Disease, and Mortality in a National Population of Veterans. *Diabetes Care*. 2003;26(8):2392–2399. doi:10.2337/diacare.26.8.2392.

60. Carter JS, Pugh JA, Monterrosa A. Non-insulin-dependent diabetes mellitus in minorities in the United States. *Ann Intern Med*. 1996;125(3):221–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8686981. Accessed April 19, 2015.

61. Gerstein HC, Waltman L. Why don’t pigs get diabetes? Explanations for variations in diabetes susceptibility in human populations living in a diabetogenic environment. *CMAJ*. 2006;174(1):25–6. doi:10.1503/cmaj.050649.

62. Anand SS, Meyre D, Pare G, et al. Genetic information and the prediction of incident type 2 diabetes in a high-risk multiethnic population: the EpiDREAM genetic study. *Diabetes Care*. 2013;36(9):2836–42. doi:10.2337/dc12-2553.

63. Anand SS, Vasudevan A, Gupta M, et al. Rationale and design of South Asian Birth Cohort (START): a Canada-India collaborative study. *BMC Public Health*. 2013;13:79. doi:10.1186/1471-2458-13-79.

64. Radha V, Mohan V. Genetic predisposition to type 2 diabetes among Asian Indians. *Indian J Med Res*. 2007;125(3):259–74. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17496355. Accessed February 8, 2014.

65. Kanaya AM, Wassel CL, Mathur D, et al. Prevalence and correlates of diabetes in South asian indians in the United States: findings from the metabolic syndrome and atherosclerosis in South asians living in america study and the multi-ethnic study of atherosclerosis. *Metab Syndr Relat Disord*. 2010;8(2):157–64. doi:10.1089/met.2009.0062.

66. Little J, Higgins JPT, Ioannidis JPA, et al. STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. *PLoS Med*. 2009;6(2):e22. doi:10.1371/journal.pmed.1000022.

67. Lewis CM, Knight J. Introduction to genetic association studies. *Cold Spring Harb Protoc*. 2012;2012(3):297–306. doi:10.1101/pdb.top068163.

68. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O’Donnell CJ, de Bakker PIW. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. 2008;24(24):2938–9. doi:10.1093/bioinformatics/btn564.

69. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods*. 2006;11(2):193–206. doi:10.1037/1082-989X.11.2.193.

70. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 2012;44(9):981–90. doi:10.1038/ng.2383.

71. Kooner JS, Saleheen D, Sim X, et al. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet*. 2011;43(10):984–9. doi:10.1038/ng.921.

72. Tabassum R, Chauhan G, Dwivedi OP, et al. Genome-wide association study for type 2 diabetes in Indians identifies a new susceptibility locus at 2q21. *Diabetes*. 2013;62(3):977–86. doi:10.2337/db12-0406.

73. Saxena R, Saleheen D, Been LF, et al. Genome-wide association study identifies a novel locus contributing to type 2 diabetes susceptibility in Sikhs of Punjabi origin from India. *Diabetes*. 2013;62(5):1746–55. doi:10.2337/db12-1077.

74. Sim X, Ong RT-H, Suo C, et al. Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. Gibson G, ed. *PLoS Genet*. 2011;7(4):e1001363. doi:10.1371/journal.pgen.1001363.

75. Haseeb A, Iliyas M, Chakrabarti S, et al. Single-nucleotide polymorphisms in peroxisome proliferator-activated receptor gamma and their association with plasma levels of resistin and the metabolic syndrome in a South Indian population. *J Biosci*. 2009;34(3):405–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19805902. Accessed February 15, 2014.

76. Anuradha S, Radha V, Mohan V. Association of novel variants in the hepatocyte nuclear factor 4A gene with maturity onset diabetes of the young and early onset type 2 diabetes. *Clin Genet*. 2011;80(6):541–9. doi:10.1111/j.1399-0004.2010.01577.x.

77. Chandak GR, Janipalli CS, Bhaskar S, et al. Common variants in the TCF7L2 gene are strongly associated with type 2 diabetes mellitus in the Indian population. *Diabetologia*. 2007;50(1):63–7. doi:10.1007/s00125-006-0502-2.

78. Chidambaram M, Radha V, Mohan V. Replication of recently described type 2 diabetes gene variants in a South Indian population. *Metabolism*. 2010;59(12):1760–6. doi:10.1016/j.metabol.2010.04.024.

79. Mukhopadhyaya PN, Acharya A, Chavan Y, Purohit SS, Mutha A. Metagenomic study of single-nucleotide polymorphism within candidate genes associated with type 2 diabetes in an Indian population. *Genet Mol Res*. 2010;9(4):2060–8. doi:10.4238/vol9-4gmr883.

80. Janipalli CS, Kumar MVK, Vinay DG, et al. Analysis of 32 common susceptibility genetic variants and their combined effect in predicting risk of Type 2 diabetes and related traits in Indians. *Diabet Med*. 2012;29(1):121–7. doi:10.1111/j.1464-5491.2011.03438.x.

81. Rees SD, Hydrie MZI, Shera AS, et al. Replication of 13 genome-wide association (GWA)-validated risk variants for type 2 diabetes in Pakistani populations. *Diabetologia*. 2011;54(6):1368–74. doi:10.1007/s00125-011-2063-2.

82. Uma Jyothi K, Jayaraj M, Subburaj KS, et al. Association of TCF7L2 gene polymorphisms with T2DM in the population of Hyderabad, India. *PLoS One*. 2013;8(4):e60212. doi:10.1371/journal.pone.0060212.

83. Rees SD, Islam M, Hydrie MZI, et al. An FTO variant is associated with Type 2 diabetes in South Asian populations after accounting for body mass index and waist circumference. *Diabet Med*. 2011;28(6):673–80. doi:10.1111/j.1464-5491.2011.03257.x.

84. Bodhini D, Radha V, Dhar M, Narayani N, Mohan V. The rs12255372(G/T) and rs7903146(C/T) polymorphisms of the TCF7L2 gene are associated with type 2 diabetes mellitus in Asian Indians. *Metabolism*. 2007;56(9):1174–8. doi:10.1016/j.metabol.2007.04.012.

85. Raza ST, Abbas S, Ahmed F, Fatima J, Zaidi ZH, Mahdi F. Association of MTHFR and PPARγ2 gene polymorphisms in relation to type 2 diabetes mellitus cases among north Indian population. *Gene*. 2012;511(2):375–9. doi:10.1016/j.gene.2012.09.072.

86. Tan JT, Ng DPK, Nurbaya S, et al. Polymorphisms identified through genome-wide association studies and their associations with type 2 diabetes in Chinese, Malays, and Asian-Indians in Singapore. *J Clin Endocrinol Metab*. 2010;95(1):390–7. doi:10.1210/jc.2009-0688.

87. Humphries SE, Gable D, Cooper JA, et al. Common variants in the TCF7L2 gene and predisposition to type 2 diabetes in UK European Whites, Indian Asians and Afro-Caribbean men and women. *J Mol Med (Berl)*. 2006;84(12):1005–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17665514. Accessed February 15, 2014.

88. Radha V, Vimaleswaran KS, Babu HNS, et al. Role of genetic polymorphism peroxisome proliferator-activated receptor-gamma2 Pro12Ala on ethnic susceptibility to diabetes in South-Asian and Caucasian subjects: Evidence for heterogeneity. *Diabetes Care*. 2006;29(5):1046–51. doi:10.2337/diacare.2951046.

89. Boodram LG, Miyake K, Hayes MG, Bell GI, Cockburn BN. Association of the KCNJ11 variant E23K with type 2 diabetes in Indo-Trinidadians. *West Indian Med J*. 2011;60(6):604–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22512215. Accessed February 15, 2014.

90. Yajnik CS, Janipalli CS, Bhaskar S, et al. FTO gene variants are strongly associated with type 2 diabetes in South Asian Indians. *Diabetologia*. 2009;52(2):247–52. doi:10.1007/s00125-008-1186-6.

91. Gupta V, Khadgawat R, Ng HKT, et al. A validation study of type 2 diabetes-related variants of the TCF7L2, HHEX, KCNJ11, and ADIPOQ genes in one endogamous ethnic group of north India. *Ann Hum Genet*. 2010;74(4):361–8. doi:10.1111/j.1469-1809.2010.00580.x.

92. Chauhan G, Spurgeon CJ, Tabassum R, et al. Impact of common variants of PPARG, KCNJ11, TCF7L2, SLC30A8, HHEX, CDKN2A, IGF2BP2, and CDKAL1 on the risk of type 2 diabetes in 5,164 Indians. *Diabetes*. 2010;59(8):2068–74. doi:10.2337/db09-1386.

93. Chauhan G, Tabassum R, Mahajan A, et al. Common variants of FTO and the risk of obesity and type 2 diabetes in Indians. *J Hum Genet*. 2011;56(10):720–6. doi:10.1038/jhg.2011.87.

94. Sanghera DK, Ortega L, Han S, et al. Impact of nine common type 2 diabetes risk polymorphisms in Asian Indian Sikhs: PPARG2 (Pro12Ala), IGF2BP2, TCF7L2 and FTO variants confer a significant risk. *BMC Med Genet*. 2008;9(1):59. doi:10.1186/1471-2350-9-59.

95. Lehtovirta M, Kaprio J, Forsblom C, Eriksson J, Tuomilehto J, Groop L. Insulin sensitivity and insulin secretion in monozygotic and dizygotic twins. *Diabetologia*. 2000;43(3):285–93. doi:10.1007/s001250050046.

96. Perry JRB, Voight BF, Yengo L, et al. Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases. *PLoS Genet*. 2012;8(5):e1002741. doi:10.1371/journal.pgen.1002741.

97. Qi Q, Li H, Loos RJF, et al. Common variants in KCNQ1 are associated with type 2 diabetes and impaired fasting glucose in a Chinese Han population. *Hum Mol Genet*. 2009;18(18):3508–15. doi:10.1093/hmg/ddp294.

98. Unoki H, Takahashi A, Kawaguchi T, et al. SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet*. 2008;40(9):1098–102. doi:10.1038/ng.208.

99. Kraft P. Curses--winner’s and otherwise--in genetic epidemiology. *Epidemiology*. 2008;19(5):649–51; discussion 657–8. doi:10.1097/EDE.0b013e318181b865.

100. Tsai F-J, Yang C-F, Chen C-C, et al. A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. *PLoS Genet*. 2010;6(2):e1000847. doi:10.1371/journal.pgen.1000847.

101. Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet*. 2008;40(5):638–45. doi:10.1038/ng.120.

102. Voight BF, Scott LJ, Steinthorsdottir V, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet*. 2010;42(7):579–89. doi:10.1038/ng.609.

103. Li H, Gan W, Lu L, et al. A genome-wide association study identifies GRK5 and RASGRP1 as type 2 diabetes loci in Chinese Hans. *Diabetes*. 2013;62(1):291–8. doi:10.2337/db12-0454.

104. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium, South Asian Type 2 Diabetes (SAT2D) Consortium, 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. doi:10.1038/ng.2897.

105. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*. 2007;316(5829):1341–5. doi:10.1126/science.1142382.

106. Di Rienzo A, Hudson RR. An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends Genet*. 2005;21(11):596–601. doi:10.1016/j.tig.2005.08.007.

107. Pickrell JK, Coop G, Novembre J, et al. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res*. 2009;19(5):826–37. doi:10.1101/gr.087577.108.

108. Chambers JC, Abbott J, Zhang W, et al. The South Asian genome. *PLoS One*. 2014;9(8):e102645. doi:10.1371/journal.pone.0102645.

109. Ayub Q, Moutsianas L, Chen Y, et al. Revisiting the Thrifty Gene Hypothesis via 65 Loci Associated with Susceptibility to Type 2 Diabetes. *Am J Hum Genet*. 2014;94(2):176–85. doi:10.1016/j.ajhg.2013.12.010.

110. Yang I V, Wade CM, Kang HM, et al. Identification of novel genes that mediate innate immunity using inbred mice. *Genetics*. 2009;183(4):1535–44. doi:10.1534/genetics.109.107540.

111. VanLiere JM, Rosenberg NA. Mathematical properties of the r2 measure of linkage disequilibrium. *Theor Popul Biol*. 2008;74(1):130–7. doi:10.1016/j.tpb.2008.05.006.

112. Rees SD, Hydrie MZI, O’Hare JP, et al. Effects of 16 genetic variants on fasting glucose and type 2 diabetes in South Asians: ADCY5 and GLIS3 variants may predispose to type 2 diabetes. Timpson NJ, ed. *PLoS One*. 2011;6(9):e24710. doi:10.1371/journal.pone.0024710.

113. De Haan HG, Bezemer ID, Doggen CJM, et al. Multiple SNP testing improves risk prediction of first venous thrombosis. *Blood*. 2012;120(3):656–63. doi:10.1182/blood-2011-12-397752.

114. Cooke JN, Ng MCY, Palmer ND, et al. Genetic risk assessment of type 2 diabetes-associated polymorphisms in African Americans. *Diabetes Care*. 2012;35(2):287–92. doi:10.2337/dc11-0957.

115. Chen H, Poon A, Yeung C, et al. A genetic risk score combining ten psoriasis risk loci improves disease prediction. Mailund T, ed. *PLoS One*. 2011;6(4):e19454. doi:10.1371/journal.pone.0019454.

116. Lin X, Song K, Lim N, et al. Risk prediction of prevalent diabetes in a Swiss population using a weighted genetic score--the CoLaus Study. *Diabetologia*. 2009;52(4):600–8. doi:10.1007/s00125-008-1254-y.

