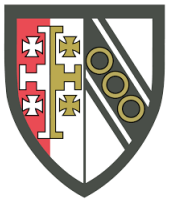
UNIVERSITY OF CAMBRIDGE

Early life phenotypes of type 1 diabetes genetic risk

Antigoni Eleftheriou

Selwyn College



August 2018

This dissertation is submitted for the degree of Doctor of Philosophy

SUMMARY

Early life phenotypes of type 1 diabetes genetic risk

Antigoni Eleftheriou

Observational studies have shown that children who develop type 1 diabetes (T1D) have larger size at birth — a marker of foetal growth — or rapid postnatal weight gains. Recent studies have explored the contribution of the T1D susceptibility human leukocyte antigen (HLA) alleles, which explain half of disease heritability, to perinatal growth but findings are inconsistent.

In the Cambridge Baby Growth Study (CBGS), a birth cohort which follows children over the first 2 years of life, I conducted genotype-phenotype correlation analyses with the objective of identifying T1D susceptibility single nucleotide polymorphisms (SNPs) and candidate genes that influence the archetypal phenotypes of the disease prodrome, which starts early in life, as a means of shedding light on the heterogeneity of data and sign-posting mechanisms. Based on prior 'biological' knowledge I selected T1D SNPs conjectured to affect growth mediated by hormones, vitamin D, or the microbiome. Genotypes were determined in ~600 children and used to derive high-risk T1D HLA haplotypes and genetic risk scores. Examination of data in the CBGS defined the contributors to physiological growth and set the framework for analyses.

My findings concurred with most prior studies on the lack of an effect of the SNP rs689 in *INS* on early growth. A genetic contribution to size at birth was detected by variants in the imprinted *DLK1* gene, the vitamin D metabolism *CYP2R1* gene and the gene encoding TYK2 known for its susceptibility to infections. Examination of longitudinal phenotypes identified that infancy growth is influenced by genes with a key role in innate immunity, e.g. *IFIH1*. The strongest associations were found between skinfold thickness, a proxy for adiposity, and variants in genes whose products play a role in immune and inflammatory pathways, exemplified by the rs653178 in *SH2B3*, previously known for its association with cardiovascular disease. Taking on the lead from my findings with anthropometry, I explored genetic-hormonal correlates and found evidence in support of IGF-1 mediating the association that I identified, more strongly in boys than in girls, between the SNP tagging the high-risk *HLA-DR3* and rapid linear growth. I made the novel finding that the rs12785878 in *DHCR7*, known to strongly associate with vitamin D levels, inversely correlated with leptin levels in girls but not in boys.

This study elucidated the effects of T1D susceptibility SNPs on infancy adiposity, which might serve as a better disease marker than weight. It identified a link between leptin and a common polymorphism in *DHCR7*, opening up the possibility that vitamin D suppresses leptin production in a sex-specific manner with protective effects against autoimmunity.

PREFACE

DECLARATION OF ORIGINALITY

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the word limit of 60,000 words set by the Degree Committee of the School of Clinical Medicine.

ABBREVIATIONS

AGA	appropriate for gestational age
AIC	Akaike information criterion
AIDS	acquired immune deficiency syndrome
AIRE	autoimmune regulator
ALS	acid-labile subunit
ALSPAC	Avon Longitudinal Study of Pregnancy and Childhood
ANCOVA	analysis of covariance
ANOVA	analysis of variance
BL	birth length
BMI	body mass index
bp	base pairs
BW	birth weight
CBGS	Cambridge Baby Growth Study
Chr	chromosome
CI	confidence interval
CTLA-4	cytotoxic T-lymphocyte-associated protein 4
DAISY	Diabetes Autoimmunity Study in the Young
DBP	vitamin D binding protein
DBS	dried blood spot
df	degrees of freedom
DHCR7	dehydrocholesterol reductase
DiPiS	The Diabetes Prevention in Skane
DIPP	Type 1 Diabetes Prediction and Prevention
DLK1	delta like non-canonical Notch ligand 1
DNA	deoxyribonucleic acid
DUPE	duplication error
EBI	European Bioinformatics Institute
EGF	epidermal growth factor
EMBT	European Group for Blood and Marrow Transplantation
ENDIA	Environmental Determinants of Islet Autoimmunity
eQTL	expressed quantitative trait loci
FFM	fat-free mass
FM	fat mass
FPIR	first phase insulin response
GADA	glutamic acid decarboxylase autoantibodies
GC	group specific component

GH	growth hormone
GHR	growth hormone receptor
GLP-1	glucagon-like peptide 1
GWAS	genome-wide association studies
HC	head circumference
HLA	human leukocyte antigen
HrBW	higher relative birth weight
HWE	Hardy-Weinberg equilibrium
IA	islet autoimmunity
IA-2A	insulinoma-associated antigen-2 autoantibodies
IAA	insulin autoantibodies
ICA	islet cell antibodies
ICP	infancy-childhood-puberty
IDDM	insulin-dependent diabetes mellitus
IFIH1	interferon-induced with helicase C domain 1
IFN	interferon
IgE	immunoglobulin E
IGF-1	insulin-like growth factor-1
IGF1R	insulin-like growth factor-1 receptor
IGF-2	insulin-like growth factor-2
IGFBP	insulin-like growth factor binding protein
IgG	immunoglobulin G
IgM	immunoglobulin M
IL	interleukin
IL2RA	interleukin 2 receptor alpha
IMD	index of multiple deprivation
IRIS	Immunogenetic Related Information Source
IUGR	intrauterine growth retardation
IVGTT	intravenous glucose tolerance test
JAK	janus kinase
JDRF	Juvenile Diabetes Research Foundation
KASP	‘kompetitive’ allele specific PCR
kDa	kilo Dalton
LD	linkage disequilibrium
LGA	large for gestational age
MAF	minor allele frequency
MEM	mixed effects model
MHC	major histocompatibility complex

MPL	myeloproliferative leukaemia virus
MRC	Medical Research Council
NCBI	National Centre for Biotechnology Information
NF-κB	nuclear factor κB
NHGRI	National Human Genome Research Institute
NTC	no template control
OGGT	oral glucose tolerance test
OR	odds ratio
PCOS	polycystic ovary syndrome
PTPN	protein tyrosine phosphatase non-receptor type
RNA	ribonucleic acid
rs	reference sequence
SD	standard deviation
SDS	standard deviation score
SE	standard error
SFT	skinfold thickness
SGA	small for gestational age
SNP	single nucleotide polymorphism
STAT	signal transducer and activator of transcription
T1D	type 1 diabetes
T1DGC	Type 1 Diabetes Genetics Consortium
T2D	type 2 diabetes
TCR	T-cell receptor
TEDDY	The Environmental Determinants of Diabetes in the Young
TH	tyrosine hydroxylase
TNF-α	tumour necrosis factor-α
Treg	T regulatory
TRIGR	Trial to Reduce IDDM in Genetically at Risk
TYK2	tyrosine kinase 2
VDR	vitamin D receptor
VNTR	variable number of tandem repeats
WTDIL	Wellcome Trust Diabetes and Inflammation Laboratory
ZnT8A	zinc transporter 8 autoantibodies

CONTENTS

Chapter 1

INTRODUCTION AND AIMS

1.1 EARLY GROWTH AND TYPE 1 DIABETES — HINT OF A LINK

Until the 1970s, the observation that diabetic boys were taller than average at diagnosis[1] guided research on the biological underpinnings of type 1 diabetes (T1D) for evidence of growth alterations at the time of clinical presentation. A prime example of this inquiry is Tatersall's and Pyke's[2] series of monozygotic twins discordant for T1D which revealed no major differences in weight or height at the time of diagnosis. A subsequent case report on monozygotic twin girls in whom diabetes developed in only one at the age of 3 years cast light on weight differences at birth and prior to disease, raising the question whether Tatersall and Pyke would have detected a correlation had weight at birth or during infancy been evaluated instead[3]. Indeed, in 1975 Baum et al.[4] promulgated that prediabetic children were significantly heavier at certain stages in the first 12 months of age compared with control infants, suggesting that 'an endocrine disorder precedes overt diabetes by a considerable time'.

How early does this critical window of aberrant physiology precede the diagnosis of T1D? Family studies broadly suggested that the disease process starts several years, even decades, before the abrupt clinical presentation[5]. Incidence data showed that T1D peaks in late childhood, leading to the speculation that early environmental exposures play an important role[6]. More recent studies reported a faster increase in T1D incidence in children with younger onset, under the age of 5 years[7], pointing to perinatal life as a critical window of foetal vulnerability that contributes to T1D[8-10]. Beyond incidence data, islet autoantibodies — the biomarker for the presymptomatic stage of islet autoimmunity — were detected from 6 months of age in prediabetic children[11,12] and in the cord blood of children who developed diabetes by the age of 15 years[13], collectively pointing to the early origins of the pathogenesis of T1D[14]. In search for the origins of the disease, observational studies found that accelerated early growth is a cardinal risk factor of childhood diabetes, compelled by evidence of a genetic link between T1D and growth, that is genes which confer susceptibility to T1D also affect early growth in the general population, exemplified by HLA[14,15] *INS*[16,17] and *DLK1*[18]. Conceivably, understanding changes of growth in early life might serve as a probe into the genetic, intrauterine and postnatal elements implicated in biological pathways that lead to the pathogenesis of T1D.

1.2 THE PATHOGENESIS OF T1D

1.2.1 Introduction

T1D is a disease most often diagnosed in children and adolescents following immune-mediated destruction of insulin-secreting pancreatic β-cells and presenting with the classical trio of symptoms (polydipsia, polyuria, polyphagia) alongside overt hyperglycaemia, necessitating exogenous insulin replacement[19]. Although the aetiology remains unclear, a trio of factors possibly drive the pathogenesis of the disorder (genetic susceptibility, environment, immune system), with 'firm' knowledge gains being most limited for the environmental facet[19].

1.2.2 Genetic susceptibility

The pathogenesis of T1D is complex and multigenic, and does not adhere to any simple pattern of inheritance[20]. The risk of T1D in siblings of patients is 15-fold higher than in the general population, pointing to a strong genetic element of disease susceptibility[21]. Family studies established that the major risk factor resides in the Human Leukocyte Antigen (HLA) region, also termed IDDM1, on chromosome 6p21, which explains half of disease heritability[22]. Fine mapping of the extended HLA region identified that the major genes conferring resistance and susceptibility to T1D map to HLA class II, specifically the DR and DQ genes[23]. The absolute risk of diabetes by the age of 15 years in children with the highest risk HLA genotypes (*DR3-DQ2/DR4-DQ8* or *DR4-DQ8/DR4-DQ8*) is 5%[24]. The prevailing view of the disease model, which revolves around the T-cell-mediated destruction of pancreatic β-cells, originates from animal studies[25]. HLA class II molecules on antigen presenting cells bind peptides from the islet autoantigens and present them to helper CD4+ T cell antigen receptors in the thymus and in the periphery (pancreatic lymph nodes and islets), allowing CD4+ T cells to help CD8+ cytotoxic T cells in killing β-cells[23]. It is also possible that genes might exert their function on immune cells by hindering the dampening effects of T regulatory (Treg) cells[26].

Next in rank of genetic susceptibility to T1D is the *INS* gene, with the Variable Number of Tandem Repeats (VNTR) class I and class III alleles conferring increased[27] and decreased[28] risk of developing T1D respectively. The mechanism by which the *INS* VNTR III allele has a dominant protective effect against T1D is likely via an increased production of insulin in the thymus during foetal and postnatal life (since foetal pancreatic insulin expression is marginally affected by allelic variation at this locus), which possibly induces negative selection of insulin-specific T-lymphocytes or better selection of Treg cells[29]. The susceptibility effects conferred by the *HLA* and *INS* genes are additive rather than interactive since stratification by *HLA* had no influence on the risk conferred by the *INS* gene[30].

Among first-degree relatives, siblings are at higher lifetime risk (8%) than offspring, and offspring of diabetic fathers are at higher risk (5%) than those of diabetic mothers (3%)[21,31]. Concordance rates in monozygotic twins is reportedly <50% and in dizygotic twins is 6%-10%[21,32]. The lack of concordance in monozygotic twins, who share identical genome, are probably attributed to somatic mutations, gene rearrangements by chance (e.g. in T cell receptors), or environmental factors[33].

1.2.3 Environmental influences

Inherited factors are determinants of an individual's risk of T1D, but the observation that >90% of subjects with an increased genetic risk remain autoantibody negative points to the critical interplay between genes and the environment[34]. In general, two types of mechanisms lead to T1D in genetically susceptible individuals[35]: an intrinsic mechanism (epigenetic) which affects gene expression[36] or an extrinsic trigger, such as an environmental influence (e.g. virus).

The effect of intrinsic factors might originate *in utero*, a critical window of 'switching on/off' genes. Recently, efforts have been directed at identifying temporal changes in gene expression in the first 2 years of life in at-risk children, identifying a genetic signature for seroconversion: near-birth expression of *ADCY9*, *PTCH1*, *MEX3B*, *IL15RA*, *ZNF714*, *TENM1*, and *PLEKHA5*, which are implicated in T-cell, B-cell and dendritic-cell immune responses[34].

On the epidemiological front, variation in T1D incidence has been observed amongst geographies with genetically proximal populations[7]. Migration studies have shown that the incidence increases as populations move from low-risk to high-risk areas[37]. For most European populations, a steady increase in the incidence of T1D has been reported[38], which is too rapid to be explained by shifts in the population gene pool[39]. The increasing incidence of the disease is composed of a temporal decline in the prevalence of high-risk HLA genotypes countered by an increase in low- to moderate-risk HLA genotypes[40-43]. These lines of evidence consistently point to a non-Mendelian risk for T1D modifiable by non-Mendelian factors[44,45], which are most likely concerned with changes in the environment rather than, or in addition to, stochastic epigenetic events.

