Fetal programming and parent-of-origin effects of type 2 diabetes and insulin secretion

Hatem, Gad

2023

Document Version:
Publisher's PDF, also known as Version of record

Fetal programming and parent-of-origin effects of type 2 diabetes and insulin secretion

GAD HATEM
DEPT OF CLINICAL SCIENCES, MALMÖ | FACULTY OF MEDICINE | LUND UNIVERSITY

Fetal programming and parent-of-origin effects of type 2 diabetes and insulin secretion

Gad Hatem

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Medicine at Lund University to be publicly defended on 2023-06-14 at 13.00 in Medelhavet, Department of Clinical Sciences, Inga Marie Nilssons gata 53, ingang 46, Skånes Universitetssjukhus, Malmö

Faculty opponent
Professor Tove Fall
Uppsala University

Organization: LUND UNIVERSITY

Document name: Doctoral dissertation **Date of issue** 2023-06-14

Author(s): Gad Hatem **Sponsoring organization:**

Title and subtitle: Fetal programming and parent-of-origin effects of type 2 diabetes and insulin secretion

Abstract Type 2 diabetes mellitus (T2DM) is a heterogeneous and a complex disease defined by hyperglycemia. The pancreas and its islets are central for glucose homeostasis and healthy adipose tissue. In turn, lipid levels in the blood are crucial for glucose level stability. Both genetic and environmental factors and their interaction play a pivotal role in the risk and development of the disease. In this thesis we aim to better understand the effect of genetic and environmental factors by investigating parental effects manifesting from early life until adulthood.

In papers I and II we examined gene expression alterations and associated epigenetic changes due to early pregnancy anemia and gestational diabetes (GDM). Moreover, we investigated associations between these changes and neonatal anthropometry. We identified several differentially expressed genes between early pregnancy anemia, GDM and controls. Most of these genes were accompanied by epigenetic changes that correlated with their expression patterns. Interestingly, we identified several differentially expressed genes associated with neonatal anthropometry indicating their possible role in fetal programming and risk of T2DM in later life due to maternal exposure to early pregnancy anemia and GDM.

In paper III we investigated whether genetic variants which were previously reported to be associated with lipid traits will exert different effects on obesity and blood lipid traits based on their parental origin. We examined These variants in two European family cohorts, where parental origin of each variant was inferred and parental-specific association with obesity and blood lipid traits was analyzed. Our results corroborated previous reports and indicated that specific genetic variants show parent-of-origin specific effects. Moreover, our results indicate possible sex-specific parental effects on some blood lipid traits.

In paper IV we questioned whether such parental specific effects observed in paper III also manifested in early life. As a result, we explored parent-of-origin effects on cardiometabolic and anthropometric traits in a birth cohort which was followed up from delivery until 18 years. Our results indicate that the parental specific effects of cardiometabolic and anthropometric traits and associated genetic variants manifested in early life. Interestingly, however, not all parental effects were found to be fixed, and they seemed to transition over time specifically during puberty.

In paper V we have examined the expression of imprinted genes to better understand their role in insulin secretion, beta-cell development, and function. First, we scrutinized gene expression data from adult pancreas, adult pancreatic islets, fetal pancreas, and single cell expression data. Next, we analyzed the association of these genes with glycemic traits. We identified imprinted genes that were specifically expressed in fetal pancreas both on a tissue and single cell level. Variants in two genes associated with indices of insulin secretion indicating their possible role in beta-cell development. Additionally, we identified imprinted genes enriched in both fetal and adult pancreas and associated with glucose and insulin traits in a parent-of-origin manner. This suggests the possible role of these genes in beta-cell function.

In summary, in this thesis we investigate paternal and maternal effects as a function of fetal programming and parent-of-origin effects to better understand their influence on type 2 diabetes and insulin secretion.

Key words: Type 2 Diabetes, fetal programming, parent-of-origin effects, transcriptopmics.

Supplementary bibliographical information

Language English **ISSN and key title**: **1652-8220**

ISBN: 978-91-8021-422-3

Recipient's notes **Number of pages**: 89

Price Security classification

Signature Date 2023-04-27

Fetal programming and parent-of-origin effects of type 2 diabetes and insulin secretion

Gad Hatem

LUND
UNIVERSITY

Figures were created with BioRender.com

‘

Faculty of Medicine
Department of Clinical Sciences, Malmö

ISBN 978-91-8021-422-3
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2023

To my family

Table of Contents

Papers included in the thesis

Paper I

Hatem G, Hjort L, Asplund O, Minja DTR, Msemo OA, Møller SL, Lavstsen T, Groth-Grunnet L, Lusingu JPA, Hansson O, Christensen DL, Vaag AA, Artner I, Theander T, Groop L, Schmiegelow C, Bygbjerg IC, Prasad RB. Mapping the Cord Blood Transcriptome of Pregnancies Affected by Early Maternal Anemia to Identify Signatures of Fetal Programming. J Clin Endocrinol Metab. 2022 Apr 19;107(5):1303-1316. doi: 10.1210/clinem/dgac010. PMID: 35021220

Paper II

Hatem G, Hjort L, Andersson J, Minja DTR, Msemo OA, Møller SL, Asplund O, Christensen DL, Theander T, Hansson O, Artner I, Lusingu JPA, Groth-Grunnet L , Bygbjerg IC, Vaag AA, Schmiegelow C, Prasad RB. Signatures of fetal programming in cord blood from mothers with gestational diabetes mellitus (*Manuscript*)

Paper III

Lessmark A, **Hatem G**, Kovacs G, Vitai M, Ahlqvist E, Tuomi T, Koranyi L, Groop L, Prasad RB. Lipid-Associated Variants near ANGPTL3 and LPL Show Parent-of-Origin Specific Effects on Blood Lipid Levels and Obesity. Genes (Basel). 2021 Dec 29;13(1):91. doi: 10.3390/genes13010091. PMID: 35052431

Paper IV

Wagh R, **Hatem G**, Kunte P, Chittaranjan SY, Prasad RB. Parent-of-origin effects in the life-course evolution of cardio-metabolic traits (*Manuscript*)

Paper V

Hatem G, Asplund O, Ahuja V, Singh T, Tuomi T, Artner I, Prasad RB. Imprinted genes in beta cell development and T2D risk (*Manuscript*)

Abbreviations

ApoA1	Apolipoprotein A1
ApoA2	Apolipoprotein A2
apoB	Apolipoprotein B
apoCIII	Apoliopprotein C-III
ASE	Allele Specific Expression Analysis
BMI	Body mass index
CDKAL1	Cdk5 Regulatory Associated Protein 1-like 1
CDKN1A	Cyclin Dependent Kinase Inhibitor 1A
CDKN1C	Cyclin Dependent Kinase Inhibitor 1C
CDKN2A/2B	Cyclin-Dependent Kinase Inhibitor 2A/2B
CETP	Cholesteryl Ester Transfer Protein
CPM	Counts per illion
CVD	Cardiovascular diseases
DEG	Differentially expressed gene
DIAGRAM	Diabetes genetics replication and meta-analysis
DLK1	Delta Like Non-Canonical Notch Ligand 1
DNMT1	DNA (cytosine-5)-methyltransferase 1
DNMT3A	DNA (cytosine-5)-methyltransferase 3A
DNMT3B	DNA (cytosine-5)-methyltransferase 3B
DOHaD	The Developmental Origins of Health and Disease
ENCODE	The Encyclopedia of DNA Elements
eQTL	Expression quantitative trait loci
FPKM	Fragment per kilobase per million
FTO	Fat mass and obesity-associated protein
GDM	Gestational diabetes mellitus
GIP	Gastric inhibitory polypeptide
GLP-1	Glucagon-like peptide 1
GLUT2	Glucose transporter 2
GWAS	Genome-wide association studies
KCNJ11	Potassium Inwardly Rectifying Channel Subfamily J Member 11
KCNQ1	Potassium Voltage-Gated Channel Subfamily Q Member 1
KLHDC1	Kelch Domain Containing 1
H19	H19 Imprinted Maternally Expressed Transcript
Hb	Hemoglobin
HDL	High-density lipoprotein

HHEX	Hematopoietically-expressed homeobox protein
IGF2	Insulin-like growth factor 2
IGF2BP2	Insulin-like growth factor 2 mRNA-binding protein 2
LADA	Latent autoimmune diabetes of adults
LCORL	Ligand Dependent Nuclear Receptor Corepressor Like
LDL	Low-density lipoprotein
MEG3	Maternally Expressed 3
MODY	Maturity-onset diabetes of the young
NUMBL	NUMB Like Endocytic Adaptor Protein
P2RX7	Purinergic Receptor P2X 7
PDE7B	Phosphodiesterase 7B
PHACTR2	Phosphatase and Actin Regulator 2
PIK3C2B	Phosphatidylinositol-4-Phosphate 3-Kinase Catalytic Subunit Type 2 Beta
PPARG	Peroxisome proliferator- activated receptor
RTL1	Retrotransposon Gag Like 1
SGLT2	Sodium/glucose cotransporter 2
SLC30A8	Zinc Transporter 8
SLC4A1	Solute Carrier Family 4 Member 1
SNP	Single nucleotide polymorphism
T1DM	Type 1 diabetes mellitus
T2DM	Type 2 diabetes mellitus
TCF7L2	Transcription factor 7-like 2
TET	Ten-eleven translocation
TPM	Transcript per million
TSHZ3	Teashirt Zinc Finger Homeobox 3
VLDL	Very low-density lipoprotein
UCB	Umbilical cord blood
WT1-AS	WT1 antisense RNA

Popular science summary

Type 2 diabetes mellitus(T2DM) is a metabolic disorder that affects how the body processes glucose, which is an essential energy source. T2DM develops gradually due to a combination of genetic and lifestyle factors. The primary risk factors for T2DM include obesity, having a sedentary lifestyle, and eating a diet high in processed and sugary foods. Genetics also plays a role, with a family history of diabetes increasing the likelihood of developing the disease. T2DM occurs when the body becomes resistant to insulin, a hormone that helps transport glucose from the bloodstream into the cells. As a result, glucose builds up in the blood, leading to high blood sugar levels. Over time, this can cause damage to organs and tissues throughout the body, increasing the risk of complications such as heart disease, stroke, kidney disease, and nerve damage. T2DM is a growing health concern worldwide, with an estimated 537 million adults living with the condition.

Genetics is the study of genes, the basic units of heredity that are passed down from parents to their offspring. Genes are made up of DNA, which contain the instructions for how to make proteins, the building blocks of life. In addition to genetics, there exists another field of study called epigenetics which is the study of how genes can be turned on or off by chemical modifications to the DNA or the proteins around it. These modifications can be influenced by various environmental factors such as diet, stress, and exposure to toxins. Epigenetic changes can alter the activity of genes and affect traits such as our risk of developing certain diseases, how we respond to medications, and even our behavior and personality. One of the most fascinating aspects of epigenetics is that some of these changes can be passed down from one generation to the next, even if they are not encoded in the DNA itself. This means that our experiences and environment can potentially influence the health and well-being of our descendants. Both genetics and epigenetics are important areas of study that help us better understand the complex interplay between our genes and our environment.

In this thesis we aim to study the effects of genetic and environmental factors on T2DM by investigating two genetic phenomena: 1) parent-of-origin effects, 2) fetal programming. Parent-of-origin effects refer to the phenomenon where genes inherited from one parent have different effects than the same genes inherited from the other parent. This is because some genes are "imprinted," meaning they are turned off or on by epigenetic means based on which parent they come from. On the other hand, fetal programming is a concept that suggests that the health and well-being of a developing fetus are influenced by its environment, including the mother's diet and lifestyle. Exposure to certain factors during pregnancy can have long-lasting effects on a child's health and development.

In two of our studies, we have investigated fetal programming and its possible involvement in the risk of T2DM after the mother's exposure to gestational diabetes

or anemia during pregnancy. Anemia is a medical condition characterized by a decrease in the amount of hemoglobin in the blood. Hemoglobin is a protein in the red blood cells that carries oxygen from the lungs to the tissues and organs of the body. Gestational diabetes on the other hand, is a type of diabetes that occurs during pregnancy in which a woman's body cannot produce enough insulin to regulate the increased blood sugar levels during pregnancy. In our studies we analyzed samples from the cord blood of pregnant mothers with either condition, since changes in the cord blood would reflect changes in the fetal environment. We were successful in identifying changes in genetic expression of hundreds of genes and we hypothesized that these are carried out by epigenetic means. Indeed, when probing changes in certain epigenetic markers, we found significant correlation between the genetic and epigenetic changes in the cord blood. Moreover, we found that these changes also correlated with changes in some biometric measurements in the offspring after delivery. Based on our observations which included changes in one gene associated with birth weight *LCORL* and *SLC4A1* gene, we propose that we have identified possible signatures of fetal programming after the exposure of pregnant mothers to anemia or gestational diabetes.

The three remaining studies were aimed to understand the parent-of-origin effects on T2DM. First, we have investigated this genetic phenomenon on obesity and blood lipid levels. High blood lipid levels can contribute to insulin resistance, as excess cholesterol and triglycerides can accumulate in the cells and interfere with insulin function. Likewise, obesity is an important risk factor in developing T2DM. Next, we investigated whether the parent-of-origin effects identified in adulthood can also be found in early life and if these effects are consistent throughout adolescence and adulthood. Finally, we investigated the expression of imprinted genes in adult pancreas, pancreatic islets, and beta-cells and compared their expression in fetal pancreatic tissues. These tissues were selected because insulin is produced and secreted from beta-cells which are located in pancreatic islets which in turn are located in the pancreas. We successfully identified different genetic variants near *ANGPTL3* and *LPL* genes that showed parent-of-origin effects on blood lipid levels and obesity. An even more interesting observation was the sex specific effects of these variants, meaning that they have showed parent of origin effects in either sons or daughters but not in both. Furthermore, we found that not all genes showing parent-of-origin effects manifested in early life and some of these effects transitioned from paternal to maternal effects or vice versa. Finally, in our endeavor studying the imprinted genes in adult and fetal pancreatic tissues, we identified the *RTL1* gene as a possible gene that plays an important role in the development of insulin producing beta-cells. Additionally, based on our observations, we suggest that two genes, *PHACTR2* and *TSHZ3*, have a possible role in beta-cell function.

With our work we show that understanding parent-of-origin effects is important not only for understanding basic genetics, but also for understanding how genetic

factors contribute to human health and disease. This is also true in understanding fetal programming because of exposure to various conditions. By better understanding how genetic and environmental factors interact we can provide a new vision and hope to develop new treatments and interventions to improve human health and prevent disease.

Abstract

Type 2 diabetes mellitus (T2DM) is a heterogeneous and a complex disease defined by hyperglycemia. The pancreas and its islets are central for glucose homeostasis and healthy adipose tissue. In turn, lipid levels in the blood are crucial for glucose level stability. Both genetic and environmental factors and their interaction play a pivotal role in the risk and development of the disease. In this thesis we aim to better understand the effect of genetic and environmental factors by investigating parental effects manifesting from early life until adulthood.

In papers I and II we examined gene expression alterations and associated epigenetic changes due to early pregnancy anemia and gestational diabetes (GDM). Moreover, we investigated associations between these changes and neonatal anthropometry. We identified several differentially expressed genes between early pregnancy anemia, GDM and controls. Most of these genes were accompanied by epigenetic changes that correlated with their expression patterns. Interestingly, we identified several differentially expressed genes associated with neonatal anthropometry indicating their possible role in fetal programming and risk of T2DM in later life due to maternal exposure to early pregnancy anemia and GDM.

In paper III we investigated whether genetic variants which were previously reported to be associated with lipid traits will exert different effects on obesity and blood lipid traits based on their parental origin. We examined These variants in two European family cohorts, where parental origin of each variant was inferred and parental-specific association with obesity and blood lipid traits was analyzed. Our results corroborated previous reports and indicated that specific genetic variants show parent-of-origin specific effects. Moreover, our results indicate possible sex-specific parental effects on some blood lipid traits.

In paper IV we questioned whether such parental specific effects observed in paper III also manifested in early life. As a result, we explored parent-of-origin effects on cardiometabolic and anthropometric traits in a birth cohort which was followed up from delivery until 18 years. Our results indicate that the parental specific effects of cardiometabolic and anthropometric traits and associated genetic variants manifested in early life. Interestingly, however, not all parental effects were found to be fixed, and they seemed to transition over time specifically during puberty.

In paper V we have examined the expression of imprinted genes to better understand their role in insulin secretion, beta-cell development, and function. First, we scrutinized gene expression data from adult pancreas, adult pancreatic islets, fetal pancreas, and single cell expression data. Next, we analyzed the association of these genes with glycemic traits. We identified imprinted genes that were specifically expressed in fetal pancreas both on a tissue and single cell level. Variants in two genes associated with indices of insulin secretion indicating their possible role in beta-cell development. Additionally, we identified imprinted genes enriched in both

fetal and adult pancreas and associated with glucose and insulin traits in a parent-of-origin manner. This suggests the possible role of these genes in beta-cell function.

In summary, in this thesis we investigate paternal and maternal effects as a function of fetal programming and parent-of-origin effects to better understand their influence on type 2 diabetes and insulin secretion.

Introduction

Diabetes

A universal concern

Diabetes is a complex metabolic disorder that poses a major global health threat. Over 643 million people are projected to suffer diabetes mellitus by 2030 with 537 million people already diagnosed in 2021. Countries and regions transitioning to middle-income standards are expected to provide the largest proportion of the newly diagnosed population [1]. The convoluted interactions between several genetic and environmental factors are the main drivers in developing diabetes mellitus. This being said, environmental factors such as obesity, a sedentary lifestyle, unhealthy diet, and aging remain some of the principal factors for the rapid growth of diabetes mellitus into global epidemic [2], [3]. Recent estimates suggest that a high percentage of diabetes mellitus goes undiagnosed and one in six pregnancies is affected with hyperglycemia. Undiagnosed or untreated diabetes mellitus can lead to several serious health complications which effects the general medical expenditure. In 2021 the global medical disbursement due to diabetes mellitus was estimated to approach one trillion USD [1]. Both China and India have witnessed an explosion in the number of people diagnosed with diabetes mellitus becoming the main epicenters of the epidemic. In these countries people are characterized with lower body mass index (BMI) upon diagnosis when compared to their western counterparts [2]. The United States has the third largest population diagnosed with diabetes mellitus, however the pacific nations have one of the highest prevalence of diabetes mellitus, approaching 30% of the population [1], [3]. Similarly, the middle east is becoming a hot spot for the global epidemic with large prevalence numbers reaching the quarter of the population in some regions [4]. Another ominous sign of the gravity of diabetes mellitus, is the growing number of childhood obesity which can directly impact the development of diabetes mellitus. Data about children and adolescents diagnosed with diabetes mellitus largely arise from developed countries. In the United States the number of children and adolescents diagnosed with diabetes mellitus rose by more than 30% [5]. It is important to accentuate that previous predictions about the number of people diagnosed with diabetes mellitus in 2021 have already been surpassed by a large margin indicating that the current

statistics and extrapolations maybe underestimations [6]. These alarming observations signify the global burden of diabetes mellitus.

Historical overview

The first written description of what is known today as diabetes was found in Egypt in the nineteenth century. Around 1500BC, the ancient Egyptians characterized this ailment by excessive urination, excessive thirst, and weight loss. About a millennium later Suchruta was able to identify diabetes by characterizing sweet urine and pointed out that this condition effected richer casts due to their diet. This however remained largely unknown to the west [7]. Sadly, our understanding did not change for the next 3000 years, and the outcome remained unaltered for the majority of human history. The ancient Greek word 'Diabianen' is the origin for the word diabetes meaning 'go through' and mellitus meaning 'sweet'. These two words describe the excessive sweet and urination and discern this condition from excessive non-sweet urination [8]. The first indication that a specific organ may cause the condition came in the late nineteenth century where pancreatectomized dogs had symptoms similar to diabetes [9]. It was not until the turn of the twentieth century and onwards where major strides of knowledge were taken in understanding diabetes. The first important breakthrough came after realizing that individuals with inadequate substance called 'insulin' produced by the pancreatic islets developed diabetes. Moreover, the condition was reversable when islet extract was provided [10]. In the mid-twentieth century a new and reliable method using radioimmunoassay was used to measure circulating insulin. This method allowed for better understanding and distinction between different forms of diabetes, one which is rare in the young and non-insulin-dependent and another common in younger subjects and is insulin-dependent [11]. These discoveries paved the way for defining two major types of diabetes which have different pathophysiologies, risk factors, complications, and management.

Types of diabetes

Type 1 diabetes mellitus (T1DM) is described as an autoimmune disorder characterized by the loss of beta-cells. beta-cells are located in the islets of Langerhans in the pancreas and act as glucose regulators by secreting insulin into the blood stream. The destruction of these cells will lead to hyperglycemia and several secondary complications such as cardiovascular disease, kidney disease and blindness [12]. The rates of T1DM changes widely and its prevalence seem to be increasing over time however, the overall frequency of the disease is estimated to be around 7.5% [1], [13]. A combination of genetic predisposition, stochastic circumstances and a variety of mostly unknown environmental factors play a role in

the development of T1DM. Generally, T1DM is developed during a young age however, given its nature as an autoimmune disorder, T1DM can manifest in a variety of age groups with different severity of auto-immune responses. This led to the identification of a new classification called latent autoimmune diabetes of adults (LADA) [14].

Contrary to T1DM, type 2 diabetes mellitus (T2DM) generally manifests in the adult population and accounts for 90% of all diabetes [1]. Nevertheless, due to negative changes in diet and lifestyle, the number of younger patients diagnosed with T2DM is on the rise [2]. T2DM is highly heterogeneous in its onset and progression. T2DM is a non-insulin-dependent diabetes mellitus implying that insulin is present in the patient's body. Regardless of the presence of insulin patients still exhibit progressive decline in their glycemic regulations. The main cause of this decline is insulin resistance, where the patient's own tissues and organs fail to respond to insulin. Initially this resistance will lead to an elevated secretion of insulin called hyperinsulinemia, however the pancreas cannot cope with the high insulin requirements. Additionally, the compensatory hyperinsulinemia will reduce insulin sensitivity which will lead to elevated blood glucose levels and usually by the time of diagnosis beta-cells are unable to secrete adequate insulin leading to hyperglycemia [15]. Though insulin resistance can be systematic both the liver and muscle tissues have been identified as the main contributors. Fat accumulation in hepatic and muscle tissue leads to defective insulin-mediated glucose uptake due to impairment of insulin signaling [16]. T2DM constitutes the largest portion of diabetes mellitus and is highly heterogenic. Several efforts have been made to sub-classify T2DM into more specific subgroups. These efforts have been relying on different phenotypes, biomarkers, and genetic data [17], [18].

With the advances in understanding the two major types of diabetes mellitus a third less common form was identified. Maturity-onset diabetes of the young (MODY) is a disorder that resembles T2DM occurring in the younger population and caused by single gene mutations. These monogenic types of diabetes exhibit inheritance pattern similar to an autosomal dominant one and are often characterized by the slow progression from mild hyperglycemia to diabetes [19]. MODY1 and MODY3 are caused by mutations in the *HNF4A* and *HNF1A* respectively. These mutations impair glucose transport into cells due to failure in producing GLUT4 transporters. Mutations in glucokinase/GCK results in MODY2 which is characterized by impaired detection of low concentrations of insulin [20].

Gestational diabetes mellitus (GDM) occurs during pregnancy and is associated with detrimental outcomes for both the mothers and offspring. GDM causes fetal macrosomia, stillbirth, and neonatal metabolic problems for the fetus and both mothers and offspring are at higher risk of developing diabetes [21].