117. Sohani ZN, Meyre D, de Souza RJ, et al. Assessing the quality of published genetic association studies in meta-analyses: the quality of genetic studies (Q-Genie) tool. *BMC Genet*. 2015;16(1):50. doi:10.1186/s12863-015-0211-2.

118. Li A, Meyre D. Challenges in reproducibility of genetic association studies: lessons learned from the obesity field. *Int J Obes (Lond)*. 2013;37(4):559–67. doi:10.1038/ijo.2012.82.

119. Li A, Meyre D. Jumping on the Train of Personalized Medicine: A Primer for Non- Geneticist Clinicians: Part 3. Clinical Applications in the Personalized Medicine Area. *Curr Psychiatry Rev*. 2014;10(2):118–132. doi:10.2174/1573400510666140630170549.

120. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet*. 2010;11(6):415–25. doi:10.1038/nrg2779.

121. Wray NR, Purcell SM, Visscher PM. Synthetic associations created by rare variants do not explain most GWAS results. Flint J, ed. *PLoS Biol*. 2011;9(1):e1000579. doi:10.1371/journal.pbio.1000579.

122. Williams AL, Jacobs SBR, Moreno-Macías H, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature*. 2014;506(7486):97–101. doi:10.1038/nature12828.

123. Estrada K, Aukrust I, Bjørkhaug L, et al. Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. *JAMA*. 2014;311(22):2305–14. doi:10.1001/jama.2014.6511.

124. Moltke I, Grarup N, Jørgensen ME, et al. A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature*. 2014;512(7513):190–3. doi:10.1038/nature13425.

125. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn’s disease. *Nature*. 2001;411(6837):599–603. doi:10.1038/35079107.

126. Bonnefond A, Clément N, Fawcett K, et al. Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nat Genet*. 2012;44(3):297–301. doi:10.1038/ng.1053.

127. Simmons D, Williams R. Dietary practices among Europeans and different South Asian groups in Coventry. *Br J Nutr*. 2007;78(01):5. doi:10.1079/BJN19970114.

128. Babakus WS, Thompson JL. Physical activity among South Asian women: a systematic, mixed-methods review. *Int J Behav Nutr Phys Act*. 2012;9(1):150. doi:10.1186/1479-5868-9-150.

129. Gerich JE. The genetic basis of type 2 diabetes mellitus: impaired insulin secretion versus impaired insulin sensitivity. *Endocr Rev*. 1998;19(4):491–503. doi:10.1210/edrv.19.4.0338.

130. Gastaldelli A. Role of beta-cell dysfunction, ectopic fat accumulation and insulin resistance in the pathogenesis of type 2 diabetes mellitus. *Diabetes Res Clin Pract*. 2011;93 Suppl 1:S60–5. doi:10.1016/S0168-8227(11)70015-8.

131. Poitout V, Robertson RP. Glucolipotoxicity: fuel excess and beta-cell dysfunction. *Endocr Rev*. 2008;29(3):351–66. doi:10.1210/er.2007-0023.

132. Sohani ZN, Deng WQ, Pare G, Meyre D, Gerstein HC, Anand SS. Does genetic heterogeneity account for the divergent risk of type 2 diabetes in South Asian and white European populations? *Diabetologia*. 2014. doi:10.1007/s00125-014-3354-1.

133. McKeigue PM, Shah B, Marmot MG. Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in South Asians. *Lancet*. 1991;337(8738):382–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1671422. Accessed November 10, 2014.

134. Chandalia M, Abate N, Garg A, Stray-Gundersen J, Grundy SM. Relationship between generalized and upper body obesity to insulin resistance in Asian Indian men. *J Clin Endocrinol Metab*. 1999;84(7):2329–35. doi:10.1210/jcem.84.7.5817.

135. Misra A. Impact of ethnicity on body fat patterning in Asian Indians and blacks: relation with insulin resistance. *Nutrition*. 2003;19(9):815–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12921898. Accessed November 20, 2014.

136. Sniderman AD, Bhopal R, Prabhakaran D, Sarrafzadegan N, Tchernof A. Why might South Asians be so susceptible to central obesity and its atherogenic consequences? The adipose tissue overflow hypothesis. *Int J Epidemiol*. 2007;36(1):220–5. doi:10.1093/ije/dyl245.

137. Gerstein HC, Yusuf S, Bosch J, et al. Effect of rosiglitazone on the frequency of diabetes in patients with impaired glucose tolerance or impaired fasting glucose: a randomised controlled trial. *Lancet*. 2006;368(9541):1096–105. doi:10.1016/S0140-6736(06)69420-8.

138. Janssens ACJW, Moonesinghe R, Yang Q, Steyerberg EW, van Duijn CM, Khoury MJ. The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases. *Genet Med*. 2007;9(8):528–535. doi:10.1097/GIM.0b013e31812eece0.

139. Keating BJ, Tischfield S, Murray SS, et al. Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PLoS One*. 2008;3(10):e3583. doi:10.1371/journal.pone.0003583.

140. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904–9. doi:10.1038/ng1847.

141. Wright S. Coefficients of Inbreeding and Relationship. *Am Nat*. 1922;56(645):330–338. Available at: http://www.jstor.org/stable/2456273?seq=1#page\_scan\_tab\_contents. Accessed August 4, 2015.

142. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75. doi:10.1086/519795.

143. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76–82. doi:10.1016/j.ajhg.2010.11.011.

144. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res*. 2011;46(3):399–424. doi:10.1080/00273171.2011.568786.

145. Le Floch JP, Escuyer P, Baudin E, Baudon D, Perlemuter L. Blood glucose area under the curve. Methodological aspects. *Diabetes Care*. 1990;13(2):172–175. doi:10.2337/diacare.13.2.172.

146. Hanley AJ, Zinman B, Sheridan P, Yusuf S, Gerstein HC. Effect of Rosiglitazone and Ramipril on {beta}-cell function in people with impaired glucose tolerance or impaired fasting glucose: the DREAM trial. *Diabetes Care*. 2010;33(3):608–13. doi:10.2337/dc09-1579.

147. Utzschneider KM, Prigeon RL, Faulenbach M V, et al. Oral disposition index predicts the development of future diabetes above and beyond fasting and 2-h glucose levels. *Diabetes Care*. 2009;32(2):335–41. doi:10.2337/dc08-1478.

148. De Koning L, Merchant AT, Pogue J, Anand SS. Waist circumference and waist-to-hip ratio as predictors of cardiovascular events: meta-regression analysis of prospective studies. *Eur Heart J*. 2007;28(7):850–6. doi:10.1093/eurheartj/ehm026.

149. De Koning L, Gerstein HC, Bosch J, et al. Anthropometric measures and glucose levels in a large multi-ethnic cohort of individuals at risk of developing type 2 diabetes. *Diabetologia*. 2010;53(7):1322–1330. doi:10.1007/s00125-010-1710-3.

150. Czernichow S, Kengne A-P, Huxley RR, et al. Comparison of waist-to-hip ratio and other obesity indices as predictors of cardiovascular disease risk in people with type-2 diabetes: a prospective cohort study from ADVANCE. *Eur J Cardiovasc Prev Rehabil*. 2011;18(2):312–9. doi:10.1097/HJR.0b013e32833c1aa3.

151. Rees SD, Hydrie MZI, O’Hare JP, et al. Effects of 16 genetic variants on fasting glucose and type 2 diabetes in South Asians: ADCY5 and GLIS3 variants may predispose to type 2 diabetes. *PLoS One*. 2011;6(9):e24710. doi:10.1371/journal.pone.0024710.

152. Takeuchi F, Katsuya T, Chakrewarthy S, et al. Common variants at the GCK, GCKR, G6PC2-ABCB11 and MTNR1B loci are associated with fasting glucose in two Asian populations. *Diabetologia*. 2010;53(2):299–308. doi:10.1007/s00125-009-1595-1.

153. Islam M, Jafar TH, Wood AR, et al. Multiple genetic variants explain measurable variance in type 2 diabetes-related traits in Pakistanis. *Diabetologia*. 2012;55(8):2193–204. doi:10.1007/s00125-012-2560-y.

154. MAGIC Investigators. MAGIC Cohorts. Available at: http://www.magicinvestigators.org/cohorts/. Accessed May 18, 2015.

155. Brady EM, Webb DR, Morris DH, et al. Investigating endothelial activation and oxidative stress in relation to glycaemic control in a multiethnic population. *Exp Diabetes Res*. 2012;2012:386041. doi:10.1155/2012/386041.

156. El-Assaad W, Buteau J, Peyot M-L, et al. Saturated fatty acids synergize with elevated glucose to cause pancreatic beta-cell death. *Endocrinology*. 2003;144(9):4154–63. doi:10.1210/en.2003-0410.

157. McKeigue PM, Shah B, Marmot MG. Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in South Asians. *Lancet*. 1991;337(8738):382–386. doi:10.1016/0140-6736(91)91164-P.

158. Jacqueminet S, Briaud I, Rouault C, Reach G, Poitout V. Inhibition of insulin gene expression by long-term exposure of pancreatic beta cells to palmitate is dependent on the presence of a stimulatory glucose concentration. *Metabolism*. 2000;49(4):532–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10778881. Accessed March 18, 2015.

159. Zhou YP, Grill V. Long term exposure to fatty acids and ketones inhibits B-cell functions in human pancreatic islets of Langerhans. *J Clin Endocrinol Metab*. 1995;80(5):1584–90. doi:10.1210/jcem.80.5.7745004.

160. Abate N, Carulli L, Cabo-Chan A, Chandalia M, Snell PG, Grundy SM. Genetic polymorphism PC-1 K121Q and ethnic susceptibility to insulin resistance. *J Clin Endocrinol Metab*. 2003;88(12):5927–34. doi:10.1210/jc.2003-030453.

161. Guettier J-M, Georgopoulos A, Tsai MY, et al. Polymorphisms in the fatty acid-binding protein 2 and apolipoprotein C-III genes are associated with the metabolic syndrome and dyslipidemia in a South Indian population. *J Clin Endocrinol Metab*. 2005;90(3):1705–11. doi:10.1210/jc.2004-1338.

162. Majithia AR, Flannick J, Shahinian P, et al. Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. *Proc Natl Acad Sci U S A*. 2014;111(36):13127–32. doi:10.1073/pnas.1410428111.

163. Keating BJ, Tischfield S, Murray SS, et al. Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PLoS One*. 2008;3(10):e3583. doi:10.1371/journal.pone.0003583.

164. Seldin MF, Shigeta R, Villoslada P, et al. European population substructure: clustering of northern and southern populations. *PLoS Genet*. 2006;2(9):e143. doi:10.1371/journal.pgen.0020143.

165. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(1):7. doi:10.1186/s13742-015-0047-8.

166. American Association of Clinical Endocrinologists. American Association of Clinical Endocrinologists/American College of Endocrinology statement on the use of hemoglobin A1c for the diagnosis of diabetes. *Endocr Pract*. 16(2):155–6. doi:10.4158/EP.16.2.155.

167. International Expert Committee. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care*. 2009;32(7):1327–34. doi:10.2337/dc09-9033.

168. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2010;33 Suppl 1:S62–9. doi:10.2337/dc10-S062.

169. Sacks DB, Arnold M, Bakris GL, et al. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. *Diabetes Care*. 2011;34(6):e61–99. doi:10.2337/dc11-9998.

170. Sacks DB. A1C versus glucose testing: a comparison. *Diabetes Care*. 2011;34(2):518–23. doi:10.2337/dc10-1546.

171. Wisdom K, Fryzek JP, Havstad SL, Anderson RM, Dreiling MC, Tilley BC. Comparison of laboratory test frequency and test results between African-Americans and Caucasians with diabetes: opportunity for improvement. Findings from a large urban health maintenance organization. *Diabetes Care*. 1997;20(6):971–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9167108. Accessed August 4, 2015.

172. Davidson M, Schriger D. Effect of age and race/ethnicity on HbA1c levels in people without known diabetes mellitus: implications for the diagnosis of diabetes. *Diabetes Res Clin Pr*. 2010;87(3):415–421. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20061043. Accessed August 4, 2015.

173. Barrera V, Peinado MA. Evaluation of single CpG sites as proxies of CpG island methylation states at the genome scale. *Nucleic Acids Res*. 2012;40(22):11490–8. doi:10.1093/nar/gks928.

174. Sofer T, Schifano ED, Hoppin JA, Hou L, Baccarelli AA. A-clustering: a novel method for the detection of co-regulated methylation regions, and regions associated with exposure. *Bioinformatics*. 2013;29(22):2884–2891. doi:10.1093/bioinformatics/btt498.

175. Bock C. Analysing and interpreting DNA methylation data. *Nat Rev Genet*. 2012;13(10):705–19. doi:10.1038/nrg3273.

176. Illingworth RS, Bird AP. CpG islands--’a rough guide'. *FEBS Lett*. 2009;583(11):1713–20. doi:10.1016/j.febslet.2009.04.012.

177. Straughen JK, Sipahi L, Uddin M, Misra DP, Misra VK. Racial differences in IGF1 methylation and birth weight. *Clin Epigenetics*. 2015;7(1):47. doi:10.1186/s13148-015-0080-6.

178. King K, Murphy S, Hoyo C. Epigenetic regulation of Newborns’ imprinted genes related to gestational growth: patterning by parental race/ethnicity and maternal socioeconomic status. *J Epidemiol Community Health*. 2015. doi:10.1136/jech-2014-204781.