Several prospective cohorts have investigated the environmental determinants of T1D from pregnancy through early life in infants at familial and/or HLA risk, including the multicentre studies The Environmental Determinants of Diabetes in the Young (TEDDY)[46] and the Trial in Reducing IDDM in Genetically at Risk (TRIGR)[47], as well as Australia's Environmental Determinants of Islet Autoimmunity (ENDIA)[48], Colorado's Diabetes Autoimmunity Study in the Young (DAISY)[49], Finland's Type 1 Diabetes Prediction and Prevention (DIPP)[50,51], and Germany's BABYDIAB[11].

The environmental component of susceptibility to T1D is poorly understand, but viral infection *in utero* or early in life is deemed a critical triggering event that insults the immune system and launches a selective autoimmune 'suicidal attack' on insulin-secreting cells[52,53]. This proposition is reinforced by experimental studies showing that viral infections caused diabetes in animals as a result of viral replication within β-cells[54]. Human studies concur on the strongest candidate diabetogenic pathogen being a member of the group of enteroviruses[37], namely coxsackie B virus, affirmed by various lines of evidence: i) a higher percentage of T1D patients were positive for coxsackie B virus-specific immunoglobulin M (IgM) responses vs. controls[55], ii) children who developed T1D were more often born to mothers who had infections with coxsackie or echovirus during pregnancy[56], iii) infections with coxsackie B virus occurred at a higher frequency in children with T1D[57] and iv) detection of an enterovirus infection in pancreatic tissue of recent onset T1D patients[58,59]. Not only are enteroviruses found in pancreatic islets of T1D patients, they demonstrate a clear tropism to β-cells based on *in vitro* studies[60]. Although research has predominantly focussed on the human enterovirus as the insult of innate immunity in T1D, recent studies in Sardinia identified that the *Mycobacterium avium* subspecies *paratuberculosis*, which is transmitted from infected cattle to humans by cow's milk, was strongly associated with T1D in both adults and children[61].

Enteroviral infections in early infancy have been suggested to enhance the development of immune responsiveness to insulin via the gut-associated lymphoid tissue, which serves as the common site of immunisation against insulin and replication of enteroviruses[62]. Early introduction of cow's milk had been suspected as a culprit for heightened risk of β-cell autoimmunity, which was thought to arise via a mechanism of exposure to complex foreign proteins. The speculation was given momentum by a study in Finland showing that infants exposed to cow's milk formula before 3 months of age carried higher immunoglobulin G (IgG) antibodies binding to bovine insulin — which also cross-reacted with human insulin — than exclusively breastfed infants, concluding that early introduction of dietary insulin induces primary immunisation to insulin that could develop into an autoimmune reaction against endogenous insulin secretion[63]. The hypothesis was recently tested by the double-blind TRIGR study which assessed whether weaning to hydrolysed casein formula — containing small protein fragments that are not recognisable by the immune system — vs. the conventional cow's milk-based formula — containing intact protein — reduces the risk of autoimmune diabetes in children with genetic and familial risk[64]. Contrary to the hypothesis, the study showed that the intervention had no effect on the progression or delay of autoimmunity[64], leaving open the possibility that the well-replicated association of infant feeding and T1D risk is mediated by some other factor, possibly accelerated growth.

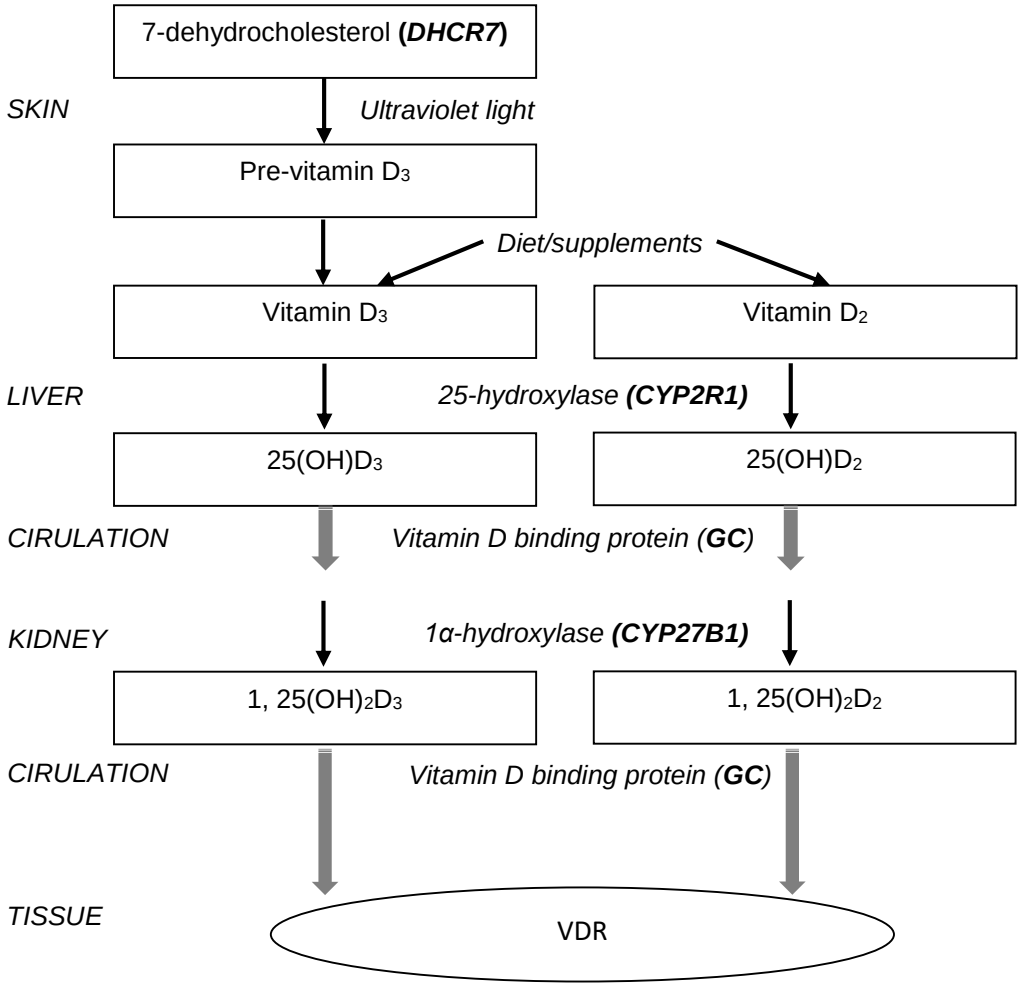
Emerging evidence suggests that the pathway between infant feeding and growth might involve vitamin D. It has been shown that exclusive breastfeeding of any duration is

associated with low serum vitamin D levels in children[65-67]. The finding informs that breastfeeding supplies insufficient amounts of vitamin D[67], which provides a justification for why these children are at increased risk of rickets[68,69], and could further explain the reduced growth rates associated with breastfeeding vs. formula-feeding. In addition to its well-established biological actions on growth via its effects on calcium metabolism in most cells of the body, vitamin D is intricately coupled with the immune system via its immunosuppressive properties[70]. This link was also shown by an experimental study which reported that long-term administration of the active form of vitamin D significantly reduced the incidence of spontaneous diabetes in non-obese diabetic mice[71].

Vitamin D is an anti-inflammatory fat-soluble steroid which, in addition to its classic effects on calcium and bone homeostasis, is believed to modulate the innate and adaptive immune responses by exerting immunomodulatory effects in a paracrine or autocrine manner, reinforced by the observation that immune cells (B cells, T cells and antigen-presenting cells) have the machinery to synthesise and respond to 1,25-hydroxyvitamin D[72]. Vitamin D, initially identified as a vitamin but now recognised as a prohormone, exists in two major forms which differ only in their side chain structure: vitamin D_2 (ergocalciferol) is sourced exogenously via animal-based foods or dietary supplements, whereas vitamin D_3 (cholecalciferol) is synthesised via the action of sunlight (ultraviolet B radiation) on 7-dehydrocholesterol in the epidermis, as well as being sourced exogenously[73]. Both D_2 and D_3 are biologically inactive until they undergo two sequential hydroxylation reactions catalysed by P450 enzymes: in the liver, 25-hydroxylase (CYP2R1) catalyses the formation of 25-hydroxyvitamin D (25(OH)D), which has a half-life of several weeks and circulates in the plasma mainly bound to a specific plasma carrier protein known as vitamin D binding protein (DBP); in the kidney, 25(OH)D is further hydroxylated by 1α-hydroxylase (CYP27B1) to the biologically active hormone 1,25-dihydroxyvitamin D ($1,25(OH)_2D$), known as calcitriol, which has a half-life of a few hours and also binds to DBP to be transported to target organs[73]. The biological actions of calcitriol (e.g. regulation of gene expression) are mediated through binding to the vitamin D receptor (VDR), which is found primarily in the nuclei of target cells[74]. Various polymorphisms of genes in the vitamin D metabolism milieu have been associated with risk of T1D[75-77] (Figure 1-1):

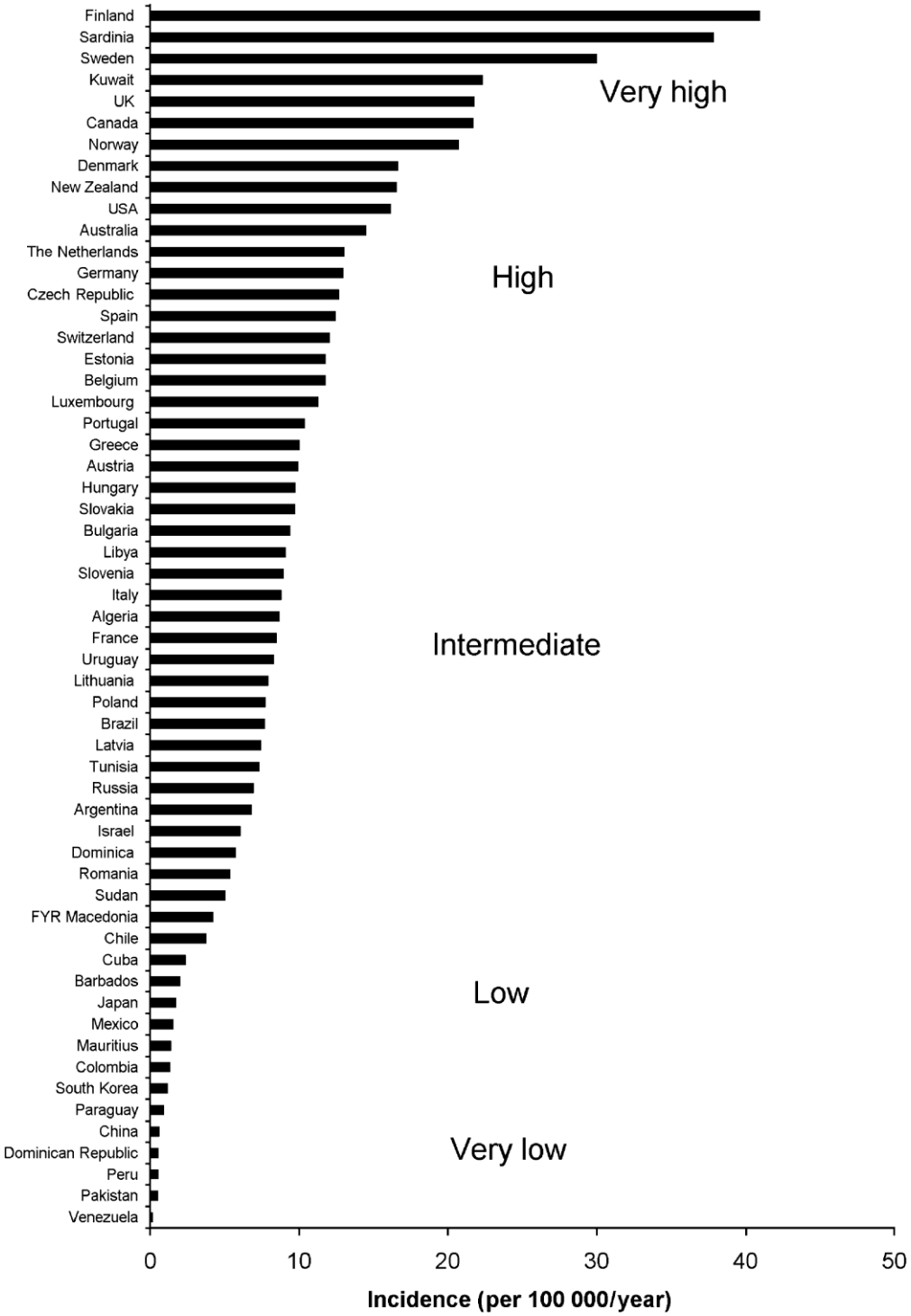
i. genes involved in dehydrocholesterol synthesis (*DHCR7*);
ii. genes involved in vitamin D transport (*GC*);
iii. genes that encode 25-hydroxylase (*CYP2R1)* and 1α-hydroxylase (*CYP27B1*).

Figure 1-1 | Vitamin D synthesis and transport pathway highlighting implicated T1D candidate genes.



Vitamin D and viral infections have been at the centre of two theories attempting to explain the seasonal pattern of T1D incidence. The association of vitamin D with T1D reconciles with the seasonality of diagnosis, higher in autumn and winter and lower in the summer[78,79], as well as the north-south geographic gradient that reflects an inverse correlation between the amount of sunshine and T1D incidence[80-82]. The north-south divide was initially explained in the context of the positive association between blood glucose and cold climate, since low temperatures increase the need for insulin[83]. However, the supposition of climatological factors fails to account for the exception of Sardinia where the incidence ranks second in the world behind Finland (Figure 1-2)[39]. The finding that climatological effects make a low contribution to the variance in T1D incidence rate suggests that other factors dominate[80].

Figure 1-2 | Global incidence of T1D. From the DIAMOND Project Group[39].