Other lesser-known forms of diabetes mellitus include secondary diabetes which is a direct result of other diseases and neonatal diabetes which is caused by genetic

and epigenetic alterations [22], [23]. Additionally, it is worth noting that due to the deterioration of eating habits and lifestyle which lead to obesity and insulin resistance more patients are being diagnosed with a combined type 1 and insulin resistance diabetes [24].

T2DM and 'The Ominous Octet'

The pathophysiology of T2DM is an intricate one. It involves different impairments and malfunctions in several networks and feedback loops governing blood glucose levels, insulin secretion and insulin action. Originally three main culprits were identified in the pathophysiology of T2DM and over the years the list of disruptions grew to eight earning the name 'the Ominous octet' [25] (figure 1).

The first member is located in the pancreas and is manifested as beta-cell disfunction. This leads to reduced insulin secretion which limits the body's capacity to maintain normal blood glucose levels [26].

The second and third members are presented as insulin resistance in the muscles and liver. Insulin resistance leads to reduced glucose uptake by muscle, liver and gut tissues and increased glucose production by the liver. It is important to note that albeit insulin resistance and beta-cell dysfunction are both necessary for the development of the disease, beta-cell disfunction usually occurs early in the progression and is usually quite severe [26].

Adipocyte dysfunction is the fourth member contributing to the pathogenesis of T2DM. Adipose tissue is an essential energy source. This energy, which is the primary energy supply pathway, is released in a process called lipolysis and insulin plays an important role in its regulation. Insulin inhibits lipolysis, however the adipocyte resistance to this action will lead to elevated levels of free fatty acids in the plasma. This will lead to reduced insulin secretion, increased gluconeogenesis in the liver and hinder insulin-induced glucose uptake in the muscle and liver [27].

The fifth member of 'the Ominous Octet' comes in the form of two hormonal dysfunctions in the gastrointestinal tissue. Oral glucose load induces considerably higher insulin response when compared to the intravenous injection. This is called the incretin effect and mainly occurs due to two hormones produced by the small intestine. Glucagon-like peptide 1 (GLP-1) is secreted by the L-cells in the small intestine and is insufficient in T2DM patients leading to elevated plasma glucagon levels. Gastric inhibitory polypeptide (GIP) secreted by the K-cells in the small intestine. Contrary to GLP-1, GIP levels are elevated, however its effect on beta-cells and insulin secretion is inadequate in T2DM patients [28].

In addition to beta-cells the pancreatic islets are also home to the alpha-cells which are the sixth member in 'the Ominous Octet'. One of the main functions of the alpha-cells is to produce glucagon which causes elevated glucose levels in the blood

stream. Interestingly, the basal plasma concentration of glucagon was observed to be elevated in T2DM patients. Notably, alpha-cell disfunction leads to elevated hepatic glucose production and fasting hyperglycemia [29].

The second to last member of this group is the kidney and its importance to T2DM pathophysiology rises from sodium/glucose cotransporter 2 (SGLT2). SGLT2, along with SGLT1, is tasked with the reabsorption of the filtered glucose in the kidneys to prevent glucosuria. It is important to note that SGLT2 is responsible for the majority of the reabsorption and due to hyperglycemia in T2DM, SGLT2 is upregulated, and renal glucose preservation is escalated exacerbating the hyperglycemic state [30].

The last, but certainly not least, member of 'the Ominous Octet' is the brain. The brain is one of the main consumers of glucose in the body. In obese individuals, the appetite inhibitory effect of insulin was found to be reduced. This may suggest that insulin resistance, found in peripheral tissue, can be similarly found in the brain. Indeed, functional magnetic resonance imaging found dysregulations at crucial hypothalamic centers in the brain which are responsible for appetite suppression [31]. Combined, these eight members play an important role in regulating blood glucose levels using a complex and intertwined signaling networks and serve as potential markers and therapeutic targets. In the work presented in this thesis, we explored how beta-cells and adipose tissue contribute to the risk and development of T2DM.

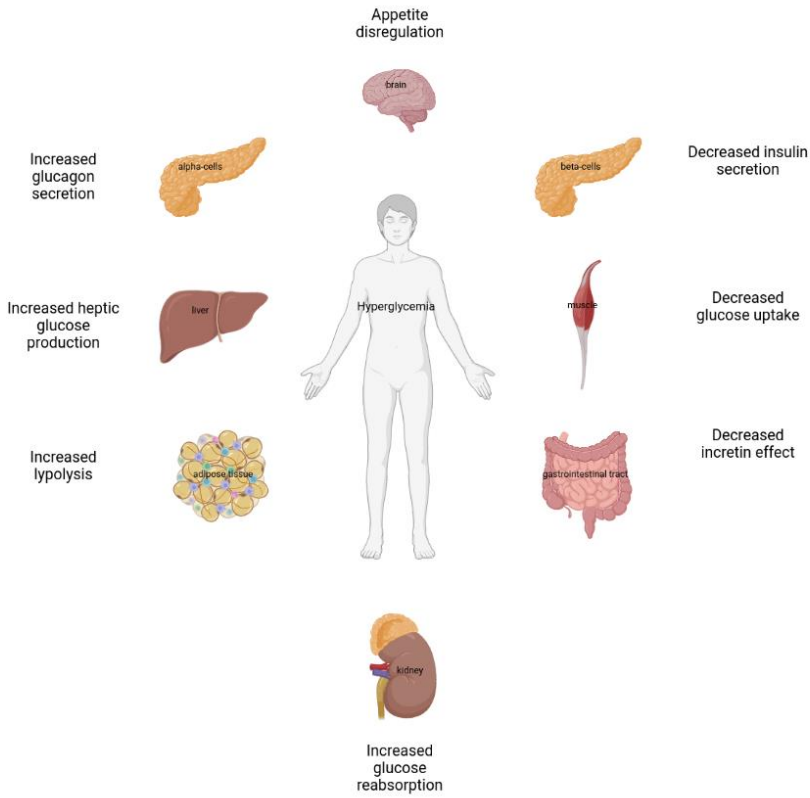

Figure 1:
The ominous octet in the pathophysiology of T2DM.

Insulin secretion

Pancreatic Islets

Around 150 years ago the pancreatic isles were discovered by Paul Langerhans. In addition to the insulin producing beta-cells, the islets are home to four other cell types each producing a different endocrine product. Glucagon producing alpha-cells, somatostatin producing delta-cells, pancreatic polypeptide producing gamma-cells and the rare ghrelin producing epsilon-cells. These cells produce their respective hormones as a response to various factors under different physiological circumstances and orchestrate the homeostasis of fuel metabolism and regulation [32]. Insulin is a vital hormone which stimulates metabolic fuel uptake, and its effects are counteracted by glucagon which promotes the release of stored fuels. Ghrelin plays a role in increasing the appetite while the pancreatic polypeptide suppresses the appetite and triggers the secretion of digestive enzymes form the

pancreas [33]. Somatostatin has an inhibitory effect of glucagon and insulin secretion [34].

Beta-cells

The fundamental physiological catalyst for insulin secretion form beta-cells is the elevated concentrations of blood glucose levels which occurs after feeding. To save the integrity of insulin secretion, there are several tightly regulated mechanisms and pathways to safeguard beta-cell function. The insulin secretion journey starts with pre-proinsulin which is first synthesized in the beta-cells and during a maturation process it is modified into proinsulin. This modification is conformational and is carried out in the endoplasmic reticulum with the help of several proteins. The proinsulin is then converted into two molecules: C-peptide and insulin. This conversion is carried out in the Golgi apparatus and the molecules are stored in immature vesicles [35]. The mature insulin is finally released from granules when beta-cells are stimulated with high glucose levels. This stimulation is carried out when glucose transporter 2 (GLUT2) sense the glucose molecules and shuttle them into the beta-cells. Within the cells the glucose is processed causing an elevation in the ATP/ADP ratio. This increase shuts the ATP-dependent potassium channels bringing about the depolarizations of the plasma membrane. Subsequently, the voltage dependent calcium channels open and take in calcium molecules which in turn induces the exocytosis of insulin [36], [37] (figure 2). It is important to note that there exist other stimuli which cause insulin secretion form beta-cells, these include other hormones, amino and fatty acids [36].

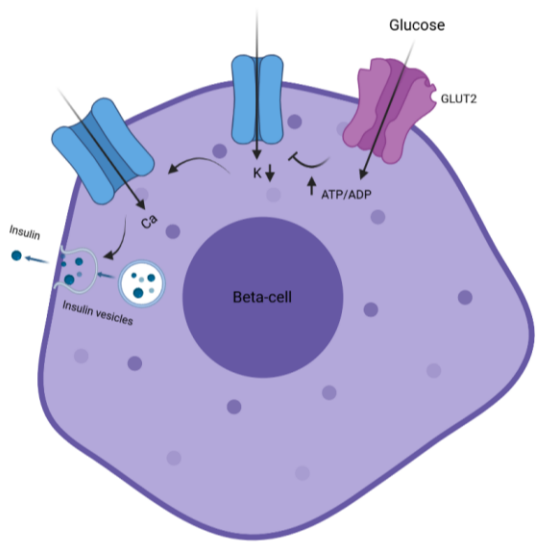

Figure 2:
Insulin secretion in beta-cells.

Historically, it was customary to believe that beta-cell dysfunction leads to beta-cell death causing T2DM, nevertheless current research suggests otherwise. As mentioned above there exists a complex network of pathways that safeguard beta-cell function, however a dysfunction may occur due to different environmental and genetic factors that lead to T2DM. A major culprit in beta-cell dysfunction is the inordinate feeding state which leads to hyperglycemia, hyperlipidemia and obesity causing insulin resistance and chronic inflammation. These conditions may eventually threaten islet integrity and beta-cell function due to the adverse pressures of inflammatory stress, endoplasmic reticulum stress and metabolic/oxidative stress [2], [38]. Endoplasmic reticulum stress can occur due to high concentrations of free fatty acids which interfere with calcium induced exocytosis. Hyperglycemia will also elevate the synthesis of proinsulin which may lead to the accumulation of misfolded insulin [39]. These irregularities in creating insulin or insulin precursors and disturbances in insulin secretion are some of the main drivers in beta-cells dysfunction and ultimately failure causing T2DM.

Dyslipidemia and Obesity

In the past several decades there has been a substantial shift in lifestyle in several parts of the world. These changes have been contributing to the growing number of the obese population globally [40], [41]. Obesity is generally associated with the increased number of adipocytes. Additionally, macrophages that enter the adipose tissue due to obesity generally become pro-inflammatory. These conditions among others will bring about different alterations in the structure and function of the adipose tissue giving rise to adiposopathy which is characterized as adipose tissue dysfunction [42], [43]. One of the main metabolic disorders associated with adiposopathy is insulin resistance/hyperinsulinemia which ultimately leads to dyslipidemia. High levels of fasting and postprandial triglyceride and low-density lipoprotein (LDL)-cholesterol and low levels of high-density lipoprotein (HDL)-cholesterol are some of the main hallmarks of dyslipidemia [44]. It is important to note that hypertriglyceridemia, low levels of HDL-cholesterol and high levels of LDL-cholesterol may be detected in individuals with regular blood glucose levels several years before T2DM diagnosis. In fact, this is usually interpreted as an early manifestation of T2DM, and insulin resistance is linked to dyslipidemia [45].

Very low-density lipoprotein (VLDL) is one of the main transporters of fat and cholesterol in the blood stream and is elevated during insulin resistance due to overproduction. Insulin affects both the liver and the adipose tissue which are critical in the production of VLDL. Lipolysis, a process by which free fatty acids are mobilized by lipase in the adipose tissue, is regulated by insulin. During insulin resistance, lipolysis is elevated leading to larger mobilization of free fatty acids. Free fatty acids are both regulators and substrates of VLDL [46]. Insulin also regulates the production of apolipoprotein B (apoB) which is one of the main surface

proteins of VLDL. Generally, the production of apoB is approximately fixed and is dependent on lipid levels. However, due to insulin resistance, the free fatty acid levels in the liver are elevated leading to an abundance of VLDL [47]. A second reason for elevated levels of VLDL is its diminished clearance. Lipoprotein kinase, which plays a crucial role in breaking down triglycerides from their transporters, is inhibited due to high levels of free fatty acids [46]. Additionally, the hepatic uptake of lipoproteins rich with triglycerides is diminished due to apolipoprotein C-III (apoCIII). apoCIII is a surface protein found in VLDL molecules and has an inhibitory effect on lipoprotein lipase. apoCIII expression was found to be elevated in T2DM patients and positively correlated with BMI [48].

From the mechanisms discussed above it is evident that insulin resistance plays a major role in elevated free fatty acid levels in the blood by reducing their absorption and enhancing lipolysis in the adipose tissue. However, the high level of free fatty acids in the blood stream itself can cause insulin resistance leading to a malicious spiral where both elevated levels of free fatty acids and insulin resistance enable each other [49]. The elevated levels of VLDL promotes the transfer of triglycerides from VLDL to HDL in exchange for cholesterol. This exchange is carried on by cholesteryl ester transfer protein (CETP) and leads to cholesterol-depleted HDL which is thought to be less protective and cholesterol-rich VLDL which can help the formation of plaques in the arteries [50]. In a similar manner, high plasma levels of VLDL and with the help of CETP, an exchange between VLDL and LDL will take place producing triglyceride-rich LDL particles. The triglyceride is removed by hydrolysis leading into smaller and denser molecules. These have a better ability to infiltrate blood vessels illustrating its higher potential in atherogenicity [51]. Diabetic dyslipidemia with its main hallmarks mentioned above is prevalent in T2DM patients and leads to elevated risk of atherosclerosis. Dyslipidemia is a critical bridge between obesity, T2DM and cardiovascular diseases (CVD) and weight loss has been shown to have a favorable effect in the development of T2DM and the lipid profile of T2DM patients [52]. Even though diabetes is a well-known risk factor for CVD, treatment for hyperglycemia may only have diminished effects on reducing the risk of CVD [53].

Diagnosis and treatment

Customarily, the assessment of glucose concentration in blood was used to diagnose diabetes patients. Today, diabetes is diagnosed based on two main criteria, the A1C criteria and plasma glucose criteria. The A1C criteria, also called hemoglobin (Hb) A1C or HbA1C test, evaluates the amount of blood sugar levels over a period of 12 weeks. This can be assessed due to the ability of sugar molecules to attach to the Hb protein on the red blood cells [54]. The plasma glucose criteria measures glucose levels in the plasma over different time points. Generally, the individual is orally administered 75g of glucose after fasting of 8 hours and glucose levels in the plasma

are measured at different time points. The fixed time points are at administration, 30 minutes, 1 hour and 2 hours after administration. This standardized metabolic stress test is called oral glucose tolerance test (OGTT) [55]. These tests are used for both diagnosis and screening purposes helping the identification of individuals with prediabetes, high-risk individuals, and symptomatic patients. The world health organization defines a fasting glucose value of 110mg/dl (6.1 mmol/l) as the cutoff for prediabetes. In the A1C criteria, values between 5.7 and 6.4% are indicative of prediabetes and individuals with these values have higher risk of developing diabetes and CVD. Fasting plasma glucose values above 126 mg/dl (7.0 mmol/l) or 2-hour plasma glucose values of 200mg/dl (11.1mmol/l) after administration or A1C above 6.5% are criteria for diagnosis with diabetes [56].

It is important to note that there exist several other molecules in the blood that can be used to track the development of diabetes. One such important molecule is insulin. The main function of insulin is to drive down blood glucose levels, therefore its measurement in the blood can provide an immense value in understanding the disease by comparing its levels to that of blood glucose. Two indices have been proposed, based on blood glucose and insulin levels, that would reflect beta-cell function (HOMA-B) and insulin resistance (HOMA-IR). These indices are relatively simple models that have been shown to provide decent prognostic value of complications related to diabetes [57].

Since OGTT provides values for glucose over different time points, this can be combined with blood insulin values at the same time points to develop different indices that can reflect insulin secretion and insulin resistance. Insulin sensitivity index (ISI) is a feasible substitute for HOMA-IR measuring insulin resistance and corrected insulin response (CIR) can be used to measure beta-cell function [58], [59]. Additionally, using the product of these two indices, it is possible to approximate beta-cell function after adjustment for insulin sensitivity.

The Treatment T2DM typically involves a combination of lifestyle modifications, medications, and sometimes insulin therapy. The primary goal of treatment is to control blood glucose levels, prevent complications, and improve quality of life. Lifestyle modifications are essential for T2DM patients and may include changes in diet, exercise, and weight loss. Dietary modifications may involve reducing intake of simple sugars, processed foods and sugar-sweetened beverages, and increasing consumption of complex carbohydrates, fiber, fruits, vegetables, whole grains, and lean protein. Exercise can help improve insulin sensitivity and glucose uptake in muscles, and weight loss can help improve glycemic control and reduce cardiovascular risk. A minimum of 150 minutes of moderate-intensity exercise is recommended per week, spread over at least three days per week, with no more than two consecutive days without exercise [56].

Pharmacological therapy is indicated when lifestyle modifications alone are not sufficient to achieve glycemic control. Metformin is the first-line therapy for T2DM.

It improves insulin sensitivity and reduces hepatic glucose production, leading to decreased blood glucose levels. Medications for type T2DM include oral agents such as metformin, sulfonylureas, meglitinides, DPP-4 inhibitors, GLP-1 receptor agonists, and SGLT2 inhibitors. Insulin therapy is usually reserved for patients who cannot achieve glycemic control with oral agents. Bariatric surgery, characterized as changes to the digestive system to help lose weight, has also been shown to have significant benefits in the treatment of T2DM, leading to remission in some cases. The choice of medication depends on various factors, including the patient's age, comorbidities, and preferences. In addition to medications, patients with T2DM may also need treatment for comorbidities such as hypertension and dyslipidemia. Regular monitoring of blood glucose levels, blood pressure, and lipids is essential to prevent complications such as cardiovascular disease, neuropathy, and retinopathy. Therefore, the treatment of T2DM involves a multifaceted approach and should be individualized based on the patient's medical history, comorbidities, and preferences [56], [60]–[64]. Ongoing research will likely lead to further advances in the treatment of T2DM, improving outcomes for patients with this condition.

Heritability of type 2 diabetes

Rationale

In recent years there has been an evident resemblance between the growing numbers of obesity and other metabolic disorders, and T2DM. The growth trajectory has been consistent in both developed and developing countries. This rapid rise of T2DM can be explained by the accelerated changes in human environment and behavior, after all changes in the genetic elements require much longer time periods to manifest in such a manner than a few decades. Nevertheless, various studies have illustrated that alterations in DNA sequences can have meaningful consequences on the risk of T2DM. Studying T2DM occurrence in monozygotic twins, who share 100% of their DNA, and dizygotic twins, who share 50% of their DNA, has shown that diabetes occurrence is higher in monozygotic twins [65]. Additionally, populations of certain ancestries have been shown to have higher risk of T2DM when compared to nearby populations despite sharing a common environment [66], [67]. Moreover, different studies have shown that diabetes in the family history is a risk factor and alterations in single pivotal genes can lead to detrimental changes in key protein functions which ultimately lead to the development of diabetes [19], [68]. These studies provide evidence that indicate the importance of genetic factors and the role they play in T2DM development. It is important to recognize that the interactions between environmental and genetic factors are essential for the development of

T2DM and highlight that the heritability estimate of T2DM goes as high as 72% [69].

Genetic studies

Historically, the genetics of monogenic forms of diabetes have been easier to identify. This was accomplished due to the high segregation of the causal genetic variants with the disease [19], [20]. In complex diseases, however, several genetic loci are involved and some of their effects are exerted due to their interaction with the environment making their identification complicated. T2DM is a complex disease and a successful approach to identify the effect of a genetic variation on the phenotype is to test whether that specific variant is significantly enriched in T2DM patients when compared to the healthy population. The approach preceding genome-wide association studies (GWAS) required prior biological insight about the given variant and the importance of a specific gene in T2DM development. However, the discovery of millions of variants in the human genome called single nucleotide polymorphisms (SNP)s together with the technological advancements paved the way for GWAS. This allowed the production of genotyping arrays that can probe hundreds of thousands of these SNPs with high accuracy. The successful understanding of the association structure between these SNPs have helped reduce the complexity when conducting GWAS, nevertheless a large sample size is usually required to find meaningful associations between variants and the phenotype investigated. In 2007 there were several GWAS on T2DM which led to the identification of several novel loci associating T2DM. These included genetic variants in Hematopoietically-expressed homeobox protein HHEX (*HHEX*), Zinc Transporter 8 (*SLC30A8*), cyclin-dependent kinase inhibitor 2A/2B (*CDKN2A/2B*), Cdk5 regulatory associated protein 1-like 1 (*CDKAL1*) and Insulin-like growth factor 2 mRNA-binding protein 2 (*IGF2BP2*). These studies also confirmed previously identified T2DM loci including Transcription factor 7-like 2 (*TCF7L2*), Peroxisome proliferator- activated receptor (*PPARG*), Potassium Inwardly Rectifying Channel Subfamily J Member 11 (*KCNJ11*) and Fat mass and obesity-associated protein (*FTO*) and demonstrated that numerous genetic variants have limited contribution to T2DM development [70]–[72]. Since populations of different ancestries can have different susceptibility to T2DM, GWAS on non-European populations uncovered loci that were only nominally implicated in European populations including Potassium Voltage-Gated Channel Subfamily Q Member 1 (*KCNQ1*) [73].

Since its conception, a relatively large cohort in GWAS is required to draw significant associations between genetic variants and the phenotype in question. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium was successful in identifying 18 new loci when larger number of European subjects were included. At first, 6 loci were identified using more than 53000 individuals and an

additional 12 were identified when the sample size was inflated to more than 100000 individuals [74], [75]. Similar efforts have been made by the Asian Genetic Epidemiology Network where the study population was East Asian and 8 new loci associated with T2DM [76]. Efforts have also been made to combine populations from different ethnicities including Europeans, East and South Asians and Mexican Americans. This approach would help in identifying additional variants and aid in better understanding the effects of previously discovered loci. The majority of variants that showed association with T2DM in these diverse study populations showed consistent directionality in all ethnic groups. Interestingly, however, several variants showed different association levels with T2DM in the different ethnic groups indicating the relevance of accounting ethnicity in GWAS [77].

Most of the genetic variants that have been identified in the previous studies are relatively common in some or all the populations that were used for their discovery. However, these variants reportedly had somewhat small effects on T2DM risk and development. In contrast, it is also reasonable to consider the existence of rather rare genetic variants with large effects on T2DM. These rare variants may explain a considerable fraction of T2DM heritability however they require more advanced methods for their identification. Advances in massive parallel sequencing and variant calling methods permitted the identification of such rare genetic variants in large cohorts. In fact, the frequency of some of these rare variants were as low as 0.2% in the population they were discovered in where generally 5% frequency of a genetic variant is considered common [78]. The endeavor of genetic studies was not exclusive to the T2DM risk and development but extended also to investigate the association of both common and rare variants with glycemic traits. These traits include quantitative measurements of blood glucose and insulin levels obtained and derived from OGTT. Interestingly, some variants that have shown association with fasting glucose concentrations were not found to be associated with the risk of T2DM and, in contrast, other genetic loci associating with fasting glucose levels associated with the increased risk of T2DM [79], [80]. Similarly, genetic studies also encompassed obesity and related measures taking into account demographic and environmental factors. These efforts included both rare and common genetic variants and were successful in identifying rare variants and gene-by-environment interactions [81], [82]. Up until 2022, the largest association analysis that included over 1 million participants from diverse ancestral backgrounds have identified more than 1000 genetic variants correlating with the risk and development of T2DM [83], [84].