179. Horikoshi M, Yaghootkar H, Mook-Kanamori DO, et al. New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nat Genet*. 2013;45(1):76–82. doi:10.1038/ng.2477.

180. Freathy RM, Weedon MN, Bennett A, 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. doi:10.1086/518517.

181. Freathy RM, Bennett AJ, Ring SM, et al. Type 2 diabetes risk alleles are associated with reduced size at birth. *Diabetes*. 2009;58(6):1428–33. doi:10.2337/db08-1739.

182. Engel SM, Joubert BR, Wu MC, et al. Neonatal genome-wide methylation patterns in relation to birth weight in the Norwegian Mother and Child Cohort. *Am J Epidemiol*. 2014;179(7):834–42. doi:10.1093/aje/kwt433.

183. Turan N, Ghalwash MF, Katari S, Coutifaris C, Obradovic Z, Sapienza C. DNA methylation differences at growth related genes correlate with birth weight: a molecular signature linked to developmental origins of adult disease? *BMC Med Genomics*. 2012;5(1):10. doi:10.1186/1755-8794-5-10.

184. Takaro TK, Scott JA, Allen RW, et al. The Canadian Healthy Infant Longitudinal Development (CHILD) birth cohort study: assessment of environmental exposures. *J Expo Sci Environ Epidemiol*. 2015. doi:10.1038/jes.2015.7.

185. Dick KJ, Nelson CP, Tsaprouni L, et al. DNA methylation and body-mass index: a genome-wide analysis. *Lancet*. 2014;383(9933):1990–8. doi:10.1016/S0140-6736(13)62674-4.

186. Woods KA, Camacho-Hübner C, Savage MO, Clark AJ. Intrauterine growth retardation and postnatal growth failure associated with deletion of the insulin-like growth factor I gene. *N Engl J Med*. 1996;335(18):1363–7. doi:10.1056/NEJM199610313351805.

187. Ong K, Kratzsch J, Kiess W, Costello M, Scott C, Dunger D. Size at birth and cord blood levels of insulin, insulin-like growth factor I (IGF-I), IGF-II, IGF-binding protein-1 (IGFBP-1), IGFBP-3, and the soluble IGF-II/mannose-6-phosphate receptor in term human infants. The ALSPAC Study Team. Avon Longitudinal Stu. *J Clin Endocrinol Metab*. 2000;85(11):4266–9. doi:10.1210/jcem.85.11.6998.

188. Abuzzahab MJ, Schneider A, Goddard A, et al. IGF-I receptor mutations resulting in intrauterine and postnatal growth retardation. *N Engl J Med*. 2003;349(23):2211–22. doi:10.1056/NEJMoa010107.

189. Chen Y, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013;8(2):203–9. doi:10.4161/epi.23470.

190. Bell JT, Pai AA, Pickrell JK, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol*. 2011;12(1):R10. doi:10.1186/gb-2011-12-1-r10.

191. Stranger BE, Stahl EA, Raj T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*. 2011;187(2):367–83. doi:10.1534/genetics.110.120907.

192. Bell CG, Finer S, Lindgren CM, et al. Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. *PLoS One*. 2010;5(11):e14040. doi:10.1371/journal.pone.0014040.

193. Turan N, Ghalwash MF, Katari S, Coutifaris C, Obradovic Z, Sapienza C. DNA methylation differences at growth related genes correlate with birth weight: a molecular signature linked to developmental origins of adult disease? *BMC Med Genomics*. 2012;5:10. doi:10.1186/1755-8794-5-10.

194. Filiberto AC, Maccani MA, Koestler D, et al. Birthweight is associated with DNA promoter methylation of the glucocorticoid receptor in human placenta. *Epigenetics*. 2011;6(5):566–72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3121971&tool=pmcentrez&rendertype=abstract. Accessed April 18, 2015.

195. Gillberg L, Jacobsen SC, Rönn T, Brøns C, Vaag A. PPARGC1A DNA methylation in subcutaneous adipose tissue in low birth weight subjects--impact of 5 days of high-fat overfeeding. *Metabolism*. 2014;63(2):263–71. doi:10.1016/j.metabol.2013.10.003.

196. Michels KB, Harris HR, Barault L. Birthweight, maternal weight trajectories and global DNA methylation of LINE-1 repetitive elements. *PLoS One*. 2011;6(9):e25254. doi:10.1371/journal.pone.0025254.

197. Chen KT, Chen CJ, Gregg EW, Engelgau MM, Narayan KM. Prevalence of type 2 diabetes mellitus in Taiwan: ethnic variation and risk factors. *Diabetes Res Clin Pract*. 2001;51(1):59–66.

198. Simpkin AJ, Suderman M, Gaunt TR, et al. Longitudinal analysis of DNA methylation associated with birth weight and gestational age. *Hum Mol Genet*. 2015:ddv119–. doi:10.1093/hmg/ddv119.

199. Gloyn AL, Braun M, Rorsman P. Type 2 diabetes susceptibility gene TCF7L2 and its role in beta-cell function. *Diabetes*. 2009;58(4):800–2. doi:10.2337/db09-0099.

200. Da Silva Xavier G, Loder MK, McDonald A, et al. TCF7L2 regulates late events in insulin secretion from pancreatic islet beta-cells. *Diabetes*. 2009;58(4):894–905. doi:10.2337/db08-1187.

201. Groenewoud MJ, Dekker JM, Fritsche A, et al. Variants of CDKAL1 and IGF2BP2 affect first-phase insulin secretion during hyperglycaemic clamps. *Diabetologia*. 2008;51(9):1659–63. doi:10.1007/s00125-008-1083-z.

202. Ohara-Imaizumi M, Yoshida M, Aoyagi K, et al. Deletion of CDKAL1 affects mitochondrial ATP generation and first-phase insulin exocytosis. *PLoS One*. 2010;5(12):e15553. doi:10.1371/journal.pone.0015553.

203. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet*. 2007;39(6):770–5. doi:10.1038/ng2043.

204. Economides DL, Proudler A, Nicolaides KH. Plasma insulin in appropriate- and small-for-gestational-age fetuses. *Am J Obstet Gynecol*. 1989;160(5 Pt 1):1091–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2658601. Accessed April 7, 2014.

205. Davey Smith G. Birth characteristics of offspring and parental diabetes: evidence for the fetal insulin hypothesis. *J Epidemiol Community Heal*. 2004;58(2):126–128. doi:10.1136/jech.58.2.126.

206. Thies HW, Nolte I, Wenk H, Mertens F, Bullerdiek J, Markowski DN. Permanent activation of HMGA2 in lipomas mimics its temporal physiological activation linked to the gain of adipose tissue. *Obesity (Silver Spring)*. 2014;22(1):141–50. doi:10.1002/oby.20137.

207. Li Z, Gilbert JA, Zhang Y, et al. An HMGA2-IGF2BP2 axis regulates myoblast proliferation and myogenesis. *Dev Cell*. 2012;23(6):1176–88. doi:10.1016/j.devcel.2012.10.019.

208. Orr AW, Pedraza CE, Pallero MA, et al. Low density lipoprotein receptor-related protein is a calreticulin coreceptor that signals focal adhesion disassembly. *J Cell Biol*. 2003;161(6):1179–89. doi:10.1083/jcb.200302069.

209. Lillis AP, Van Duyn LB, Murphy-Ullrich JE, Strickland DK. LDL receptor-related protein 1: unique tissue-specific functions revealed by selective gene knockout studies. *Physiol Rev*. 2008;88(3):887–918. doi:10.1152/physrev.00033.2007.

210. Liu Q, Zhang J, Zerbinatti C, et al. Lipoprotein receptor LRP1 regulates leptin signaling and energy homeostasis in the adult central nervous system. *PLoS Biol*. 2011;9(1):e1000575. doi:10.1371/journal.pbio.1000575.

211. Okada Y, Kamatani Y, Takahashi A, et al. A genome-wide association study in 19 633 Japanese subjects identified LHX3-QSOX2 and IGF1 as adult height loci. *Hum Mol Genet*. 2010;19(11):2303–12. doi:10.1093/hmg/ddq091.

212. Dupuis J, Langenberg C, Prokopenko I, 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. doi:10.1038/ng.520.

213. Teh AL, Pan H, Chen L, et al. The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. *Genome Res*. 2014;24(7):1064–74. doi:10.1101/gr.171439.113.

214. Fatemi M, Pao MM, Jeong S, et al. Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. *Nucleic Acids Res*. 2005;33(20):e176. doi:10.1093/nar/gni180.

215. Ueki M, Cordell HJ. Improved statistics for genome-wide interaction analysis. *PLoS Genet*. 2012;8(4):e1002625. doi:10.1371/journal.pgen.1002625.

216. Whincup PH, Kaye SJ, Owen CG, et al. Birth weight and risk of type 2 diabetes: a systematic review. *JAMA*. 2008;300(24):2886–97. doi:10.1001/jama.2008.886.

217. Hattersley AT, McCarthy MI. What makes a good genetic association study? *Lancet*. 2005;366(9493):1315–23. doi:10.1016/S0140-6736(05)67531-9.

218. Gordis L. *Epidemiology*. 5th editio. Elsevier; 2014. Available at: https://elsevier.ca/product.jsp?isbn=9781455737338. Accessed May 11, 2015.

219. Wax JR, Lucas FL, Lamont M, Pinette MG, Cartin A, Blackstone J. Maternal and newborn outcomes in planned home birth vs planned hospital births: a metaanalysis. *Am J Obstet Gynecol*. 2010;203(3):243.e1–8. doi:10.1016/j.ajog.2010.05.028.

220. Reading R, Raybould S, Jarvis S. Deprivation, low birth weight, and children’s height: a comparison between rural and urban areas. *BMJ*. 1993;307(6917):1458–1462. doi:10.1136/bmj.307.6917.1458.

221. Ray JG, Jiang D, Sgro M, Shah R, Singh G, Mamdani MM. Thresholds for small for gestational age among newborns of East Asian and South Asian ancestry. *J Obstet Gynaecol Can*. 2009;31(4):322–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19497151. Accessed May 5, 2015.

222. Aschard H, Vilhjálmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am J Hum Genet*. 2015;96(2):329–39. doi:10.1016/j.ajhg.2014.12.021.

223. Mefford J, Witte JS. The Covariate’s Dilemma. *PLoS Genet*. 2012;8(11):e1003096. doi:10.1371/journal.pgen.1003096.

224. Meyre D. Is FTO a type 2 diabetes susceptibility gene? *Diabetologia*. 2012;55(4):873–6. doi:10.1007/s00125-012-2478-4.

225. Zeggini E, Morris A, eds. *Analysis of Complex Disease Association Studies: A Practical Guide*. Academic Press; 2010. Available at: https://books.google.com/books?id=jD8ST0\_kwakC&pgis=1. Accessed May 5, 2015.

226. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet*. 2003;361(9357):598–604. doi:10.1016/S0140-6736(03)12520-2.

227. Barfield RT, Almli LM, Kilaru V, et al. Accounting for population stratification in DNA methylation studies. *Genet Epidemiol*. 2014;38(3):231–41. doi:10.1002/gepi.21789.

228. Adkins RM, Krushkal J, Tylavsky FA, Thomas F. Racial differences in gene-specific DNA methylation levels are present at birth. *Birth Defects Res A Clin Mol Teratol*. 2011;91(8):728–36. doi:10.1002/bdra.20770.

229. Heyn H, Moran S, Hernando-Herraez I, et al. DNA methylation contributes to natural human variation. *Genome Res*. 2013;23(9):1363–72. doi:10.1101/gr.154187.112.

230. Pembrey ME, Bygren LO, Kaati G, et al. Sex-specific, male-line transgenerational responses in humans. *Eur J Hum Genet*. 2006;14(2):159–66. doi:10.1038/sj.ejhg.5201538.

231. Kerkel K, Spadola A, Yuan E, et al. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet*. 2008;40(7):904–8. doi:10.1038/ng.174.

232. Schalkwyk LC, Meaburn EL, Smith R, et al. Allelic skewing of DNA methylation is widespread across the genome. *Am J Hum Genet*. 2010;86(2):196–212. doi:10.1016/j.ajhg.2010.01.014.

233. Bittles AH. Population stratification and genetic association studies in South Asia. *J Mol Genet Med*. 2005;1(2):43–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2702070&tool=pmcentrez&rendertype=abstract. Accessed July 29, 2015.

234. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*. 2008;9(6):477–85. doi:10.1038/nrg2361.

235. Wall JD, Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet*. 2003;4(8):587–97. doi:10.1038/nrg1123.

236. Ma X, Wright J, Dou S, et al. Ethnic divergence and linkage disequilibrium of novel SNPs in the human NLI-IF gene: evidence of human origin and lack of association with tuberculosis susceptibility. *J Hum Genet*. 2002;47(3):140–5. doi:10.1007/s100380200016.

237. Tai ES, Corella D, Deurenberg-Yap M, et al. Differential effects of the C1431T and Pro12Ala PPARgamma gene variants on plasma lipids and diabetes risk in an Asian population. *J Lipid Res*. 2004;45(4):674–85. doi:10.1194/jlr.M300363-JLR200.

238. Sanghera DK, Nath SK, Ortega L, et al. TCF7L2 polymorphisms are associated with type 2 diabetes in Khatri Sikhs from North India: genetic variation affects lipid levels. *Ann Hum Genet*. 2008;72(Pt 4):499–509. doi:10.1111/j.1469-1809.2008.00443.x.