An alternative candidate mechanism was drawn from the seasonal correlation between infections and onset of T1D[84]. As long ago as 1969, Gamble and Taylor[85] reported that the trend of T1D incidence follows the coxsackie B virus epidemic, echoing the general view that diseases with a seasonal pattern are caused by infection. However, the evidence that injurious processes on β-cells precede overt diabetes by years casts doubt on this hypothesis[36]. Recent studies found no dramatic changes in the gut virome (stool sample) of T1D children[86], nor

higher viral load in the plasma of rapid progressors vs. control subjects during the period surrounding seroconversion[87]. The absence of viraemia shortly before seroconversion leaves open the possibility that viral infections occur early in life. Is T1D caused by a virus? A recent review makes the case that the link between coxsackie B virus infection and T1D is causal and explained by a combination of mechanisms: in the acute phase of the infection, the defensive immune response to the viral amplification in an attempt to destroy infected cells is deleterious for the β-cells, which have a limited capacity to proliferate and compensate for their loss; subsequently, progressive destruction of β-cells is driven by autoreactive T cells[88]. Notably, a study found that gestational infections increased the risk of higher birth weight, in association with T1D in the child, which was speculated to arise from either an increased insulin resistance that ensues from the infection, or a direct adipogenic effect of the pathogen on the foetus[89].

Central to the view that a viral infection is at the root of T1D is the role of the gut in the pathogenesis of T1D. Traditionally perceived as an organ whose function was limited to the digestion and absorption of nutrients and homeostasis of electrolytes, the gastrointestinal tract emerged as a barrier of the trafficking of macromolecules (proteins and viruses) between the environment and the host via the intestinal epithelium, alongside the gut-associated lymphoid tissue and neuroendocrine network[90]. In 2000 Fasano et al.[91] identified a novel protein, zonulin, which increases intestinal permeability by disassembling the intercellular tight junctions of the intestinal epithelium. Zonulin can be measured in the plasma and its serum levels were elevated in a subgroup of patients with T1D compared with their relatives and controls[92]. These findings fuelled the promulgation of the mechanism of a 'leaky gut'[93], which allows entry of antigens that, upon presentation to T lymphocytes, elicit an immune reaction.

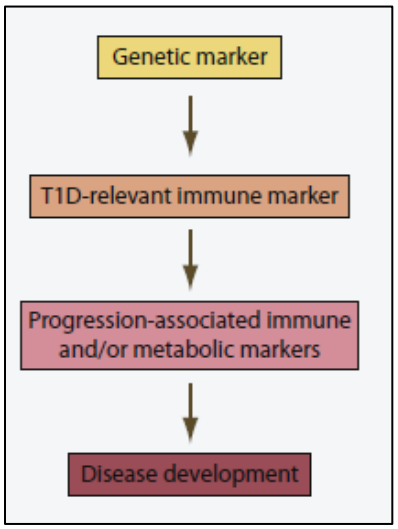
Activated intestinal immunity is another facet of the gut that is indicative of its involvement in the pathogenesis of T1D. Intestinal biopsies from children with T1D but free from Coeliac disease showed i) increased expression of HLA class II molecules and ii) greater density of α4β7-integrin cells in the lamina propria[93-95]. The α4β7-integrin, a gut-associated homing receptor of lymphocytes that directs them to mucosal tissue[96], is implicated in T1D based on genetic and functional data: it is encoded by the T1D susceptibility gene *ITGB7*[97], and studies showed that depletion of α4β7-integrin cells in peripheral blood taken from T1D patients was associated with a reduction in the reactivity of lymphocytes to islet antigens[98].

1.2.4 Autoimmunity

The binary outcome of progression or not to clinical T1D represents the endpoint of genetic-environmental interactions that result or not in the destruction of β-cells by infiltrating immune cells and the cytokines they release[99-101]. The immunological nature of T1D is manifested at each stage of the disease process (Figure 1-3): i) the genetic susceptibility is largely conferred

by loci shared by other autoimmune diseases[102], ii) the presence of islet autoantibodies in the serum of prediabetic children during the disease prodrome[19] and iii) the development of insulitis — the inflammatory infiltration of pancreatic islet cells by T lymphocytes (CD8+ cells outnumber CD4+ cells) and fewer counts of other cells of immune response — at around the time of clinical presentation[103], which is reinforced by the experimental finding that the immunomodulatory products (cytokines) of lymphocytes and other cells of the immune system directly mediate synergistic cytotoxic effects on pancreatic islet cells[104].

Figure 1-3 | Main stages of T1D progression. Adapted from Ziegler and Nepom[105].



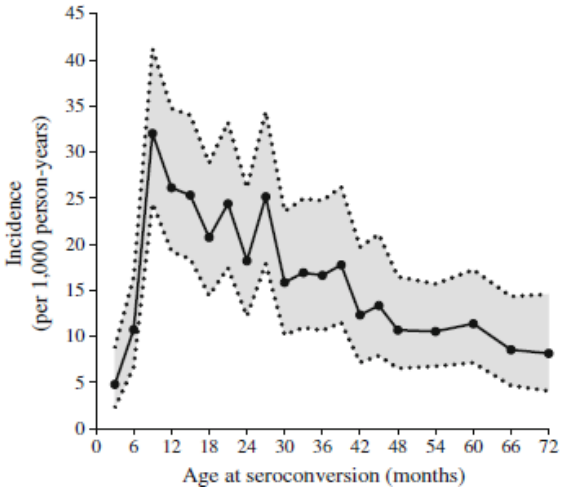
The temporal progression to islet autoimmunity and, thereafter, to T1D is characterised by distinct immunoinflammatory markers. The DAISY study reported that progression to islet autoimmunity in early childhood was predicted by serum C-reactive protein levels[106]. In turn, persistent islet autoimmunity is highly predictive of T1D among first-degree relatives[107], so that it has been under intense investigation and used as an anchor for the recent staging classification of T1D: the initial presymptomatic stage of β-cell autoimmunity is judged by the presence of two or more islet autoantibodies (stage 1), which progresses at variable rates to the development of dysglycaemia (stage 2) and culminates in the onset of symptomatic disease (stage 3)[108].

Autoantibodies are useful for detecting developing T1D as they provide a readout of β-cell autoimmunity and are easily sampled in venous blood[109]. Previously described in terms of islet cell antibodies (ICA) immunofluorescence assay on pancreatic sections, autoantibodies are now described in terms of defined ICA target antigens (e.g. insulin)[109]. Standardised immunohistochemical assays have been established as diagnostic immune markers for 'islet autoantibody positivity' to represent an estimate of T1D risk, a combined measure of the probability of disease development and the rate at which it develops, i.e. average probability[105].

More than 90% of children with newly diagnosed T1D have at least one autoantibody against the islet cell antigen groups regarded as predictive biomarkers for T1D[110]: insulin (IAA), glutamic acid decarboxylase 65 (GADA), insulinoma-associated antigen-2 (IA-2A) which is otherwise known as ICA512, and the three variants of the zinc transporter 8 (ZnT8A) protein that were later incorporated because they increase the sensitivity of detection alongside the three primary autoantibodies[111]. The presence of islet autoimmunity is considered when the individual has one or more autoantibodies persistent for at least 3-6 months[33].

Unless acquired via transplacental transfer, islet autoantibodies most commonly develop from the age of 6 months[11,12] with a peak incidence around 1-2 years of age[112,113]. The median age at seroconversion was estimated as 2.1 years based on data pooled from three large international cohorts[114]. The multi-centre TEDDY study has recently shown that the incidence of islet autoantibodies rose sharply until 9 months of age and declined thereafter (Figure 1-4)[115]. It is, therefore, apparent that factors that influence the immune response to β-cells in the first year of life are important for programming the initiation of islet autoimmunity[81,116], with the implication that the period of escalating autoimmunity offers a window for interventions to halt progression to T1D[117].

Figure 1-4 | Incidence of antibodies in children (0 to 6 yr) in the TEDDY study by age of seroconversion. From Krischer et al.[115].



The risk of progression from islet autoimmunity to clinical disease is multifactorial; it varies in relation with the type of autoantibody present (IA-2A confers the highest risk), and is incremental with the number of detected autoantibodies (risk is increased with antibodies to two or more antigen groups), higher titre, and broadness of epitope reactivity[105]. Most at-risk children with multiple autoantibody seroconversion will progress to diabetes within 15 years[114]. In the offspring of parents with T1D, disease progression was accelerated by an earlier age of appearance of multiple autoantibodies and was fastest in children who presented with the first autoantibody by the age of 2 years[118]. Rapid progression was defined, based on observations

in young and early pubertal children of the DIPP cohort, as progression to clinical T1D within 1.5 years of autoantibody seroconversion, and characterised by younger age, higher autoantibody titres and multiple islet autoantibodies[119]. The median time from seroconversion to diagnosis was 0.5 years in rapid progressors vs. 5.4 years in slow progressors[119].

The order of autoantibody appearance influences the degree of pathological consequences: the appearance of IAA first may trigger GADA and IA-2A as subsequent autoantibodies leading to an aggressive pathogenesis, i.e. insulitis and earlier age at onset; whereas GADA, more common to children ≥2 years of age, is followed by a delayed appearance of ensuing autoantibodies and results in a less aggressive pathogenesis[120]. The constitution of autoantibodies is not static but shows a dynamic interplay with disease progression. The DIPP study unveiled that in the presymptomatic stage, IAA were the most frequent primary antibodies followed by GADA, IA-2A and ZnT8A; at diagnosis, IA-2A was the most frequent, owing to the appearance of secondary antibodies and disappearance of IAA[121].

Whereas HLA class II influence the initiation of autoimmunity, the rate of progression to clinical disease is a class I effect[122], exemplified by HLA-A*24, -B*18, -B*39 which are linked to more rapid progression[123,124]. Based on these lines of evidence, it is expected that appearance of islet autoantibodies is intricately coupled with HLA types. This field is under ongoing investigation, but the TEDDY study has shown that the *DR4-DQ8/DR4-DQ8* is predominantly linked to the appearance of IAA as the first autoantibody in young children and the *DR3-DQ2/DR3-DQ2* is strongly linked to the appearance of GADA as the first autoantibody in older children[115]. However, these associations are perplexed by intrauterine interactions; the associations were not observed in the presence of a gestational respiratory infection, pointing to a bidirectional trigger for appearance of IAA or GADA as the first autoantibody[120].

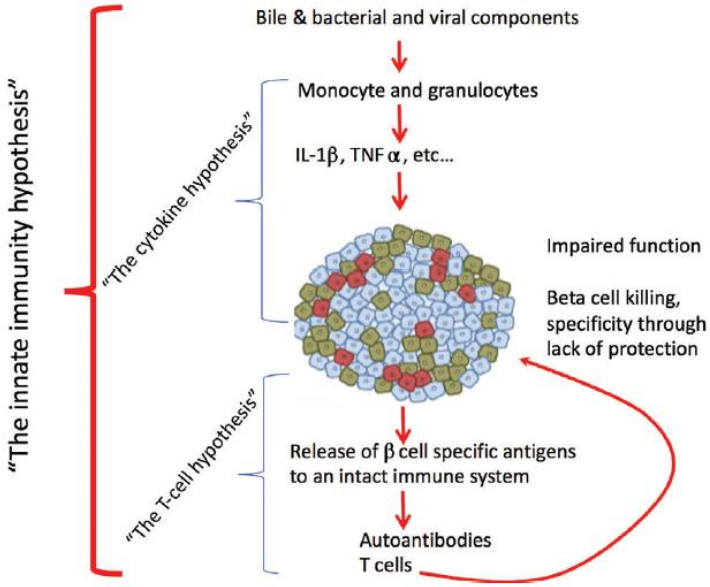
Mechanistically, autoantibodies and HLA types are linked via the proposed scenario that the autoantigen is processed and presented to HLA class II molecules on antigen-presenting cells (macrophage, dendritic cell, B-cell); for example, insulin peptides are presented on DR4-DQ8 and GADA peptides on DR3-DQ2[120]. The ensuing complex would be recognised by the T-cell receptor (TCR) on $CD4^+$ T cells, which activate the $CD8^+$ cytotoxic cells that infiltrate pancreatic islets and recognise antigens presented by HLA class I molecules[120,122]. Insights into the mechanisms of events that occur in the islets thereafter are drawn from findings that progressors to T1D possess a distinct molecular signature vs. non-progressors, which are characterised by crescendos in induced proinflammatory transcription and a reduction in induced regulatory transcription[125]. These molecular profiles align with the proposition that a combination of immune-mediated mechanisms might bring about β-cell death, which probably comprise: i) production of proinflammatory cytokines and chemokines, ii) decline of the immune regulatory function, iii) activation of reactive T cells against β-cells[126].

1.2.5 Genetic — Environmental — Immune link

The genetic, environmental and immune elements that orchestrate the pathogenesis of T1D do not operate in isolation. Consideration of their intricate act in tandem provides a glimpse into the complex biological pathways at play. Nevertheless, our evidence-based understanding of the multitude of pathways that lead to complete loss of insulin secretion remains poor and is tackled by the promulgation of hypotheses that make speculative links.

There is compelling evidence for the involvement of proinflammatory cytokines and autoreactive T cells in the injurious processes of the islets, yet there is no clear explanation for the reasons why they accumulate in the vicinity of the islets[127]. A possible scenario is proposed by the 'innate immunity hypothesis', according to which dying or dead bacteria or viruses in the pancreatic juice and bile mixture[128] become entrapped within the pancreatic ducts and release toxins and cell components that trigger a proinflammatory microenvironment, which drive antigen-presenting cells to elicit islet autoimmunity[129] and recruit autoreactive T cells[5], resulting in an immune-mediated attack on β-cells (Figure 1-5)[127].

Figure 1-5 | A tentative model for the initiation of T1D. From Skog and Korsgren[127].



The interface between the immune system and the islets was epitomised by the pancreatic islet transcriptome project, which revealed that >60% of T1D susceptibility genes — previously believed to be mostly expressed in immune cells — are also expressed in human pancreatic islets[101]. The experiment simulated conditions that possibly prevail in early T1D by exposing the islets to proinflammatory cytokines[101]. As a testimony, the chronic effect of proinflammatory cytokines — interferon-α (IFN-α), interferon-β (IFN-β), interferon-γ (IFN-γ),

tumour necrosis-α (TNF-α), interleukin-1β (IL-1β) — on insulin-secreting cells is evidenced by the clinical observation that long-term treatment with IFN-α of patients with hepatitis C increased the risk of T1D by 10- to 18-fold[130].