Gene expression

The main information flow in the mammalian cell starts in the nucleus and results, generally, in proteins. Genes, which are considered basic units of heredity, are made up of DNA in mammalian cells. Proteins are considered the building blocks of life

and are a result of translating the genetic information into polypeptides. The genome contains most of the hereditary information of the cell and is selectively transcribed into RNA. This selective transcription is dependent on various internal factors and external signals. Similarly, not all RNA molecules are translated into proteins and this selective translation is also dependent on various factors. The process of transcribing DNA molecules into RNA molecules is called gene expression or transcription and is a complex process. Gene expression is dependent on the interplay between different DNA, RNA, protein molecules and extracellular signals. Moreover, this process is highly variable not only between species but also within populations and even within an organism. Additionally, there is a random aspect of gene transcription with random fluctuations of the expression of individual genes which can be explained by the involvement of intrinsically random biochemical reactions [85].

Gene expression has been extensively studied specially after the advent of microarray and massive parallel sequencing technologies. Furthermore, due to its significance in cellular function, efforts have been increasingly focused on surveying genetic expression in various human conditions and diseases including obesity and diabetes. The unique gene expression profile in a cellular population is what gives rise to different tissue types in the body. Moreover, high expression of certain genes in specific tissues generally suggest their importance in the function of that specific tissue [86]. This concept can be extrapolated in time, meaning that genes expressed in a given time span play an important role in the ongoing cellular function. These functions can vary and encompass the complete capacity of cellular functions including development. Therefore, gene expression can be thought of as a snapshot in tissue and time. Given this unique aspect of gene expression, several studies have analyzed gene expression profiles in different tissues involved in T2DM risk and development. For example, the expression of different genes affecting lipogenesis and adipogenesis were found to be changed during obesity due to the inflammatory response in adipose tissue. This is a contributing factor to the insulin resistance accompanying obesity [87], [88]. Moreover, the genetic expression of specific enzymes involved in muscle respiratory function was also found to be changed. Additionally, byproduct of fatty acids impairing insulin signaling due to altered mitochondrial function were also detected [89], [90]. Similarly, different alterations in gene expression patterns were observed in the liver and the pancreas. For instance, high-fat diet, which is a major risk factor for obesity and subsequently T2DM, was found to increase the expression of different transcription factors and gluconeogenic enzymes in liver. This resulted in nonalcoholic fatty liver disease and, interestingly, beta-cell failure and hyperglycemia. Overall, the expression of more than 600 genes were observed to be altered in pancreatic islets effecting beta-cell viability due to high-fat diet [91], [92].

Gene expression can be cellular specific and can change between cells belonging even to the same cell type. Some beta-cells were found to be more stimulated by

glucose signaling in contrast to others which were glucose non-responsive and necessitate additional catalyst for insulin secretion. Moreover, the responsiveness of beta-cells to the different stimuli is also diverse, where a small population of beta-cells were found responsible of secreting the majority of insulin [93]. A major shortcoming in the approaches depending on bulk RNA expression analysis, such as microarrays and RNA-seq, is that they cannot capture this heterogeneity of expression. These technologies can only provide a snapshot of the mean expression in the cellular population in comparison to single-cell RNA-seq approaches. The single-cell gene expression approach can successfully discern the unique expression of different cells and identify subtypes of beta-cells based on cellular expression profiles [94]. Moreover, this approach can generate a large amount of data that can be used to mine for mutations and alternative splicing that can be subsequently used in numerous ways in downstream studies [95]. Nevertheless, gene expression, whether based on bulk or single-cell approaches, can provide important insights of cellular function by identifying specific genes and signaling pathways and these approaches were used in understanding the development of T2DM and insulin secretion.

Epigenetics

Epigenetics is described as the changes that effect genetic function and are heritable but do not entail changes in the genetic sequence. Even though these changes are both mitotically and meiotically heritable, they can be reversable. Epigenetic changes can occur due to normal development during a lifespan of an organism but also due to interactions with environmental factors. There are several epigenetic determinants that can affect gene expression irrespective of the underlying DNA sequence. The combination of these determinants is called the epigenome and its main factors are DNA methylation, histone modifications and non-coding RNA molecules.

DNA methylation is described by the addition of a methyl group to the cytosine residues on the DNA strands and this addition can have a direct effect on gene expression. There are several enzymes, called methyltransferases, that are responsible for adding methyl groups to cytosine residues such as DNA (cytosine-5)-methyltransferase 1 (DNMT1), DNA (cytosine-5)-methyltransferase 3A (DNMT3A) and DNA (cytosine-5)-methyltransferase 3B (DNMT3B). However, the demethylation of the DNA can occur either in a passive way with the inactivity of methyltransferases during replications or actively by enzymes such as ten-eleven translocation (TET). The methylation can occur on all cytosine residues in the genome however, it has been mainly observed when a cytosine nucleotide is followed by a guanine nucleotide. This pattern is called a CpG site and genomic regions with high frequency of CpG sites are called CpG islands. Most gene

promoters in the human genome contain CpG islands and methylation patterns in these islands can have a direct effect on the gene expression.

The DNA double helix is generally wrapped around proteins called histones that are susceptible to different chemical modifications. These modifications occur in different parts of the protein and include among others acetylation, methylation, and phosphorylation figure 3). Depending on the modification patterns, the DNA will be either tightly or loosely wrapped around the histones. This, in turn, will have an effect on the ability of other proteins to access the DNA strands and essentially effect genetic expression. The combination of these modification patterns is called "histone code" and in addition to its role in gene expression it plays a vital role in genomic stability and DNA repair.

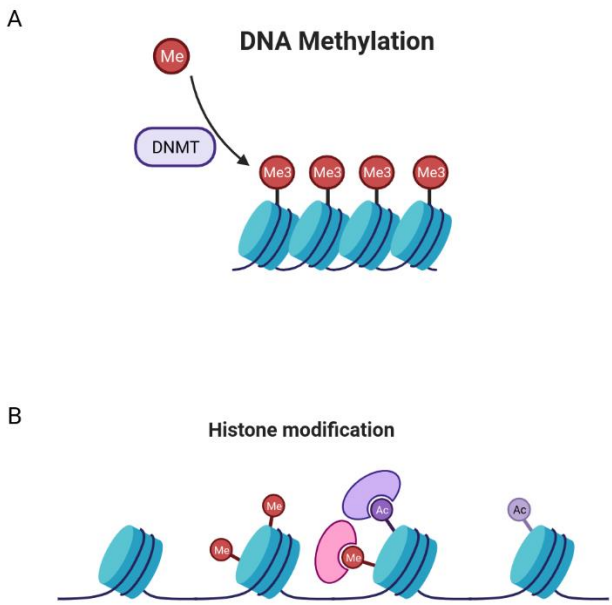

Figure 3:
A. methylation of DNA strands. B. modification of histones.

Non-coding RNA molecules exert their effects on gene expression in different ways based on the type of the molecule. Some non-coding RNA molecules can attach to their complementary coding RNA molecules and cause them to decay effectively stopping the protein synthesis of that specific gene leading to gene silencing. Other non-coding RNA molecules have the ability to recruit different proteins that can change the chemical modification patterns of histones which will affect gene expression in the manner mentioned above. Therefore, epigenetic patterns and

alterations in these patterns can play an important role in cell-specific gene expression, differentiation and genomic stability and structure among other things [96]. To better understand the effects of epigenetics in development and disease, two pivotal public databases have been created that provide essential information about gene expression, genomic structure, DNA methylation, histone modification and regulatory elements. The Encyclopedia of DNA Elements (ENCODE) accommodates information about functional elements in the genome based on several cell lines [97]. The Roadmap Epigenomics project provides references to the human epigenome based on different tissues and cell types [98]. These public databases have been used as a reference point in investigating the effects of epigenetic changes in T2DM.

Epigenetic studies

Due to the known effects of environmental changes and individual behavior on T2DM development, it is reasonable to take the epigenetic factors into account when conducting genetic studies on T2DM. This may indeed increase the complexity level of the genetic studies but at the same time provides useful information in understanding the interactions between genetic and environmental factors. It is important to keep in mind that epigenetic patterns are tissue and cell specific therefore studies concerning the epigenetics of T2DM are usually conducted in relevant tissues such as islet of Langerhans, adipose tissue, liver, skeletal muscle, and blood due to ease of access. Early epigenetic studies were conducted on the DNA methylation level where specific genes of interest were the subjects of the investigation. Diminished expression of some genes including *INS*, which encodes insulin, were found in the islets of T2DM donors. Increased DNA methylation accompanied the diminished expression of *INS* which also associated with defective insulin secretion. Moreover, high levels of plasma glucose and HbA1C appeared to have a direct effect on the elevated methylation levels [99], [100]. In light of recent advances in technology and methodology it is now possible to interrogate more than a handful of CpG sites at time. In fact, methylation arrays, similar to genotyping arrays, allow the investigation of more than 450 000 methylation sites which span the majority of CpG islands and their flanking regions in the human genome [101]. Nevertheless, it is still important to keep in mind that this only covers a small fraction of methylated sites in the human genome and the remaining sites may still play important roles in gene expression and regulation. With methylation arrays, it was possible to perform epigenome-wide associations in pancreatic islets from T2DM and non-diabetic donors. Indeed, genes such as *HHEX*, *KCNQ1*, *FTO*, *TCF7L1* and *PPARG* that were implicated in T2DM and obesity using GWAS were found to have divergent methylation patterns between T2DM patients and controls. Genes such as Cyclin Dependent Kinase Inhibitor 1A (*CDKN1A*) and Phosphodiesterase 7B (*PDE7B*) showed decreased methylation in T2DM islets. Methylation of the promoter regions of these genes were found to have a detrimental

effect on their expression. When replicating T2DM conditions in vitro by elevating the expression of these genes, glucose induced insulin secretion was diminished. *CDKN1A* plays a role in cell cycle regulation and its overexpression was shown to reduce beta-cell replication [102]. Due to the complexity of T2DM and the several tissues involved in its pathogenesis including adipose tissue, liver and skeletal muscle, DNA methylation was investigated in these tissues as well. DNA methylation patterns within several CpG sites seem to have been changed in these tissues form T2DM patients and healthy individuals. This indicates that the epigenome may play a role in the pathogenesis of T2DM by effecting changes in these tissues however, similar to GWAS, each CpG site seems to have a small effect on T2DM [103]–[105].

Advances in massive parallel sequencing techniques have helped overcome the limitation of methylation arrays used to investigate CpG sites in the genome. Using bisulfite conversion, whole-genome bisulfite sequencing allows the investigation of more than 80% of the CpG sites in the human genome. Bisulfate treatment of the DNA converts unmethylated cytosines into uracil residues introducing SNPs which can be analyzed and provide single nucleotide resolution of the methylation status of the analyzed DNA. Using whole-genome bisulfate sequencing researchers have successfully identified genomic regions with altered methylation patterns that overlapped with genes that have been implicated by DIAGRAM in T2DM risk. Genes including *TCF7L2* and *KCNQ1* had altered methylation and expression in T2DM islets when compared to controls. These changes were simulated in in vitro environments and led to defective insulin secretion which suggests that islet dysfunction and insulin secretion can be mediated by epigenetic mechanisms [106]. It is important to note that epigenetics as a field is relatively new and most efforts investigating epigenetic changes in the context of T2DM research used DNA methylation. There have been efforts in studying histone modifications in the human genome however, most of these were conducted in islets from non-diabetic individuals. Nevertheless, peripheral blood form T2DM patients have been used to investigate histone modifications and found that changes in histone modification impact vascular dysfunction in T2DM patients [107].

Similar to environmental changes, changes in the DNA sequence can also lead to changes in DNA methylation. The simplest form of this would be the introduction or the removal of a CpG site in a given position. However, the effect of other changes in the DNA sequence on DNA methylation, whether in close proximity or not, can be harder to identify and investigate. Several genetic variants that associated with T2DM either introduced or removed CpG sites and effected the methylation of surrounding sites. These changes in the DNA sequence and alterations in methylation patterns were also shown to be associated with gene expression and insulin secretion [86]. Interestingly, a direct implication between SNPs, DNA methylation and gene expression was found in the human leukocyte antigen (*HLA*) region which is firmly involved in T1DM. Functional studies of genes in this region

showed their involvement in pivotal biological mechanisms such as cell replication and programmed cell death. Additionally, genetic variations, in the form of SNPs, which associated with CpG sites also showed association with several diabetic genes including *INS*. These variations were shown to have a causal effect on gene expression and insulin secretion [108].

Epigenetics is viewed as heritable elements that can change over time and effect genetic function. However, as a field it is still relatively new and epigenetic studies investigating T2DM and related diseases are on the rise. All the evidence indicate that the epigenome plays an important role in T2DM and further understanding the epigenetic mechanisms provides new opportunities in understanding diabetes. Environmental factors that can affect the epigenome such as nutrition, physical activity and even the intrauterine environment can be crucial in the risk of T2DM by effecting beta-cell development and function. Importantly, similar to genetic studies, large cohorts are required to uncover the effects of epigenetic alterations in T2DM risk and development.

Fetal Programming

Perhaps the importance of epigenetic and environmental effects and their consequences on gene expression become more evident when their contribution towards the individual's health is considered from the time of conception. The intrauterine environment is the first and most important environment that plays a fundamental role in fetal development. However, this environment may also influence infancy and adulthood. Assaults and disruption to the ideal conditions during the development of the fetus and its essential organs can have detrimental effects in the future. Fetal programing refers to these changes that occur during the development due to alterations to the optimal intrauterine environmental conditions. One of the early descriptions, of how in utero fetal development is affected by changes in the environment, was the effect of malnutrition on birthweight. Individuals who were born to mothers who lived during the Dutch famine during World War II were significantly underweight at birth. The consequences of this manifested in adulthood when these individuals showed impaired glucose metabolism and effectively had higher risk to develop T2DM. Interestingly, in the cases where the famine occurred during a later stages of gestation the risk of insulin resistance and T2DM was shown to be higher [109]. This observation suggests that assaults to the in utero conditions amid development is "memorized" by the body and may end up expressed in the form of future health problems. However, it is important to note that programming can occur at even earlier stages than conception. During the formation and production of gametes most of the epigenetic signature of the parents which include DNA methylation and histone modification, among others, are removed. Nevertheless, during fetal development epigenetic imprinting of the DNA, which leads to adjusted gene expression, is established [110].

Of the different environmental factors that can have an effect on fetal programming and gene expression, nutrition is considered the most important. Studies that were conducted on the mothers and offspring from the Dutch famine showed the negative effect of inadequate nutrition in fetal programming. A “thrifty phenotype” hypothesis was proposed by David Barker, now called the “Barker hypothesis”, which proposes that under-nutrition of mothers during pregnancy can affect the offspring’s health in later life. Furthermore, the hypothesis suggests that inadequate nutrition may hinder fetal growth by focusing the blood supply to the brain and away from other organs such as muscle, liver, kidney, and pancreas. Moreover, hormones that are important for fetal growth such as insulin seem to be less effective and are produced with lower quantity [111], [112]. Based on this, The Developmental Origins of Health and Disease (DOHaD) hypothesis was proposed which suggests that the fetus is capable to adapt to different in utero conditions such that these adaptations will enhance its survivability after birth. The discrepancy between the prenatal programming and postnatal conditions is a major reason for the elevated risk of disease in later life [112], [113] (figure 4).

There have been several observations that connect the maternal nutritional state during pregnancy to the epigenome of the fetus. Nutritional elements such as vitamins, proteins, fat, and carbohydrates have an impact on genetic expression. This impact is augmented during the early stages of development and infancy and is manifested as alterations in the epigenome and gene expression. The metabolism of nutrients such as carbohydrates can affect DNA methylation and during malnutrition low levels of carbohydrates will cause diminished levels of methyl donors leading to changes in DNA methylation [114]. However, it is important to consider that not only malnutrition can lead to fetal programming and changes in the epigenome. Overnutrition is also responsible to the increased risk of several diseases such as obesity, hypertension, and diabetes mellitus and this is also realized due to fetal programming [115]. During pregnancy, it is commonly accepted that the basal energy levels are elevated due to the demands of the pregnancy. This is also typically accompanied by elevated insulin levels. However, this can be intensified due to over-nutrition but also due metabolic abnormalities. One such abnormality that can have a profound effect on the fetus during development and later life is GDM.

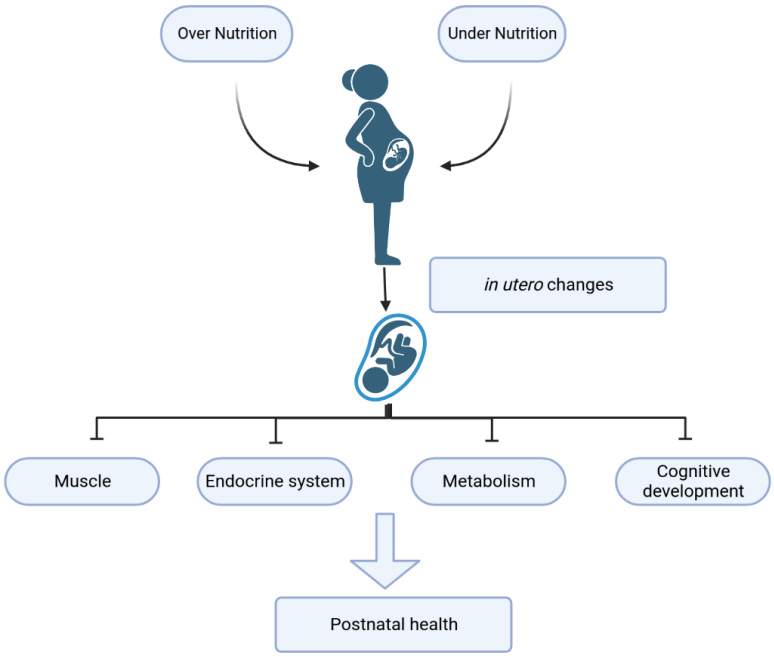

Figure 4:
Effects of over/under nutrition during pregnancy on the fetus.

GDM

Prior to the diagnosis with GDM, metabolic changes such as elevated insulin resistance and possible deficient insulin secretion could be present. These defects in pregnant women are on the rise due to the global increase of obesity. During a normal pregnancy, baseline energy levels are elevated due to increased hepatic production of glucose which is accompanied by increased insulin levels. This elevation is needed to accommodate the energy needs of the fetus, nevertheless, fasting glucose levels in the plasma is decreased. A possible reason for this decrease, in early gestation, is the inflation in plasma volume. Another reason for the continuation of low glucose level in the plasma is because of the higher energy needs of the fetus during late gestation. Additionally, during normal pregnancy, maternal sensitivity to insulin is decreased where the glucose uptake by muscle and fat tissue is reduced by around 50%. This leads to elevated levels of insulin production and secretion to maintain normal glycemic levels in the blood. Due to reduced insulin sensitivity, the metabolism of fats and lipid is also affected and can have negative effects on the fetus such as elevated adiposity and fetal growth [116]. Generally, most pregnant women who get diagnosed with GDM initially have normal blood

glucose levels and their decreased insulin sensitivity is not detected. During early gestation, beta-cells in the pancreas can keep up with the increasing demands of insulin secretion. However, at later stages of gestation, and due to pregnancy induced increased insulin resistance, insulin production becomes inadequate leading to hyperglycemia [117]. One reason for the elevated insulin resistance during pregnancy and GDM is disruption in the insulin signaling cascade in the body. This disruption can occur due to elevated levels of inflammation during obesity [118]. Additionally, the interaction of insulin with peripheral muscle cells is impaired leading to lower glucose uptake by these cells. This defective interaction between insulin and muscle cells is due to reduction in signaling molecules during pregnancy which is augmented in women with GDM [119].

One of the fundamental roles of the placenta is to transport nutrients from the mother to the fetus. During normal metabolism, sufficient quantity of glucose, fat and amino acids are delivered to the fetus however, the placenta has different efficiency levels in transporting the different metabolites. Glucose transportation is the most efficient and a saturation is reached only when the difference between glucose concentration of the mother and the fetus is larger than 25mmol/l. Therefore, the placenta is usually ineffective in protecting the fetus from high glucose levels [120]. The difference between maternal and fetal glucose concentration levels is the foremost factor effecting the glucose flow from the mother to the fetus. However, this difference is not only determined by maternal glucose concentration but also by fetal insulin levels. Insulin produced by the fetus induces the glucose uptake by the fetal tissue and fetal hyperinsulinemia increases the difference of fetal and maternal glucose concentration, therefore the amount of maternal glucose reaching the fetus is elevated [121]. The placenta possesses the capability to adapt to the changes in the environment however limited this capacity maybe. An example of this capacity is the increased vascularization in the placenta as a response to fetal hyperinsulinemia. Nevertheless, GDM represents an acute environmental disturbance eclipsing the capacity of the placenta to adapt to environmental changes therefore leading to possible pathological outcomes [122].

Children who are born to mothers that were diagnosed with different diabetes types have developed a wide range of detrimental outcomes. These outcomes are shared with children born to mothers who were diagnosed with GDM and include increased risk of diabetes mellitus. This was accompanied by reduced insulin sensitivity and secretion, increased fasting glucose levels, obesity, and risk of cardiovascular disease [123]. The consequences of GDM on the infant includes fetal macrosomia or “large for gestational age”. Consequently, this can cause several adverse outcomes including higher birth weight. Generally, infants born to mothers with GDM have higher risk of becoming obese and have an elevated risk of developing T2DM [124]. Studies investigating the outcomes of GDM on the fetus uncovered different transcriptional discrepancies between GDM fetuses when compared to non-exposed controls. Maternal nutrition during this period is a pivotal risk factor

for GDM and several studies have indicated the involvement of different epigenetic mechanisms in impaired gene expression and identified dissimilar DNA methylation profiles [125]. Animal studies that mimicked GDM conditions have replicated these observations and the normalization of maternal blood glucose levels during pregnancy have alleviated some of these pathological outcomes [126]. In contrast to the immediate benefits of GDM treatment the long-term benefits are yet to be established since most follow-up studies are conducted on rather short time span. Sadly, without any solid interference a GDM diagnosis is likely to lead to diabetes of both mother and offspring and is a start of a vicious circle of diabetes mellitus and related comorbidities such as obesity and cardio metabolic disorders.

Anemia

Another major insult to the optimal developmental conditions in the intrauterine environment is low oxygen levels as a result of undernutrition. This can have significant implications on the development of the fetus and leads to fetal programming preparing the fetus for this adverse condition. Anemia is a condition where the blood cannot carry enough levels of oxygen to meet the physiological needs of the individual. It is one of most important health problems effecting a large proportion of the population and is overrepresented in women. Anemia rates are higher in low- and middle-income countries and can have significant detrimental effects on maternal and fetal health [127]. Anemia can be a result of simple arithmetic, where the number of produced red blood cells are less than the number lost. Additionally, low concentrations of Hb, which is an iron containing oxygen transporter, will reduce the ability of the red blood cells to carry oxygen. These circumstances occur due to deficient nutrition, genetic disorders, and blood loss among others [128]. Importantly, anemia is diagnosed based on Hb levels which changes based on different factors such as age, sex, pregnancy, genetic and environmental factors.