239. Rees SD, Bellary S, Britten AC, et al. Common variants of the TCF7L2 gene are associated with increased risk of type 2 diabetes mellitus in a UK-resident South Asian population. *BMC Med Genet*. 2008;9(1):8. doi:10.1186/1471-2350-9-8.

240. Sanghera DK, Been L, Ortega L, et al. Testing the association of novel meta-analysis-derived diabetes risk genes with type II diabetes and related metabolic traits in Asian Indian Sikhs. *J Hum Genet*. 2009;54(3):162–8. doi:10.1038/jhg.2009.7.

241. Sanghera DK, Demirci FY, Been L, et al. PPARG and ADIPOQ gene polymorphisms increase type 2 diabetes mellitus risk in Asian Indian Sikhs: Pro12Ala still remains as the strongest predictor. *Metabolism*. 2010;59(4):492–501. doi:10.1016/j.metabol.2009.07.043.

242. Vimaleswaran KS, Radha V, Jayapriya MG, et al. Evidence for an association with type 2 diabetes mellitus at the PPARG locus in a South Indian population. *Metabolism*. 2010;59(4):457–62. doi:10.1016/j.metabol.2009.07.034.

243. Chavali S, Mahajan A, Tabassum R, et al. Association of variants in genes involved in pancreatic β-cell development and function with type 2 diabetes in North Indians. *J Hum Genet*. 2011;56(10):695–700. doi:10.1038/jhg.2011.83.

244. Been LF, Ralhan S, Wander GS, et al. Variants in KCNQ1 increase type II diabetes susceptibility in South Asians: a study of 3,310 subjects from India and the US. *BMC Med Genet*. 2011;12(1):18. doi:10.1186/1471-2350-12-18.

245. Ramya K, Radha V, Ghosh S, Majumder PP, Mohan V. Genetic variations in the FTO gene are associated with type 2 diabetes and obesity in south Indians (CURES-79). *Diabetes Technol Ther*. 2011;13(1):33–42. doi:10.1089/dia.2010.0071.

246. Been LF, Hatfield JL, Shankar A, et al. A low frequency variant within the GWAS locus of MTNR1B affects fasting glucose concentrations: genetic risk is modulated by obesity. *Nutr Metab Cardiovasc Dis*. 2012;22(11):944–51. doi:10.1016/j.numecd.2011.01.006.

247. Ali S, Chopra R, Manvati S, et al. Replication of type 2 diabetes candidate genes variations in three geographically unrelated Indian population groups. Dewan A, ed. *PLoS One*. 2013;8(3):e58881. doi:10.1371/journal.pone.0058881.

248. Tariq K, Malik SB, Ali SHB, et al. Association of Pro12Ala polymorphism in peroxisome proliferator activated receptor gamma with proliferative diabetic retinopathy. *Mol Vis*. 2013;19:710–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3611930&tool=pmcentrez&rendertype=abstract. Accessed February 16, 2014.

# Appendices

**Supplementary Table 2.1** Full search strategy for systematic review

|  |
| --- |
| 1. diabetes.mp. or exp Diabetes Mellitus, Type 2/  2. exp Phenotype/ or exp Genetic Variation/ or exp Genetic Predisposition to Disease/ or exp Genotype/ or genetic variants.mp. or exp Polymorphism, Single Nucleotide/  3. exp Asian Continental Ancestry Group/ or south asian.mp.  4. ASIA$.mp.  5. INDIA$.mp. or India/  6. BANGLADESH$.mp. or Bangladesh/  7. PAKISTAN$.mp. or Pakistan/  8. SRI LANKA$.mp. or Sri Lanka/  9. CEYLON$.mp. or Sri Lanka/  10. exp India/ or india\*.mp.  11. bengali\*.mp.  12. indo\*.mp.  13. gujarat\*.mp.  14. sikh\*.mp.  15. sind\*.mp.  16. genome wide association study.mp. or Genetic Markers/ or exp Genome-Wide Association Study/ or Disease Susceptibility/ or Pedigree/ or Genetic Linkage/ or Genome, Human/  17. candidate gene\*.mp.  18. exp Mutation/ or mutation.mp.  19. (south adj2 asia\*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]  20. niddm.mp.  21. type 2 diabetes.mp.  22. genetic predisposition.mp. or exp Genetic Predisposition to Disease/  23. polymorphism.mp.  24. genotype.mp.  25. allelic frequenc\*.mp.  26. genetic linkage.mp.  27. genetic marker\*.mp.  28. phenotype.mp.  29. genetic varia\*.mp.  30. single nucleotide polymorphism.mp.  31. punjab\*.mp.  32. tamil.mp.  33. goa.mp.  34. kerala.mp.  35. sikkim.mp.  36. haryana.mp.  37. 2 or 16 or 17 or 18 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30  38. rajasthan\*.mp.  39. bihar\*.mp.  40. kashmir\*.mp.  41. maharashtra\*.mp.  42. pradesh\*.mp.  43. bharat\*.mp.  44. 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 19 or 31 or 32 or 33 or 34 or 35 or 36 or 38 or 39 or 40 or 41 or 42 or 43  45. type 2 diabet\*.mp.  46. 1 or 20 or 21 or 45  47. 37 and 44 and 46  48. limit 47 to animals  49. 47 not 48 |

**Supplementary Table 2.2** Reported SNPs and their proxies in LD (*r*2 > 0.8) determined using SNAP

|  |  |  |  |
| --- | --- | --- | --- |
| **Gene** | **Reported SNP** | **Lead SNP** | ***r*2** |
| *CDKAL1* | rs6931514 | rs7756992 | 0.917 |
| rs10946398 | rs7754840 | 1 |
| rs9295474 | rs7754840 | 0.885 |
| *FTO* | rs1421085 | rs9939609 | 0.901 |
| rs7193144 | rs9939609 | 0.967 |
| rs8050136 | rs9939609 | 1 |
| rs9930506 | rs9939609 | 0.839 |
| rs3751812 | rs9939609 | 1 |
| *HHEX* | rs1111875 | rs5015480 | 0.966 |
| *HNF4A* | rs1884613 | rs4812829 | 0.848 |
| rs2144908 | rs4812829 | 1 |
| rs1884614 | rs4812829 | 1 |
| *KCNJ11* | rs5215 | rs5219 | 0.935 |
| *KCQN1* | rs2283228 | rs2337892 | 0.867 |
| *SLC30A8* | rs11558471 | rs13266634 | 0.960 |
| *TCF7L2* | rs4506565 | rs7903146 | 0.892 |
| rs7901695 | rs7903146 | 0.892 |
| *IGF2BP2* | rs1470579 | rs4402960 | 1 |
| *CHCHD9* | rs4295736 | rs13292136 | 1 |
| *WFS1* | rs1801214 | rs10010131 | 1 |