The islet transcriptome project made two critical findings: i) cytokines modify alternative splicing in islets (thus generating novel RNAs which translate to novel proteins recognised by the immune system); and ii) induce a local release of cytokines and chemokines (proteins that attract immune cells), allowing for a 'molecular dialogue' between the β-cell and the immune system that may amplify or dampen insulitis and determine development of diabetes or not[101]. These results give credence to the thesis postulated in the 1980s that the β-cell 'contributes to its own demise'[104], and override the conventional view which regards 'β-cells as passive victims of a process that starts and is regulated elsewhere'; instead β-cells deliver immunogenic signals and release inflammatory mediators that affect their own survival[101].

Even though the expression analyses were performed on whole islets, it is speculated that T1D candidate genes are expressed in β-cells, or their products affect β-cells in a paracrine fashion if expressed in neighbouring islet cells, paralleled by the glucagon-like peptide 1 (GLP-1), which is synthesised and released by α-cells but exerts cytoprotective effects on β-cells[26]. The question that persists is: why are the neighbouring α-cells spared from autoimmune attack? The resistance of α-cells to autoimmune-mediated destruction is best explained in the framework of the infectious basis of the disease: the α-cells might possess the machinery of viral amplification but show a more efficient antiviral response compared with β-cells, i.e. they are better at clearing infections, evidenced by the presence of viral markers in α-cells during acute but not chronic infections[88].

Interestingly, the outcome of an infection is dependent on both the pathogen and the host response. The latter is modulated by the individual's genetic background and/or early education of the immune system, a concept which is encapsulated in the 'hygiene hypothesis' that attributes the increased risk of allergic and autoimmune diseases to a reduced microbial exposure in early life[131]. On these grounds, Op de Beeck and Eizirik[88] suggested that divergent environmental factors account for T1D in different parts of the world; in populations with high genetic risk, removal of a protective factor (e.g. non-specific infections in early life) is sufficient to increase disease risk, whereas in populations with low genetic risk both protection removal and increased infections by coxsackie viruses are required for increased incidence of T1D[88]. The hygiene hypothesis resonates with the paradox of the inverse relationship observed between the incidence of enterovirus infections and incidence of T1D[132], which spawned the 'polio hypothesis' — in reflection of the 19th century decreased polio circulation accompanied by an increase in polio paralysis in children due to lack of protection from maternal poliovirus antibodies[133] — to suggest that shifts in population dynamics of virus infections account for

trends in the incidence of T1D[134]. The polio hypothesis is further underpinned by the observation that the levels of enterovirus antibodies in pregnant mothers decreased over the last three decades in Finland contrary to an increase in T1D incidence in this country[132].

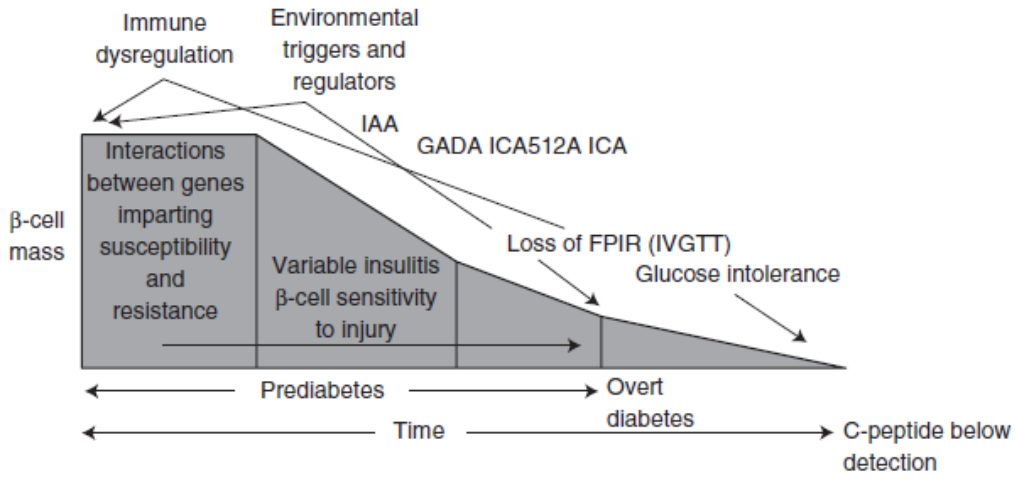
The credibility of this hypothesis lies in the possibility that coxsackie B virus, a member of the group of enteroviruses, is central to the pathogenesis of T1D. Coxsackie B virus is thought to contribute to the initiation of insulin-driven autoimmunity based on evidence that infections with this virus associate with appearance of IAA[135]. Aside infections, feeding in infancy is the alternative environmental culprit of the pathogenesis of T1D, with recent evidence linking early introduction of gluten (before 4 months of age) with reduced risk of islet autoimmunity in children at risk[136]. Findings pertaining to elements of coeliac disease should probably be interpreted with cation, owing to the genetic overlap between coeliac disease and T1D via the HLA DQ2 and DQ8. Perhaps, of central relevance to T1D is the capacity of surplus glucose to elicit an immune response. Skog and Korsgren[127] point to a collection of experimental studies that concur on exposure to high glucose inducing immune-cell activation in CD4+ and CD8+ T cells, and in monocytes[137-141]. I posit that these lines of evidence, i.e. the possible effect of coxsackie B virus on triggering insulin-driven autoimmunity which leads to glucose excess that might further elicit an immune response, collectively close the loop between systemic viral infection and islet inflammation, and suggest that synergistic factors bring about the immune-mediated impaired insulin release — the hallmark of T1D.

1.2.6 The endocrine microenvironment

The interplay between genetics, immunoinflammatory and environmental aspects is encapsulated in the modern model of the pathogenesis and natural history of T1D (Figure 1-6). The slope depicts the rate of β-cell loss in the prediabetic period leading to symptomatic T1D, which according to the traditional model presents when >80%-90% of β-cells have been destroyed, but recently revised to suggest that up to 40% of β-cell viability might be present[142], which probably explains why insulin secretion is stable for periods of time in some patients despite persistent islet autoimmunity[19]. However, it is accepted that a loss in the first phase of insulin response (FPIR), which represents early insulin secretion and is measured by intravenous glucose tolerance test (IVGTT) or oral glucose tolerance test (OGTT), is followed by a period of glucose excursions (glucose intolerance)[19]. C-peptide is a standard method of assessing the insulin secretory capacity (owing to its slower degradation rate than that of insulin), with a concentration of less than 0.2 nmol/L predictive of poor β-cell reserve and insulin requirement[143]. In the early stages of the disease some patients experience ‘the honeymoon period’, a transient independence from insulin therapy, which exhibits morphological and functional signs of a returning insulin production, possibly in an attempt of the reduced β-cell

mass to meet the insulin demand[144]. Nevertheless, within months of diagnosis, there is a marked reduction of almost 30% of the total pancreas volume[145].

Figure 1-6 | Natural history of T1D. From Atkinson[19].



Insulin is stored in its active form inside pancreatic islet secretory vesicles and released into the circulation upon stimulation by increased plasma glucose or amino acid levels[146]. The FPIR, the Ca^{2+} influx-triggered steep secretory peak of insulin generated by a small pool of readily releasable granules[147], is perturbed as early as 6 years prior to the onset of T1D, whereas insulin sensitivity remains unaffected, 'implying an intrinsic defect in β-cell mass and/or function'[148]. In children of the DIPP cohort, the FPIR inversely correlated with multiple autoantibodies, but not directly correlated with HLA-risk genotypes[149].

The lingering question about the reasons for the vulnerability of the β-cell has also been approached through the lens of the severe fall in β-cell mass. It has been remarked that the large variation in β-cell mass (0.2 to 1.5 g) in non-diabetic adults is suggestive of subclinical assaults of the islets, which might be explained by: i) the low replication rate of β-cell, rendering it unable to compensate for massive loss[150], and ii) the high islet blood perfusion, which provides optimal conditions for immune surveillance[127]. The observation that hyperactive β-cells have a heightened expression of islet autoantigens[151] strengthens the possibility that β-cell stress, as a result of increased insulin demands during rapid growth, might be the culprit.

In conclusion, two polarising views prevail around the pathogenesis of T1D: one scenario claims that it is a consequence of a severe fall in β-cell mass resulting directly from T-cell mediated attack (that might ensue from a viral infection); another view posits that glucose intolerance and increasing insulin demands in a growing child are not adequately countered by an increase in β-cell volume resulting in major loss of β-cell mass or insulin secretory capacity[127]. On the premise that 'genetics unlock the biology of a disease', delving into the genetic architecture of T1D is essential to shed light on the debate.

1.3 THE GENETIC ARCHITECTURE OF T1D

1.3.1 Genetic studies

The genetic architecture of T1D has been extensively unveiled by genome-wide association studies (GWAS) in large patient and control cohorts, which expanded the set of 6 loci identified by family (linkage studies) and candidate gene (association studies) approaches in the 1970s, to 40 loci by 2006, and 60 loci by 2012[152]. In the pre-GWAS era, the HLA was the first marker identified, which explained 50% of genetic risk[22]. *INS* mapped on chromosome 11p15 was the second locus established with an Odds Ratio (OR) of ~2.5, followed in chronological order by the identification of cytotoxic T-lymphocyte-associated protein 4 (*CTLA-4*); the protein tyrosine phosphatase non-receptor type 22 (*PTPN22*); the interleukin 2 receptor alpha (*IL2RA*); and the interferon-induced with helicase C domain 1 (*IFIH1*)[152]. According to the current state of knowledge, susceptibility to T1D is conferred by a set of 59 chromosomal regions (www.immunobase.org, accessed 2015).

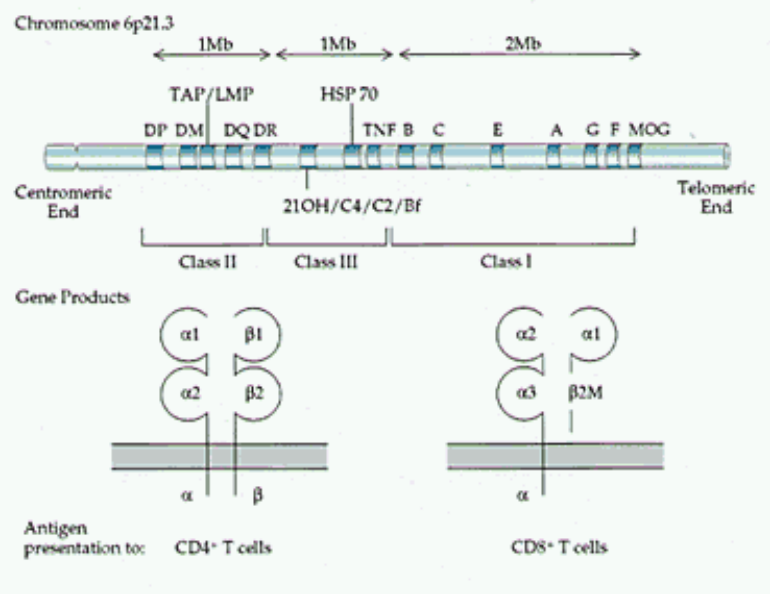
The fundamental principle of GWAS is the measurement and analysis of DNA sequence variations in the form of single nucleotide polymorphisms (SNPs) from across the human genome[153], which are mapped to candidate disease causal genes for the ultimate purpose of identifying genetic risk factors and elucidating biological pathways. Albeit increasing the pace and efficiency of the discovery of T1D susceptibility loci by a factor of ten[152], GWAS pose two inherent challenges. First, a particular locus might contain multiple genes, as is often the case for T1D, and GWAS vaguely suggest potential causative genes based on interpretations that rely on either physical proximity to the genetic marker or biological relevance[26], thus discounting that SNPs may affect expression of remote genes located thousands of base pairs (bp) away by modulation of chromatin-loop formation[154]. Second, a key characteristic of GWAS is linkage disequilibrium (LD) — a property related to a contiguous stretch of genomic sequence and defined as the degree to which one SNP is correlated with another SNP — underpinned by the concept of 'chromosomal linkage' whereby two markers remain physically joined on a chromosome through generations[153]. LD is usually quantified by r^2, a statistical measure of correlation, which denotes the extent of similarity captured by two SNPs (high r^2 values suggest high degree of similarity). LD is exploited by GWAS via the selection of tag SNPs, which capture the genetic variation in the surrounding stretch, thus obviating the need to genotype SNPs that convey redundant information[153]. The drawback of this approach is the further complexity added to the identification of causal SNPs[26].

1.3.2 The prominence of the HLA

The HLA is part of the genetic region known as Major Histocompatibility Complex (MHC), a long contiguous DNA stretch of approximately 3.5 million bp, which was sequenced and analysed with extremely high quality in the late 1990s[155]. The MHC is located on Chr6p21 and is renowned for extensive linkage[156,157]. Aside its well-known role in the regulation of self- vs. nonself-recognition, which is integral to antigen presentation and initiating an immune response, genes in the MHC region influence a variety of reproductive parameters including protection of the foetus from attack by the maternal immune system[158], abortions[159], preeclampsia[160] and size at birth[14,15,89,161].

Within this gene-dense region — counting over 220 genes that encode homologous cell-surface proteins which are the most polymorphic known to date — the HLA class I or telomeric region (A, B, Cw) and HLA class II or centromeric region (DR, DQ, DP) each spread over about one third of the MHC stretch (Figure 1-7)[155]. Class I molecules are found on all nucleated cells and present antigens originating from inside the cell to cytotoxic T lymphocytes ($CD8^+$); class II molecules are only found on immune competent cells (B cells, antigen-presenting cells) and present antigens that enter the cell by endocytosis to helper T cells ($CD4^+$) which set up a generalised immune response[162].

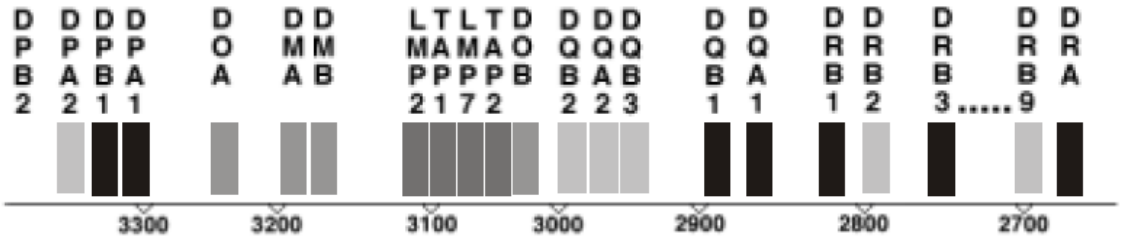
Figure 1-7 | Overview of the HLA region. From the EBMT HLA manual[163].