A major global cause of anemia during pregnancy is nutritional deficiency specifically in iron, folate, and B vitamins. “Anemia” during pregnancy is usually a normal physiological process. From the early weeks of gestation, the maternal plasma volume inflates however the red blood cell count does not increase with the same rate. Since anemia is diagnosed based on Hb levels or red blood cell counts this increased plasma volume will lead to “anemia”. A proposed hypothesis for this plasma expansion, is that this circumstance will expedite the transportation of nutrients and oxygen by facilitating placental angiogenesis. It is important to note that throughout a normal pregnancy the number of red blood cell increase as well, and this increase is dependent on iron availability. However, the elevated count of erythrocytes is masked due to the increased plasma volume [129]. In addition to the maternal plasma inflation, increasing levels of iron is needed by the fetus and the placenta. Throughout the pregnancy the fetus will require elemental iron for its intrinsic metabolic needs and the placenta functions as a buffer with iron stores

against possible stages of low iron availability [130]. A mild and a short-lived iron deficiency most likely will not lead to anemia where stored iron molecules are used to compensate for this shortage. However, when iron deficiency is consistent, Hb synthesis is threatened, and generally available iron molecules will be directed to red blood cells on behalf of other tissues. For example, during iron deficiency, the available iron molecules are used by the red blood cells instead of the brain leading to abnormal brain function. This observation has been replicated using genetic models accentuating the importance of iron during pregnancy [131]. In addition to iron deficiency, deficiencies in other nutrients can cause anemia such as folic acid and B vitamins. B vitamins play an important role in Hb synthesis and low levels have been shown to be associated with anemia [132]. B12 deficiency is usually due to malabsorption whereas folic acid deficiency is usually due to insufficient dietary intake. These nutritional deficiencies will lead to an impaired erythropoiesis [133]. Moreover, genetic conditions such as sickle cell, where the red blood cells have an abnormal shape resembling a sickle, and thalassemias can cause anemia.

Another major cause of anemia is malaria which is caused among others by *Plasmodium falciparum* and *Plasmodium vivax*. These parasites are mainly present in Africa and cause major health complications that can lead to death. Pregnant women, infants and young children in Africa have higher risk of contracting malaria and the majority of death caused by malaria are among young children [134]. Malaria can cause both anemia and severe anemia and its effects are manifested in different ways. Malaria leads to decreased production of red blood cells and diminished survival of erythrocytes. This shortened red blood cell survival is due to the destructive effects of the Plasmodium parasites [135]. Interestingly, malaria and iron deficiency normally coincide, nevertheless, low iron levels can have a protective effect. Iron is a necessary nutrient for the growth and the survival of the Plasmodium parasites, and these parasites have a disruptive role in iron metabolism. The expression levels of hepcidin, which plays a role in regulating iron distribution, is altered by these parasites and plays an important role in reducing anemia and severe anemia due to malaria [136].

Regardless of the cause, anemia during pregnancy is a hard condition that leads to major changes in the intrauterine environment and can cause sever effects on fetal health. Anemia has been shown to be associated with reduced birth weight and this has been shown to play a role in increased risk of diabetes mellitus and CVD [111], [137].

Parent-of-origin effects

Changes in the maternal environment during pregnancy is manifested as alterations in the intrauterine environment and transferred to the fetus via the placenta. This means that the main effects of fetal programming are of maternal origin. As discussed above these are manifested in changes in the epigenome and genetic

expression of the fetus such that the fetus is ready for these environmental changes after delivery. However, epigenetic changes that effect gene expression are not exclusively of maternal origin and it can also be of paternal origin in what is called parent-of-origin effects. In general, autosomal chromosomes have two alleles at each locus where one is inherited from the father and the other from the mother. The exception to this in humans is alleles on sex chromosomes. Male offspring will obtain the X chromosome from the mother and Y form the father whereas, female offspring will obtain an X chromosome from each parent. Essentially, nearly all genes are expressed in similar quantity regardless of their parental origin and this can have beneficial effects. For example, if an allele form one of the parents has a defect and is not expressed, the expression of the other allele from the other parent can be beneficial. However, in certain circumstances, the expression of selected genes can be dependent on their parental origin. For example, the intrauterine environment provided by the mother normally has optimal conditions for the benefit of the offspring and alleles that favor the transfer of resources from the mother to the offspring can be over expressed in the fetus. Generally, maternal commitment in reproduction is much larger than the paternal one and this does not only include the transfer of resources during gestation but also includes months or years after delivery such as lactation. This differential demand between the maternal and paternal provisions leads to the differential expression of certain genes [138]. A "Parental conflict" theory was proposed suggesting that the commitment of paternal and maternal genomes in terms of resource allocation can differ. The paternally expressed genes predominantly favor the transfer of resources from the mother whereas the maternally expressed genes play a more balancing role of resource allocation [139].

The frequency of paternal and maternal alleles in the next generation can be used to explain this discrepancy in the expression. It is important to note that, during a lifetime of mating, maternal alleles are always represented in the offspring with a frequency of 0.5. However, this is not the case with paternal alleles since females can mate with several males. Therefore, paternal alleles are represented with a frequency of 0.5 in their offspring and 0 in the half siblings. Hence there can be a selection of allele specific expression based on the parent of origin. For example, when both parents contribute towards the survival of a given offspring on behalf of another and when these offsprings are half siblings. The increased maternal investment towards a given offspring will enhance the survival of the paternal alleles however the survival of the maternal allele will stay the same regardless of the offspring. Therefore, under this selection, paternal alleles can have higher expression which can be also accompanied by lower maternal allele expression. When alleles in a given gene have a complete differential expression, meaning one allele is not expressed at all, this gene is then called an "imprinted" gene. This is achieved by changes in epigenetic patterns and in mammals more than 100 genes have been discovered to be imprinted [140], [141] (figure 5).

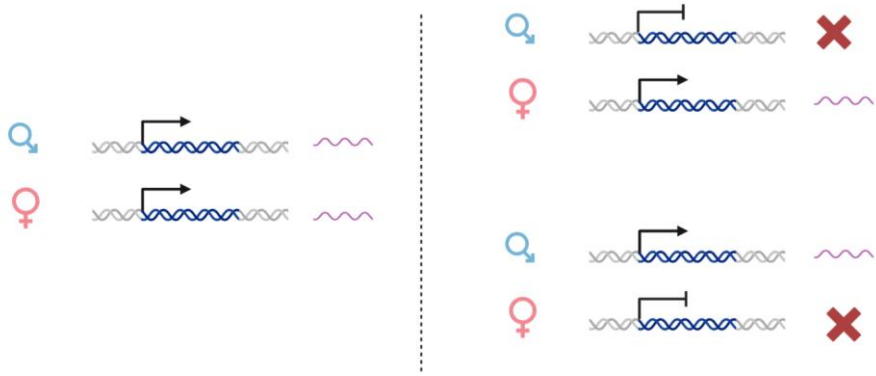

Figure 5:
Genetic imprinting. Left panel: bi-allelic expressed gene. Right panel: Paternal imprinted gene (top), and maternal imprinted gene (bottom).

Genomic imprinting is the main process in which parent-of-origin effects are manifested. This phenomenon has been observed in several taxa and is involved in various pivotal functions such as embryonic and placental development. The discovery of genomic imprinting was accelerated due to genomic abnormalities that affected imprinted genes such as translocations, loss-of-function mutations and uniparental disomies and these genomic abnormalities led to complex disorders such as Prader-Willi and Angelman syndromes [142]. Imbalanced genetic expression in reciprocal heterozygote individuals can also occur in other conditions where genomic imprinting is not present. These conditions, such as environmental changes that cause epigenetic alterations leading to under-expression of certain alleles and random monoallelic expression, will not necessarily give rise to parent-of-origin effects. On the other hand, conditions that lead to the appearance of parent-of-origin effects in the absence of genomic imprinting can include the effects of parental genetic variants that can have a phenotypic effect on the offspring and mitochondrial maternal effects [143]. To better understand how parent-of-origin effects are manifested through genomic imprinting it is important to consider the reciprocal heterozygote genotype. Generally, the reciprocal genotypes are considered phenotypically equivalent however this is not the case during genomic imprinting. In genomic imprinting it is important to ascertain the parent of origin of the two alleles and in populations with genotypes in Hardy-Weinberg equilibrium the allelic parent of origin of the offspring can be determined using the genotype of the parents. However, heterozygotic parents for a given locus will not be informative in directly inferring the offspring genotype and additional information may be needed for inferring allelic parent of origin. In such cases linkage information can be useful if the locus in question is linked to informative loci or if the haplotype information containing the locus of interest can be obtained using certain algorithms [144].

Genomic imprinting generally occurs on a cluster of genes and the regulation of these genes is carried on by a shared imprinting control region. Various common regulatory elements such as long non-coding RNAs and DNA methylation are utilized to achieve the regulation of an imprinted cluster. Generally, the DNA methylation is divergent between the paternal and maternal alleles giving rise to differentially methylated regions. These regions also contain distinctive chromatin confirmation and histone modifications leading to the recruitment of dissimilar transcription factors that play a role in the imprinted expression of these genes [145]. These contrasting patterns in the epigenetic marks allow genomic imprinting to occur in different and complex constellations. In the *H19-IGF2* locus, 19 Imprinted Maternally Expressed Transcript (*H19*) is maternally expressed whereas insulin-like growth factor 2 (*IGF2*) is paternally expressed and this locus was found to be associated with Beckwith-Wiedemann syndrome [146]. Furthermore, genetic alterations may also lead to a complex imprinting pattern. A SNP in an intergenic region in the callipyge locus was shown to have molecular effects altering DNA methylation patterns, transcription factor binding and transcription of the intergenic region. This mutation leads to the inversion of parent-of-origin expression from the maternal and paternal chromosomes which is also achieved by employing microRNAs [147]. The scope of complex imprinting constellations can be further extended when considering intricate environmental and genetic interactions occurring in complex traits or diseases. Imprinting patterns can be dependent on various factors such as environment and sex where the paternal or maternal expression of an imprinted gene changes. An example of this context-dependent imprinting was found in mice where the paternal and maternal expression of imprinted loci was dependent on diet [148]. Moreover, tissue-specific imprinted expression was also observed in different embryonic and extra-embryonic tissues. These imprinted genes were mainly controlled by epigenetic mechanisms and the expression pattern of most was concordant with "parental conflict" theory mentioned above [149].

Several imprinted genes are involved in beta-cell function and development highlighting the importance of considering genomic imprinting and parent-of-origin effects in T2DM studies. A delicate control of imprinted gene expression is essential for the regular function of pancreatic islets, beta-cell function, and glucose homeostasis. Indeed, differences in methylation patterns and imprinted gene expression was found between T2DM and non-diabetic islet donors. The expression of Delta Like Non-Canonical Notch Ligand 1 (*DLK1*) was overexpressed in beta-cells of T2DM patients in contrast to non-diabetic donors. On the other hand, the differentially methylated region of the long non-coding RNA Maternally Expressed 3 (*MEG3*) was found to be hypermethylated in T2DM donor islets leading to under expression. Moreover, methylation levels in the same region were found to be correlated with GDM and fetal growth. *DLK1* is a paternally expressed gene and *MEG3* is a maternally expressed transcript and both have an important role in insulin sensitivity and beta-call stability [150]. Furthermore, mutations found in the

imprinted loci hosting the Cyclin Dependent Kinase Inhibitor 1C (*CDKN1C*), a maternally expressed imprinted gene that plays a role in cellular replication and differentiation, and *KCNQ1C,* which is pivotal in insulin secretion, associated with T2DM and insulin secretion [66], [124]. The above-mentioned imprinted genes, among others, play an immense role in beta-call development and function. Moreover, the expression of imprinted genes can be modulated based on cell/tissue type and different environmental factors. Similarly, Complex diseases such as T2DM include inter-organ crosstalk, convoluted genotype-phenotype relationship and intricate genetic and environmental interactions compelling the inclusion of parent-of-origin effects and genomic imprinting while conducting T2DM genetic studies [151].

Materials and methods

Study cohorts

Given the complex nature of T2DM and considering the role of genetics and environmental factors in its development, several population cohorts have been established. These cohorts serve as the foundation upon which genetic studies are conducted to investigate T2DM and related phenotypes and metabolic disorders. In this thesis several population cohorts were used for this purpose, which vary in their size, ancestry, and genetic structure. These include:

1. The Botnia cohort.
2. FOETAL for NCD (Foetal exposure and Epidemiological Transition: the role of Anaemia in early Life for Non-Communicable Diseases in later life).
3. Hungarian Transdanubian biobank.
4. Pune Maternal Nutrition Study.

Different subsets of these cohorts were utilized in the different projects included in this thesis work. These subsets were selected based on the aim of the projects and the analysis approach used to fulfil these aims. Below is a general description of these cohorts and detailed cohort characteristics are further provided for each cohort in their respective projects.

The Botnia cohort

To study non-insulin dependent diabetes mellites the Botnia study was founded in 1990. The study population is predominantly European with Finish and Swedish individuals being the main participants. Initially, the participants were recruited from 5 different primary healthcare units in the Botnia region in Finland and subsequently was expanded by inviting their family members. The recruitment is based on patients with T2DM diagnosis and their family members where different test and biomarkers were measured. However, to study relevant metabolic disorders and DNA alterations associated with them additional biomarkers and phenotypes

were also collected from the study participants. Due to the nature of this cohort and the recruitment of its participants, this cohort can be characterized as a large family study [152]. A subset of this cohort was selected for genetic analysis and association studies that was included in different projects within this thesis work.

FOETAL for NCD

The FOetal Exposure and Epidemiological Transitions: the role of Anaemia in early Life for Non-Communicable Diseases in later life was created for the study of non-communicable diseases including GDM and anemia. This population-based study initiated in 2014 and recruitment was continued until 2016. The study participants were recruited through mobile clinics from two different districts in northeast Tanzania. This study recruited women either before pregnancy or after pregnancy but before 14 gestational weeks. The pregnant women were divided into three different categories based on their blood Hb levels. The categories were defined as follows. Severe anaemia: Hb ≤ 8 g/dL [4.96 mmol/L], mild-moderate anaemia: Hb 8.1–10.9 g/dL [5.02–6.81 mmol/L] or without anaemia: Hb ≥ 11 g/dL [6.82 mmol/L]. Different types of data were collected at different time points ranging from the recruitment until after delivery. These included general health screening, socioeconomic status, anthropometry, ultrasound, blood samples, cord-blood samples, placenta biopsy and neonatal anthropometry [153]. Data collected from subjects in the FOETAL for NCD were primarily used to investigate the effects of both anemia and GDM on fetal programming.

Hungerian Transdanumbian biobank

The Hungarian Transdanubian biobank (HTB) includes data and biological samples from patients with T2DM and their families. The biobank includes data from over 9000 individuals and was initiated in 1992 at the Hungarian Heart Center in Balatonfüred.

Pune Maternal Nutrition Study

To study the effects of maternal nutrition and physical activity on fetal growth, risk of diabetes and other metabolic disorders in later life the Pune Maternal Nutrition Study was initiated. The study participants in this cohort were recruited from six villages around Pune city in 1994. This region was selected due to its heavy reliance on agriculture for subsistence. Moreover, the Pune area lacked proper irrigation making it drought prone and leading to malnutrition of the population. However, the study area is continuously developing since 1994 and witnessed improved irrigation and urbanisation. Initially the study included pregnant women (F0) who

experienced malnutrition and were followed up with their offspring (F1). Currently the study has expanded even more to include the offspring of F1 (F2). In addition to fetal growth, different anthropometry was performed throughout childhood with follow-ups at 6,12,18 and 24 years. During this period different phenotypes and cognitive functions were measured. The study follow-up is still ongoing and includes different epigenetic markers. This prospective cohort provides a unique opportunity to study different outcomes throughout the life course of an individual.

Tissues

Given the intricate pathophysiology of T2DM and the involvement of different organs and tissues in its development, various tissues were investigated through the different projects in this thesis. Moreover, to further understand the genetics and the development of T2DM, different tissues were also assessed at different developmental stages. In addition, specific cells within specific tissues were also investigated to have a more detailed understanding of T2DM and its development. Importantly, blood was the main tissue from which different phenotypes and genotypes were obtained. However, genetic expression data was obtained from relevant tissues such as liver, kidney, pancreas, and skin in addition to blood. Three main tissue data sources were used:

1. Genotype-Tissue Expression (GTEx) project.
2. Pancreatic islets from the Nordic Network for Clinical Islet Transplantation.
3. Fetal tissues (liver, kidney, pancreata and skin).

Below is a general description of these data sources and specific characteristics of data used for the different projects are available in their respective papers.

Genotype-Tissue Expression (GTEx) project

The Genotype-Tissue Expression (GTEx) project was initiated in the late 2000s to better understand the statistical associations found between different genetic variations with various human traits. Most of these genetic variations were found outside of the protein-coding regions indicating their involvement in expression regulation. Gene expression patterns are unique across different tissues and developmental stages within the same tissues therefore investigating the genetic variants in the context of expression can help understand the genetic basis of human traits. The GTEx project provides a platform where large scale genetic variation and expression data is available across different tissues making it a powerful tool to use in understanding the genetics of T2DM [86, p.].

Pancreatic islets

The Nordic Islet Transplantation Programme involves different clinical and transplant surgery units in Sweden, Norway, Finland, and Denmark. The main focus of this collaborative effort is islet transplantation, nevertheless, it also has a cooperative effort with the excellence of diabetes research network in Sweden. Islets from the transplantation program that are not used in transplantation are employed in research efforts in Scandinavia after consent. Both diabetic and non-diabetic islets are collected in this dataset where their genetic and expression data are available. This dataset provides a unique opportunity to study diabetes since it mainly focuses on pancreatic islets and not the full pancreas as it is in the GTEx project.

Fetal tissues

Conversely, studying fetal tissue provides a unique opportunity from the temporal perspective. Even though the pancreas is not fully developed before 14 gestational weeks, the genetic expression patterns during development can be different from the fully developed pancreas. Therefore, this dataset provides a rare opportunity to understand beta-cell and pancreatic development and how this can translate in T2DM risk. In addition to fetal pancreas, this dataset also includes genetic and expression data from fetal liver, kidney and skin which can be used in identifying genes that are uniquely expressed in a specific fetal tissue and compare it to its adult counterparts from the GTEx project.

Bioinformatics

Irrespective of the origin of the biological material obtained from the different cohorts and the different tissue types described above, several bioinformatic tools have been developed and/or applied for analysing this material. Different aspects of this material were investigated using different tools and different approaches such that both genetic and environmental factors were covered as much as possible given the data availability and financial constraints. The biological material has been probed on several levels using different technologies and approaches covering the main information flow pathway in the living cell. DNA constitutes the starting point of this pathway, and both the DNA sequence and the methylation marks associated with it were examined. The combination of these two factors is essential for the DNA transcription into RNA which in turn is translated into protein. The pathway ends with a final protein product which gives rise to phenotypes. In the projects covered in this thesis specific proteins were analysed, nevertheless, the genetic and the transcriptomic material were examined genome wise. Below is a general

description of the technologies and approaches used in this thesis and the specifics of each is found in their respective manuscripts.

Genotyping

To identify the genetic makeup of a given individual their genome can be analysed either by sequencing the whole genome or probing specific parts of the DNA sequence. Information obtained from both approaches can provide enough ground for studying the association between genetic variants and phenotypes of interest. Genotyping has been a corner stone in GWASs where thousands of genetic variants are probed in a large number of individuals. Nevertheless, conducting appropriate quality control is the first step in performing a successful association between genotypes and phenotypes. Given the large numbers of individuals and the genetic variants involved in a GWAS, several steps must be taken in to overcome factors that can negatively influence the analysis. Generally, these steps can be divided under two categories:

Sample quality

Sample quality control describes the effort of ensuring that the biological material obtained form an individual has the desired characteristics and that the probing of this material was conducted with high standards. Having a high-quality DNA is pivotal in a successful subsequent analysis. The quality of the DNA sample can be inferred form the success rate in identifying the genetic variants. This is called call rate, and generally samples with call rate lower than 99% are deemed of low quality and are disregarded from the analysis. Additionally, poor DNA quality can be inferred from the heterozygosity rate of the given sample. Divergent heterozygosity rate compared to the population mean suggests poor genotyping, however it is important to keep in mind That it can also be due to sample contamination or inbreeding. Sample mix-up can also pose a potential problem in GWAS however comparing the recorded sex against the predicted one based on the genetic data can help in identifying such problems. Moreover, the genetic data can predict chromosomal anomalies and can be used to remove individuals from subsequent analysis if desired. Sample mix-up can also be identified by comparing the reported relationship with the genetic data. Pairwise kinship estimates calculated using a subset of the probed variants can be used in identifying inconsistencies of reported relationship between individuals or to identify samples of bad quality. Moreover, this can also identify nonpaternity, adoption and duplicate samples therefore further investigation may be required. Since GWAS relies on large cohorts, it is possible that individuals in these cohorts belong to groups of different ancestries which can give rise to spurious associations. One explanation to this is the correlation between the genetic difference between the groups and the phenotype of interest. Therefore, the genetic ancestry should be taken into account either by excluding individuals or

by compensating for this systematic genetic difference. Principle component analysis is used to compensate for the population stratification where significant eigenvectors are included in subsequent association analysis.

Genetic variant quality

Genetic variant quality describes the quality of the genetic marker probed across all samples. Similar to the sample call rate, marker call rate can be used as an indicator to identify high quality markers. Generally, markers missing from more than 2% of the study population are removed, however this threshold can change across different studies. Moreover, genetic markers which produce mendelian errors given the known pedigree information are also removed. Monomorphic markers and markers with extremely low minor allele frequency potentially do not provide additional information and can burden both the computational time and multiple testing during the subsequent statistical analysis. However, some rare variants can have deleterious effects and can be of interest therefore removing such markers should be evaluated according to the goals of the analysis. Hardy-Weinberg equilibrium is also used in assessing the quality of a given marker. Deviation from this equilibrium can indicate genotyping errors in the technical sense, however, natural deviation can occur under selection. Therefore, removing variants not satisfying this equilibrium may be counterproductive. To overcome this, the deviations of markers are evaluated only in control samples and different thresholds can be applied to different markers based on their minor allele frequency.

Modern assays used in genotyping examine hundreds of thousands of genetic variants scattered across the human genome, nevertheless, this may not be enough in finding disease variants. In a process called imputation, millions of genetic variants, not probed by the genotyping assays, are inferred by leveraging haplotype blocks and linkage disequilibrium information provided by reference genome information. Two main references are used for imputation purposes the 1000 genome project and the haplotype reference consortium [154], [155]. Moreover, public imputation servers such as Michigan imputation server and Sanger Institute imputation server are freely available for imputation. These have access to both reference panels and employ binaries performing the imputation of the genetic variants in a fast and secure manner.

RNA-seq

Studying the transcriptional patterns of genes with respect to the genetic variants in the genome provides us with an opportunity to understand the effects of these variants on gene function. Extracting the transcriptome i.e., the total RNA, provides a tissue specific snapshot of the genetic activity frozen in time. However, depending on the purpose of the study certain types of RNA are removed whereas the rest are retained. After obtaining the target RNA molecules these are sequenced using high

throughout platforms after which subsequent analysis is performed. In this thesis several different analyses were performed using the sequenced RNA molecules.

1. Differential expression.
2. Expression quantitative trait loci (eQTL).
3. Allele specific expression analysis.

However, as it is the case with genotyping, a robust quality control steps are required for a successful analysis of RNA-seq data.

Quality control

The quality of the initial RNA sample is a pivotal starting point for a fruitful RNA-seq analysis. This can be determined using the RNA integrity number metric which measures RNA degradation. This degradation can occur naturally, however, it is inflated in poorly processed and/or preserved samples. Another metric that is used to determine the quality of RNA-seq data is the abundance of sequencing yield which is a function of the input RNA. Moreover, assessing the percentage of reads that are mapped correctly to the reference genome is an additional metric which is commonly used as a quality control metric. Additionally, over-represented sequences and adapter contamination are evidence of poor quality requiring either a re-run or exclusion from subsequent analysis.

Sequencing and alignment

Generally, RNA samples are first fragmented into small molecules which are ligated to adapter sequences and amplified. This leads to an increased number of RNA molecules that have specific lengths. A paired-end sequencing approach is used where a given molecule is read both from ends decreasing the likelihood of a given sequence to be mapped into multiple sites in the genome. Due to the special features that RNA molecules can possess, read alignment to the reference genome requires additional care. Introns and alternative splicing mean that aligned reads are not always continuous and software used for RNA read alignment, such as STAR, have to take such features into account [156]. Moreover, additional information of the reference genome, such as exon and intron positions, gene structure and variants and other genomic annotations, are taken into account for more accurate alignment.