**Supplementary Table 2.3** Characteristics of studies included in the systematic review

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Reference** | **Region of origin** | **N cases** | **N controls** | **Mean age cases** | **Mean age controls** | **Mean fasting glucose cases (mmol/l)** | **Mean fasting glucose controls (mmol/l)** | **Mean BMI cases** | **Mean BMI controls** | **Mean WC cases** | **Mean WC controls** | **Mean HC cases** | **Mean HC controls** | **WHR cases** | **WHR controls** | **% males cases** | **% males controls** |
| Tai J Lipid Res 2004 237 | Singaporean Indians | 108 | 305 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| Radha Diabetes Care 2006 88 | Not specified (US cohort) | 81 | 616 | 56 (10) | 42 (13) | 8.93 (3.6) | 5.33 (1.3) | 25.6 (3.7) | 24.9 (3.7) | - | - | - | - | - | - | - | - |
| South India | 799 | 820 | 52 (11) | 41 (13) | 8.99 (3.8) | 4.72 (0.4) | 25.1 (4.2) | 23.4 (4.7) | - | - | - | - | - | - | - | - |
| Humphries J Mol Med 2006 87 | Not specified | 841 | 302 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| Chandak Diabetologia 2007 77 | North India | 955 | 399 | 47.2 (9.3) | 30.9 (5.1) | - | 5.1 (4.7–5.5)\* | Men: 25.4 (3.6) Women: 27.1 (3.9) | Men: 20.5 (3.1)  Women: 19.1 (2.5) | - | - | - | - | Men: 0.98 (0.06), women: 0.89 (0.06) | Men: 0.90 (0.06), women: 0.76 (0.05) | 53.7 | 46.6 |
| Bodhini Clin Exp Met 2007 84 | South India | 1031 | 1038 | 49 (10) | 41 (11) | 9.2 (4.0) | 4.6 (0.4) | 25.1 (4.3) | 23.6 (4.6) | 90.4 (10.0) | 83.7 (11.6) | - | - | - | - | - | - |
| Sanghera BMC Med Genet 2008 94 | North India | 532 | 386 | Men: 53.2 (11.1), Women: 55.2 (11.0) | Men: 51.7 (15.6), Women: 50.8 (13.0) | Men: 10.1 (3.7) Women: 10.0 (3.5) | Men: 5.4 (0.6) Women: 5.4 (0.5) | Men: 26.6 (4.3); Women: 29.0 (5.5) | Men: 26.6 (4.4), Women: 27.5 (4.9) | - | - | - | - | Men: 0.99 (0.06), Women: 0.93 (0.07) | Men: 0.97 (0.07), Women: 0.91 (0.07) | 56.2 | 47.7 |
| Sanghera Ann Hum Genet 2008 238 | North India | 556 | 537 | Men: 56.22 (10.9), Women: 55.05 (11.0) | Men: 52.58 (15.7), Women: 51.63 (13.0) | Men: 9.89 (3.62), Women: 10.0 (3.45) | Men: 5.36 (0.56), Women: 5.40 (0.57) | Men: 26.42 (4.3), Women: 29.10 (5.5) | Men: 26.65 (4.4), Women: 27.40 (4.9) | - | - | - | - | Men: 0.99 (0.06), Women: 0.93 (0.07) | Men: 0.97 (0.07), Women: 0.90 (0.07) | 55.6 | 48.4 |
| Rees BMC Med Genet 2008 239 | Pakistan (Punjab) | 831 | 437 | 56.9 (12.1) | 55.0 (11.8) | - | - | 28.3 (4.7) | 28.1 (4.9) | 102.4 (10.7) | 99.8 (13.1) | - | - | - | - | - | - |
| Sanghera J Human Genet 2009 240 | North India (Punjab) | 680 | 637 | men: 55.8 (11.2) women: 56.9 (10.8) | men: 51.1 (15.1), women: 50.2 (13.2) | men: 10.3 (3.7), women: 10.6 (3.8) | men: 5.4 (0.7), women: 5.4 (0.7) | men: 26.5 (4.7), women: 28.0 (5.0) | men: 26.3 (4.5), women: 27.0 (5.1) | men: 94.8 (10.5), women: 92.8 (11.4) | men: 92.3 (11.5), women: 87.6 (10.8) | men: 96.0 (8.4), women: 99.5 (10.3) | men: 96.2 (8.8), women: 97.0 (10.2) | men: 0.99 (0.31), women: 0.93 (0.15) | men: 0.96 (0.09), women: 0.90 (0.16) | - | - |
| Yajnik Diabetologia 2009 90 | North India | 1453 | 1361 | 46.6 (9.3) | 34.5 (6.1) | 8.50 (6.89-11.28)\* | 5.06 (4.61-5.56)\* | Men: 25.4 (4.0), Women: 27.2 (4.3) | Men: 21.9 (3.6), Women: 20.9 (4.1) | Men: 95.4 (10.7), Women: 92.4 (10.3) | Men: 83.2 (10.5), Women: 69.9 (10.4) | Men: 97.3 (7.4), Women: 103.2 (10.2) | Men: 91.6 (7.5), Women: 90.9 (9.3) | Men: 0.98 (0.06), Women: 0.89 (0.06) | Men: 0.91 (0.06), Women: 0.77 (0.06) | 56.3 | 53.6 |
| Haseeb J Biosci 2009 75 | South India | 350 | 349 | 61.8 (11.25) | 61.93 (10.40) | - | - | - | - | - | - | - | - | - | - | - | - |
| Chauhan Diabetes 2010 92 | North India (Pune) | 1467 | 1672 | 46 (40–52)\* | 33 (29–37)\* | 8.50 (6.90–11.30)\* | 5.11 (4.67–5.56)\* | Men: 24.90 (22.80–27.70), Women: 26.90 (24.40–29.60)\* | Men: 21.18 (19.15–23.62), Women: 19.53 (17.60–22.74)\* | - | - | - | - | Men: 0.97 (0.94-1.02), Women: 0.89 (0.85-0.94)\* | Men: 0.91 (0.86-0.95), Women: 0.76 (0.73-0.80)\* | 56.3 | 52.9 |
| North India (Delhi) | 1019 | 1006 | 53 (45–62)\* | 50 (44–60)\* | 7.90 (6.40–10.30)\* | 4.90 (4.50–5.30)\* | Men: 23.80 (22.00–26.00), Women: 26.70 (24.20–29.20)\* | Men: 23.20 (20.20–25.70), Women: 24.90 (21.10–28.60)\* | - | - | - | - | Men: 1.00 (0.97–1.03), Women: 1.00 (0.97–1.03)\* | Men: 0.97 (0.92–1.00), Women: 0.86 (0.82–0.92)\* | 58.1 | 60.2 |
| Gupta Ann Hum Genet 2010 91 | North India | 219 | 184 | Men: 59.12 (10.34), Women: 55.65 (10.44) | Men: 55.52 (10.61), Women: 51.99 (9.37) | - | - | Men: 27.18 (5.75), Women: 30.08 (4.91) | Men: 27.48 (7.00), Women: 29.64 (5.27) | - | - | - | - | Men: 1.01 (0.09), Women: 0.92 (0.08) | Men: 1.00 (0.08), Women: 0.89 (0.07) | 66.2 | 51.6 |
| Chidambaram Metabolism 2010 78 | South India | 926 | 812 | 52 (11) | 38 (12) | 8.97 (4.13) | 4.68 (0.44) | 25 (4) | 23 (4) | - | - | - | - | - | - | - | - |
| Mukhopadhyaya Genet Mol Res 2010 79 | North India | 40 | 40 | Men: 46.0 (14.6), Women: 48.6 (9.0) | Men: 42.29 (13.41), Women: 45 (8.83) | Men: 10.25 (1.88), Women: 10.28 (1.47) | Men: 5.12 (0.94); Women: 5.14 (0.73) | Men: 31.9 (5.0), Women: 31.9 (4.0) | Men: 27.9 (4.9), Women: 28.4 (4.3) | - | - | - | - | Men: 0.99 (0.06), Women: 1.01 (0.08) | Men: 0.97 (0.06), Women: 0.99 (0.08) | 52.5 | 52.5 |
| Sanghera Metabolism 2010 241 | North India | 554 | 527 | Men: 56.1 (11.1), Women: 55.2 (11.1) | Men: 51.8 (15.6), Women: 51.5 (13.3) | Men: 9.4 (3.3), Women: 9.6 (3.1) | Men: 5.4 (0.6), Women: 5.5 (0.6) | Men: 26.6 ± 4.4; Women: 29.0± 5.4 | Men: 26.7 (4.2), Women: 27.5 (4.8) | - | - | - | - | Men: 0.99 (0.08), Women: 0.93 (0.07) | Men: 0.97 (0.08), Women: 0.91 (0.07) | 55.8 | 49.0 |
| Vimaleswaran Met 2010 242 | South India | 1000 | 1000 | 52 (11) | 46 (12) | - | - | 26.1 (4.2) | 24.0 (4.7) | 92.3 (9.4) | 87.2 (11.4) | - | - | - | - | - | - |
| Tan J Clin End 2010 86 | Singaporean Indians | 246 | 364 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| Rees Diabet Med 2011 83 | Pakistan | 385 | 1281 | 53.5 (10.7) | 51.1 (10.7) | 10.6 (4.0) | 5.3 (0.6) | 26.7 (5.6) | 25.2 (5.2) | 93.2 (11.7) | 88.1 (12.0) | - | - | - | - | 40.0 | 46.3 |
| North India and Pakistan | 1568 | 1177 | 55.8 (11.9) | 56.4 (10.7) | - | 5.4 (0.6) | 27.5 (4.8) | 25.7 (5.2) | 99.8 (11.4) | 93.8 (13.0) | - | - | - | - | 51.6 | 50.0 |
| Rees PloS One 2011 112 | Pakistan | 857 | 417 | 56.9 (12.0) | 54.9 (11.7) | - | - | 28.6 (4.6) | 28.0 (4.9) | - | - | - | - | - | - | 45.3 | 52.0 |
| Pakistan | 821 | 1167 | 54.6 (11.7) | 56.3 (10.8) | - | 5.5 (0.6) | 26.1 (4.7) | 24.3 (5.0) | - | - | - | - | - | - | 52.4 | 52.9 |
| Chavali J Human Genet 2011 243 | North India | 1019 | 1006 | 53 (45-62)\* | 50 (44-60)\* | 7.9 (6.4-10.3)\* | 4.9 (4.5-5.3)\* | Men: 23.80 (22.00-26.00), Women:26.70 (24.20-29.20)\* | median: men=23.10 (20.10-25.70), women=25.00 (21.10-28.50) | - | - | - | - | Men: 1.00 (0.97-1.03), Women: 1.00 (0.97-1.03)\* | Men: 0.97 (0.92-1.00), Women: 0.86 (0.82-0.92)\* | 58.1 | 60.2 |
| Boodram West Indian Med J 2011 89 | Indo-Trinidadian | 168 | 61 | 45.11 (12.03) | 51.68 (17.22) | - | - | 26.05 (4.47) | 25.08 (5.03) | - | - | - | - | - | - | 27.4 | 40.5 |
| Rees Diabetologia 2011 81 | Pakistan | 1678 | 1584 | UKADS: 56.90 (12.30), DGP: 54.62 (11.67) | UKADS: 54.92 (11.75), DGP: 56.27 (10.81) | - | DGP: 4.75 (0.47) | UKADS: 28.56 (4.61), DGP: 26.07 (4.72) | UKADS: 28.01 (4.86), DGP: 24.30 (5.03) | UKADS: 102.31 (10.48), DGP: 96.78 (11.88) | UKADS: 99.75 (11.52), DGP: 91.91 (13.13) | - | - | - | - | UKADS: 54.7%, DGP: 47.6% | UKADS: 47.96%, DGP: 47.13% |
| Sim PLoS Genet 2011 74 | Singaporean Indians | 977 | 1169 | 60.71 (9.85) | 55.73 (9.72) | - | - | 27.06 (5.10) | 25.33 (4.40) | - | - | - | - | - | - | 54.4 | 48.4 |
| Chauhan J Hum Genet 2011 93 | North India (replication cohort) | 1401 | 1848 | 55 (48–62)\* | 52 (45–62)\* | 8.05 (6.39–10.60)\* | 4.82 (4.41–5.20)\* | Men: 25.40 (22.94–28.36), Women: 27.34 (24.46–31.00)\* | Men: 24.69 (22.15–27.35), Women: 26.30 (23.11–29.23)\* | Women: 86.68 (40.00-98.00), Men: 90.00 (36.00-98.00)\* | Women: 88.00 (81.00-95.50), Men: 93.00 (86.00-100.00)\* | - | - | Women: 0.95 (0.89-0.98), Men: 0.98 (0.96-1.03)\* | Men: 0.97 (0.92-1.00), Women: 0.87 (0.82-0.91)\* | 60.7 | 55.0 |
| North India | 1019 | 1006 | 53 (45-62)\* | 50 (44-60)\* | 7.90 (6.40–10.30)\* | 4.90 (4.50–5.30)\* | Men: 23.80 (22.00–26.00), Women: 26.70 (24.20–29.20)\* | Men: 23.20 (20.20–25.70), Women: 24.90 (21.10–28.60)\* | Men: 86.36 (86.36–91.44), Women: 91.44 (86.36–96.52)\* | Men: 88.50 (80.64–95.00), Women: 85.00 (75.60–93.00)\* | - | - | Men: 1.0 (0.97-1.03), Women: 1.0 (0.97-1.03)\* | Men: 0.97 (0.92–1.00), Women: 0.86 (0.82–0.92)\* | 58.1 | 60.2 |
| Anuradha Clin Genet 2011 76 | South India (Chennai) | 792 (505 Early onset) (287 Late onset) | 247 | Early onset: 34 (4), Late onset: 52 (7) | 58 (7) | Early onset: 9.5 mmol (3.9), Late onset: 8.5 (3.6) | 4.7 (0.4) | Early onset: 25.7 (4.2), Late onset: 24.8 (4.3) | 22.4 (0.7) | - | - | - | - | - | - | - | - |
| Kooner Nat Genet 2011 71 |  | 18731 | 39856 | LOLIPOP 610: 59.4 (9.2); LOLIPOP 317: 54.1 (10.1); SINDI: 60.7 (9.9); PROMIS: 55.0 (9.4) | LOLIPOP 610: 53.9 (10.7); LOLIPOP 317: 46.8 (10.1); SINDI: 55.7 (9.7); PROMIS: 52.9 (10.5) | LOLIPOP 610: 8.6 (3.1); LOLIPOP 317: 5.2 (0.6); SINDI: 8.9 (2.9); PROMIS: 9.71 (4.44) | LOLIPOP 610: 5.2 (0.6); LOLIPOP 317: 5.1 (0.6); SINDI: 5.38 (1.06); PROMIS: 6.89 (2.91) | LOLIPOP 610: 28.1 (4.6); LOLIPOP 317: 27.6 (4.7); SINDI: 27.1 (5.1); PROMIS: 26.0 (4.0) | LOLIPOP 610: 26.8 (4.2); LOLIPOP 317: 26.6 (4.2); SINDI: 25.3 (4.4); PROMIS: 25.3 (3.9) | LOLIPOP 610: 100.8 (11.5); LOLIPOP 317: 100.0 (12.2); SINDI: - ; PROMIS: 92.0 (12.0) | LOLIPOP 610: 96.6 (10.9); LOLIPOP 317: 96.3 (11.4); SINDI: - ; PROMIS: 90.1 (11.7) | - | - | LOLIPOP 610: 0.99 (0.07); LOLIPOP 317: 0.99 (0.07); SINDI: - ; PROMIS: 0.95 (0.06) | LOLIPOP 610: 0.95 (0.07); LOLIPOP 317: 0.95 (0.07); SINDI: - ; PROMIS: 0.94 (0.07) | LOLIPOP 610: 82.9; LOLIPOP 317: 100.0; SINDI: 54.4; PROMIS: 76.5 | LOLIPOP 610: 84.8; LOLIPOP 317: 100.0; SINDI: 48.4; PROMIS: 83 |
| Been BMC Med Genet 2011 244 | Not specified (US cohort) | 139 | 557 | - | - | 8.10 (2.4) | 5.47 (0.7) | - | - | - | - | - | - | - | - | - | - |
| North India | 1290 | 1019 | - | - | 9.03 (3.5) | 5.27 (0.74) | - | - | - | - | - | - | - | - | - | - |
| Ramya Diabetes Technol Ther 2011 245 | South India | 851 | 1001 | 51 (11) | 41 (13) | 9.0 (3.9) | 4.7 (0.4) | 25.3 (4.3) | 23.4 (4.7) | 91.0 (10.1) | 83.5 (12.1) | 98.0 (10.0) | 94.0 (10.0) | - | - | 44.3 | 41.8 |
| Janipali Diabetic Med 2012 80 | North India | 1808 | 1549 | 51.8 (45.6-58.6) | 40.0 (36.1-44.4) | 8.2 (6.7-11.0) | 5.1 (4.8-5.5) | 25.7 (23.4-28.6) | 19.4 (17.7-21.7) | - | - | - | - | 0.90 (0.88-0.99) | 0.84 (0.77-0.91) | 55.8 | 53.2 |
| Been Nutr Metab 2012 246 | North India | 1201 | 1021 | Men: 54.1 (10.2), Women: 53.7 (9.8) | Men: 51.3 (15.2), Women: 50.1 (13.3) | Men: 9.09 (3.4), Women: 9.06 (3.6) | Men: 5.29 (0.6), Women: 5.20 (0.6) | Men: 26.6 (4.4), Women: 28.4 (5.4) | Men: 25.7 (4.8), Women: 27.1 (6.7) | - | - | - | - | - | - | 52.3 | 52.4 |
| Raza Gene 2012 85 | North India | 87 | 88 | 48.47 (12.16) | 33.79 (12.07) | - | - | 25.80 (3.97) | 23.35 (3.02) | - | - | - | - | - | - | 67.8 | 53.4 |
| Anand Diabetes Care 2013 62 | Not specified | 638 | 2125 | 47.72 (9.40) | 44.14 (9.22) | 7.35 (2.87) | 4.80 (0.78) | 27.0 (4.1) | 26.3 (4.4) | 100.9 (10.2) | 100.1 (10.4) | - | - | - | - | 40.5 | 50.8 |
| Ali PLoS One 2013 247 | North India | 1583 | 1317 | 53.28 (10.13) | 51.47 (12.40) | 8.82 (3.36) | 4.64 (0.66) | 25.86 (4.76) | 24.82 (4.35) | - | - | - | - | - | - | - | - |
| Sexena Diabetes 2013 73 |  | 19482 | 27821 |  |  |  |  |  |  | - | - | - | - | - | - | - | - |
| Tabassum Diabetes 2013 72 | South India | 1545 | 1304 |  |  |  |  |  |  | - | - | - | - | - | - | - | - |
| North India | 5193 | 4493 |  |  |  |  |  |  | - | - | - | - | - | - | - | - |
| Uma Jyothi PLoS One 2013 82 | South India | 758 | 621 | 52.5 (9.08) | 52.2 (7.55) | 6.41 (1.8) | - | 27.07 (4.63) | 24.72 (4.62) | - | - | - | - | 1.69 (0.43) | 0.94 (0.03) | 58.4 | 63.6 |
| Tariq Mol Vis 2013 248 | Pakistan | 373 | 200 | - | - | - | - | - | - | - | - | - | - |  |  | 48.5 | 50.0 |