In T1D, the critical interaction between the HLA class II molecule, antigen and T-cell receptor evokes T-cell activation and ensuing immune response to the antigen[164]. It follows that the extent of immune response in T1D is partly a function of the polymorphic amino acids of the HLA class II molecules[165]. The class II molecules are heterodimeric glycoproteins, consisting of two transmembrane polypeptides, α-chain and β-chain. Within the class II

complex, the DR3 and DR4 alleles have a major contribution to T1D[166,167], in conjunction with the strongly-linked DQ molecule whose function in susceptibility to T1D is determined by the amino acid residue at position 57 of the DQβ chain[165]. It has been shown that Asp-57-negative homozygosity on DQβ alleles, especially if they are on the DR4 and/or DR3 haplotypes, is essential but not sufficient for development of T1D in 90% of Caucasians; in contrast, the possession of Asp-57-positive DQβ alleles confers almost complete resistance to T1D[165]. For instance, Alanine 57 — encoded by the DQB1*03:02 allele on DR4 haplotypes — is strongly predisposing, whereas Aspartic acid 57 — encoded by the DQB1*06:02 allele on DRB1*15:01 haplotype — is strongly protective[168]. In summary, the extent of susceptibility or resistance to T1D is best determined by three genes, *HLA-DRB1*, *-DQA1* and *-DQB1* (Figure 1-8), which show the greatest allelic diversity in exon 2 and encode the outer domain of DRβ, DQα and DQβ polypeptide chains respectively[30,168,169]. It follows that the variable DR-DQ haplotype associates with a spectrum of risk for T1D, ranging from high-risk, to neutral, to protective[21].

Figure 1-8 | The HLA class II gene complex Adapted from Shankarkumar[170].



The mode of inheritance of MHC genes is believed to be simple recessive[36]. In Caucasians, the haplotypes DR4 (containing DRB1*04 alleles, except for 04:03 and 04:06, and carrying DQB1*03:02) and DR3 (containing DRB1*03:01) are the most T1D–predisposing haplotypes, whereas the DR2 haplotype is strongly protective (found in <1% of T1D children and >20% in the general population)[171]. Specifically, the highly conserved DRB1*04-DQA1*03:01-DQB1*03:02 (abbreviated DR4-DQ8) and DRB1*03-DQA1*05:01-DQB1*02:01 (abbreviated DR3-DQ2) are the strongest risk alleles for T1D[30,168], whereas the DQA1*01:02-DQB1*06:02 confers dominant protection even in the presence of islet autoantibodies[172].

The complexity of the IDDM1 region is aggravated by the dramatic effect of the DR4 haplotype on disease susceptibility even if the DQ molecule is fixed; for example the strong risk allele DQB1*03:02 (DQ8) on subtype DRB1*04:05 loses its predisposing effect when found on the protective DRB1*04:03 instead[168]. This illustrates that the effect of a single HLA molecule is mitigated or augmented by its HLA context[168]. Furthermore, in estimating individual T1D risk it is essential to consider the identity of the HLA alleles on the other chromosome[168]. As such, children with the heterozygote DR3/DR4 genotype have the highest risk and comprise almost 50% of children who develop islet autoimmunity by 5 years of age[21]. Of interest, the high-risk DR3/DR4 was found to be absent from individuals with childhood T1D

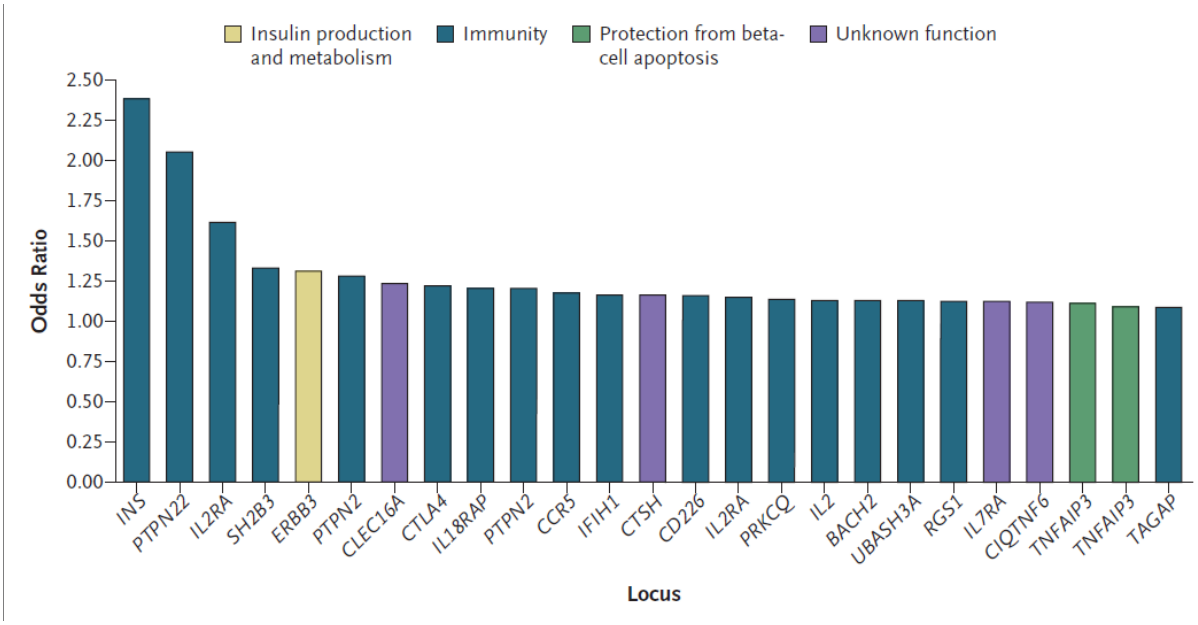
who carried a polymorphism in *TCF7L2*, a well-established type 2 diabetes (T2D) gene, highlighting the genetic heterogeneity of juvenile diabetes[173].

The rate of concordance of T1D in HLA-identical siblings (12% to 16%) is sufficient evidence that HLA is a T1D susceptibility locus, but the comparatively higher rate of concordance amongst monozygotic twins (40% to 50%) suggests that non-HLA loci are required for T1D to develop[36]. In comparison, the rate of concordance in dizygotic twins falls abruptly (5% to 10%) but is 10-fold higher than the prevalence in the general population (0.2% to 0.4%)[174]. The decline in risk by degree of relatedness indicates that a combination of alleles at multiple loci with small effects contribute to the disease[175].

1.3.3 Beyond the HLA

Beyond the HLA region, the extent of genetic contributions to T1D by individual genes follows the law of diminishing returns[19] as illustrated by the decaying OR of respective genes (Figure 1-9). Pociot et al.[176] stated that 'the principal value of the newly discovered SNPs would be to increase our understanding of the disease pathogenesis, rather than increase our ability to predict disease development on an individual level'. The fact that most T1D candidate genes are shared with other autoimmune diseases[102] underscores the immune nature of the disease.

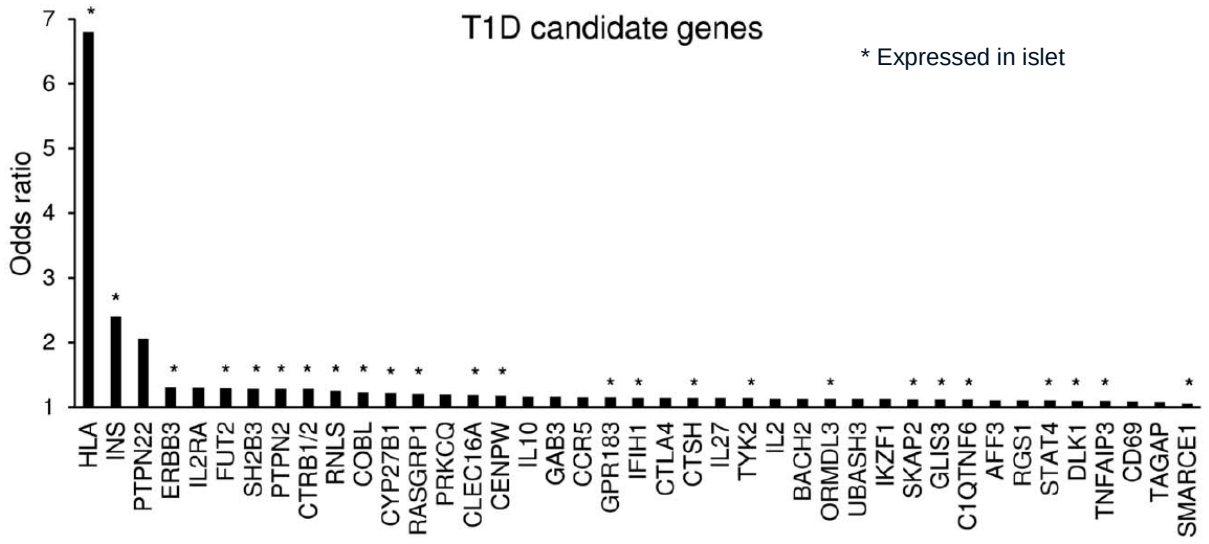
Figure 1-9 | Ranking of effect size of non-HLA T1D susceptibility genes. From Concannon et al.[117].



It has been suggested that whereas the HLA influences the development of autoimmunity, other T1D candidate genes regulate the anatomical location of the autoimmune attack and the rate of progression to disease[88]. This proposition is given momentum by the association of the high-risk DR4-DQ8 and DR3-DQ2 haplotypes with autoimmunity across the

board, which is attributed to the evolutionary selected ability of these proteins to present a range of pathogen-derived peptides to T cells[177]. The observation that families with a history of T1D are characterised by an increased innate inflammatory state independent of HLA status provides strong evidence that non-MHC genetic variants also potentiate immune pathways[125]. Specifically, the expression of numerous non-MHC genes in pancreatic islets (Figure 1-10)[101] informs of complex biological pathways in both immune cells and β-cells, thus cementing that T1D is a disease of the pancreas as much as of the immune system[26].

Figure 1-10 | Islet expression or not of T1D candidate genes ranked by OR. From Eizirik et al.[101].



Clearly, the genetic architecture sheds light on the biology of the disease. Yet, the identification of causal genes represents a major undertaking in the genetics of T1D[176]. This challenge necessitates the interrogation of candidate genes by interweaving the GWAS output with biological knowledge. To illustrate the point, the speculation that *CD45* or *PTPRC* — a transmembrane protein tyrosine phosphatase receptor type — is a T1D candidate gene, since it is expressed on all leukocytes and its phosphatase activity is essential for T cell activation, was invalidated for T1D[178] and other autoimmune diseases, including Hashimoto's disease[179] and myasthenia gravis[180]. Another candidate gene whose relevance to T1D is disputed is *ADAD1* on Chr4 on the grounds of its testis-specific expression and function of its product being involved in RNA packaging, alternative splicing and transport of mRNAs[181]. Last but not least, the *RNLS* gene — which encodes renalase, a protein that is secreted from the kidney and has a key role in cardiac function and systolic blood pressure — is questionably relevant to T1D due to a lack of functional data demonstrating its role in the immune system[102].

Population studies are instrumental for illuminating key candidate genes. I propose that the four genetic variants in *PTPN22, SH2B3, ERBB3* and *INS*, which were associated with

progression from seroconversion to islet autoimmunity in the TEDDY cohort[182] are beginning to paint out a biologically meaningful picture of possible mechanisms in the disease prodrome that involve genes i) with a key role in adaptive immunity (*PTPN22*), or ii) expressed in the anatomical location of autoimmunity (*INS*, *SH2B3*, *ERBB3)*. In the aggregate, the genetic architecture of T1D includes genes that play a key role in systemic immunity (e.g. *PTPN22*, *UBASH3A*). Predominantly, it comprises genes which exert effects on β-cells relating to antigen presentation, inflammation, antiviral response, innate immunity, apoptosis, IFN-γ signalling, nuclear factor kB (NF-κB) regulation, and modulation of chemokines/cytokines which influence a range of β-cell functions (glucose and lipid metabolism, protein synthesis, transcription factors)[101].

1.4 SIZE AT BIRTH

1.4.1 Overview

The β-cell and its rate of destruction possibly underpin the link between perinatal growth and childhood diabetes, which has spawned a large literature on the association between size at birth and risk of T1D. The minimum neonatal anthropometrics recorded at birth are weight, length and head circumference[183], which are introduced here.

1.4.2 Birth weight

Foetal growth rate serves as a sensitive indicator of a newborn's ability to survive and thrive in life[184] and is crudely represented by size at birth, which is most simply measured by birth weight[185]. Naturally, birth weight features amongst the most widely-studied variables in epidemiological research, owing to its accessibility (it is precisely measured and readily available), objectivity, and its association with perinatal survival and health outcomes later in life[186]. Low birth weight is a barometer of survival in childhood; it correlates with risk of infectious disease mortality[187,188] and hospitalisation[189]. Prominently, birth weight is emerging as a rough predictor of adult disease. The association between birth weight and risk of disease in adulthood was epitomised by the 'thrifty phenotype hypothesis' by Barker and Hales[190], who postulated that factors operating *in utero*, such as poor nutrition, could program the foetus at a time of organ plasticity and alter development in a way that insidiously leads to adult disease. The concept of foetal programming referred to events *in utero* that induced changes in metabolic characteristics, namely reduced development of pancreatic β-cell, which played a key role in the origins of T2D and cardiovascular disease in adulthood[191]. The thrifty phenotype

hypothesis echoes the idea previously explored by Kermack et al.[192], according to which the long-term effects of health shocks might remain latent, with the implication that foetal shocks induce later-life impacts[193]. Mechanistically, foetal programming might operate via epigenetic factors — interindividual variation in DNA methylation and chromatin remodelling that induce genes to be expressed or not — whose malleability allows them to react to environmental signals, critically during the period that the foetus is *in utero*[193,194].