Read Counting and normalization

Read alignment results in identifying which reads, per sample, belong to which genes in the genome. This information can be subsequently used in quantifying the expression levels of these genes using specific software such as htseq_counts [157] and featureCounts [158]. However, it is important to keep in mind that the sequencing yield, or more generally known as sequencing depth, is not uniform across samples. For a more accurate comparison between samples these differences need to be accounted for. This is achieved by normalizing the depth with the library

size resulting in counts per million (CPM). Additional normalization approaches can be also used depending on the purpose of subsequent analyses. For example, fragment per kilobase per million (FPKM) or transcripts per million (TPM) can be used to normalize for both gene size and library size. These can be useful when the expression between different genes needs to be examined.

Differential expression analysis

RNA-seq data is integral in identifying genes that are differentially expressed between different groups such as T2DM vs controls. Identifying differentially expressed genes can provide a better understanding of the molecular basis of the phenotype of interest, nevertheless, performing such comparison requires additional care. As mentioned above the variance between samples have to be accounted for and sample specific normalization factors can be used for this purpose. Since RNA-seq data provides non-negative integer count data it follows a discrete distribution. Generally, two main distributions have been proposed and are commonly used. Poisson distribution, which is relatively simple, with only one parameter to be estimated. The constraint of the variance to be equal to the mean, however, can sometime be a limiting factor which depends on the study design. The negative binomial distribution, on the other hand, provides a more flexible model with two parameters to be estimated, the mean and the dispersion. This permits greater comprehensive modelling of the mean-variance relationship which is well suited for the analysis of biological replicates that necessitate the incorporation of overdispersion. When the dispersion of the genes is diminished to zero the negative binomial distribution is regressed to a Poisson distribution which is the more convenient alternative in analysing technical replicates [159].

eQTL analysis

The large number of genetic variants being probed during GWASs and simultaneous availability of transcriptome wide data provides a unique opportunity to study how genetic variants across the whole genome can effect expression of different genes. eQTL is a general term that describes the significant statistical association between a genetic variant and the expression of a given gene. The approach in eQTL analysis is similar to the one followed GWASs. GWASs examine whether certain genetic variants in the genome are statistically associated with a given phenotype. In an eQTL analysis the same genetic variants can be examined to see if they associate with the expression of different genes. Since the discovery of the effects of genetic variants residing outside of coding sequences on genetic expression, eQTL analysis proved to be relatively straightforward technique to study such effects. Theoretically, this analysis can be performed on any genetic variant and gene pair in the genome thus providing the opportunity to detect variants with effects on

genetic expression without prior knowledge into a specific gene or locus in the genome. Generally, variants in the eQTL analysis are described either by cis or trans acting. This is commonly defined based on the linear distance between the variant and the transcription start site of a given gene. Commonly, variants within 1 MB from the transcription start site are called *cis* and those outside of this region are called *trans*. However, this definition can change depending on the study design and analysis goals. It is important to keep in mind that the number of *cis* eQTLs identified are relatively larger compared to *trans* eQTLs. This enrichment in *cis* eQTL can be attributed to the relatively intimidating computational task and statistical burden accompanying the *trans* eQTL analysis. Nonetheless relying on linear models and on the efficiency of algorithms running large matrix operations can expedite such analysis and an example of this application is Matrix eQTL which was used in the projects covered in this thesis [160].

Allele specific expression (ASE) analysis

ASE is a term that described the unequal expression state of a given gene in multiple-ploidy organisms. This difference in relative expression between the paternal and maternal alleles in diploid organisms leads to allelic imbalance. Such differences can be due to genetic variants in the genome and/or epigenetic effects which influence the chromatin structure, gene transcription and post-transcription processes. ASE analysis can be performed by exploiting RNA-seq data, where read counts of specific alleles in transcribed regions can be obtained from heterozygous loci. Moreover, this information can also be obtained from several individuals since ASE is generally present across individuals. However, ASE analysis, specifically across individuals, is challenging and several steps are taken to assure a good data quality and appropriate statistical inference.

Sequencing

RNA-seq data can be used to differentiate the expression levels of alternative alleles, however, higher read counts spanning sites of interest is often needed. This increases the demand of high sequencing depth, nevertheless, similar approaches used in RNA-seq quality control is used with specific changes and additional steps.

Alignment and mapping

Read mapping is generally computationally fast and simpler to achieve when compared to read alignment. This is achieved by waiving the validation step of predicted fragments during mapping [161]. In order to limit the number of misidentified sites showing allelic imbalance, mapping and alignment errors have to be minimized and accounted for. Several approaches can be used for this purpose such as the use of specific genomes or transcriptomes, using masked references,

aligning sequences while allowing for genetic variations and using different remapping strategies [162].

Duplicates

It is possible that a given DNA fragment is sequenced several times resulting in duplicate reads during library preparations. These reads are identified and marked such that they are not accounted for during subsequent analysis. This also applies during ASE analysis however some care needs to be taken during duplicates removal such that the retained reads are selected at random and not based on mapping score.

Variant discovery and read counting

Prior to genetic variant discovery, bases form the sequencing data are recalibrated using different statistical methods that take into account all samples in the dataset. This is applied to counteract biases arising from library preparation and/or faulty instruments. After a successful recalibration, genetic variants are identified by local *de novo* assembly of haplotypes, and these are subsequently filtered such that only high-quality variants are retained in the dataset. Once the final set of genetic variants are identified, the RNA-seq data can be used to calculate the number of reads spanning a given variant. However, mapping quality, depth and overlapping reads can be used to obtain a more accurate estimation of reads covering a genetic variant.

ASE statistical analysis

Statistical tests to infer ASE and interpret the results are not trivial. Binomial test can be used to determine if the expressed ratio between two alleles is statistically significant. However, overdispersion in allelic imbalance is higher when compared to the expected variance in binomial random variable convoluting such approach [163]. Bayesian methods can also be employed for ASE analysis which have a higher capacity in inferring allelic imbalance [164]. Moreover, a generalized linear mixed models can be used for successful inference of allelic imbalance, where individual effects are taken into account and allow correlation of multiple SNPs with a given gene [165].

DNA methylation

Similar to genotyping, it is possible to investigate the methylation make up (methylome) of a given individual either by sequencing or by probing specific sites in the methylome using methylation arrays. Treating the genome with bisulphate will convert non methylated cytosine residues into uracil making it possible to differentiate between the methylated and unmethylated sites. In this thesis we have utilized Illumina produced methylation arrays to interrogate the DNA methylation. The methylation is reported as a ratio between the methylated and unmethylated

alleles. To ensure reliable methylation analysis several subsequent steps are taken to remove problematic samples or sites in different consecutive steps described below.

Sample and probe quality control

In addition to the GenomeStudio software [166], which helps detecting low quality samples, control probes can be used in identifying poorly performing samples where samples outside three standard deviations from the median can be discarded. Similarly, probes which were not detected in a certain proportion of the total samples can be discarded. Moreover, probes which provide a detection below a certain value can also be filtered out.

Background correction

In addition to the GenomeStudio software which corrects for nonspecific signals, different packages have been developed to perform this function. Additionally, it is possible to use negative control probes to estimate intensities measured by each probe.

Normalization

In order to remove noise and artifacts, normalization is performed which addresses the removal of artifacts of both between samples on different arrays and within the same array. GenomeStudio software provides internal normalization control, however, other approaches exist which rely on fluorescence signal intensity.

Probe scaling

Illumina methylation arrays use two different types of probes, type I and type II, where the methylation signals are measured with two different approaches necessitating appropriate scaling. A popular approach is to use a beta-mixture model where type II values are transformed to be comparable with type I values.

Batch effects

In order to overcome effects that are confounding and are not relevant for the biological question but rather are due to technical application, batch effect correction should be applied. A popular method for batch effect correction is ComBat which uses empirical Bayes procedures which can remove other confounders along with batch. Nevertheless, it is important to keep in mind that it follows the normality assumption and methylation data has to be handled accordingly [167].

Statistical analysis

Generally, in mathematical terms, it is very common to find functional relationship between two or more real valued random variables. Such functions, describing these relationships, can be of different types such as linear, polynomial or any other functions fulfilling specific restrictions. In biological and medical sciences, linear functions are often used to characterize the relationship between a dependant variable and one or more independent variables in what is called a linear regression. A regression coefficient beta is used to describe the changes of the dependant variable with respect to changes in the independent variable(s). moreover, when the dependant variable follows some arbitrary distribution as opposed to a normal distribution, an arbitrary link function can be used for transformation such that the relationship between the variables remain linear. Models that are used to study such variables are called general linear models. Logistic regression, which is an example of general linear model, is often used to study the effects of different variables on a binary outcome such as having T2DM or not. Importantly however, when there is dependency in the collected data, such as familial data or repeated data, the generalized linear models may provide noisy and inaccurate results. This dependency between collected data can be overcome by using linear mixed models which uses both fixed and random effects. This linear model allows for parameters to vary and treats them as random variables themselves. In other words, the slopes and intercept of the regression have variations around their mean. Furthermore, if the relationship between the variables is cryptic or unknown generalized estimating equations can be used to obtain parameter estimates. An additional advantage of this approach is that it can be computationally less demanding when compared to linear mixed models, making it a viable approach to study parental effects in genetic associations using large familial datasets. Additionally, quantitative transmission-disequilibrium test can be used to study the parental effects of alleles in familial setting. This test is constructed by taking into account all offspring information, between and within-family components and can support different sized nuclear families [168]. Of importance to mention is that, when estimating the effects of a given independent variable on the dependant one others are included in such analysis to account for their confounding effect. This will allow the linear models to give an unbiased estimation of the effect size of the dependant variable of interest, however, this fails to provide a numerical value describing the relationship between the two variables of interest. This can be achieved by conducting partial correlation which controls for the confounding variables in a similar manner as the generalized linear models.

Pipelines

Almost all the analyses conducted in the projects covered in this thesis were run on calculation servers using the HTCondor job management system. This was achieved

using the UNIX shell and utilizing proprietary pipelines created using bash command language. The pipelines were created such that row genotyping and RNA-seq data in addition to quality-controlled methylation data were provided as inputs. Moreover, clinical, and phenotypical data was also used as inputs to the pipelines in addition to a set of arguments. The combination of data provided, and the arguments used determined the type of analysis to be run. In addition to bash, the pipelines utilized AWK, R, python, and different set of command line driven binaries to perform all the required analysis and produce the results. Such pipelines were created to make the analysis reproducible and easily rerunnable when a new set of data was provided. Moreover, logs were kept for transparency and traceability of the conducted analysis.

Ethical considerations

In the projects covered by the thesis we mainly studied human samples which demand different ethical considerations compared to animal and cell-line based research. All cohorts used were approved by their respective ethical committees and written informed consent was obtained from all cohort participants. Importantly all data used were anonymized for the protection of individual privacy and genetic data that can be used for identification of study participants are presented only in low resolution and summary level. Moreover, all the data used is stored on computational servers where access is gained after establishing double encrypted connection minimizing any data leakage.

Aims and results

Paper I

Aim

The aim of this paper was to understand the effects of under-nutrition as a result of anemia during pregnancy on the fetus. Specifically, we aimed to understand the molecular mechanisms of fetal programming and how they can affect adult cardiometabolic diseases.

Results

Gene expression and epigenetic profiles

To identify genes with altered expression due to anemia exposure we conducted differential expression of RNA extracted from umbilical cord blood (UCB) of anemia cases and controls. We identified 137 differentially expressed genes 41 of which had methylation patterns associated with their expression. We found that the DNA methylation of 27 genes associated with maternal Hb levels in early pregnancy indicating potential epigenetic signature of fetal programming. Importantly, we found that the methylation at cg12884187 CpG site correlated positively with maternal Hb levels and negatively with Ligand Dependent Nuclear Receptor Corepressor Like (*LCORL*) expression.

Gene expression correlation with phenotypes

Next, we have investigated whether genes identified above also correlated with different phenotypes. We found that *LCORL* expression correlated with length of neonate at birth, placental size, and weight. Moreover, several CpG sites correlated either nominally or significantly with these phenotypes. Additionally, Kelch Domain Containing 1 (*KLHDC1*) correlated with birthweight, neonatal weight-length ratio, and abdominal circumference. Importantly *KLHDC1* was correlated with insulin levels and was found to be downregulated in T2DM donor islets. We additionally investigated the effects of glucose-treatment of pancreatic islets form hyperglycemic and normoglycemic donors on the expression of these genes and found *LCORL* to be down regulated upon glucose treatment.

Gene expression in fetal and adult tissues

We next assessed if the differentially expressed genes were enriched in fetal or adult tissues. We found 43 genes to be upregulated in fetal pancreas relative to the adult pancreas suggesting a possible developmental role of these genes. We found NUMB Like Endocytic Adaptor Protein (*NUMBL*) and Phosphatidylinositol-4-Phosphate 3-Kinase Catalytic Subunit Type 2 Beta (*PIK3C2B*) to be enriched and beta-cell specific. However, *LCORL* was reported to be as alpha-cell enriched gene [169]. Moreover, we found the expression of Purinergic Receptor P2X 7 (*P2RX7*) and *PIK3C2B* to be enriched in fetal beta-cells when compared to adult beta-cells and alpha-cells. Finally, we used immunohistochemical staining of fetal pancreas to identify the expression of *P2XR7*, *NUMBL* and *PIK3C2B*. *P2XR7* was expressed in endocrine progenitors, *NUMBL* and *PIK3C2B* were expressed in the epithelium *NUMBL* was also expressed in insulin and glucagon producing cells.

Paper II

Aim

The aim of this paper was to understand the effects of over-nutrition as a result of GDM during pregnancy on the fetus. We have investigated the sex-specific effects of GDM and aimed to understand the molecular mechanisms of fetal programming and how they can affect adult cardiometabolic diseases.

Results

Gene expression and epigenetic profiles in male, female, and all offspring

We have performed differential expression analysis on UCB samples from mothers with GDM and controls. First, we have identified 36 differentially expressed genes in all offspring, however, we only identified 4 DEGs in sons whereas 273 DEGs in daughters. We next assessed the epigenetic relationship of these genes with their expression levels. Several methylation sites correlated with the expression of their respective genes in all offspring including at Solute Carrier Family 4 Member 1 (*SLC4A1*). Interestingly, 95 CpG sites correlated with 31 genes in daughters whereas non showed such associations in sons. Moreover, we identified 24 CpG sites in 17 genes to be associated with fasting glucose in daughters.

Gene expression correlation with different phenotypes

We then investigated whether DEGs associated with different glycemic and anthropometric phenotypes. First, we found *SCL4A1* to be associated with chest

circumference and subscapular skinfolds using data from all offspring. Next, we found that several CpG sites in the same loci associated with chest circumference, subscapular skinfolds, birth weight and abdominal circumference. Our sex-specific analysis, however, provided a different view. We identified several DEGs that associated with at least 1 phenotype in daughters and CpG sites from these genes showed nominal associations with head circumference and umbilical circumference. Moreover, various CpG sites associated with insulin and c-peptide.

Gene expression in adult and fetal pancreas

To relate the expression of the DEGs to insulin secretion and islet function we investigated their expression in adult pancreatic islets. We found 19 DEGs to be expressed in adult islets with different genes being either up or downregulated in diabetic islets and correlated either negatively or positively with glucagon on *INS* expression. We observed similar results when analyzing sex specific DEGs in adult pancreatic islets. On the other hand, to examine the possible role of DEGs in development and/or function we investigated their expression in fetal pancreas and compared them to adult pancreas. We found *SLC4A1* expressed mainly in fetal tissue where its expression in fetal alpha- and beta-cells was significantly higher when compared to the adult cells. Moreover, the expression of different developmental genes correlated significantly with *SCL4A1* in fetal pancreas.

DEGs in vivo

We next examined the DEGs from son, daughters, and all offspring in the IMPC database to investigate their possible effects on metabolism. We have identified 6 DEGs from all offspring that when knocked-out in mice had effects on different glycemic measures including blood insulin and fasting blood glucose concentration. We found similar results for DEGs in daughters with some knock-out genes showing effects only in female mice. Furthermore, knock-out models of 4 DEGs from all offspring showed altered body composition including fat mass and total lean mass, whereas 9 DEGs found in daughters had similar effects.

Paper III

Aim

Given the relationship of dyslipidemia with T2DM, CVD and obesity, we aimed in this study to further explore the genetics affecting dyslipidemia. Specifically, we focused on understanding whether genetic variants will exert different effects based on their parental origin.

Results

Parental effects of lipid traits

To study the parental effects of lipid traits, we analyzed specific loci that have been previously associated with lipid traits using family cohorts (Botnia study and HTB study). Expectedly, we have shown that a general correlation in lipid traits between father-offspring and mother-offspring existed. However, we found that mother-offspring correlation was stronger than the father-offspring correlation. Additionally, we have shown that the rs2131925 SNP at the *DOCK7/ANGPLT3* locus exhibits a trend of parent-of-origin effects on Apolipoprotein A1 (ApoA1) and Apolipoprotein A2 (ApoA2). Moreover, the rs12272004 SNP near *APOA* gene showed parental effects on ApoB and ApoB/ApoA1 ratio.

Parental effects on Obesity

We wanted to assess whether the variants showing parental effects of lipid traits also showed the same effects on obesity traits given the strong relationship between the two. Expectedly, the rs2131925 SNP at the *DOCK7/ANGPLT3* locus showed parent-of-origin effects on waist-hip ratio even when adjusting for BMI. Similarly, the rs10503669 SNP at the *LPL* locus showed parental effect on waist-hip and waist-height ratio.

Sex-specific effects

We identified stronger correlation of lipid traits between mothers-daughters when compared to fathers-daughters however we did not observe this discrepancy between mothers-sons and fathers-sons prompting us to study the parental effects on sons and daughters separately. We found that the paternal G allele of the rs2131925 SNP had ApoA1 lowering effect in daughters only and associated with ApoB/ApoA1 ratio in daughters but not sons. On the other hand, the parental effects of rs10503669 and rs12027135 on ApoA2 were observed only in sons. Moreover, given the strong relationship between obesity and lipid levels, we assessed the sex-specific effects on obesity. rs10503669 showed parental effect on BMI, waist-hip ratio, and waist-height ratio in daughters. Additionally, rs2131925 showed parent-of-origin effects on waist-hip ration even after adjusting for BMI in daughters.

Paper IV

Aim

In this study we aimed to study if genetic variants produced different effects on metabolic and anthropometric traits based on their parental origin. Moreover, we

investigated if these effects manifested in early life and whether they were consistent throughout adulthood.

Results

Metabolic and anthropometric heritability

We first investigated the effects of parental phenotypes of different metabolic and anthropometric traits on the offspring phenotypes at different time points in the PMNS birth cohort using data from 700 parent-offspring trios. Expectedly, most measured traits showed significant association between parental and offspring trait values. Interestingly, coefficient for weight, height and BMI showed an upward trend when moving forward through the different timepoints. Furthermore, the observed association trend was consistent when separately assessed for sons and daughters with the exception of HDL levels at 12 years in daughters.

Parental effects on metabolic traits

We next investigated the parental effects on fasting glucose and insulin levels. Our results indicated that there was a positive association between the glucose levels of the offspring and those of both the mother and father at ages 6 and 12. The maternal effect was stronger than the paternal effect, and there was a significant parent-of-origin effect at age 12 in sons. In contrast, we observed a significant negative maternal association with fasting insulin and HOMA-S at age 6, which became a significant positive association at age 12. For HOMA-B, there was a stronger paternal effect at age 6 that became a stronger maternal effect at age 12. There is also a significant negative maternal association with insulin and its indices in sons at age 6, which becomes a positive association at age 12, although the parent-of-origin effects are only significant at age 6. On the other hand, we observed stronger positive maternal associations in all offspring and sons in total and HDL cholesterol. Similarly, positive maternal association was present in sons for triglyceride levels and daughters showed strong paternal association only at 12 years.

Genetics of parental effects

We then studied the effects of genetic variants in *KCNQ1* on insulin secretion, fasting glucose, BMI, cholesterol, and HDL at 6- and 12-years. The rs2237892 SNP showed significant maternal, paternal, and POE specific associations with insulin secretion and HOMA-B at age12 and with opposite directions of effect at 6-years. It also showed a significant POE on fasting glucose, BMI, cholesterol, and HDL at different ages. Another SNP at the *KCNQ1* locus, rs231362, showed significant maternal effects on HOMA-B and fasting insulin at 12-years. The two SNPs differed in their POE on birth weight, with only rs2237892 showing an effect.

Paper V

Aim

The aim of this paper was to understand how imprinting plays a role in insulin secretion and beta-cell development. Specifically, we investigated imprinted genes in different tissues with different resolution ranging from the whole pancreas to specific cells in the pancreatic islets.

Results

Imprinted gene expression in fetal and adult tissues

We have identified 173 imprinted gene to be expressed in fetal pancreatic tissue using RNA-seq data. We also compared the expression of these imprinted gene to adult pancreas and identified 75 imprinted genes that were either specifically expressed and/or enriched in the fetal pancreas when compared to the adult pancreas. Interestingly we found 15 of these gene expressed in fetal kidney and skin and 1 was expressed in fetal liver, kidney, and skin in addition to fetal pancreas. Finally, we have checked the expression of the 75 imprinted gene in human pancreatic islets and found only 36 to be expressed.

Imprinted gene expression in pancreatic islet cells.

We have utilized publicly available data from Blodgett et al. (2015) to check the expression of the imprinted genes which we found to be fetal specific/enriched in different adult and fetal cell types. Based on this data we found 14 imprinted genes to be specific to fetal, fetal alpha and fetal beta clusters. Moreover, we assessed the expression of these gene in the different pancreatic cell types using another publicly available resource provided by Segerstolpe et al. (2016), where single cell RNA sequencing was used. We found most fetal-enriched genes had modest expression across the different cell types and handful of genes to be enriched in the alpha-, beta-, and delta- cell types.

Imprinted gene in beta-cell development and function

We next checked the allelic imbalance in the expression of the 14 genes which showed fetal pancreas-specific expression and identified two genes, Retrotransposon Gag Like 1 (*RTL1*) and WT1 Antisense RNA *(WT1-AS*). Since these showed fetal pancreas specific expression and had no discernible expression in the adult pancreatic islets or in the adult pancreatic cell types, we hypothesized that these genes are more likely to play a developmental role of the pancreas and

possibly beta-cells. Two variants in *RTL1* associated significantly with HOMA-B in the Botnia study. Additionally, we found the expression of *RTL1* to be positively correlated with *DLK1* and both genes showed higher expression during later staged of the differentiation of induced pluripotent stem cells to beta-cells. On the other hand, we found 7 genes to be enriched in both adult and fetal beta-cells. We hypothesized their possible involvement both in the development and function of beta-cells and assessed if they also portrayed allelic imbalance in their expression. Of these, only Teashirt Zinc Finger Homeobox 3 (*TSHZ3*) showed allele-specific expression in both fetal pancreas and adult pancreatic islets. Whereas we found 4 genes to be expressed in an allele specific manner in the fetal pancreata and 2 in adult islets. we identified two variants in the Phosphatase And Actin Regulator 2 (*PHACTR2*) gene that associated significantly in parent-of-origin manner with HOMA-B and variant in *TSHZ3* associated with corrected insulin response.