**Supplementary Table 2.4** Risk of bias assessment for studies included in the systematic review

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reference** | **HWE tested in controls** | **Genotyping call rate > 95%** | **Source of cases** | **Source of controls** |
| Tai J Lipid Res 2004 237 | Y | not reported; 20% re-genotyped with 100% concordance | general population | general population |
| Radha Diabetes Care 2006 88 | Y | not reported | general population | general population |
| Humphries J Mol Med 2006 87 | Y | not reported | WHSS: recruited from general practice clinics, PREDICT: recruited from the diabetes clinics, UCS: recruited from the diabetes clinics | WHSS: recruited from general practice clinics |
| Chandak Diabetologia 2007 77 | Y | not reported; 15% were re-genotyped with 99.9% concordance | recruited from Diabetes Clinic of the King Edward  Memorial Hospital and Research Centre in Pune | parents of children in the Pune Maternal Nutrition Study |
| Bodhini Clin Exp Met 2007 84 | Y | not reported; 20% re-genotyped with 99% concordance | general population | general population |
| Sanghera BMC Med Genet 2008 94 | Y | Y | endogamous Khatri Sikh population | endogamous Khatri Sikh population with no family history of type 2 diabetes; 262 were non-diabetic spouses |
| Sanghera Ann Hum Genet 2008 238 | Y | Y | endogamous Khatri Sikh population | endogamous Khatri Sikh population with no family history of type 2 diabetes; 262 were non-diabetic spouses |
| Rees BMC Med Genet 2008 239 | Y | Y | not specified | same geographical areas through community screening |
| Sanghera J Human Genet 2009 240 | Y | Y | endogamous Khatri Sikh population | endogamous Khatri Sikh population with no family history of type 2 diabetes; 262 were non-diabetic spouses |
| Yajnik Diabetologia 2009 90 | Y | not reported | recruited from Diabetes Clinic of the King Edward  Memorial Hospital and Research Centre in Pune | parents of children in the Pune Maternal Nutrition Study from urban and rural regions, parents in Pune Children Study, and Coronary Risk of Insulin Sensitivity in Indian Subjects study |
| Haseeb J Biosci 2009 75 | Y | not reported; subset re-genotyped | recruited from Mediciti hospital | recruited from Mediciti hospital |
| Chauhan Diabetes 2010 92 | Y | >90%, 10% with concordance of 99.7% | Delhi: consecutively recruited from the Endocrinology clinic of All India Institute of Medical Sciences; Pune: general population in Pune and surrounding areas | Delhi: urban dwellers of Indo-European ethnicity with no family history of diabetes in ﬁrst and/or second degree relatives; Pune: parents of children in the Pune Maternal Nutrition Study and Coronary Risk of Insulin Sensitivity in Indian Subjects study |
| Gupta Ann Hum Genet 2010 91 | Y | not reported | Aggarwal population (60% had family history of type 2 diabetes) | same geographic region as cases |
| Chidambaram Metabolism 2010 78 | Y | not reported; 20% re-genotyped with 99.6% concordance | general population | general population |
| Mukhopadhyaya Genet Mol Res 2010 79 | Y | Y | pre-specified sub-population | same population as cases |
| Sanghera Metabolism 2010 241 | Y | Y | endogamous Khatri Sikh population | endogamous Khatri Sikh population with no family history of type 2 diabetes; 262 were non-diabetic spouses |
| Vimaleswaran Met Clin Exp 2010 242 | Y | not reported; 20% re-genotyped with 100% concordance | general population | general population |
| Tan J Clin End 2010 86 | Y (rs2237892 and rs2237897 deviated from HWE p<0.01) | >90% call rate; 30 samples re-genotyped with 100% concordance | SDCS: recruited from Singapore National Healthcare Group Polyclinics; NHS98: general population | NHS98: general population |
| Rees Diabet Med 2011 83 | Y | Y | Cobra: community sampling; UKADS: not specified; DGP: hospitals in Mirpur | Cobra: community sampling; UKADS: same geographic region; DGP: community screening |
| Rees PloS One 2011 112 | Y | Y | UKADS: not specified; DGP: hospitals in Mirpur | UKADS: same geographic region; DGP: community screening |
| Chavali J Human Genet 2011 243 | Y | Y | consecutively recruited from the Endocrinology clinic of All India Institute of Medical Sciences | urban dwellers of Indo-European ethnicity with no family history of diabetes in ﬁrst and/or second degree relatives |
| Boodram West Indian Med J 2011 89 | Y | not reported | recruited from diabetic outpatient clinics | recruited from Chest Clinic at Mt. Hope Hospital (Champs Fleurs, Trinidad and Tobago) and from San Fernando General Hospital |
| Rees Diabetologia 2011 81 | Y | Y | UKADS: not specified; DGP: hospitals in Mirpur | UKADS: same geographic region; DGP: community screening |
| Sim PLoS Genet 2011 74 | Y | Y | randomly sampled from general population | randomly sampled from general population |
| Chauhan J Hum Genet 2011 93 | Y | >90%, 5% with concordance of 99.99% | consecutively recruited from the Endocrinology clinic of All India Institute of Medical Sciences | recruited from diabetes awareness camps |
| Anuradha Clin Genet 2011 76 | Y | not reported; 20% re-genotyped with 99% concordance | general population | general population |
| Kooner Nat Genet 2011 71 | Y | Y | LOLIPOP: recruited from lists of GPs in West London; PROMIS: out-patient departments; SINDI: general population; COBRA: general population; DGP: hospitals or Diabetes awareness camps; CURES: general population; Mauritius cohort: population based survey; RHS: population based survey used to participate in study; SDS: endogamous Khatri Sikh population; SCCS: recruited from hospitals and polyclinics | LOLIPOP: recruited from lists of GPs in West London; PROMIS: matched cases from visitors in out-patient; SINDI: general population; COBRA: general population; DGP: recruited from community screening camps; CURES: random sampling from general population; Mauritius cohort: population based survey; RHS: population based survey; SDS: endogamous Khatri Sikh population with no family history of type 2 diabetes, 262 were non-diabetic spouses; SCCS: general population |
| Been BMC Med Genet 2011 244 | Y | Y | SDS: endogamous Khatri Sikh population; US cohort: not specified | SDS: endogamous Khatri Sikh population with no family history of type 2 diabetes, 262 were non-diabetic spouses; US cohort: public advertisement for free health screening |
| Ramya Diabetes Technol Ther 2011 245 | Y | not reported; 20% re-genotyped with 99% concordance | general population | general population |
| Janipali Diabetic Med 2012 80 | Y | Y | general population in Pune and surrounding areas | parents of children in the Pune Maternal Nutrition Study from urban and rural regions and Coronary Risk of Insulin Sensitivity in Indian Subjects study |
| Been Nutr Metab 2012 246 | Y | Y | endogamous Khatri Sikh population | endogamous Khatri Sikh population with no family history of type 2 diabetes; 262 were non-diabetic spouses |
| Raza Gene 2012 85 | not reported | not reported | recruited from a diabetic clinic | recruited from same diabetic clinic with no history of type 2 diabetes |
| Anand Diabetes Care 2013 62 | Y | Y | Individuals at risk for dysglycemia recruited from 191 centres around the world via a variety of methods | Sample population as cases |
| Ali PLoS One 2013 247 | Y | Y | - | - |
| Saxena Diabetes 2013 73 | Y | Y | LOLIPOP: recruited from lists of GPs in West London; PROMIS: out-patient departments; SINDI: general population; DGP: hospitals or Diabetes awareness camps; CURES: general population; SDS: endogamous Khatri Sikh population; RACE: recruited from six hospital centres in Pakistan; UKADS: general population; SLDS: general population | LOLIPOP: recruited from lists of GPs in West London; PROMIS: matched cases from visitors in out-patient; SINDI: general population; DGP: recruited from community screening camps; CURES: random sampling from general population; SDS: endogamous Khatri Sikh population with no family history of type 2 diabetes, 262 were non-diabetic spouses RACE: recruited from six hospital centres in Pakistan; UKADS: general population (same geographical area); SLDS: general population |
| Tabassum Diabetes 2013 72 | Y | Y | INDICO: consecutively recruited from the Endocrinology clinic of All India Institute of Medical Sciences; CURES: general population | INDICO: recruited from diabetes awareness camps; CURES: general population |
| Uma Jyothi PLoS One 2013 82 | Y | Y | recruited from J.P. Endocrine center | community diabetic camp |
| Tariq Mol Vis 2013 248 | Y | not reported; 10% replicated with 100% concordance | recruited from hospitals | not specified |

**Supplementary Table 3.1** Risk allele frequencies and HWE-P values for SNPs available in the EpiDREAM cohort

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **White Caucasians (n=2,651)** | | | | | **South Asians (n=2,651)** | | | | |
| **Gene** | **SNP** | **Proxy SNP** | **Risk allele** | **RAF** | **Genotype distribution** | **HWE**  ***P*** | **Proxy SNP** | **Risk allele** | **Genotype distribution** | **RAF** | **HWE**  ***P*** |
|  |  |  |  |  |  |  |  |  |  |  |  |
| *GIPR* | rs10423928 |  | T | 0.80 | 118/848/1685 | 0.403 |  | T | 51/666/1934 | 0.86 | 0.530 |
| *SLC2A2* | rs11920090 | rs10513685 | G | 0.86 | 50/639/1962 | 0.871 | rs10513688 | G | 33/570/2048 | 0.88 | 0.407 |
| *MTNR1B* | rs10830963 |  | G | 0.69 | 259/1109/1283 | 0.386 |  | G | 486/1263/902 | 0.58 | 0.232 |
| *SLC30A8* | rs13266634 |  | C | 0.72 | 217/1073/1361 | 0.775 |  | C | 126/899/1625 | 0.78 | 0.909 |
| *FADS1* | rs174550 | rs174535 | T | 0.67 | 292/1189/1169 | 0.727 |  | T | 30/512/2109 | 0.89 | 1.000 |
| *MTNR1B* | rs2166706 | rs10830962 | C | 0.43 | 481/1309/856 | 0.634 |  | G | 624/1261/766 | 0.47 | 0.019 |
| *PDX1* | rs2293941 |  | A | 0.23 | 128/946/1577 | 0.376 |  | A | 83/661/1906 | 0.16 | 0.008 |
| *FOXA2* | rs6048205 |  | A | 0.95 | 9/226/2415 | 0.177 |  | A | 2/146/2460 | 0.97 | 1.000 |
| *TCF7L2* | rs7903146 |  | T | 0.32 | 282/1141/1228 | 0.477 |  | T | 261/1128/1262 | 0.31 | 0.717 |
| *GCK* | rs4607517 | rs6975024 | C | 0.18 | 97/762/1789 | 0.168 | rs2908282 | T | 49/583/2019 | 0.13 | 0.341 |
| *PCSK1* | rs13179048 | rs17085675 | A | 0.72 | 213/1063/1375 | 0.701 | rs4869272 | T | 320/1150/1181 | 0.66 | 0.129 |
|  |  |  |  |  |  |  |  |  |  |  |  |

**Supplementary Table 3.2** Effect of 11 SNPs included in the genotype score on fasting glucose from published GWA-studies

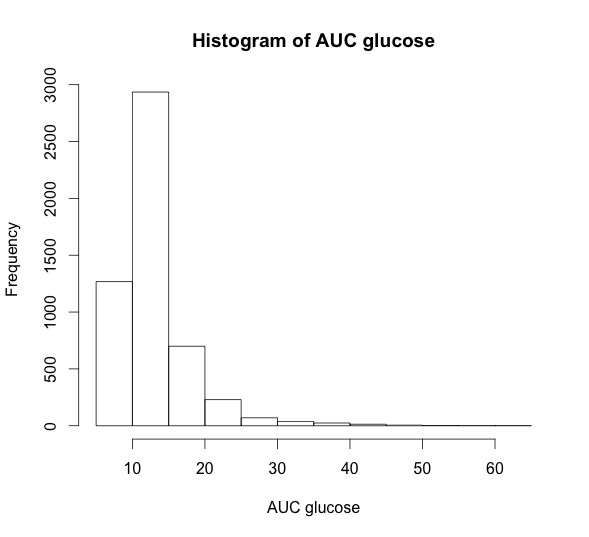
|  |  |  |
| --- | --- | --- |
| **Gene** | **SNP** | **Effect size (in mmol/l)** |
| *FADS1* | rs174550 | 0.017 |
| *FOXA2* | rs6048205 | 0.040 |
| *GCK* | rs4607517 | 0.062 |
| *GIPR* | rs10423928 | 0.090 |
| *IPF1/PDX1* | rs2293941 | 0.019 |
| *MTNR1B* | rs2166706 | 0.070 |
| *MTNR1B* | rs10830963 | 0.067 |
| *PCSK1* | rs13179048 | 0.022 |
| *SLC2A2* | rs11920090 | 0.020 |
| *SLC30A8* | rs13266634 | 0.027 |
| *TCF7L2* | rs7903146 | 0.023 |