Birth weight is sensitive to gestational age and sex. Nevertheless, the earlier literature on perinatal growth and T1D omitted gestational age in the calculation of birth weight[195,196]. This practice probably stemmed from the interchangeable use of 'premature birth' (<37 weeks of gestation) and low birth weight (<2,500 g) until the 1960s — a distinction made possible by the introduction of the terms 'preterm' to denote a baby born early and 'intrauterine growth retardation' (IUGR) to describe a disease that starts during pregnancy (antenatally) and attaches a condition to small babies that are growth retarded and could be born at term[186]. This differentiation required the introduction of 'birth weight for gestational age', which defines normal birth weight by applying percentiles on a reference population. Under the new conventions of birth anthropometry adopted by neonatal and obstetric research, 'small for gestational age' (SGA) and 'large for gestational age' (LGA) are commonly defined as <10th and >90th percentile birth weight respectively in each gestational age and sex stratum, with the in-between spectrum corresponding to 'appropriate for gestational age' (AGA). Even though the terms IUGR and SGA have been used interchangeable, the Royal College of Obstetricians and Gynaecologists states that they are not synonymous: 'foetal growth restriction' is defined as the 'pathologic restriction of genetic growth potential (…) that may manifest evidence of foetal compromise', whereas SGA foetuses are constitutionally small, i.e. normal in size, with foetal growth appropriate for maternal size and ethnicity[197]. IUGR implies slow growth *in utero* that is evidenced by two ultrasound measurements and necessitates ongoing surveillance of the infant irrespective of birth weight[198].

The comparison of a child's weight against the distribution of the weight of a reference sample of healthy children informs of the 'normality or otherwise of the process of growth'[199]. The availability of more than one comparator calls for a distinction between growth references vs. growth standards: growth reference charts are based on cross-sectional data and are descriptive of growth 'as is'; growth standards are based on longitudinal data in which the source sample has been selected based on predefined criteria and reflect growth 'as it ought to be'[199,200]. The growth charts commonly used for the assessment of infants in the UK are:

i. the British 1990 growth *reference* (UK90) based on 37,700 children from England, Scotland and Wales[201], who were not exclusively breastfed in their majority for the recommended duration of 4 months[199];

ii. the WHO 2006 international growth *standard*, which illustrates optimal physiological growth for all children from birth to 5 years of age based on information from approximately 8,500 breastfed children pooled from six countries (Brazil, Ghana, India, Norway, Oman and the USA)[202].

Growth charts are expressed in centile lines with the distance between two lines corresponding to ~2/3 of SD[200].

Relative birth weight is represented by standard deviation (SD) scores (SDS) or *z*-scores, calculated using the subject's actual age in days. The SDS is defined as the number of standard deviations that a baby's anthropometric measures deviate from the mean of the reference population; a SDS of zero denotes a growth trajectory similar with the reference, whilst SDS>0 suggests accelerated (higher than normal) gains and vice versa. However, the use of standards to produce relative values for comparison comes at the cost of narrowing the variation of birth weight[203].

The concurrent use of both definitions of birth weight, absolute and relative, in assessing its association with risk of T1D has cast light on confounding factors. It has been observed that studies reported higher rates of T1D for children weighing ≥80th percentile vs. children with absolute birth weight ≥4,000 g[204]. The largest-to-date study in 3.6 million children found no associations between T1D and birth weight when unadjusted except for a protective effect of very low birth weight (<1,500 g) against normal weight[205]. Such results underscore the importance of using size-for-gestational age as the exposure variable, instead of birth weight alone, which introduces random variation at the cost of underestimating relative risk[204].

1.4.3 Birth length

Height represents a crude measure of bone at the whole body level[206]. In contrast to birth weight, birth length is difficult to measure in neonates with accuracy. As a result, length at birth has received less attention than weight at birth, and often delayed by study protocols (e.g. DAISY) to ages which allow for standing height to be recorded, i.e. 2 years[207].

1.4.4 Weight to length indices

A composite index of relative weight is the weight-to-length proportionality, represented by i) simple weight to height ratio, ii) body mass index (BMI), calculated as weight (kg)/height squared (m^2), and iii) ponderal index, calculated as weight (kg)/length cubed (m^3). On the grounds of the best 'relative weight index' having the lowest correlation with height and highest correlation with body fatness, the BMI lends itself as a surrogate measure of adiposity

applicable to any population[208]. A study assessing all three indices of relative weight in men, reported that the ponderal index was the poorest index[208]. In neonates, the downside of calculating proportionality using the cube of length is the amplification of inaccuracies in the measurement of length[209]. Nevertheless, ponderal index is the preferred index in paediatric clinics over BMI, whose distribution is skewed and the degree of skewness is age-dependent[210,211]. It has been suggested that whereas body weight is a measure of the aggregate gains of soft tissue and length, ponderal index is thought to reflect soft tissue only[212], an interpretation which has lent itself to findings according to which either weight or ponderal index, but not both, associated with T1D risk.

1.4.5 Head circumference

Head circumference reflects intrauterine neurodevelopment[213]. As such, it has received little to no attention by studies investigating associations between size at birth and risk of T1D. Brain growth is marked by two growth spurts: the initial growth spurt occurs from 12 to 18 weeks of gestation and features neuronal multiplication; the second growth spurt begins at 28 weeks of gestation and extends until the third year of life[213]. Davies[214] suggested that head circumference is influenced mostly by genetic factors and, compared with weight and length, is less sensitive to maternal influences; yet, the growth of the brain might be hindered by intrauterine (spacing) constraints.

1.4.6 Influences on size at birth

Observational studies have adjusted for factors that could explain an association between foetal growth and risk of T1D to exclude them as primary explanatory mechanisms. Genetic associations with size at birth also require that the maternal-uterine environment be carefully considered[215]. Intrauterine effects are defined as the physical constraints or nutritional factors that influence the foetal growth environment[216]. The critical importance of these factors is encapsulated in the 'foetal programming' or 'foetal origins hypothesis', which postulates that the intrauterine environment can 'program' or influence pregnancy and neonatal outcomes[217].

Birth weight is affected by multiple factors, both genetic and non-genetic. To illustrate the point, the clinical observation that sibling birth weights highly correlate[218] has been attributed to similarities in both the parental gene pool and *in utero* maternal influences[219]. The genetic constitution is believed to be less important for neonatal anthropometrics vs. the cluster of environmental influences, which may be classified into i) preestablished maternal influences, ii) pregnancy-related influences and iii) paternal factors[220].

1.4.6.1 Pre-pregnancy maternal influences

Birth weight is an outcome of intrauterine growth rate[89], hence it is predominantly under the influence of gestational length[221]. At a particular gestational age, preestablished biological influences include maternal parity (birth order). Birth weight increases along birth order in singleton pregnancies[222], implying that infants of primiparous pregnancies (first-born infants) are more likely to be restrained *in utero* and born smaller than subsequent offspring, but likely to show compensatory 'catch-up' growth, becoming heavier and taller after infancy[217].

Maternal anthropometry before pregnancy exerts its effect on foetal growth, arguably via the intrauterine environment and/or genetic factors. A recent systematic review has shown that pre-pregnancy overweight or obesity increased the risk of higher birth weight, macrosomia and LGA[223], and a prospective study found that pre-pregnancy BMI positively correlated with foetal macrosomia (birth weight >90th centile)[224]. Maternal height has been shown to positively associate with birth weight[225,226] and birth length[227]. The correlation between maternal height and birth size — traditionally explained upon the mechanistic assumption that shorter women have a smaller uterus, which imposes a physical constraint on foetal growth[228] — has been recently attributed, by means of Mendelian randomisation, to foetal genetic effects[229].

1.4.6.2 Pregnancy-related influences

Pregnancy-related influences include maternal glucose levels, smoking, maternal alcohol consumption, maternal weight gain, cord blood flow, maternal disease and drugs[89,219,221,230]. Amongst these, the universal finding is that elevated blood glucose levels during pregnancy give birth to babies with larger size[231], testified by the macrosomic babies of pregnancies with gestational diabetes[232]. Even in the absence of diabetes in pregnancy, there is well-replicated evidence of a continuous effect of maternal glucose levels on offspring's birth weight[233-235]. Ong et al.[224] showed that the mother's fasting glucose correlated with foetal macrosomia as well as adiposity in infancy up to 3 months of age. Another study showed that post-prandial glucose levels exerted a stronger effect than fasting glucose, corroborating that poor glycaemic control poses a pregnancy-specific risk factor[219]. Maternal weight gain during pregnancy emerged as an important risk factor for higher birth weight[219] and greater neonatal fat mass[236]. In contrast, elevated maternal blood pressure associated with lower infant birth weight[237].

The caveat of such associations, namely their confounding by socioeconomic status and behaviours, was addressed by a recent Mendelian randomisation study in >30,000 women, which found that elevated fasting blood glucose was causal for high birth weight, whereas elevated maternal systemic blood pressure was causal for low birth weight[238].

Lifestyle habits have an established role in influencing size at birth. Smoking during pregnancy is considered a causal factor of small birth size as it detrimentally restricts intrauterine growth, with a study using biochemical measurements of thiocyanate concentration in the mothers to precisely and objectively measure extent of smoking[221]. Alcohol in the absence of smoking increases birth weight, albeit marginally[221]. The TEDDY study showed that the offspring of mothers who consumed alcohol in moderation during pregnancy, defined as one to two drinks per month, were longest at birth[230]. A very recent investigation in Poland has shown that the combined effect of alcohol and tobacco use during pregnancy increased the risk of preterm birth and exacerbated the adverse effect on size at birth[239]. Aside smoking and alcohol consumption, maternal nutrition in the peri-conceptional period and pregnancy plays a crucial role in foetal development[184]. A prospective cohort study in 44,612 Danish women surprisingly showed that a high-calorie, high-protein diet during pregnancy carried an increased risk of giving birth to SGA newborns compared with women who consumed a low-calorie plant-based diet[240], concluding that a high animal-protein and fat diet might impair foetal growth, or trace elements in plants induce proper foetal development[184].

1.4.6.3 Paternal influences

Whereas the intrauterine environment is largely under the maternal influence, parental height reflects the influence of both parents. A study on *paternal* influences on birth weight found that neither paternal age, BMI nor weight were independent predictors of birth weight, and paternal height explained less than 2%[241]. Such findings single out height as the potential domain of influence from the father. Expressed as mid-parental height, its effect on growth has been a contentious topic. It has been suggested that mid-parental height correlates with both size at birth and postnatal growth in the child[242,243]. The TEDDY study reported that parental height and weight associated with the child's weight and length at birth[230]. Others contend that birth weight is influenced only by the mother, but later length gain reflects the size of both parents[185].

The dichotomy regarding the age of offspring at which parental height switches on its effect perpetuates with studies on T1D that detect growth differences between cases and controls. Analysis of cohort data in Sweden showed that weight and length at birth of prediabetic children was higher than controls when corrected for mid-parental height[244]. In contrast, the DIABIMMUNE study showed that mid-parental height exerted a modest effect on linear growth in infancy, thus obviating the need to correct growth for mid-parental height[245]. Finally, a study in France identified strong associations between the mother's height and the child's height growth velocity in the first months of life, and strong associations with the father's height after two years, thus 'mirroring an alternative influence of maternal and paternal genes'[246].

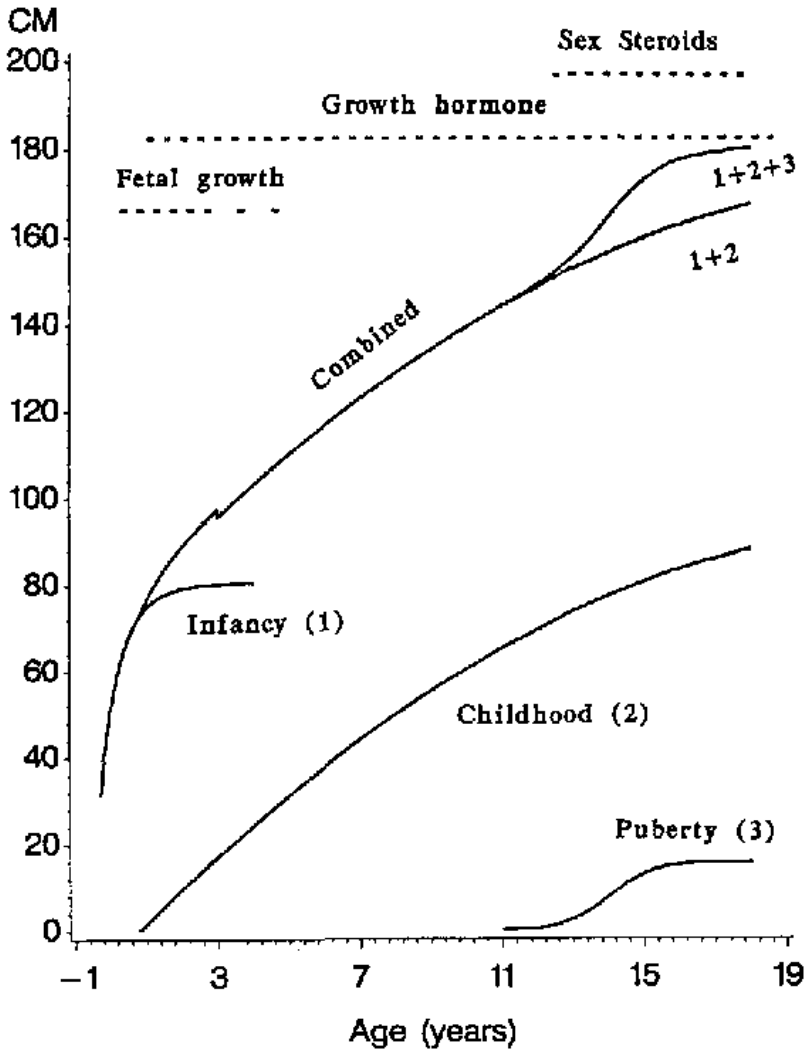
1.5 GROWTH IN INFANCY

1.5.1 Description of postnatal growth

Growth is defined as the increase in the number, the size and complexity of cells, resulting in an increase in body weight and height[247]. Growth originates at the moment the ovum is fertilised and involves many processes, with protein synthesis being the most fundamental[248].