Discussion

The work presented in this thesis aims to enhance the basic understanding of T2DM by focusing on genetic and epigenetic phenomena and consequent changes in gene expression related to fetal programming and parent-of-origin effects and how they impact insulin secretion. It is well established that T2DM is a complex disease and is influenced by both genetic and environmental factors. Moreover, T2DM is a polygenic disease, meaning that it is affected by multiple genes with relatively small effect sizes. This, in turn, provides another dimension to the complexity to be considered while investigating the genetics of T2DM. Nevertheless, GWASs were successful in identifying hundreds of genetic variants showing association with T2DM [83]. In our work, we exploit these previous findings and employ GWAS using different cohorts to further expand our understanding of the genetics of T2DM. In addition to genetic factors, T2DM is also influenced by environmental factors such as diet, physical activity, and obesity. A sedentary lifestyle and a diet high in calories, saturated fats, and sugars have been shown to increase the risk of T2DM. Obesity is also a major risk factor for T2DM, as excess body fat can lead to insulin resistance and impaired glucose tolerance [170]. Moreover, the relationship between genetics and environmental factors in the development of T2DM is complex and multifactorial. While genetic factors may increase an individual's susceptibility to T2DM, environmental factors play a critical role in the onset and progression of the disease. To better understand the environmental effect and the complex interaction between genetic and environmental factors, we have evaluated the transcriptome from different cohorts, tissue sources and developmental stages. During this process we also examined DNA methylation and investigated which alterations can possibly affect gene expression. Utilizing the different cohorts, methods and statistical analysis approaches mentioned above we aimed to build a better picture of the processes involved in the development and function of insulin secreting beta-cells and possibly shed a brighter light on the risk and development of T2DM.

Fetal programming

Since the 1980s it was proposed that environmental factors during fetal development can have lasting effects on the health and development of an individual throughout their life. The mechanisms underlying fetal programming are complex and not yet fully understood. However, several mechanisms have been proposed and one of the most widely studied mechanisms is epigenetic modification, which can be induced by a variety of environmental factors, including maternal nutrition, stress, and exposure to toxins. In this thesis we have studied two such exposures i.e., GDM and anemia, and their effect on the offspring.

GDM represents and overnutrition state due to the increased baseline of glucose levels in the blood of mothers which passes through the placenta. In our study we have investigated the fetal programming signatures related to metabolism, anthropometry, and T2DM by analyzing gene expression, DNA methylation, and genotyping data from UCB obtained from offspring of Tanzanian mothers who had GDM compared to controls. Unsurprisingly, we found 36 genes with altered expression in the UCB from offspring of GDM mothers. Interestingly however, we observed sex-specific effects by identifying only 4 genes with altered expression in sons vs 273 in daughters of GDM mothers. Sex-specific gene expression has been studied before, and tissue-specific sex differences have also been reported [171]. Expectedly, several CpG sites correlated with the expression of their respective genes, however, several CpG sites in various genes including *SLC4A1* also correlated with fasting and 2-hour glucose levels. We hypothesized that the alteration in the expression level due to GDM is possibly mediated by DNA methylation. Additionally, we observed significant differences in neonatal anthropometry in daughters of GDM mothers only. Similarly, gene expression and DNA methylation at several CpG sites associated with these neonatal anthropometries in daughters but not sons. These observations are not surprising since sex-differences in growth and metabolism and potential difference in response to detrimental exposures during pregnancy have been previously reported [172]. Furthermore, we found that altered gene expression and associated DNA methylation were related to insulin, and c-peptide levels in daughters. Some of these genes were also associated with insulin and glucagon levels in human pancreatic islets from T2D donors and altered glucose and insulin response levels in knockout mice models. Based on these observations, we proposed that these genes represent fetal programming signatures that alter body composition and increase susceptibility to cardiometabolic disorders in later life. Interestingly, we also found that *SLC4A1* was expressed specifically in fetal pancreas and significantly associated with the expression of fundamental pancreatic development genes leading us to hypothesize that it may play an important role pancreas development.

Additionally, we have studied the effects of early pregnancy anemia and its possible effects on fetal programing. Anemia during pregnancy can be caused by various

factors including nutritional deficiencies, genetic disorders, chronic diseases, and infections. In our selection criteria for our cohort, we have considered mothers who were malaria- and HIV-negative and approached the anemia exposure as an under-nutritional state. We have used a similar approach to the one we used to investigate the effect of GDM on fetal programming where we used global gene expression, DNA methylation, and genotyping on UCB obtained from offspring of mothers who had anemia during early pregnancy. First, we demonstrated that maternal anemia during early pregnancy is associated with low birth weight, small for gestational age, and preterm delivery. Expectedly, our differential expression analysis identified 137 gene with altered expression. Moreover, the expression of some of these genes was associated with maternal Hb levels at delivery. We hypothesize that these expression changes initiated in early pregnancy are retained throughout pregnancy and are also modulated by maternal Hb levels at a later stage. We found that epigenetic mechanisms could be involved in the association between maternal anemia in early pregnancy and low birth weight. The *LCORL* gene is a component of the polycomb repression complex-2, which mediates methylation of histone H3lys27, and is important for determining methylation status and regulation of cellular identity during fetal development [173], [174]. We found that DNA methylation at CpG sites in the *LCORL* gene associated with *LCORL* expression, maternal Hb levels in early pregnancy, the length of the newborn, placental size, and placental weight. The epigenetic signature is a potential consequence of maternal anemia in early pregnancy and could, therefore, represent an example of a signature of fetal programming. This provides further evidence for the mechanism by which altered fetal programming could affect the outcomes for the newborn with potential for long-term consequences. Additionally, we observed lower insulin levels in UCB from neonates of early pregnancy anemic mothers compared to those without. Undernutrition associated with early pregnancy anemia may cause impaired fetal pancreatic development and maturation, resulting in reduced in vivo insulin secretion and subsequently increased lifetime risk of T2DM. Insulin is an important growth hormone, which has a significant effect on fetal growth, mediated through altered gene expression. *KLHDC1* gene expression was increased in UCB from early anemic mothers, and was negatively correlated with birth weight, weight to length ratio of the neonate, and with UCB insulin levels. Finally, of the genes with altered expression, several associated with indices of insulin secretion and T2DM risk in previous GWAS studies [83] and some were involved in the different signaling pathways that are important for beta-cell development, and defects in these pathways could potentially affect the pathogenesis of diabetes [175], [176].

We have utilized UCB in our efforts since it is relatively easy to obtain. Moreover, cord blood contains a wealth of information about fetal development and can provide valuable insights into the mechanisms underlying fetal programming. Additionally, cord blood contains a variety of biomarkers that can be used to assess fetal exposure to various environmental factors. For example, levels of certain hormones, growth factors, cytokines, and other signaling molecules in cord blood

can provide information about fetal nutrition, stress, and inflammation, among other things. Moreover, cord blood can also be used to identify potential biomarkers of later health outcomes. For example, several studies have identified biomarkers in cord blood that are associated with an increased risk of obesity, cardiovascular disease, and other chronic conditions later in life [177], [178].

Our efforts to understand the effects of fetal programming after exposure to both over and under-nutrition involved the study of a population from Tanzania. Both studies are pilot studies and contain some limitations. The cohort sizes we employed may be considered a limiting factor to identify all possible alterations in methylation and expression levels. Moreover, different replication cohorts will provide a more vigorous examination of our findings. The prospective replication cohort should possibly be with larger number of participants with different ancestries. Additionally, to further understand the effect of fetal programming after the exposure to GDM or anemia during early pregnancy, long term follow up will be pivotal and provide a better understanding of lifelong effects as opposed to our use of surrogate measures. Moreover, we have conducted sex-specific analysis to study the effects of GDM only, future efforts should also investigate the sex-specific effects of early pregnancy anemia on fetal programming.

Parent-of-origin effects

Parent-of-origin effects can occur due to different mechanisms, including genomic imprinting, mitochondrial DNA inheritance, and sex chromosome-linked effects. This genetic phenomenon leads to the manifestation of different phenotypes depending on the maternal or paternal inheritance of the genetic material. Moreover, these effects may play a critical role in the development of various health conditions, including obesity, blood lipid levels, and T2DM. Therefore, studying these effects can give important insights into the molecular mechanisms underlying these complex traits and diseases.

Obesity is a complex trait that is influenced by both genetic and environmental factors. Moreover, parental effects may play a role in the development of obesity. Previous efforts in studying these effects have indicated that two genes, *MEG3* and *H19*, can play a role in obesity and related disorders. Both genes are imprinted genes and showed altered expression in visceral and subcutaneous adipose tissues in obese women vs controls. Moreover, their expression correlated with different obesity related indices, insulin resistance and key genes involved in adipogenesis and lipogenesis [179].

On the other hand, blood lipid levels, including levels of cholesterol and triglycerides, are important risk factors for CVD. Similar to obesity, these are complex traits that are governed by intricate meshing between environmental and

genetic factors. Generally, blood lipid levels are reported to have a high variance of heritability that could reach to over 90% in some traits [180]. This high variance in heritability could be explained when sex of the parents is taken into account. Moreover, the sex of the offspring has been also indicated to be a relevant factor when studying the genetics of blood lipid levels [181], [182].

The substantial relationship between obesity and blood lipid levels motivated us to investigate the genetic basis of sex-specific parental effects on plasma lipid levels and obesity. We proceeded by examining whether common variants associated with lipid levels from previous GWAS studies associate with lipid traits and obesity in a parent-of-origin-specific manner. Indeed, we found several variants at the *LPL* and *DOCK7/ANGPTL3* loci showing parental effects. Both loci play an important role in regulating HDL in plasma in addition to the regulatory influence of *ANGPTL3* on *LPL*. This explains our observation of the association between the variants in these genes with ApoA2 which is an important element of HDL [183], [184]. Additionally, we illustrated a consistent opposite effect of some of these variants across our cohorts on obesity measures, where the maternal variant had a negative effect in contrast to the paternal variant. A possible explanation of our observation is that the expression level of certain genes can have different fitness effects on different individuals based on their parent-of-origin. Specifically, genes that are inherited from the father and are expressed, favor prenatal growth and energy transfer from the mother where the expression of maternally inherited genes favor the constraint of energy transfer. In our study we have focused on a handful of genetic variants based on prior knowledge. We chose this approach in a bid to increase the power of our study since parental effects require relatively large cohorts. Moreover, some of our observations using our data can possibly be due to retained structure in our cohorts which should be addressed by using larger study cohorts that can corroborate our findings. Additionally, our study establishes association between different genetic variants and blood lipid levels and obesity traits however subsequent studies should investigate the causal effects of our findings and confirm whether the effects of the genetic variants on these phenotypes are independent and not mediated through one another.

In our efforts to study parental effects on blood lipid levels and cardio-metabolic traits we also investigated whether these effects manifested in early life. Additionally, we have evaluated parent-of-origin effects on anthropometric traits utilizing serial measurement of these phenotypes from birth. Initially we hypothesized that these effects should be observed in early life since the parental effects influencing the development of these traits already take effect during conception and pregnancy [185]. First, we corroborated previous findings of parental specific influences on metabolic traits including beta-cell response to oral glucose, insulin action in target tissues, and lipid levels [181], [186]. Moreover, we have also evaluated parent-of-origin effects on anthropometric traits and our observations implied the concealment of paternal effects on fetal growth, especially

weight and soft tissues but not skeleton. This phenomenon has been previously reported [187] and the parental-offspring association of these traits in our cohort were similar after birth. Interestingly, our results show a differential maternal and paternal effects on lipid levels in sons at 6 years and daughters at 12 years respectively. Furthermore, the parental effects we observed in our cohort at 12 years seem consistent with the effects reported in other studies when measuring such effects in adults. Nonetheless, our results suggest a different direction of effect at 6 years. We hypothesize that such a change of effects is mediated by epigenetic mechanisms. Parent-of-origin effects are primarily a consequence of genomic imprinting, which alters gene dosage, modifies the resemblance of the offspring to its two parents, and possibly increases the probability of expressing the fitter of the two alleles at a given locus. Epigenetic mechanisms, such as DNA methylation, which can alter the way the DNA is read and expressed, play a crucial role in mediating the selective expression of alleles and the consequent impact on the phenotype. Early life exposures can bring about such epigenetic reprogramming and alter the development and function of organs, which, in later life, can increase susceptibility to cardiometabolic disorders [188]. We observed a similar shift in parental effects on glycemic traits where, paternal effects on insulin secretion at age 6 transitioned into maternal one at age 12 and negative maternal effects on fasting insulin and insulin sensitivity shifted into positive ones at age 12. We reason that these shifts are necessary and coincide with pubertal age for maintaining energy homeostasis which is important for natural development however this hypothesis needs to be further tested. Moreover, similar to our study above, our consideration of the parental phenotypes on offspring phenotypes are associative in nature and do not assess the causal factor of the parental phenotypes. We suggest consequent investigations of the observations we report in our work, where a causal analysis can be performed in similar and dissimilar cohorts. Moreover, considering investigating epigenetic changes in relevant tissues in a subsequent analysis may prove helpful in uncovering possible epigenetic mechanisms, involved in the parental effects.

Given the importance of imprinted genes in paternal effects and their involvement in different metabolic pathways we hypothesized that imprinted genes may play a role in pancreatic development and T2DM. We have utilized RNA-seq data from fetal tissues (pancreas, liver, kidney, and skin), adult pancreas, adult pancreatic islets, and different pancreatic cell types to assess the expression of imprinted genes. We successfully identified imprinted genes specifically expressed in fetal pancreas and hypothesized their role in pancreatic and possibly beta-cell development. Moreover, we identified allele specific expression in both *RTL1* and *WT1-AS*. Two variants in RTL1 associated with HOMA-B which is a measurement used in quantifying beta-cell function and insulin secretion. Based on our observations, we propose that *RTL1* may play a role in beta-cell mass due to its specific expression in fetal tissues. Interestingly, *RTL1* is located in an imprinted region with maternally expressed microRNAs regulating its expression and was observed to be implicated

to play a role in Temple syndrome [189, p.], [190]. Evidently, 10% of patient with Temple syndrome were found to develop T2DM [191]. Furthermore, the expression of *RTL1* correlated with the expression of *DLK1* during the differentiation of induced pluripotent stem cells into beta-cells. This observation may be relevant since genes in the *DLK1-MEG3* region were found to be repressed in islets from T2DM [150]. Separately, we found seven genes expressed in both fetal and adult pancreas with an enrichment in beta-cells. Only *TSHZ3* showed allele-specific expression in both fetal pancreas and adult pancreatic islets, suggesting temporal imprinting of the remaining six genes. Surprisingly, we did not observe associations of genes showing allele specific expression in both fetal and adult tissue with HOMA-B or corrected insulin response. We hypothesized that this may be due to parental effects and indeed when considering parental origin, *PHACTR2* variants were associated with HOMA-B, while *TSHZ3* variants were associated with CIR. Our observations regarding the potential involvement of several imprinted genes in beta-cell development and function is based on expression data and further in vitro and in vivo validation is necessary to elucidate how *RTL1*, *TSHZ3*, and *PHACTR2* are involved in beta-cell development and function. Additionally, despite the uniqueness of our dataset, we are limited by the small number of fetal pancreas samples, and the results of the enrichment of gene expression between fetal and adult pancreas should be interpreted with care due to the difficulty in accounting for batch effects.

Conclusions and future perspectives

Diabetes mellitus is a lifelong, incapacitating disease affecting multiple organs. The number of people suffering from the disease is continuously increasing and the reported prevalence varies between different countries. In spite of identifying hundreds of genetic loci contributing to the susceptibility of T2DM the explained heritability remains low. A portion of the missing heritability can be explained when examining distinct parent-of-origin transmission of different variants so that only one parent transmits the risk allele to the offspring, whereas transmission from the other parent may be neutral or even protective. Under such circumstances, studies performed in outbred case-control populations would miss the association and genetic studies considering from which parent an offspring has inherited a susceptibility allele were successful in identifying some of the missing heritability [192], [193]. Parent-of-origin effects represent an important and complex aspect of genetics that can have significant implications for human health and behavior. Moreover, by unraveling the complexities of parent-of-origin effects, the implications for the future of medicine and genetics can be significant. Parent-of-origin effects could complicate genetic testing and make it more difficult to accurately diagnose certain conditions. However, a more sophisticated genetic tests that considers which parent a particular variant came from can be developed leading to more accurate diagnoses and personalized treatment plans for patients. Furthermore, since many diseases are caused by multiple genetic and environmental factors, understanding parent-of-origin effects could shed light on the role that specific genetic variants play in disease development. This could help identifying new targets for treatments and prevention strategies. Additionally, our understanding of these effects provides us with improved reproductive options which could have implications on genetic counseling and in vitro fertilization. By understanding these effects, doctors and patients may be able to make more informed decisions about fertility treatments and family planning.

Similarly, fetal programing is another complex phenomenon which could have life-long implications and further understanding of it could be beneficial. Awareness how early-life experiences can affect an individual's health can help policymakers to design interventions and preventative measures that can reduce the risk of chronic disease and improve overall health outcomes. Moreover, by understanding an individual's early-life experiences, doctors may be able to develop personalized

treatment plans that take into account the unique genetic and environmental factors that contribute to that person's health. For example, if a person has a higher risk of developing a certain disease based on their fetal programming experiences, doctors could recommend targeted preventative measures or more frequent screening for that individual. Finally, fetal programming has important implications for our understanding of the intergenerational transmission of health and disease. Environmental factors experienced by a fetus can have effects that are passed down through multiple generations [194]. This suggests that addressing health disparities and improving health outcomes for vulnerable populations may require a multigenerational approach that takes into account the experiences of past generations.

This thesis attempts to address the effects of two complex genetic phenomena and their effects on complex traits. We have corroborated that early life exposures, represented as under and over-nutrition during early pregnancy may have detrimental effects in offspring and potentially increase their risk for T2DM in later life in an understudied population such as Tanzania. Additionally, due to the manifestation of parent-of-origin effects on lipid, glycemic and cardiometabolic traits both in early and later life, we show that is it important to include such specific effects when conducting genetic studies. Finally, we have identified imprinted genes which can be of specific interest in better understanding the development and risk of T2DM. Further studies expanding on our findings with larger and diverse cohorts and scrutinizing causality of genetic variants and environmental effects would be the next step in exploring fetal programming and parent-of-origin effects of type 2 diabetes and insulin secretion.

Acknowledgements

Rashmi – I am very grateful for all the efforts and time you have dedicated to me as your PhD student. Moreover, your grounded patience and view of things have been very helpful to counteract my somewhat pessimistic view of things. You have been very supportive and understanding through the years with all the injuries I have managed to accumulate. I am very grateful for everything, I truly am! Thank you.

I would like to thank my co-supervisors **Leif** and **Ola**. Thank for all the support and the discussions. It has been a pleasure to know you and interact with you.

Nikolay - thank you very much for your help, guidance, and mentorship during the PhD years. I have more than enjoyed our talks meet ups and discussions.

Olof and **Dina -** thank you for all the help, support, and discussions during the project work.

The projects that I was involved in during my time as a PhD student were all a combined effort involving researchers in the **Genomics, Diabetes and Endocrinology unit**. I am thankful to all the researchers, guest-researchers and collaborators that were involved in the unit and provided very helpful discussion. Last but not least, **Mattias**, **Johan**, **Jasmina** and **Ulrika,** thanks you for all the help and support with the technical work, the database and logistic assistance.

Iram, **Bushra**, **Esther**, **Ruchi** and **Pradeep** it has been great to get to know you! Thank you for all the fun and hilarious times and for all the support during my time as a PhD student. It has been fantastic!

Jon, **Marcus**, and all other friends I appreciate all the patience and support throughout the years.

Mayssoun – you have been a great support during my writing, even though you may not know it. I am more than grateful for our time and talks.

To my **Family** I owe everything to you! I can never do justice in describing what you have given me and in expressing my appreciation towards you. For all that I am, and I have I thank you!

References

[1] International Diabetes Federation, “IDF Diabetes Atlas 10th edn. Brussels, Belgium.” https://diabetesatlas.org/atlas/tenth-edition/

[2] D. R. Leitner *et al.*, “Obesity and Type 2 Diabetes: Two Diseases with a Need for Combined Treatment Strategies - EASO Can Lead the Way,” *OFA*, vol. 10, no. 5, pp. 483–492, 2017, doi: 10.1159/000480525.

[3] B. Zhou *et al.*, “Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4·4 million participants,” *The Lancet*, vol. 387, no. 10027, pp. 1513–1530, Apr. 2016, doi: 10.1016/S0140-6736(16)00618-8.

[4] K. Al-Rubeaan *et al.*, “Epidemiology of abnormal glucose metabolism in a country facing its epidemic: SAUDI-DM study,” *J Diabetes*, vol. 7, no. 5, pp. 622–632, Sep. 2015, doi: 10.1111/1753-0407.12224.

[5] D. Dabelea *et al.*, “Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009,” *JAMA*, vol. 311, no. 17, pp. 1778–1786, May 2014, doi: 10.1001/jama.2014.3201.

[6] P. Z. Zimmet, “Diabetes and its drivers: the largest epidemic in human history?,” *Clin Diabetes Endocrinol*, vol. 3, p. 1, 2017, doi: 10.1186/s40842-016-0039-3.

[7] M. Karamanou, A. Protogerou, G. Tsoucalas, G. Androutsos, and E. Poulakou-Rebelakou, “Milestones in the history of diabetes mellitus: The main contributors,” *World J Diabetes*, vol. 7, no. 1, pp. 1–7, Jan. 2016, doi: 10.4239/wjd.v7.i1.1.

[8] *Principles of Diabetes Mellitus*. Accessed: Oct. 29, 2022. [Online]. Available: http://link.springer.com/book/10.1007/978-0-387-09841-8

[9] J. v. Mering and O. Minkowski, “Diabetes mellitus nach Pankreasexstirpation,” *Archiv f. experiment. Pathol. u. Pharmakol*, vol. 26, no. 5, pp. 371–387, Jan. 1890, doi: 10.1007/BF01831214.

[10] F. G. Banting, C. H. Best, J. B. Collip, W. R. Campbell, and A. A. Fletcher, “Pancreatic extracts in the treatment of diabetes mellitus: preliminary report. 1922,” *CMAJ*, vol. 145, no. 10, pp. 1281–1286, Nov. 1991.

[11] R. S. Yalow and S. A. Berson, “Immunoassay of endogenous plasma insulin in man,” *J Clin Invest*, vol. 39, pp. 1157–1175, Jul. 1960, doi: 10.1172/JCI104130.

[12] D. M. Maahs and M. Rewers, “Mortality and Renal Disease in Type 1 Diabetes Mellitus—Progress Made, More to Be Done,” *The Journal of Clinical Endocrinology & Metabolism*, vol. 91, no. 10, pp. 3757–3759, Oct. 2006, doi: 10.1210/jc.2006-1730.

[13] D. Daneman, “Type 1 diabetes,” *The Lancet*, vol. 367, no. 9513, pp. 847–858, Mar. 2006, doi: 10.1016/S0140-6736(06)68341-4.

[14] M. A. Atkinson, G. S. Eisenbarth, and A. W. Michels, “Type 1 diabetes,” *Lancet*, vol. 383, no. 9911, pp. 69–82, Jan. 2014, doi: 10.1016/S0140-6736(13)60591-7.