**Supplementary Table 3.3** Genotypic effects for SNPs stratified by glucose traits and ethnicity

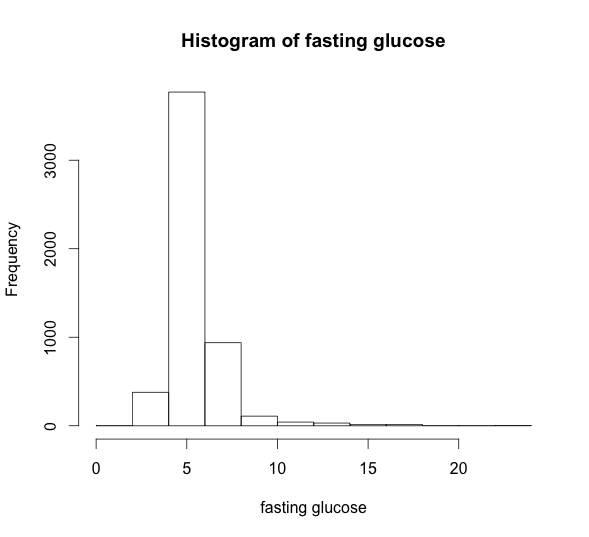
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Gene** | **SNP** | **Glucose trait** | **White Caucasians (n=2,651)** | | | **South Asians (n=2,651)** | | |
| **Proxy SNP** | **β-coefficient** | ***P*** | **Proxy SNP** | **β-coefficient** | ***P*** |
| *GIPR* | rs10423928 | AUC glucose |  | 0.29 | 2.28x10-2 |  | 0.39 | 2.28x10-2 |
| *SLC2A2* | rs11920090 | AUC glucose | rs10513685 | 0.08 | 5.71x10-1 | rs10513688 | -0.11 | 6.40x10-1 |
| *MTNR1B* | rs10830963 | AUC glucose |  | -0.04 | 7.30x10-1 |  | -0.25 | 1.06x10-1 |
| *SLC30A8* | rs13266634 | AUC glucose |  | 0.08 | 4.80x10-1 |  | -0.05 | 8.11x10-1 |
| *FADS1* | rs174550 | AUC glucose | rs174535 | -0.02 | 8.42x10-1 |  | 0.01 | 9.67x10-1 |
| *MTNR1B* | rs2166706 | AUC glucose | rs10830962 | -0.08 | 4.21x10-1 |  | -0.10 | 5.30x10-1 |
| *PDX1* | rs2293941 | AUC glucose |  | -0.01 | 9.29x10-1 |  | 0.18 | 4.08x10-1 |
| *FOXA2* | rs6048205 | AUC glucose |  | 0.43 | 7.64x10-2 |  | -0.34 | 4.77x10-1 |
| *TCF7L2* | rs7903146 | AUC glucose |  | 0.38 | 4.43x10-4 |  | -0.02 | 8.88x10-1 |
| *GCK* | rs4607517 | AUC glucose | rs6975024 | 0.15 | 2.52x10-1 | rs2908282 | -0.08 | 7.29x10-1 |
| *PCSK1* | rs13179048 | AUC glucose | rs17085675 | <0.01 | 9.99x10-1 | rs4869272 | -0.31 | 6.20x10-2 |
| *GIPR* | rs10423928 | Fasting glucose |  | 0.04 | 3.31x10-1 |  | 0.11 | 1.18x10-1 |
| *SLC2A2* | rs11920090 | Fasting glucose | rs10513685 | 0.01 | 8.14x10-1 | rs10513688 | -0.07 | 4.06x10-1 |
| *MTNR1B* | rs10830963 | Fasting glucose |  | -0.06 | 5.86x10-2 |  | -0.07 | 1.93x10-1 |
| *SLC30A8* | rs13266634 | Fasting glucose |  | 0.04 | 2.94x10-1 |  | 0.00 | 9.36x10-1 |
| *FADS1* | rs174550 | Fasting glucose | rs174535 | -0.04 | 2.81x10-1 |  | 0.02 | 8.34x10-1 |
| *MTNR1B* | rs2166706 | Fasting glucose | rs10830962 | -0.04 | 2.48x10-1 |  | -0.04 | 4.51x10-1 |
| *PDX1* | rs2293941 | Fasting glucose |  | 0.03 | 3.87x10-1 |  | 0.03 | 6.88x10-1 |
| *FOXA2* | rs6048205 | Fasting glucose |  | 0.14 | 5.61X10-2 |  | -0.11 | 4.82x10-1 |
| *TCF7L2* | rs7903146 | Fasting glucose |  | 0.10 | 3.18x10-3 |  | -0.02 | 7.02x10-1 |
| *GCK* | rs4607517 | Fasting glucose | rs6975024 | 0.08 | 3.72x10-2 | rs2908282 | 0.02 | 7.71x10-1 |
| *PCSK1* | rs13179048 | Fasting glucose | rs17085675 | 0.03 | 3.33x10-1 | rs4869272 | -0.11 | 4.63x10-2 |

**Supplementary Figure 3.1** Histograms of AUC glucose (a) and fasting glucose (b)

(a)

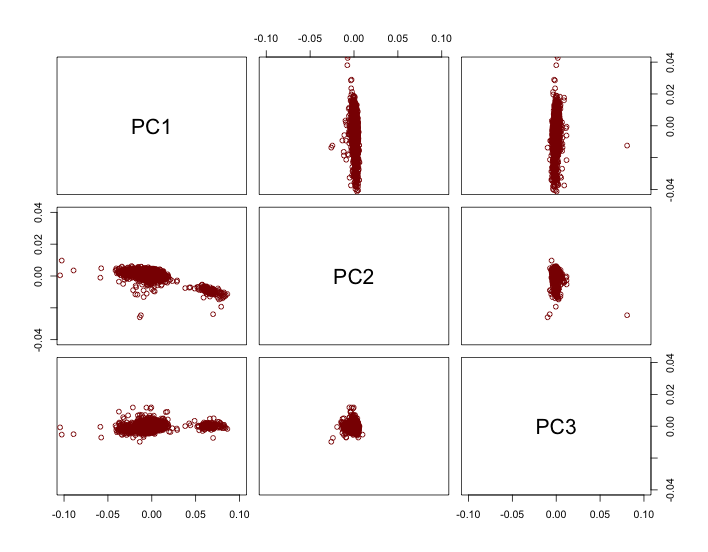
****

(b)

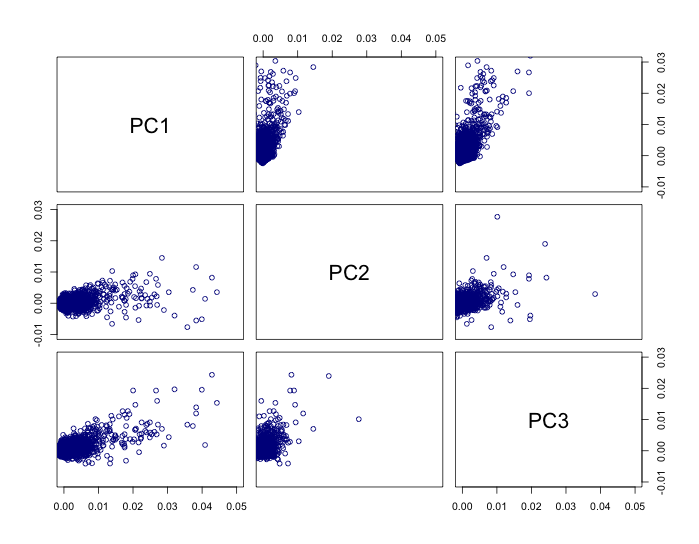
****

**Supplementary Figure 3.2** First three principle components for (a) South Asians and (b) white Caucasians

(a)

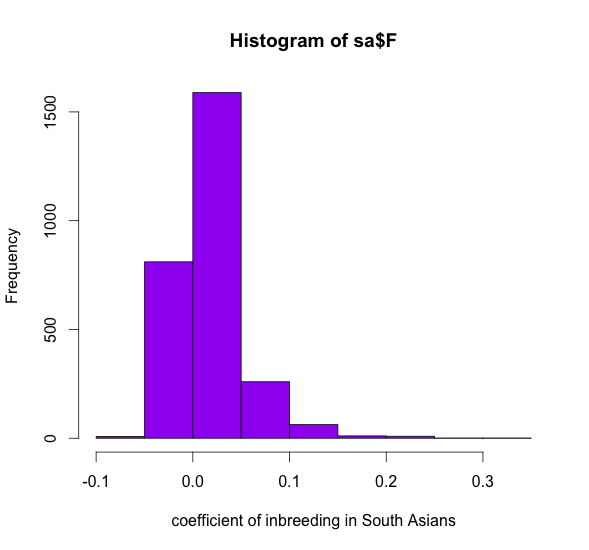


(b)

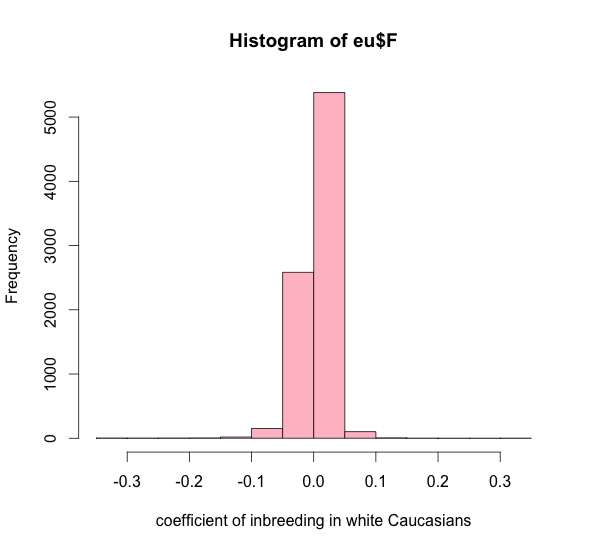


**Supplementary Figure 3.3** Histograms of coefficients of inbreeding in (a) South Asians and (b) white Caucasians

(a)

****

(b)

****

**Supplementary Table 4.1** Inclusion and exclusion criteria for the START and CHILD cohorts

|  |  |
| --- | --- |
| **Inclusion Criteria** | **Exclusion Criteria** |
| ***START cohort*** | |
| * Women between the ages of 18 and 40 * South Asian origin * Pregnant with a single fetus | * Expected multiple births * Artificial / Assisted conception of fetus * >4 live births * Surrogate mothers * Living in Canada <9 months * Father of the baby is not South Asian * Following chronic medication conditions:   - Active cancers  - HIV  - Hepatitis B or C  - VDRL Positive  - Rheumatic Heart Disease  - Seizure Disorder |
| ***CHILD cohort*** | |
| * Pregnant women aged 18 years and older (19 in Vancouver) * Residence in reasonable proximity to the delivery hospital * Able to read, write, and speak English * Willing to provide informed consent * Willing to consent to cord blood collection for the study * Planning to give birth at a designated recruitment centre participating hospital * Infants born at or after 35 weeks * Able to provide name, address and telephone numbers of two alternate contact individuals | * Children born with major congenital abnormalities or respiratory distress syndrome (RDS) * Expectation of moving away from a recruitment area within 1 year * Children of multiple births * Children resulting from in vitro fertilization * Children who will not spend at least 80% of nights in the index home * Children born before 35 weeks gestation |

**Supplementary Table 4.2** Allele frequencies and HWE-P values for SNPs available in the START and CHILD cohorts.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Gene** | **Proxy SNP** | **Trait** | **Allele** | **South Asians** | | **White Caucasians** | |
| **AF** | **HWE-*P*** | **AF** | **HWE-*P*** |
| *ADCY5* | rs2877716 | Birth weight | A | 0.249 | 0.736 | 0.270 | 0.860 |
| *ZBTB38* | rs9846396 | Birth length | A | 0.321 | 0.191 | 0.446 | 0.035 |
| *CCNL1* | rs17451107 | Birth weight | G | 0.270 | 0.522 | 0.373 | 0.051 |
| *LCORL* | rs724577 | Birth weight / length | A | 0.196 | 0.224 | 0.257 | 1.000 |
| *CDKAL1* | rs9368222 | Birth weight | A | 0.252 | 0.736 | 0.273 | 0.727 |
| *GPR126* | rs155259 | Birth length | A | 0.342 | 1.000 | 0.282 | 0.002 |
| *JAZF1* | rs849141 | Birth length | A | 0.290 | 0.014 | 0.259 | 0.856 |
| *CALCR* | rs6968642 | Birth weight | A | 0.377 | 0.499 | 0.449 | 1.000 |
| *HHEX* | rs10882099 | Birth weight | G | 0.563 | 0.302 | 0.387 | 0.057 |
| *TCF7L2* | rs4132670 | Birth weight | A | 0.302 | 0.763 | 0.340 | 0.876 |
| *ADRB1* | rs740746 | Birth weight | G | 0.268 | 0.421 | 0.297 | 0.402 |
| *HMGA2* | rs10784502 | Birth weight | G | 0.276 | 1.000 | 0.471 | 0.402 |
| *HMGA2* | rs8756 | Birth length | C | 0.272 | 0.873 | 0.473 | 0.676 |
| *ADAMTSL3* | rs4842838 | Birth length | C | 0.408 | 0.895 | 0.471 | 0.482 |
| *ANKRD13B* | rs565977 | Birth length | A | 0.438 | 0.197 | 0.309 | 0.033 |
| *ACBD4* | rs4986172 | Birth length | A | 0.343 | 0.402 | 0.306 | 1.000 |
| *GDF5* | rs6087704 | Birth length | G | 0.557 | 0.248 | 0.404 | 0.662 |

**Supplementary Table 4.3** Genes (and SNPs) chosen for investigation in the primary analysis from a literature search

|  |  |  |
| --- | --- | --- |
| **Gene** | **SNP** | **Trait** |
| *LCORL* | rs724577 | birth length |
| *PTCH1* | rs473902 | birth length |
| *GPR126* | rs7763064 | birth length |
| *HMGA2* | rs1351394 | birth length |
| *DCST2* | rs905938 | birth length |
| *SF3B4* | rs11205277 | birth length |
| *PTPDC1* | rs1257763 | birth length |
| *HHIP* | rs7689420 | birth length |
| *ADAMTSL3* | rs11259936 | birth length |
| *ZBTB38* | rs724016 | birth length |
| *HMGA1* | rs2780226 | birth length |
| *IGF1R* | rs2871865 | birth length |
| *GDF5* | rs143384 | birth length |
| *DTL* | rs10863936 | birth length |
| *JAZF1* | rs1708299 | birth length |
| *ACBD4* | rs4986172 | birth length |
| *ANKRD13B* | rs3110496 | birth length |
| *PML* | rs5742915 | birth length |
| *CCNL1* | rs900400 | birth weight |
| *CENPM* | rs5758511 | birth weight |
| *ADCY5* | rs9883204 | birth weight |
| *HMGA2* | rs1042725 | birth weight |
| *CDKAL1* | rs6931514 | birth weight |
| *CALCR* | rs7780752 | birth weight |
| *ACTBL2* | rs4432842 | birth weight |
| *LCORL* | rs724577 | birth weight |
| *ADRB1* | rs1801253 | birth weight |
| *SLC2A4* | rs5415 | birth weight |
| *TCF7L2* | rs7903146 | birth weight |
| *HHEX-IDE* | rs1111875 | birth weight |
| *IGF-1* | - | Birth weight / length |
| *IGF-2* | - | Birth weight / length |

**Supplementary Table 4.4** Global methylation differences between START and CHILD cohorts

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Comparison groups** | **N South Asians** | **N white Caucasians** | **Estimated heterogeneity variance** | **P heterogeneity** |
| START vs. CHILD | 234 | 250 | 0.00022 | <0.01 |

A 1000 CpG sites across the genome were randomly selected and the average methylation level was meta-analysed to estimate global methylation. Heterogeneity estimates for the meta-analysis is presented above.