Conceptualisation of the complex phenomenon of growth and development in humans has been made possible by mathematical modelling of individual linear growth. Karlberg's infancy-childhood-puberty (ICP) model breaks down human growth from the second half of intrauterine life to maturity into three additive and partly superimposed components — infancy, childhood and puberty — which reflect distinct hormonal phases[249] (Figure 1-11). The infancy component is nutrition dependent, the childhood component is mainly dependent on growth hormone (GH), and the pubertal component relies on the synergism of sex steroids and GH[250].

Figure 1-11 | Karlberg's ICP model of childhood growth. From Karlberg[249].



The model implies the partition between prenatal and postnatal growth, which divide further[251]:

I. Prenatal
   1. Ovum: cell division (first 2 weeks);
   2. Embryo: steady growth and cell differentiation (to 8 weeks);
   3. Foetus: rapid growth and development of tissues and organs (to 40 weeks).

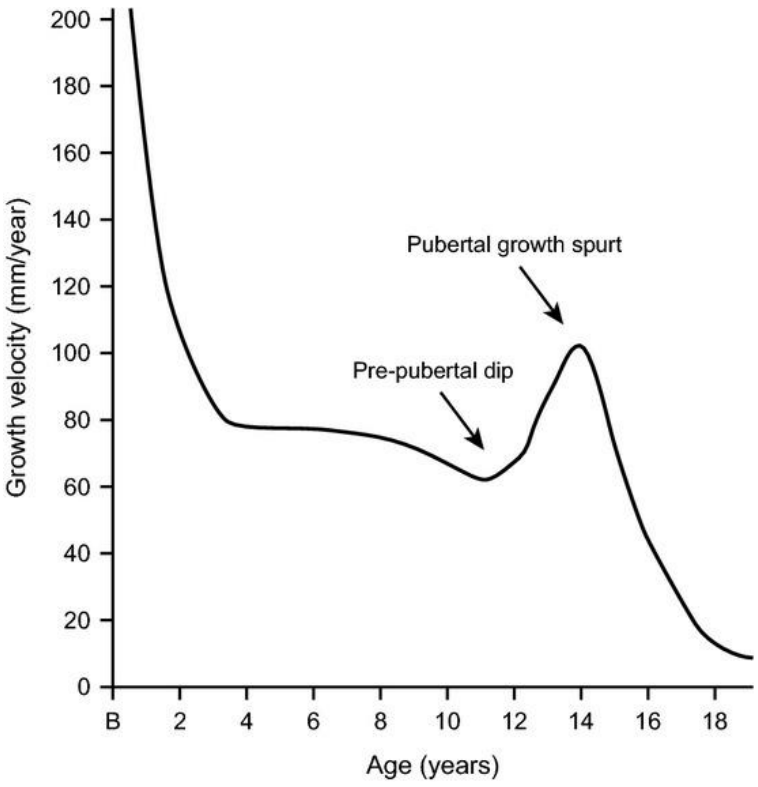
II. Postnatal
   1. Infancy (birth to end of weaning at 24-26 months);
   2. Early childhood (to 7 years of age);
   3. Later childhood (to puberty);
   4. Adolescence (to sexual and physical maturation at about 20 years of age).

The usefulness of the ICP model lies in constituting 'an improved instrument for detecting and understanding growth failure'[249].

1.5.2 Trajectory of postnatal growth

Growth curves are described by three parameters: size, velocity and tempo. Growth velocity is highest at birth and progressively declines until the pubertal growth spurt, which is marked by an increase in height followed by a sudden deceleration to a zero growth velocity as epiphyseal fusion occurs[252] (Figure 1-12).

Figure 1-12 | A typical growth velocity curve in children. From Haymond et al.[252].



Growth velocity and age at peak velocity are indicators of growth tempo. Tanner[253] described tempo as 'a set that runs through the whole of childhood and refers to rate of (skeletal) maturation, not rate of growth', and suggested that differences before puberty are of a limited degree. That the tempo of growth is under genetic control was shown by a study of longitudinal growth in families of twins[254]. Recently, a study in families of twins in the UK found that infants who were younger at the peak weight velocity had larger size at all ages, and ascribed a lower genetic effect but stronger environmental influence to their tempo[255]. The authors of this study concluded that tempo varies amongst infants and, in fact, its acceleration in infancy serves an indicator of adverse long-term health outcomes[255].

Normal growth in the first months of life is characterised by rapid changes: it starts off with an initial drop in weight after birth, followed by an increasing weight gain that peaks at 6 weeks, which later declines to a plateau at around six months[255,256]. The first six months of life represents the fastest growth in the entire lifespan, during which postnatal weight gain is largely composed of fat[246]. The deposition of large amounts of fat *in utero* is unique to human, who is classified as the fattest species at birth[257]. Kuzawa[257] hypothesised that the reason that newborns expend roughly 70% of energy to fat deposition during early postnatal life is in response to 'nutritional disruption', which occurs twice:

i. at birth and until lactation commences — a concept recently corroborated by a meta-analysis which concluded that infants who initiated breastfeeding ≥24 hours after birth had an 85% higher risk of mortality than newborns who initiated <24 hours[258];
ii. at weaning, the transition from breast milk to complementary foods, which is accompanied by the transition from maternal to endogenous immune protection and maturation of the gastrointestinal tract.

These changes are compounded by hormonal fluctuations and expression of paternal alleles, rendering the first six months of life a critical passage from the intrauterine environment.

Satisfactory growth is suggestive of energy demands being adequately met[256]. Energy requirements for maintenance take precedence over protein synthesis, which is an energy-demanding process, so that in the event of energy intake falling below maintenance needs, growth slows down or ceases[256]. This adjustment in growth rate allows the child to cope with the physiological and metabolic needs in sub-optimal conditions[259]. Prader[260,261] observed that upon removing the cause of growth deficit, such as illness or malnutrition, growth acceleration ensues at abnormally high velocities, coined 'catch-up', which is then followed by deceleration to restore the normal growth channel.

1.5.3 Body composition

Persistent accelerated growth in infancy has qualified this period as a window of obesity risk, which can be evaluated by body composition assessment[262]. Body composition analysis quantifies the amount and relative proportions of body tissue compartments, often considering their cellular and molecular components, as a means to describe populations or assess individual disease risk[206]. At its simplest, a two-compartment model divides body mass into fat mass (FM) and fat-free mass (FFM), whereas the three-compartment model addresses the tissue heterogeneity of the FFM — which comprises skeletal muscle, organs, bone and connective tissues — by partitioning the body into FM, non-osseous lean body mass and bone mass[206]. The four-compartment model partitions weight into fat, protein, water and mineral[263].

Simple body composition indices commonly used in the paediatric population include[206,264]:

i. BMI, a global index of nutritional status and obesity, whose predictive value is less clear for children and fails at distinguishing between fat and lean mass;
ii. waist circumference, an estimate of central adiposity, which is more predictive of insulin resistance and lipid profile than total body fat;
iii. skinfold thickness, an indicator of regional (subcutaneous) fatness, which is limited by the lack of contemporary paediatric skinfold reference data;
iv. upper arm fat and muscle areas, calculated from measurements of upper arm circumference and triceps skinfold thickness.

Butte et al.[265] reported normative values of body composition during the first two years of life, which changes with age and is sex-specific. Per cent body FM is higher in infancy peaking between 3 to 6 months of age, and longitudinal changes in the components of FFM include a gradual increase in protein content and a decline in total body water[265]. At birth, girls have proportionally more adipose tissue than boys with the differences persisting in infancy[266]. Although FM accumulates rapidly in the first six months of life, subsequent growth of FFM makes the greatest contribution[266,267]. A recent study found that the greater adiposity in the first months of life is composed of increases in subcutaneous vs. internal adipose tissue, and intrahepatocellular lipid content substantially increases irrespective of feeding[268].

The reasons for these dynamic changes in body composition are not clear but might reflect substrate selection. Foetal development is reliant on a high rate of placental transfer of glucose but null transfer of triglycerides[269]. These conditions possibly program the newborn with a low capacity for fat oxidation and high capacity for de novo lipogenesis, evidenced by the suppressed oxidation of fat in the first months of life despite a high dietary fat intake, in contrast to adequate levels of non-fat oxidation following dietary intake[267].

1.6 REGULATION OF EARLY GROWTH

1.6.1 Nutrition

Growth in infancy is nutrition dependent[250], and has pronounced and long-term effects on size and body composition. The importance of the first 1,000 days of life — the period from conception to age two — is an important time for body and brain development that represents a critical window for prevention of later metabolic abnormalities[183], casting light on the benefits of breastfeeding. Although the WHO recommendation for the duration of exclusive breastfeeding is the first 6 months, some countries recommend the introduction of complementary feeding between 4 and 6 months[270]. In the short-term, breastfeeding confers immediate protection against infections that persist throughout the first year of life[271], and moulds the infant's acceptance of a wider food repertoire including a higher fruit and vegetable intake[272,273]. The long-range impact of early nutrition on later health is thought to be intricately coupled with the immediate effects on infancy growth. Formula-fed infants gain more in weight and out of proportion to linear growth[274], more in BMI[275] and more in adiposity[276]. It is currently admitted that formula-fed infants gain more FFM in the first months of life[275,277,278], and have lower FM than their breastfed counterparts, but beyond the first year a reversal trend is observed with formula-fed infants acquiring higher FM[277].

One possible reason for the patterned body composition by feeding lies in the total energy intake, which is dependent on feeding mode and delivery; bottle feeding led to higher weight gain in the first year of life irrespective of whether it contained breast milk or formula milk[279], and larger bottle contributed to greater weight gain[280]. The chemical composition of milk has also been explored. Prentice et al.[281] showed that irrespective of milk intake, human milk composition had a functional relevance to growth. Growth in the first two years of life is particularly sensitive to protein intake[282], reflecting the 'early protein hypothesis', which postulates that excess protein intake via infant formula feeding induces elevated circulating levels of insulin-releasing amino acids that stimulate the secretion of insulin and insulin-like growth factor-1 (IGF-1), which result in increased adipogenic activity[283]. Human milk provides more than half of its energy from fat[184], and even though total fat content is not influenced by maternal nutrition[284] the composition of milk lipids is sensitive to maternal diet[285]. Specifically, mono-unsaturated fatty acids have a dual role in providing nutritional benefit for the development of the nervous system[285] and conferring the proper liquidity for breast milk formation[284]. The spectrum of components of human milk that influence growth is expanding: fructose was the only sugar in human milk, albeit its very low concentration, that was positively associated with body composition (weight, lean mass, fat mass, mineral content) at 6 months

post-delivery[286] and ghrelin positively correlated with weight gains from birth[287]. These findings point to a complex molecular network affecting the biological pathway of growth.

1.6.2 Hormones

A recent study unveiled the cord blood hormones at play in relation to timing of increased gestational weight gain: i) 1st trimester weight gain associated with higher cord blood insulin and C-peptide, and lower adiponectin, i.e. hormones relating to glucose/insulin regulation, ii) 2nd trimester weight gain associated with higher cord blood IGF-1, IGF-2 and leptin, i.e. hormones central to growth and adiposity, iii) 3rd trimester showed no associations[288].

1.6.2.1 Insulin

Insulin is the most important foetal growth factor in late pregnancy[232]. Foetal development is regulated by the insulin-dependent processes of amino acid transport and protein synthesis[232]. The placenta is impermeable to insulin[289], forcing the foetus to rely on its own insulin reserves. Insulin starts to synthesise in the human foetal pancreas from as early as the 10th week of gestation and becomes detectable in the foetal plasma from the 12th week of gestation, beyond which insulin levels progressively rise until the 24th week of gestation[290]. Insulin release remains insensitive to glucose until around the 28th week of gestation[232] when preadipocytes mature to insulin-sensitive cells that are able to store fat[291]. The metabolic disturbances observed during diabetic pregnancy, manifested by accelerated foetal somatic growth in the last trimester of pregnancy, are explained in terms of surplus nutrients (glucose and amino acids) rapidly crossing the placenta down a concentration gradient from mother to foetus and having the ability to influence β-cell ontogenesis and/or stimulate insulin release, affecting protein synthesis and cell proliferation[291,292]. At birth, plasma insulin levels positively correlate with weight[293] and dramatically fall over the first 48 hours after birth[294]. Human β-cells are functionally mature by the age of one year, evidenced by their comparable responsiveness to stimuli with that of adult β-cells, but secreting smaller proportions in line with lower demands[295].

The identification of clinical defects in insulin secretion has enhanced our understanding of its critical role in regulating foetal growth. The infant of a diabetic mother is heavy, long, with enlarged β-cells and low blood sugar[296]. The uncontrolled development of foetal pancreas that is not arrested in pregnancy but continues during infancy, known as nesidioblastosis, gives rise to macrosomic infants[297]. Similarly, the Beckwith-Wiedemann syndrome, attributed to hyper-insulinaemia, is characterised by gigantism, macroglossia, and visceromegaly that includes the pancreas[298]. Whereas intrauterine insulin abnormalities lead to severe growth faltering, GH-deficient newborns are on average only 2 to 4% shorter than

healthy infants at birth[299]. This reflects that foetal growth is independent of GH[232], possibly due to immature GH-receptors (GHR) in the growth plate, but initiates its effect in childhood[299].