[15] D. Jallut *et al.*, “Impaired glucose tolerance and diabetes in obesity: a 6-year follow-up study of glucose metabolism,” *Metabolism*, vol. 39, no. 10, pp. 1068–1075, Oct. 1990, doi: 10.1016/0026-0495(90)90168-c.
[16] R. A. DeFronzo, “Pathogenesis of type 2 diabetes mellitus,” *Med Clin North Am*, vol. 88, no. 4, pp. 787–835, ix, Jul. 2004, doi: 10.1016/j.mcna.2004.04.013.
[17] E. Ahlqvist *et al.*, “Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables,” *The Lancet Diabetes & Endocrinology*, vol. 6, no. 5, pp. 361–369, May 2018, doi: 10.1016/S2213-8587(18)30051-2.
[18] M. S. Udler *et al.*, “Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis,” *PLOS Medicine*, vol. 15, no. 9, p. e1002654, Sep. 2018, doi: 10.1371/journal.pmed.1002654.
[19] S. S. Fajans and G. I. Bell, “MODY: History, genetics, pathophysiology, and clinical decision making,” *Diabetes Care*, vol. 34, no. 8, pp. 1878–1884, Jul. 2011, doi: 10.2337/dc11-0035.
[20] T. Urakami, “Maturity-onset diabetes of the young (MODY): current perspectives on diagnosis and treatment,” *Diabetes Metab Syndr Obes*, vol. 12, pp. 1047–1056, 2019, doi: 10.2147/DMSO.S179793.
[21] J. B. O’Sullivan, S. S. Gellis, R. V. Dandrow, and B. O. Tenney, “The potential diabetic and her treatment in pregnancy,” *Obstet Gynecol*, vol. 27, no. 5, pp. 683–689, May 1966.
[22] N. Chen, R. Unnikrishnan I, R. M. Anjana, V. Mohan, and C. S. Pitchumoni, “The complex exocrine-endocrine relationship and secondary diabetes in exocrine pancreatic disorders,” *J Clin Gastroenterol*, vol. 45, no. 10, pp. 850–861, Dec. 2011, doi: 10.1097/MCG.0b013e31822a2ae5.
[23] J. Beltrand *et al.*, “Neonatal Diabetes Mellitus,” *Front Pediatr*, vol. 8, p. 540718, Sep. 2020, doi: 10.3389/fped.2020.540718.
[24] P. Pozzilli, C. Guglielmi, S. Caprio, and R. Buzzetti, “Obesity, Autoimmunity, and Double Diabetes in Youth,” *Diabetes Care*, vol. 34, no. Supplement_2, pp. S166–S170, Apr. 2011, doi: 10.2337/dc11-s213.
[25] R. A. DeFronzo, “From the Triumvirate to the Ominous Octet: A New Paradigm for the Treatment of Type 2 Diabetes Mellitus,” *Diabetes*, vol. 58, no. 4, pp. 773–795, Apr. 2009, doi: 10.2337/db09-9028.
[26] R. A. DeFronzo, “The Triumvirate: β-Cell, Muscle, Liver: A Collusion Responsible for NIDDM,” *Diabetes*, vol. 37, no. 6, pp. 667–687, Jun. 1988, doi: 10.2337/diab.37.6.667.
[27] S. Kashyap *et al.*, “A sustained increase in plasma free fatty acids impairs insulin secretion in nondiabetic subjects genetically predisposed to develop type 2 diabetes,” *Diabetes*, vol. 52, no. 10, pp. 2461–2474, Oct. 2003, doi: 10.2337/diabetes.52.10.2461.
[28] J. J. Meier and M. A. Nauck, “Incretins and the development of type 2 diabetes,” *Curr Diab Rep*, vol. 6, no. 3, pp. 194–201, Jun. 2006, doi: 10.1007/s11892-006-0034-7.
[29] A. D. Baron, L. Schaeffer, P. Shragg, and O. G. Kolterman, “Role of Hyperglucagonemia in Maintenance of Increased Rates of Hepatic Glucose Output in

Type II Diabetics," *Diabetes*, vol. 36, no. 3, pp. 274–283, Mar. 1987, doi: 10.2337/diab.36.3.274.
[30] M. A. Abdul-Ghani and R. A. DeFronzo, "Inhibition of Renal Glucose Reabsorption: A Novel Strategy for Achieving Glucose Control in Type 2 Diabetes Mellitus," *Endocrine Practice*, vol. 14, no. 6, pp. 782–790, Sep. 2008, doi: 10.4158/EP.14.6.782.
[31] M. Matsuda *et al.*, "Altered hypothalamic function in response to glucose ingestion in obese humans," *Diabetes*, vol. 48, no. 9, pp. 1801–1806, Sep. 1999, doi: 10.2337/diabetes.48.9.1801.
[32] Q. Zhou and D. A. Melton, "Pancreas regeneration," *Nature*, vol. 557, no. 7705, pp. 351–358, May 2018, doi: 10.1038/s41586-018-0088-0.
[33] M. F. Brereton, E. Vergari, Q. Zhang, and A. Clark, "Alpha-, Delta- and PP-cells," *J Histochem Cytochem*, vol. 63, no. 8, pp. 575–591, Aug. 2015, doi: 10.1369/0022155415583535.
[34] P. Rorsman and M. O. Huising, "The somatostatin-secreting pancreatic δ-cell in health and disease," *Nat Rev Endocrinol*, vol. 14, no. 7, pp. 404–414, Jul. 2018, doi: 10.1038/s41574-018-0020-6.
[35] Z. Fu, E. R. Gilbert, and D. Liu, "Regulation of insulin synthesis and secretion and pancreatic Beta-cell dysfunction in diabetes," *Curr Diabetes Rev*, vol. 9, no. 1, pp. 25–53, Jan. 2013.
[36] B. B. Boland, C. J. Rhodes, and J. S. Grimsby, "The dynamic plasticity of insulin production in β-cells," *Mol Metab*, vol. 6, no. 9, pp. 958–973, Sep. 2017, doi: 10.1016/j.molmet.2017.04.010.
[37] S. Seino, T. Shibasaki, and K. Minami, "Dynamics of insulin secretion and the clinical implications for obesity and diabetes," *J Clin Invest*, vol. 121, no. 6, pp. 2118–2125, Jun. 2011, doi: 10.1172/JCI45680.
[38] A. A. Christensen and M. Gannon, "The Beta Cell in Type 2 Diabetes," *Curr Diab Rep*, vol. 19, no. 9, p. 81, Aug. 2019, doi: 10.1007/s11892-019-1196-4.
[39] "Endoplasmic reticulum stress alters ryanodine receptor function in the murine pancreatic β cell - PMC." https://www-ncbi-nlm-nih-gov.ludwig.lub.lu.se/pmc/articles/PMC6322901/ (accessed Nov. 03, 2022).
[40] Prospective Studies Collaboration, "Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies," *The Lancet*, vol. 373, no. 9669, pp. 1083–1096, Mar. 2009, doi: 10.1016/S0140-6736(09)60318-4.
[41] J. M. Chan, E. B. Rimm, G. A. Colditz, M. J. Stampfer, and W. C. Willett, "Obesity, Fat Distribution, and Weight Gain as Risk Factors for Clinical Diabetes in Men," *Diabetes Care*, vol. 17, no. 9, pp. 961–969, Sep. 1994, doi: 10.2337/diacare.17.9.961.
[42] L. A. Muir *et al.*, "Adipose tissue fibrosis, hypertrophy, and hyperplasia: Correlations with diabetes in human obesity," *Obesity*, vol. 24, no. 3, pp. 597–605, 2016, doi: 10.1002/oby.21377.
[43] A. Castoldi, C. Naffah de Souza, N. O. S. Câmara, and P. M. Moraes-Vieira, "The Macrophage Switch in Obesity Development," *Frontiers in Immunology*, vol. 6, 2016, Accessed: Nov. 05, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fimmu.2015.00637

[44] B. Klop, J. Elte, and M. Cabezas, “Dyslipidemia in Obesity: Mechanisms and Potential Targets,” *Nutrients*, vol. 5, no. 4, pp. 1218–1240, Apr. 2013, doi: 10.3390/nu5041218.
[45] C. Lorenzo *et al.*, “Impaired fasting glucose and impaired glucose tolerance have distinct lipoprotein and apolipoprotein changes: the insulin resistance atherosclerosis study,” *J Clin Endocrinol Metab*, vol. 98, no. 4, pp. 1622–1630, Apr. 2013, doi: 10.1210/jc.2012-3185.
[46] M. Adiels, S.-O. Olofsson, M.-R. Taskinen, and J. Borén, “Overproduction of very low-density lipoproteins is the hallmark of the dyslipidemia in the metabolic syndrome,” *Arterioscler Thromb Vasc Biol*, vol. 28, no. 7, pp. 1225–1236, Jul. 2008, doi: 10.1161/ATVBAHA.107.160192.
[47] M. Arca, G. Pigna, and C. Favoccia, “Mechanisms of diabetic dyslipidemia: relevance for atherogenesis,” *Curr Vasc Pharmacol*, vol. 10, no. 6, pp. 684–686, Nov. 2012, doi: 10.2174/157016112803520864.
[48] A. Kawakami and M. Yoshida, “Apolipoprotein CIII Links Dyslipidemia with Atherosclerosis,” p. 6.
[49] E. Sivan and G. Boden, “Free fatty acids, insulin resistance, and pregnancy,” *Curr Diab Rep*, vol. 3, no. 4, pp. 319–322, Jul. 2003, doi: 10.1007/s11892-003-0024-y.
[50] R. M. Krauss, “Atherogenicity of Triglyceride-Rich Lipoproteins,” *The American Journal of Cardiology*, vol. 81, no. 4, Supplement 1, pp. 13B-17B, Feb. 1998, doi: 10.1016/S0002-9149(98)00032-0.
[51] H. R. Superko, “Small, dense, low-density lipoprotein and atherosclerosis,” *Curr Atheroscler Rep*, vol. 2, no. 3, pp. 226–231, May 2000, doi: 10.1007/s11883-000-0024-1.
[52] T. D. P. P. O. S. R. G. prepared on behalf of the D. R. Group *et al.*, “Long-term effects of the Diabetes Prevention Program interventions on cardiovascular risk factors: a report from the DPP Outcomes Study,” *Diabetic Medicine*, vol. 30, no. 1, pp. 46–55, 2013, doi: 10.1111/j.1464-5491.2012.03750.x.
[53] Action to Control Cardiovascular Risk in Diabetes Study Group *et al.*, “Effects of intensive glucose lowering in type 2 diabetes,” *N Engl J Med*, vol. 358, no. 24, pp. 2545–2559, Jun. 2008, doi: 10.1056/NEJMoa0802743.
[54] “International Expert Committee Report on the Role of the A1C Assay in the Diagnosis of Diabetes,” *Diabetes Care*, vol. 32, no. 7, pp. 1327–1334, Jul. 2009, doi: 10.2337/dc09-9033.
[55] “Diagnosis and Classification of Diabetes Mellitus | Diabetes Care | American Diabetes Association.” https://diabetesjournals-org.ludwig.lub.lu.se/care/article/37/Supplement_1/S81/37753/Diagnosis-and-Classification-of-Diabetes-Mellitus (accessed Nov. 06, 2022).
[56] American Diabetes Association, “2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2021,” *Diabetes Care*, vol. 44, no. Supplement_1, pp. S15–S33, Dec. 2020, doi: 10.2337/dc21-S002.
[57] T. M. Wallace, J. C. Levy, and D. R. Matthews, “Use and abuse of HOMA modeling,” *Diabetes Care*, vol. 27, no. 6, pp. 1487–1495, Jun. 2004, doi: 10.2337/diacare.27.6.1487.

[58] C. Ghosh, P. Mukhopadhyay, S. Ghosh, and M. Pradhan, “Insulin sensitivity index (ISI0, 120) potentially linked to carbon isotopes of breath CO2 for pre-diabetes and type 2 diabetes,” *Sci Rep*, vol. 5, p. 11959, Jul. 2015, doi: 10.1038/srep11959.
[59] R. L. Hanson *et al.*, “Evaluation of simple indices of insulin sensitivity and insulin secretion for use in epidemiologic studies,” *Am J Epidemiol*, vol. 151, no. 2, pp. 190–198, Jan. 2000, doi: 10.1093/oxfordjournals.aje.a010187.
[60] M. P. Gilbert and R. E. Pratley, “GLP-1 Analogs and DPP-4 Inhibitors in Type 2 Diabetes Therapy: Review of Head-to-Head Clinical Trials,” *Front Endocrinol (Lausanne)*, vol. 11, p. 178, Apr. 2020, doi: 10.3389/fendo.2020.00178.
[61] E. Brown, H. J. L. Heerspink, D. J. Cuthbertson, and J. P. H. Wilding, “SGLT2 inhibitors and GLP-1 receptor agonists: established and emerging indications,” *Lancet*, vol. 398, no. 10296, pp. 262–276, Jul. 2021, doi: 10.1016/S0140-6736(21)00536-5.
[62] B. Zinman *et al.*, “Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes,” *New England Journal of Medicine*, vol. 373, no. 22, pp. 2117–2128, Nov. 2015, doi: 10.1056/NEJMoa1504720.
[63] S. E. Nissen and K. Wolski, “Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes,” *New England Journal of Medicine*, vol. 356, no. 24, pp. 2457–2471, Jun. 2007, doi: 10.1056/NEJMoa072761.
[64] P. R. Schauer *et al.*, “Bariatric Surgery versus Intensive Medical Therapy in Obese Patients with Diabetes,” *New England Journal of Medicine*, vol. 366, no. 17, pp. 1567–1576, Apr. 2012, doi: 10.1056/NEJMoa1200225.
[65] P. Poulsen, K. Ohm Kyvik, A. Vaag, and H. Beck-Nielsen, “Heritability of Type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance – a population-based twin study,” *Diabetologia*, vol. 42, no. 2, pp. 139–145, Jan. 1999, doi: 10.1007/s001250051131.
[66] P. Bennett, T. Burch, and M. Miller, “DIABETES MELLITUS IN AMERICAN (PIMA) INDIANS,” *The Lancet*, vol. 298, no. 7716, pp. 125–128, Jul. 1971, doi: 10.1016/S0140-6736(71)92303-8.
[67] W. C. Knowler, D. J. Pettitt, M. F. Saad, and P. H. Bennett, “Diabetes mellitus in the Pima Indians: incidence, risk factors and pathogenesis,” *Diabetes Metab Rev*, vol. 6, no. 1, pp. 1–27, Feb. 1990, doi: 10.1002/dmr.5610060101.
[68] P. W. F. Wilson, J. B. Meigs, L. Sullivan, C. S. Fox, D. M. Nathan, and R. B. D’Agostino, “Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study,” *Arch Intern Med*, vol. 167, no. 10, pp. 1068–1074, May 2007, doi: 10.1001/archinte.167.10.1068.
[69] G. Willemsen *et al.*, “The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium,” *Twin Res Hum Genet*, vol. 18, no. 6, pp. 762–771, Dec. 2015, doi: 10.1017/thg.2015.83.
[70] Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research *et al.*, “Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels,” *Science*, vol. 316, no. 5829, pp. 1331–1336, Jun. 2007, doi: 10.1126/science.1142358.

[71] L. J. Scott *et al.*, “A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants,” *Science*, vol. 316, no. 5829, pp. 1341–1345, Jun. 2007, doi: 10.1126/science.1142382.
[72] E. Zeggini *et al.*, “Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes,” *Science*, vol. 316, no. 5829, pp. 1336–1341, Jun. 2007, doi: 10.1126/science.1142364.
[73] H. Unoki *et al.*, “SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations,” *Nat Genet*, vol. 40, no. 9, pp. 1098–1102, Sep. 2008, doi: 10.1038/ng.208.
[74] E. Zeggini *et al.*, “Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes,” *Nat Genet*, vol. 40, no. 5, pp. 638–645, May 2008, doi: 10.1038/ng.120.
[75] B. F. Voight *et al.*, “Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis,” *Nat Genet*, vol. 42, no. 7, pp. 579–589, Jul. 2010, doi: 10.1038/ng.609.
[76] Y. S. Cho *et al.*, “Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians,” *Nat Genet*, vol. 44, no. 1, pp. 67–72, Dec. 2011, doi: 10.1038/ng.1019.
[77] DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium *et al.*, “Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility,” *Nat Genet*, vol. 46, no. 3, pp. 234–244, Mar. 2014, doi: 10.1038/ng.2897.
[78] V. Steinthorsdottir *et al.*, “Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes,” *Nat Genet*, vol. 46, no. 3, Art. no. 3, Mar. 2014, doi: 10.1038/ng.2882.
[79] N. Bouatia-Naji *et al.*, “A polymorphism within the G6PC2 gene is associated with fasting plasma glucose levels,” *Science*, vol. 320, no. 5879, pp. 1085–1088, May 2008, doi: 10.1126/science.1156849.
[80] V. Lyssenko *et al.*, “A common variant in the melatonin receptor gene (MTNR1B) is associated with increased risk of future type 2 diabetes and impaired early insulin secretion,” *Nat Genet*, vol. 41, no. 1, pp. 82–88, Jan. 2009, doi: 10.1038/ng.288.
[81] P. Akbari *et al.*, “Sequencing of 640,000 exomes identifies GPR75 variants associated with protection from obesity,” *Science*, vol. 373, no. 6550, p. eabf8683, Jul. 2021, doi: 10.1126/science.abf8683.
[82] Q. Qi *et al.*, “Fried food consumption, genetic risk, and body mass index: gene-diet interaction analysis in three US cohort studies,” *BMJ*, vol. 348, p. g1610, Mar. 2014, doi: 10.1136/bmj.g1610.
[83] A. Mahajan *et al.*, “Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps,” *Nature Genetics*, vol. 50, no. 11, Art. no. 11, Nov. 2018, doi: 10.1038/s41588-018-0241-6.
[84] A. Mahajan *et al.*, “Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation,” *Nat Genet*, vol. 54, no. 5, pp. 560–572, May 2022, doi: 10.1038/s41588-022-01058-3.
[85] A. Raj and A. van Oudenaarden, “Stochastic gene expression and its consequences,” *Cell*, vol. 135, no. 2, pp. 216–226, Oct. 2008, doi: 10.1016/j.cell.2008.09.050.

[86] J. Lonsdale *et al.*, “The Genotype-Tissue Expression (GTEx) project,” *Nat Genet*, vol. 45, no. 6, Art. no. 6, Jun. 2013, doi: 10.1038/ng.2653.
[87] A. Soukas, P. Cohen, N. D. Socci, and J. M. Friedman, “Leptin-specific patterns of gene expression in white adipose tissue,” *Genes Dev*, vol. 14, no. 8, pp. 963–980, Apr. 2000.
[88] H. Xu *et al.*, “Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance,” *J Clin Invest*, vol. 112, no. 12, pp. 1821–1830, Dec. 2003, doi: 10.1172/JCI19451.
[89] V. K. Mootha *et al.*, “PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes,” *Nat Genet*, vol. 34, no. 3, pp. 267–273, Jul. 2003, doi: 10.1038/ng1180.
[90] T. R. Koves *et al.*, “Mitochondrial overload and incomplete fatty acid oxidation contribute to skeletal muscle insulin resistance,” *Cell Metab*, vol. 7, no. 1, pp. 45–56, Jan. 2008, doi: 10.1016/j.cmet.2007.10.013.
[91] C. E. McCurdy *et al.*, “Maternal high-fat diet triggers lipotoxicity in the fetal livers of nonhuman primates,” *J Clin Invest*, vol. 119, no. 2, pp. 323–335, Feb. 2009, doi: 10.1172/JCI32661.
[92] S.-F. Ng, R. C. Y. Lin, D. R. Laybutt, R. Barres, J. A. Owens, and M. J. Morris, “Chronic high-fat diet in fathers programs β-cell dysfunction in female rat offspring,” *Nature*, vol. 467, no. 7318, pp. 963–966, Oct. 2010, doi: 10.1038/nature09491.
[93] A. Wojtusciszyn, M. Armanet, P. Morel, T. Berney, and D. Bosco, “Insulin secretion from human beta cells is heterogeneous and dependent on cell-to-cell contacts,” *Diabetologia*, vol. 51, no. 10, pp. 1843–1852, Oct. 2008, doi: 10.1007/s00125-008-1103-z.
[94] Å. Segerstolpe *et al.*, “Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes,” *Cell Metab*, vol. 24, no. 4, pp. 593–607, Oct. 2016, doi: 10.1016/j.cmet.2016.08.020.
[95] M. Enge *et al.*, “Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns,” *Cell*, vol. 171, no. 2, pp. 321-330.e14, Oct. 2017, doi: 10.1016/j.cell.2017.09.004.
[96] P. Peixoto, P.-F. Cartron, A. A. Serandour, and E. Hervouet, “From 1957 to Nowadays: A Brief History of Epigenetics,” *Int J Mol Sci*, vol. 21, no. 20, p. E7571, Oct. 2020, doi: 10.3390/ijms21207571.
[97] “An Integrated Encyclopedia of DNA Elements in the Human Genome,” *Nature*, vol. 489, no. 7414, pp. 57–74, Sep. 2012, doi: 10.1038/nature11247.
[98] A. Kundaje *et al.*, “Integrative analysis of 111 reference human epigenomes,” *Nature*, vol. 518, no. 7539, pp. 317–330, Feb. 2015, doi: 10.1038/nature14248.
[99] B. T. Yang *et al.*, “Insulin promoter DNA methylation correlates negatively with insulin gene expression and positively with HbA1c levels in human pancreatic islets,” *Diabetologia*, vol. 54, no. 2, pp. 360–367, Feb. 2011, doi: 10.1007/s00125-010-1967-6.
[100] E. Hall *et al.*, “The effects of high glucose exposure on global gene expression and DNA methylation in human pancreatic islets,” *Molecular and Cellular Endocrinology*, vol. 472, pp. 57–67, Sep. 2018, doi: 10.1016/j.mce.2017.11.019.

[101] M. Bibikova *et al.*, “Genome-wide DNA methylation profiling using Infinium® assay,” *Epigenomics*, vol. 1, no. 1, pp. 177–200, Oct. 2009, doi: 10.2217/epi.09.14.
[102] T. Dayeh *et al.*, “Genome-Wide DNA Methylation Analysis of Human Pancreatic Islets from Type 2 Diabetic and Non-Diabetic Donors Identifies Candidate Genes That Influence Insulin Secretion,” *PLOS Genetics*, vol. 10, no. 3, p. e1004160, Mar. 2014, doi: 10.1371/journal.pgen.1004160.
[103] “Increased Hepatic PDGF-AA Signaling Mediates Liver Insulin Resistance in Obesity-Associated Type 2 Diabetes | Diabetes | American Diabetes Association.” https://diabetesjournals.org/diabetes/article/67/7/1310/35346/Increased-Hepatic-PDGF-AA-Signaling-Mediates-Liver (accessed Nov. 20, 2022).
[104] E. Nilsson *et al.*, “Altered DNA Methylation and Differential Expression of Genes Influencing Metabolism and Inflammation in Adipose Tissue From Subjects With Type 2 Diabetes,” *Diabetes*, vol. 63, no. 9, pp. 2962–2976, Aug. 2014, doi: 10.2337/db13-1459.
[105] “Impact of an Exercise Intervention on DNA Methylation in Skeletal Muscle From First-Degree Relatives of Patients With Type 2 Diabetes | Diabetes | American Diabetes Association.” https://diabetesjournals.org/diabetes/article/61/12/3322/33725/Impact-of-an-Exercise-Intervention-on-DNA (accessed Nov. 20, 2022).
[106] P. Volkov *et al.*, “Whole-Genome Bisulfite Sequencing of Human Pancreatic Islets Reveals Novel Differentially Methylated Regions in Type 2 Diabetes Pathogenesis,” *Diabetes*, vol. 66, no. 4, pp. 1074–1085, Jan. 2017, doi: 10.2337/db16-0996.
[107] F. Paneni *et al.*, “Adverse Epigenetic Signatures by Histone Methyltransferase Set7 Contribute to Vascular Dysfunction in Patients With Type 2 Diabetes Mellitus,” *Circulation: Cardiovascular Genetics*, vol. 8, no. 1, pp. 150–158, Feb. 2015, doi: 10.1161/CIRCGENETICS.114.000671.
[108] A. H. Olsson *et al.*, “Genome-Wide Associations between Genetic and Epigenetic Variation Influence mRNA Expression and Insulin Secretion in Human Pancreatic Islets,” *PLOS Genetics*, vol. 10, no. 11, p. e1004735, Nov. 2014, doi: 10.1371/journal.pgen.1004735.
[109] A. Ravelli *et al.*, “Glucose tolerance in adults after prenatal exposure to famine,” *The Lancet*, vol. 351, no. 9097, pp. 173–177, Jan. 1998, doi: 10.1016/S0140-6736(97)07244-9.
[110] D. J. Palmer, R.-C. Huang, J. M. Craig, and S. L. Prescott, “Nutritional Influences on Epigenetic Programming: Asthma, Allergy, and Obesity,” *Immunology and Allergy Clinics of North America*, vol. 34, no. 4, pp. 825–837, Nov. 2014, doi: 10.1016/j.iac.2014.07.003.
[111] C. N. Hales and D. J. Barker, “The thrifty phenotype hypothesis,” *Br Med Bull*, vol. 60, pp. 5–20, 2001, doi: 10.1093/bmb/60.1.5.
[112] C. H. Fall, “Fetal Programming and the Risk of Non-communicable Disease,” *Indian J Pediatr*, vol. 80, no. 0 1, pp. S13–S20, Mar. 2013, doi: 10.1007/s12098-012-0834-5.