**Supplementary Table 4.5** SNP associations for birth weight genes in South Asians and white Caucasian newborns

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Gene** | **Lead SNP** | **Proxy SNP** | **Risk allele** | **South Asian** | |  | **White Caucasians** | |  |
| **β-coefficient** | ***P*** | **Power\*** | **β-coefficient** | ***P*** | **Power\*** |
| *ADCY5* | rs9883204 | rs2877716 | C | 0.017 | 6.99 x10-1 | 0.05 | -0.042 | 3.91 x10-1 | 0.05 |
| *CALCR* | rs7780752 | rs6968642 | C | 0.0003 | 9.93 x10-1 | 0.05 | 0.015 | 7.27 x10-1 | 0.05 |
| *HHEX* | rs1111875 | rs10882099 | T | -0.007 | 8.58 x10-1 | 0.05 | -0.008 | 8.55 x10-1 | 0.05 |
| *TCF7L2* | rs7903146 | rs4132670 | G | -0.047 | 2.66 x10-1 | 0.05 | -0.101 | 2.27 x10-2 | 0.05 |
| *ADRB1* | rs1801253 | rs740746 | G | 0.004 | 9.37 x10-1 | 0.05 | 0.030 | 5.30 x10-1 | 0.05 |
| *HMGA2* | rs1042725 | rs10784502 | T | -0.074 | 7.66 x10-2 | 0.05 | -0.063 | 1.52 x10-1 | 0.05 |
| *CCNL1* | rs900400 | rs17451107 | G | -0.037 | 3.79 x10-1 | 0.05 | -0.040 | 3.26 x10-1 | 0.05 |
| *LCORL* | rs724577 | rs724577 | C | -0.089 | 7.52 x10-2 | 0.05 | -0.062 | 1.99 x10-1 | 0.05 |
| *CDKAL1* | rs6931514 | rs9368222 | A | -0.089 | 3.98 x10-2 | 0.05 | -0.022 | 6.46 x10-1 | 0.05 |
| Genotype score |  |  |  | -0.021 | 2.05 x10-1 |  | -0.042 | 1.36 x10-2 |  |

\* Calculated using a β-coefficient of 0.07 gram change in birth weight, congruent with estimates reported from GWA-studies, and allele frequency and sample size in our sample

**Supplementary Table 4.6** SNP associations for birth length genes in South Asian and white Caucasian newborns

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Gene** | **Lead SNP** | **Proxy SNP** | **Risk allele** | **South Asian** | |  | **White Caucasians** | | |
| **β-coefficient** | ***P*** | **Power\*** | **β-coefficient** | ***P*** | **Power\*** |
| *ZBTB38* | rs724016 | rs9846396 | A | 0.05 | 8.42 x10-1 | 0.101 | 0.03 | 9.24 x10-1 | 0.108 |
| *GPR126* | rs7763064 | rs155259 | A | 0.39 | 1.35 x10-1 | 0.103 | -0.29 | 4.43 x10-1 | 0.097 |
| *JAZF1* | rs1708299 | rs849141 | G | 0.10 | 6.82 x10-1 | 0.098 | -0.54 | 1.10 x10-1 | 0.095 |
| *ADAMTSL3* | rs11259936 | rs4842838 | C | 0.18 | 4.95 x10-1 | 0.107 | 0.27 | 3.40 x10-1 | 0.109 |
| *ANKRD13B* | rs3110496 | rs565977 | A | -0.20 | 4.49 x10-1 | 0.108 | -0.21 | 4.88 x10-1 | 0.100 |
| *ACBD4* | rs4986172 | rs4986172 | A | -0.15 | 5.59 x10-1 | 0.103 | 0.27 | 3.98 x10-1 | 0.106 |
| *HMGA2* | rs1351394 | rs8756 | A | 0.01 | 9.72 x10-1 | 0.097 | -0.40 | 1.93 x10-1 | 0.109 |
| *LCORL* | rs724577 | rs724577 | C | -0.15 | 6.55 x10-1 | 0.087 | -0.43 | 2.12 x10-1 | 0.095 |
| *GDF5* | rs143384 | rs6087704 | G | -0.15 | 5.24 x10-1 | 0.108 | -0.26 | 4.01 x10-1 | 0.106 |
| Genotype score |  | |  | -0.06 | 5.39 x10-1 |  | -0.11 | 3.23 x10-1 |  |

\* Calculated using a β-coefficient of 0.1 SD change in birth length, congruent with estimates reported from GWA-studies, and allele frequency and sample size in our sample

**Supplementary Table 4.7** Overview of key differences in blood collection protocols of the START and CHILD cohorts

|  |  |  |
| --- | --- | --- |
|  | **START** | **CHILD** |
| **Time of cord-blood collection** | Before placenta | After placenta, except Toronto |
| **Processing time** | Less than 2 hours | Less than 24 hours – mean=19.8 hours; range=0.33-321.70 hours |

**Supplementary Table 4.8** Variation in leukocyte and erythrocyte cell type composition between South Asians and white Caucasians from the CHILD cohort

|  |  |  |  |
| --- | --- | --- | --- |
| **Cell type** | **South Asian**  **Mean** | **White Caucasian Mean** | **P-value** |
| Lymphocytes | 4.63 | 4.83 | 6.24x10-1 |
| Monocytes | 1.21 | 1.48 | 1.72x10-1 |
| Neutrophils | 7.14 | 8.08 | 2.15x10-1 |
| Nucleated Red Blood Cells | 1.11 | 1.26 | 7.42x10-1 |
| Eosinophils | 0.33 | 0.46 | 6.36x10-2 |
| Basophils | 0.10 | 0.10 | 9.22x10-1 |

**Supplementary Table 4.9** Association between CpG sites showing ethnic heterogeneity for birth weight in South Asians from the CHILD cohort (n=18)

|  |  |  |  |
| --- | --- | --- | --- |
| **Gene** | **CpG site** | **β-coefficient\*** | **P** |
| *TCF7L2* | cg09022607 | -0.0045 | 3.33 x10-1 |
| *CALCR* | cg23061150 | -0.0044 | 3.15 x10-1 |
| *HMGA2* | cg24892571 | -0.0119 | 5.23 x10-2 |
| *HMGA2* | cg24776736 | -0.0029 | 4.02 x10-1 |
| *TCF7L2* | cg11748187 | -0.0067 | 1.32 x10-1 |
| *IGF1* | cg01305421 | 0.0104 | 9.86 x10-1 |
| *CDKAL1* | cg06512263 | 0.0002 | 9.63 x10-1 |

**\*** β-coefficient adjusted for gestational age and sex

**Supplementary Table 4.10** Effect of processing time on methylation level in the CHILD cohort

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Gene** | **CpG site** | **% methylation**  **Processing time < 2 hours**  **N=14** | **% methylation**  **Processing time > 2 hours**  **N=227** | **P-value\*** |
| *TCF7L2* | cg09022607 | 28.5 | 29.2 | 1.28x10-1 |
| *CALCR* | cg23061150 | 82.0 | 81.9 | 5.46x10-1 |
| *HMGA2* | cg24892571 | 85.9 | 85.8 | 8.77x10-1 |
| *HMGA2* | cg24776736 | 59.3 | 59.2 | 4.61x10-1 |
| *TCF7L2* | cg11748187 | 36.0 | 37.5 | 4.86x10-5 |
| *IGF1* | cg01305421 | 18.6 | 19.8 | 7.73x10-1 |
| *CDKAL1* | cg06512263 | 26.4 | 27.2 | 5.92x10-1 |

\* *P*-value for the relationship between processing time and methylation at the designated CpG site

**Supplementary Table 4.11** Effect of processing time on methylation level in the CHILD cohort

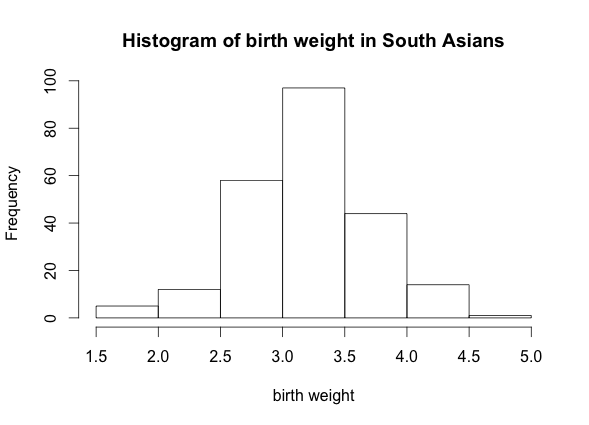
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Gene** | **CpG site** | **β-coefficient N=14** | **P-value** | **β-coefficient N=227** | **P-value** |
| *TCF7L2* | cg09022607 | 0.003 | 8.93x10-1 | -0.035 | 1.31x10-4 |
| *CALCR* | cg23061150 | -0.013 | 3.74x10-1 | -0.027 | 3.57x10-3 |
| *HMGA2* | cg24892571 | -0.016 | 5.40x10-1 | -0.022 | 7.13x10-2 |
| *HMGA2* | cg24776736 | -0.010 | 4.77x10-1 | -0.022 | 7.82x10-5 |
| *TCF7L2* | cg11748187 | 0.017 | 2.59x10-1 | -0.017 | 1.34x10-3 |
| *IGF1* | cg01305421 | 0.006 | 7.76x10-1 | -0.024 | 1.33x10-4 |
| *CDKAL1* | cg06512263 | -0.003 | 8.80x10-1 | -0.016 | 2.65x10-3 |

**Supplementary Table 4.12** Percent of variance explained by clinical variables and CpG sites on birth weight in white Caucasians

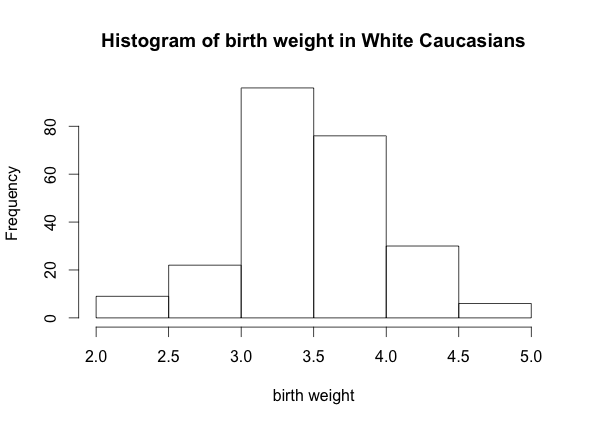
|  |  |
| --- | --- |
| **Variables in the model** | **Variance explained (%)** |
| Gestational age | 24.1 |
| + Sex | 24.9 |
| + Gestational diabetes | 25.0 |
| + Mother’s pre-pregnancy BMI | 32.5 |
| + Smoking exposure | 34.1 |
| + Mother’s hypertension during pregnancy | 34.5 |
| + cg09022607 | 35.0 |
| + cg24776736 | 35.3 |
| + cg23061150 | 36.5 |
| + cg01305421 | 36.8 |
| + cg06512263 | 36.9 |
| + cg11748187 | 37.4 |
| + cg24892571 | 37.7 |

**Supplementary Figure 4.1** Histograms of birth weight in South Asians (a) and white Caucasians (b); birth length in South Asians (c) and fasting glucose (d)

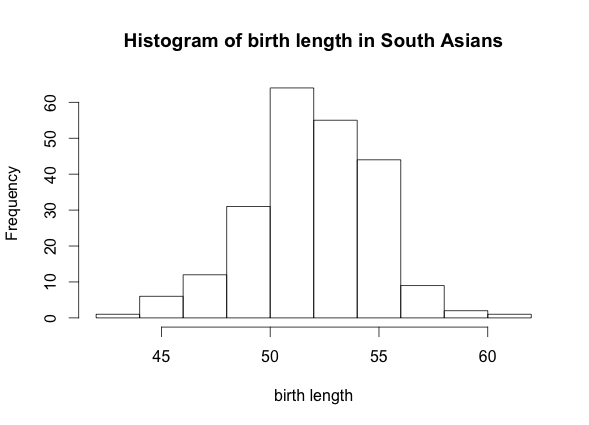
(a)



(b)



(c)



(d)