[113] M. H. Vickers, “Early life nutrition, epigenetics and programming of later life disease,” *Nutrients*, vol. 6, no. 6, pp. 2165–2178, Jun. 2014, doi: 10.3390/nu6062165.
[114] A. Chango and I. P. Pogribny, “Considering maternal dietary modulators for epigenetic regulation and programming of the fetal epigenome,” *Nutrients*, vol. 7, no. 4, pp. 2748–2770, Apr. 2015, doi: 10.3390/nu7042748.
[115] J. Zheng, X. Xiao, Q. Zhang, and M. Yu, “DNA methylation: the pivotal interaction between early-life nutrition and glucose metabolism in later life,” *Br J Nutr*, vol. 112, no. 11, pp. 1850–1857, Dec. 2014, doi: 10.1017/S0007114514002827.
[116] P. M. Catalano, E. D. Tyzbir, R. R. Wolfe, N. M. Roman, S. B. Amini, and E. A. Sims, “Longitudinal changes in basal hepatic glucose production and suppression during insulin infusion in normal pregnant women,” *Am J Obstet Gynecol*, vol. 167, no. 4 Pt 1, pp. 913–919, Oct. 1992, doi: 10.1016/s0002-9378(12)80011-1.
[117] T. A. Buchanan, “Pancreatic B-cell defects in gestational diabetes: implications for the pathogenesis and prevention of type 2 diabetes,” *J Clin Endocrinol Metab*, vol. 86, no. 3, pp. 989–993, Mar. 2001, doi: 10.1210/jcem.86.3.7339.
[118] G. S. Hotamisligil, P. Peraldi, A. Budavari, R. Ellis, M. F. White, and B. M. Spiegelman, “IRS-1-mediated inhibition of insulin receptor tyrosine kinase activity in TNF-alpha- and obesity-induced insulin resistance,” *Science*, vol. 271, no. 5249, pp. 665–668, Feb. 1996, doi: 10.1126/science.271.5249.665.
[119] J. E. Friedman, T. Ishizuka, J. Shao, L. Huston, T. Highman, and P. Catalano, “Impaired glucose transport and insulin receptor tyrosine phosphorylation in skeletal muscle from obese women with gestational diabetes,” *Diabetes*, vol. 48, no. 9, pp. 1807–1814, Sep. 1999, doi: 10.2337/diabetes.48.9.1807.
[120] S. Hauguel, V. Desmaizieres, and J. C. Challier, “Glucose uptake, utilization, and transfer by the human placenta as functions of maternal glucose concentration,” *Pediatr Res*, vol. 20, no. 3, pp. 269–273, Mar. 1986, doi: 10.1203/00006450-198603000-00015.
[121] G. Desoye and C. J. Nolan, “The fetal glucose steal: an underappreciated phenomenon in diabetic pregnancy,” *Diabetologia*, vol. 59, no. 6, pp. 1089–1094, Jun. 2016, doi: 10.1007/s00125-016-3931-6.
[122] M. Gauster *et al.*, “Dysregulation of placental endothelial lipase in obese women with gestational diabetes mellitus,” *Diabetes*, vol. 60, no. 10, pp. 2457–2464, Oct. 2011, doi: 10.2337/db10-1434.
[123] L. Kelstrup *et al.*, “Insulin resistance and impaired pancreatic β-cell function in adult offspring of women with diabetes in pregnancy,” *J Clin Endocrinol Metab*, vol. 98, no. 9, pp. 3793–3801, Sep. 2013, doi: 10.1210/jc.2013-1536.
[124] K. Kc, S. Shakya, and H. Zhang, “Gestational diabetes mellitus and macrosomia: a literature review,” *Ann Nutr Metab*, vol. 66 Suppl 2, pp. 14–20, 2015, doi: 10.1159/000371628.
[125] D. F. Dłuski, E. Wolińska, and M. Skrzypczak, “Epigenetic Changes in Gestational Diabetes Mellitus,” *Int J Mol Sci*, vol. 22, no. 14, p. 7649, Jul. 2021, doi: 10.3390/ijms22147649.

[126] L. Aerts and F. A. Van Assche, “Animal evidence for the transgenerational development of diabetes mellitus,” *Int J Biochem Cell Biol*, vol. 38, no. 5–6, pp. 894–903, 2006, doi: 10.1016/j.biocel.2005.07.006.
[127] G. A. Stevens *et al.*, “Global, regional, and national trends in haemoglobin concentration and prevalence of total and severe anaemia in children and pregnant and non-pregnant women for 1995-2011: a systematic analysis of population-representative data,” *Lancet Glob Health*, vol. 1, no. 1, pp. e16-25, Jul. 2013, doi: 10.1016/S2214-109X(13)70001-9.
[128] N. J. Kassebaum *et al.*, “A systematic analysis of global anemia burden from 1990 to 2010,” *Blood*, vol. 123, no. 5, pp. 615–624, Jan. 2014, doi: 10.1182/blood-2013-06-508325.
[129] A. Stangret, M. Skoda, A. Wnuk, M. Pyzlak, and D. Szukiewicz, “Mild anemia during pregnancy upregulates placental vascularity development,” *Med Hypotheses*, vol. 102, pp. 37–40, May 2017, doi: 10.1016/j.mehy.2017.03.007.
[130] H. J. McArdle, L. Gambling, and C. Kennedy, “Iron deficiency during pregnancy: the consequences for placental function and fetal outcome,” *Proceedings of the Nutrition Society*, vol. 73, no. 1, pp. 9–15, Feb. 2014, doi: 10.1017/S0029665113003637.
[131] M. K. Georgieff, “Iron assessment to protect the developing brain,” *The American Journal of Clinical Nutrition*, vol. 106, no. suppl_6, pp. 1588S-1593S, Dec. 2017, doi: 10.3945/ajcn.117.155846.
[132] S. M. Fishman, P. Christian, and K. P. West, “The role of vitamins in the prevention and control of anaemia,” *Public Health Nutr*, vol. 3, no. 2, pp. 125–150, Jun. 2000, doi: 10.1017/s1368980000000173.
[133] W. B. Castle, “Megaloblastic anemia,” *Postgrad Med*, vol. 64, no. 4, pp. 117–122, Oct. 1978, doi: 10.1080/00325481.1978.11714952.
[134] W. World Health Organization, “Fact sheet about malaria.” https://www.who.int/news-room/fact-sheets/detail/malaria (accessed Dec. 10, 2022).
[135] N. J. White, “Anaemia and malaria,” *Malar J*, vol. 17, no. 1, p. 371, Oct. 2018, doi: 10.1186/s12936-018-2509-9.
[136] N. Spottiswoode, P. E. Duffy, and H. Drakesmith, “Iron, anemia and hepcidin in malaria,” *Front Pharmacol*, vol. 5, p. 125, 2014, doi: 10.3389/fphar.2014.00125.
[137] N. Kozuki, A. C. Lee, J. Katz, and Child Health Epidemiology Reference Group, “Moderate to severe, but not mild, maternal anemia is associated with increased risk of small-for-gestational-age outcomes,” *J Nutr*, vol. 142, no. 2, pp. 358–362, Feb. 2012, doi: 10.3945/jn.111.149237.
[138] B. G. Campbell, “Sexual selection and the descent of man, 1871-1971.”
[139] T. Moore and D. Haig, “Genomic imprinting in mammalian development: a parental tug-of-war,” *Trends Genet*, vol. 7, no. 2, pp. 45–49, Feb. 1991, doi: 10.1016/0168-9525(91)90230-N.
[140] D. Haig, “The Kinship Theory of Genomic Imprinting,” *Annual Review of Ecology and Systematics*, vol. 31, no. 1, pp. 9–32, 2000, doi: 10.1146/annurev.ecolsys.31.1.9.

[141] M. S. Bartolomei and A. C. Ferguson-Smith, “Mammalian genomic imprinting,” *Cold Spring Harb Perspect Biol*, vol. 3, no. 7, p. a002592, Jul. 2011, doi: 10.1101/cshperspect.a002592.
[142] W. Reik and J. Walter, “Genomic imprinting: parental influence on the genome,” *Nat Rev Genet*, vol. 2, no. 1, Art. no. 1, Jan. 2001, doi: 10.1038/35047554.
[143] R. Hager, J. M. Cheverud, and J. B. Wolf, “Maternal Effects as the Cause of Parent-of-Origin Effects That Mimic Genomic Imprinting,” *Genetics*, vol. 178, no. 3, pp. 1755–1762, Mar. 2008, doi: 10.1534/genetics.107.080697.
[144] J. Li and T. Jiang, “Efficient inference of haplotypes from genotypes on a pedigree,” *J. Bioinform. Comput. Biol.*, vol. 01, no. 01, pp. 41–69, Apr. 2003, doi: 10.1142/S0219720003000204.
[145] V. Riso *et al.*, “ZFP57 maintains the parent-of-origin-specific expression of the imprinted genes and differentially affects non-imprinted targets in mouse embryonic stem cells,” *Nucleic Acids Res*, vol. 44, no. 17, pp. 8165–8178, Sep. 2016, doi: 10.1093/nar/gkw505.
[146] K. Sha, “A Mechanistic View of Genomic Imprinting,” *Annual review of genomics and human genetics*, vol. 9, pp. 197–216, Feb. 2008, doi: 10.1146/annurev.genom.122007.110031.
[147] H. Takeda *et al.*, “The callipyge mutation enhances bidirectional long-range DLK1-GTL2 intergenic transcription in cis,” *Proc Natl Acad Sci U S A*, vol. 103, no. 21, pp. 8119–8124, May 2006, doi: 10.1073/pnas.0602844103.
[148] H. A. Lawson, J. E. Cady, C. Partridge, J. B. Wolf, C. F. Semenkovich, and J. M. Cheverud, “Genetic effects at pleiotropic loci are context-dependent with consequences for the maintenance of genetic variation in populations,” *PLoS Genet*, vol. 7, no. 9, p. e1002256, Sep. 2011, doi: 10.1371/journal.pgen.1002256.
[149] A. R. Prickett and R. J. Oakey, “A survey of tissue-specific genomic imprinting in mammals,” *Mol Genet Genomics*, vol. 287, no. 8, pp. 621–630, Aug. 2012, doi: 10.1007/s00438-012-0708-6.
[150] V. Kameswaran *et al.*, “Epigenetic regulation of the DLK1-MEG3 microRNA cluster in human type 2 diabetic islets,” *Cell Metab*, vol. 19, no. 1, pp. 135–145, Jan. 2014, doi: 10.1016/j.cmet.2013.11.016.
[151] A. Kong *et al.*, “Parental origin of sequence variants associated with complex diseases,” *Nature*, vol. 462, no. 7275, pp. 868–874, Dec. 2009, doi: 10.1038/nature08625.
[152] L. Groop *et al.*, “Metabolic Consequences of a Family History of NIDDM (The Botnia Study): Evidence for Sex-Specific Parental Effects,” *Diabetes*, vol. 45, no. 11, pp. 1585–1593, Nov. 1996, doi: 10.2337/diab.45.11.1585.
[153] L. Hjort *et al.*, “FOETAL for NCD—FOetal Exposure and Epidemiological Transitions: the role of Anaemia in early Life for Non-Communicable Diseases in later life: a prospective preconception study in rural Tanzania,” *BMJ Open*, vol. 9, no. 5, p. e024861, May 2019, doi: 10.1136/bmjopen-2018-024861.
[154] A. Auton *et al.*, “A global reference for human genetic variation,” *Nature*, vol. 526, no. 7571, Art. no. 7571, Oct. 2015, doi: 10.1038/nature15393.

[155] S. McCarthy *et al.*, “A reference panel of 64,976 haplotypes for genotype imputation,” *Nat Genet*, vol. 48, no. 10, Art. no. 10, Oct. 2016, doi: 10.1038/ng.3643.
[156] A. Dobin *et al.*, “STAR: ultrafast universal RNA-seq aligner,” *Bioinformatics*, vol. 29, no. 1, pp. 15–21, Jan. 2013, doi: 10.1093/bioinformatics/bts635.
[157] S. Anders, P. T. Pyl, and W. Huber, “HTSeq—a Python framework to work with high-throughput sequencing data,” *Bioinformatics*, vol. 31, no. 2, pp. 166–169, Jan. 2015, doi: 10.1093/bioinformatics/btu638.
[158] Y. Liao, G. K. Smyth, and W. Shi, “featureCounts: an efficient general purpose program for assigning sequence reads to genomic features,” *Bioinformatics*, vol. 30, no. 7, pp. 923–930, Apr. 2014, doi: 10.1093/bioinformatics/btt656.
[159] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data,” *Bioinformatics*, vol. 26, no. 1, pp. 139–140, Jan. 2010, doi: 10.1093/bioinformatics/btp616.
[160] A. A. Shabalin, “Matrix eQTL: ultra fast eQTL analysis via large matrix operations,” *Bioinformatics*, vol. 28, no. 10, pp. 1353–1358, May 2012, doi: 10.1093/bioinformatics/bts163.
[161] A. Srivastava *et al.*, “Alignment and mapping methodology influence transcript abundance estimation,” *Genome Biol*, vol. 21, no. 1, Art. no. 1, Dec. 2020, doi: 10.1186/s13059-020-02151-8.
[162] B. van de Geijn, G. McVicker, Y. Gilad, and J. K. Pritchard, “WASP: allele-specific software for robust molecular quantitative trait locus discovery,” *Nat Methods*, vol. 12, no. 11, pp. 1061–1063, Nov. 2015, doi: 10.1038/nmeth.3582.
[163] S. E. Castel, A. Levy-Moonshine, P. Mohammadi, E. Banks, and T. Lappalainen, “Tools and best practices for data processing in allelic expression analysis,” *Genome Biol*, vol. 16, no. 1, Art. no. 1, Dec. 2015, doi: 10.1186/s13059-015-0762-6.
[164] D. A. Skelly, M. Johansson, J. Madeoy, J. Wakefield, and J. M. Akey, “A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data,” *Genome Res.*, vol. 21, no. 10, pp. 1728–1737, Oct. 2011, doi: 10.1101/gr.119784.110.
[165] J. Fan *et al.*, “ASEP: Gene-based detection of allele-specific expression across individuals in a population by RNA sequencing,” *PLOS Genetics*, vol. 16, no. 5, p. e1008786, May 2020, doi: 10.1371/journal.pgen.1008786.
[166] M. Bibikova *et al.*, “High density DNA methylation array with single CpG site resolution,” *Genomics*, vol. 98, no. 4, pp. 288–295, Oct. 2011, doi: 10.1016/j.ygeno.2011.07.007.
[167] W. E. Johnson, C. Li, and A. Rabinovic, “Adjusting batch effects in microarray expression data using empirical Bayes methods,” *Biostatistics*, vol. 8, no. 1, pp. 118–127, Jan. 2007, doi: 10.1093/biostatistics/kxj037.
[168] G. R. Abecasis, L. R. Cardon, and W. O. Cookson, “A general test of association for quantitative traits in nuclear families,” *Am J Hum Genet*, vol. 66, no. 1, pp. 279–292, Jan. 2000, doi: 10.1086/302698.

[169] D. M. Blodgett *et al.*, “Novel Observations From Next-Generation RNA Sequencing of Highly Purified Human Adult and Fetal Islet Cell Subsets,” *Diabetes*, vol. 64, no. 9, pp. 3172–3181, Sep. 2015, doi: 10.2337/db15-0039.
[170] S. E. Kahn, R. L. Hull, and K. M. Utzschneider, “Mechanisms linking obesity to insulin resistance and type 2 diabetes,” *Nature*, vol. 444, no. 7121, pp. 840–846, Dec. 2006, doi: 10.1038/nature05482.
[171] I. Kassam, Y. Wu, J. Yang, P. M. Visscher, and A. F. McRae, “Tissue-specific sex differences in human gene expression,” *Hum Mol Genet*, vol. 28, no. 17, pp. 2976–2986, Sep. 2019, doi: 10.1093/hmg/ddz090.
[172] J. M. Schoonejans *et al.*, “Sex-specific effects of maternal metformin intervention during glucose-intolerant obese pregnancy on body composition and metabolic health in aged mouse offspring,” *Diabetologia*, vol. 65, no. 12, pp. 2132–2145, Dec. 2022, doi: 10.1007/s00125-022-05789-0.
[173] D. F. Gudbjartsson *et al.*, “Many sequence variants affecting diversity of adult human height,” *Nat Genet*, vol. 40, no. 5, Art. no. 5, May 2008, doi: 10.1038/ng.122.
[174] E. Conway *et al.*, “A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities,” *Mol Cell*, vol. 70, no. 3, pp. 408-421.e8, May 2018, doi: 10.1016/j.molcel.2018.03.005.
[175] T. Maffucci, F. T. Cooke, F. M. Foster, C. J. Traer, M. J. Fry, and M. Falasca, “Class II phosphoinositide 3-kinase defines a novel signaling pathway in cell migration,” *J Cell Biol*, vol. 169, no. 5, pp. 789–799, Jun. 2005, doi: 10.1083/jcb.200408005.
[176] P. H. Petersen, K. Zou, J. K. Hwang, Y. N. Jan, and W. Zhong, “Progenitor cell maintenance requires numb and numblike during mouse neurogenesis,” *Nature*, vol. 419, no. 6910, pp. 929–934, Oct. 2002, doi: 10.1038/nature01124.
[177] K. M. Godfrey and D. J. Barker, “Fetal programming and adult health,” *Public Health Nutr*, vol. 4, no. 2B, pp. 611–624, Apr. 2001, doi: 10.1079/phn2001145.
[178] J. A. Martin, B. E. Hamilton, M. J. K. Osterman, A. K. Driscoll, and P. Drake, “Births: Final Data for 2016,” *Natl Vital Stat Rep*, vol. 67, no. 1, pp. 1–55, Jan. 2018.
[179] J. Daneshmoghadam, A. Omidifar, N. Akbari Dilmaghani, Z. Karimi, S. Emamgholipour, and M. shanaki, “The gene expression of long non-coding RNAs (lncRNAs): MEG3 and H19 in adipose tissues from obese women and its association with insulin resistance and obesity indices,” *J Clin Lab Anal*, vol. 35, no. 5, p. e23741, Feb. 2021, doi: 10.1002/jcla.23741.
[180] A. Iliadou, H. Snieder, X. Wang, F. A. Treiber, and C. L. Davis, “Heritabilities of lipids in young European American and African American twins,” *Twin Res Hum Genet*, vol. 8, no. 5, pp. 492–498, Oct. 2005, doi: 10.1375/183242705774310187.
[181] P. Almgren *et al.*, “Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study,” *Diabetologia*, vol. 54, no. 11, pp. 2811–2819, Nov. 2011, doi: 10.1007/s00125-011-2267-5.

[182] S. C. Lee *et al.*, “Diabetes in Hong Kong Chinese: evidence for familial clustering and parental effects,” *Diabetes Care*, vol. 23, no. 9, pp. 1365–1368, Sep. 2000, doi: 10.2337/diacare.23.9.1365.
[183] L. Shan *et al.*, “The angiopoietin-like proteins ANGPTL3 and ANGPTL4 inhibit lipoprotein lipase activity through distinct mechanisms,” *J Biol Chem*, vol. 284, no. 3, pp. 1419–1424, Jan. 2009, doi: 10.1074/jbc.M808477200.
[184] M. Shimamura *et al.*, “Angiopoietin-like protein3 regulates plasma HDL cholesterol through suppression of endothelial lipase,” *Arterioscler Thromb Vasc Biol*, vol. 27, no. 2, pp. 366–372, Feb. 2007, doi: 10.1161/01.ATV.0000252827.51626.89.
[185] J. Stephenson *et al.*, “Before the beginning: nutrition and lifestyle in the preconception period and its importance for future health,” *The Lancet*, vol. 391, no. 10132, pp. 1830–1841, May 2018, doi: 10.1016/S0140-6736(18)30311-8.
[186] I. M. Predazzi *et al.*, “Sex-Specific Parental Effects on Offspring Lipid Levels,” *J Am Heart Assoc*, vol. 4, no. 7, p. e001951, Jun. 2015, doi: 10.1161/JAHA.115.001951.
[187] P. D. Gluckman and M. A. Hanson, “Maternal constraint of fetal growth and its consequences,” *Semin Fetal Neonatal Med*, vol. 9, no. 5, pp. 419–425, Oct. 2004, doi: 10.1016/j.siny.2004.03.001.
[188] R. A. Waterland and K. B. Michels, “Epigenetic epidemiology of the developmental origins hypothesis,” *Annu Rev Nutr*, vol. 27, pp. 363–388, 2007, doi: 10.1146/annurev.nutr.27.061406.093705.
[189] E. Davis *et al.*, “RNAi-Mediated Allelic trans-Interaction at the Imprinted Rtl1/Peg11 Locus,” *Current Biology*, vol. 15, no. 8, pp. 743–749, Apr. 2005, doi: 10.1016/j.cub.2005.02.060.
[190] M. Kitazawa, A. Sutani, T. Kaneko-Ishino, and F. Ishino, “The role of eutherian-specific RTL1 in the nervous system and its implications for the Kagami-Ogata and Temple syndromes,” *Genes to Cells*, vol. 26, no. 3, pp. 165–179, 2021, doi: https://doi.org/10.1111/gtc.12830.
[191] M. Kagami *et al.*, “Temple syndrome: comprehensive molecular and clinical findings in 32 Japanese patients,” *Genetics in Medicine*, vol. 19, no. 12, Art. no. 12, Dec. 2017, doi: 10.1038/gim.2017.53.
[192] R. L. Hanson *et al.*, “Strong parent-of-origin effects in the association of KCNQ1 variants with type 2 diabetes in American Indians,” *Diabetes*, vol. 62, no. 8, pp. 2984–2991, Aug. 2013, doi: 10.2337/db12-1767.
[193] R. B. Prasad *et al.*, “Excess maternal transmission of variants in the THADA gene to offspring with type 2 diabetes,” *Diabetologia*, vol. 59, no. 8, pp. 1702–1713, Aug. 2016, doi: 10.1007/s00125-016-3973-9.
[194] A. J. Drake, B. R. Walker, and J. R. Seckl, “Intergenerational consequences of fetal programming by in utero exposure to glucocorticoids in rats,” *Am J Physiol Regul Integr Comp Physiol*, vol. 288, no. 1, pp. R34-38, Jan. 2005, doi: 10.1152/ajpregu.00106.2004.

Supplementary data

Supplementary data are available at:
https://lu.box.com/s/50g40zf60bf857163xkq4x08scoinghj

Printed by Media-Tryck, Lund 2023 NORDIC SWAN ECOLABEL 3041 0903

FACULTY OF MEDICINE

Department of Clinical Sciences, Malmö

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2023:82
ISBN 978-91-8021-422-3
ISSN 1652-8220