1.6.2.2 IGF

Longitudinal growth is influenced by the GH/IGF axis, which is a mixed endocrine-paracrine-autocrine system[300]. IGF-1 and IGF-2 are peptide hormones consisting of 70 and 67 amino acids respectively and share circa 50% homology with insulin[301,302]. Systemic IGF-1 and IGF-2 are produced mainly in the liver but most cells can synthesise them[303]. IGF-1 circulates at nanomolar levels with a half-life of minutes that can be extended up to 15 hours when complexed with one of seven IGF-binding proteins (IGFBP); in comparison, insulin circulates at picomolar concentrations and has a half-life of minutes[304,305]. IGFBP-3 binds 80-95% of the circulating IGF-1 in a 150-kDa ternary complex, which includes the 85-kDa acid-labile subunit (ALS) that is synthesised in the liver and has a role in releasing and storing IGF-1. Less than 1% of IGF-1 circulates as free hormone. IGF-1 is detectable from the first trimester of gestation but is more important in the later months of foetal development[306]. Umbilical cord IGF-1 levels correlate with birth weight and are lower in preterm and IUGR births[306]. Normal IGF-1 bioactivity — local and endocrine IGF-1 availability, IGF-1 receptor (IGF1R) function and signalling — is required for normal early longitudinal and cranial growth[307]. Association studies of IGF-1 polymorphisms have reported a consistent correlation with head circumference, but results with height have been variable[307]. In contrast, the role of IGF-2 is limited postnatally.

Insights into the role of IGF-1 in postnatal growth have been primarily gathered from rare human syndromes. Disorders on the GH/IGF-1 axis may be classified as: i) GH deficiency, ii) GHR defects, iii) defects in the GH signal transduction, iv) IGF-1 defects, v) IGF1R defects and vi) ALS defects[300]. The first three are characterised by near-normal prenatal growth, defects of IGF-1 and IGF1R cause circa 10% of intrauterine and postnatal growth retardation[300,308], and ALS mutations result in growth hormone insensitivity[309]. For example, mutations of IGF1R are clinically manifested by IUGR and postnatal growth failure[309]. In comparison, newborns with the Laron syndrome, defined as GHR mutations that lead to primary IGF-1 deficiency, are shorter (42 to 47 cm), have a very slow postnatal growth rate, develop organomicria and retardation of skeletal maturation[310]. The profile of the Laron syndrome suggests that postnatal regulation of IGF-1 production is GH dependent[309], but possibly not until 9 months of age when GH starts to be active; until then, IGF-1 is regulated by insulin and nutrition. In addition to its critical role in brain development and early growth, IGF-1 is a sensitive indicator of nutritional status reflected by the finding that its levels were increased in formula-fed vs. breastfed infants, which correlated with faster growth rates[311].

1.6.2.3 Leptin

Leptin has a central position amongst the repertoire of indicators of nutritional status and early growth. A member of the family of adipokines (adipose-derived cytokines), leptin is a 167-amino-acid peptide hormone — the product of the *ob* gene — that is synthesised and secreted by adipose tissue[312] in proportion to fat stores[313,314]. Leptin is known for its effect as an acute regulator of food intake (by signalling nutritional status to other physiological systems) and energy homeostasis, which are mediated by various neuronal pathways[315,316]. Currently, leptin is being investigated for its role in immune function and bone metabolism[317]. Similar to other hormones, leptin's secretion is pulsatile[318] with diurnal variation, exhibiting higher nocturnal levels to suppress appetite during sleep[319]. Leptin deficiency due to rare mutations have been reported in obese individuals and addressed by replacement therapy[320,321], but resistance to leptin, defined as failure of high leptin levels to suppress appetite[322], accounts for most cases.

Of relevance to this thesis, accruing evidence points to leptin's role in reproduction. Leptin is synthesised in the fetoplacental unit, testified by its higher levels during pregnancy, and regulates foetal growth[323] independent of the insulin/IGF-1 axis[324]. Leptin was found in the cord blood in concentrations comparable with those in adults (0.6 to 55.7 ng/ml)[325], spawning a growing literature on the subject. Cord blood levels of leptin positively correlated with gestational age[326], birth weight (explaining 21% of variation), birth length, ponderal index[327], and inversely correlated with weight gain in infancy[328,329] due to slower changes in adiposity[330]. Umbilical cord blood leptin is currently accepted as a biomarker of neonatal FM[323] but not lean mass[331], addressing the limitation of birth weight as an anthropometric index that does not differentiate between FM and FFM. Finally, the higher levels of leptin in female vs. male neonates[326,328] underscore the foetal origins of sexual dimorphism of adiposity.

A study reported that cord blood leptin positively correlated with neonatal bone mineral contents[331]. An inverse relationship exists in children between bone mineral density and bone marrow adipose tissue, explained by the fate of mesenchymal stem cells to differentiate into adipocytes or osteoblasts[332]. The above finding, in conjunction with the unexpected presence of leptin receptors in human haematopoietic stem cells[333], has been interpreted as the involvement of leptin in haemotopoiesis and lymphopoiesis[333,334]. This notion underpins the role of leptin in foetal life, best put in the words of Sivan et al.[325]: 'In adults, the close correlation between adiposity and circulating leptin levels supports the generally held concept that leptin is a signal from the body fat mass to the brain, where it affects satiety. It would, however, be difficult to apply this concept to the foetal situation, (where) the mother is in complete control of the energy supply to the foetus, who has no need for feelings of hunger and satiety.'

1.6.3 Genetics

Beyond the deleterious mutations of the GH/IGF-1 axis and leptin that cause pronounced growth failure, genetic loci have been identified to exert modest effects on birth weight. The hereditary nature of birth weight was hinted by work on the British 1958 cohort, which found that parental foetal growth made the largest anthropometric contribution to offspring's birth weight[335]. In the GWAS era, it has been made possible to disentangle the maternal and foetal influences on birth weight, and establish that the infant's own genotype made a larger contribution to its birth weight than the mother's genotype or the covariance between the two[336]. The recent multi-ancestry GWAS meta-analysis in 153,781 subjects identified 60 loci that associated with birth weight (P<0.5 x10^{-8}), with the lead SNPs being frequently occurring (MAF≥5%), mapping to non-coding sequences, and exerting modest effects (10 to 26 g) per allele[336]. Of the 60 loci, three signals fell within imprinted regions: *INS-IGF2*, *RB1* and *DLK1*.

Imprinted genes have been described in rare disorders, e.g. Prader-Willi, and are central to regulating foetal growth as well as having a role in early life functions including feeding, sleep, temperature regulation and metabolism[337]. Genomic impriting is an epigenetic modification of mammalian placentas that results in monoallelic gene expression according to parental origin by silencing one allele and without altering the DNA sequence[338,339]. Chr11 houses a large cluster of imprinting genes, including the contiguous reciprocally imprinted *IGF2/H19:* the paternally-expressed *IGF2* encodes IGF-2, which is the main foetal growth factor[340], and the maternally-expressed *H19* encodes a non-coding transcript whose function is unclear[341]. Size at birth was found to associate with common polymorphisms in foetal *H19*[342] and *IGF2* alleles in the normal paediatric population, which are possibly controlled by DNA methylation[343]. This is in line with the hypothesis that gain in methylation at *H19* decreases its expression, allowing for higher expression of its reciprocal *IGF2* that leads to foetal overgrowth, exemplified by the Beckwith-Wiedemann syndrome; conversely, loss of methylation at *H19* is characterised by feotal growth restriction, as implicated in the Russel-Silver syndrome[341,343]. Predictors for varying degrees of DNA methylation at this locus are poorly understood. A recent study has cast light on maternal anxiety during pregnancy, which associated with decreased methylation in *IGF2/H19* particularly in female low birth weight neonates[344].

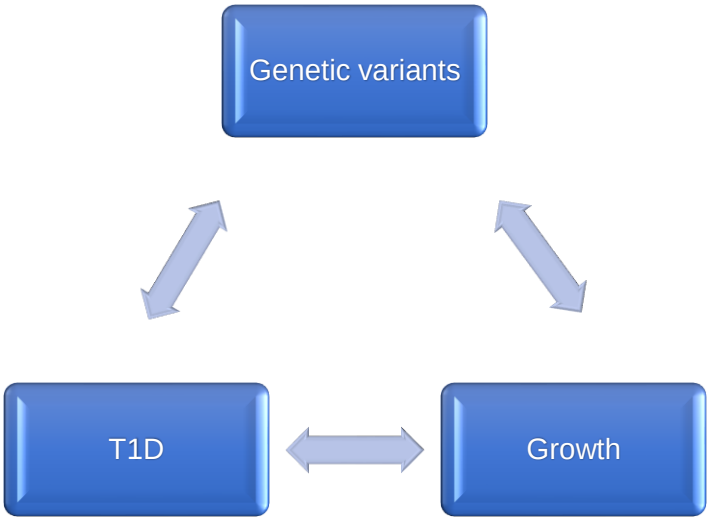
The maternally-expressed *DLK1-MEG3* imprinting region, which is a T1D susceptibility locus, also influences foetal growth: the paternal transmission of the T1D protective allele associated with lower birth weight and head circumference[18]. This finding puts the long-standing debate on the link between growth and T1D risk in new perspective, and hints at a genetic link between the two closely observed phenotypes.

1.7 RESEARCH GAP AND AIMS

1.7.1 Need for genetic studies

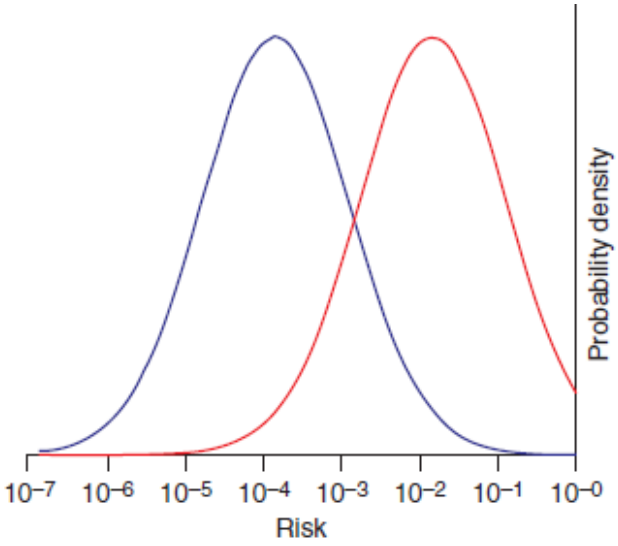
The literature on the associations between foetal or postnatal growth and T1D is beset with contradiction. Genetic analyses could shed light on the heterogeneity of observational data on early growth and T1D risk[345] traditionally obtained from cohorts of diabetic or prediabetic children. First, introducing genes with pleiotropic effects on growth and T1D (Figure 1-13) would circumvent confounding factors (e.g. social conditions). Second, on the grounds of early growth in disease operating via a different mechanism vs. health[346], testing hypotheses in a disease-free population of infants would eliminate the variable effects of autoimmunity on physiological growth.

Figure 1-13 | The putative interlink between T1D and early growth.



Existing population genetic studies on the subject have limited their scope to the effects of *HLA* and *INS*. However, in a polygenic disease with many variants independently contributing, the logarithm of risk is normally distributed in the population (Figure 1-14) and an 'average disease risk' is thought of as a mix of predisposing and protective alleles weighed by their effect on disease susceptibility[347]. Clearly, there is wider scope for loci implicated in early growth, which could be tested not only in isolation, but also in the aggregate expressed as genetic risk scores. Genetic risk scores — alternatively known as 'genotype scores' or 'allele scores' — represent a convenient way of summarising a number of genetic variants associated with a common risk factor and are used for modelling polygenic traits, particularly when the score consists of multiple frequently-occurring variants with small effects, or rare variants[348].

Figure 1-14 | Genetic risk distribution in cases (red) and controls (blue). From Goris and Liston[347].



Counting 59 chromosomal regions conferring susceptibility to T1D, the task of selecting candidate genes with a suspected link to growth is not trivial. Prioritising variants and genes that are implicated in growth-related phenotypes is a key selection strategy. In complementing this approach, the network of shared genetic loci amongst immune-mediated diseases[349] may be exploited. The high degree of genetic similarity across autoimmune diseases becomes clearer by categorising their genetic variants into three biologically distinct classes[347]:

i. HLA locus, which shows the strongest association and is specific to a disease;
ii. non-HLA loci, which are specific to an autoimmune disease;
iii. non-HLA loci, which are the majority and involved in multiple autoimmune diseases.

The co-occurrence of diseases informs of the underlying network biology of multifunctional genes and pathways[350]. I turned this around to suggest that the co-occurrence, or not, of phenotypes inform of the presence or absence of shared genetic factors. Hence, if alterations in early growth are specific to T1D amongst immune-mediated diseases, it is likely that T1D candidate genes that are not shared by other autoimmune conditions are accountable.

1.7.2 Implications

This topic is important because of the increasing incidence of T1D particularly in the youngest age groups in European populations[39]. The EURODIAB study projected that the incidence rate in European populations increases by 4% annually but new cases of children under 5 years of age will double between 2005 and 2020[7]. The International Diabetes Federation estimated that circa 79,000 children aged 14 years and under are diagnosed with T1D every year worldwide, with Europe bearing the highest burden[351]. The increase in the incidence of

childhood T1D over the last 60 years is poorly understood[352], but could be countered by duly identifying at-risk infants.

Stratification of T1D risk is far from being the staple of clinical practice but has been experimented by prospective studies. Current screening relies on the predictive capacity of family history and genetic markers on the HLA locus, capturing up to 5% risk from the general population[353]. However, almost 9 out of 10 new cases lack a family history of T1D[105] and progressively fewer diagnoses present the classical high-risk HLA genotypes[354], thus rendering current screening schemes inadequate. Recent combination approaches to improve the specificity of familial and HLA information using an expanded set of genetic risk loci could improve prediction of T1D[353]. Identifying variants that associate with postnatal endocrine processes has the potential to fine-tune predictive loci and elucidate mechanisms.

1.7.3 Aims

The overarching aim of this thesis is the investigation of the presence and degree of a link between selected T1D susceptibility genetic variants, mostly in immune genes, and early growth. This investigation is conducted on the basis of existing information and novel work by setting the following objectives:

1. Review and critically assess the existing literature on associations between prenatal or postnatal growth and T1D risk to decipher prevailing patterns and inconsistencies.
2. Evaluate the debated association between *INS* and perinatal growth in a new cohort.
3. Explore novel genotype-phenotype associations by expanding the gamut of anthropometric indices and probing for endocrine regulators (IGF-1, C-peptide, leptin).

The Cambridge Baby Growth Study (CBGS) — a prospective, observational pregnancy and birth cohort from Cambridge, UK — provides access to hundreds of newborns and detailed phenotypes of postnatal determinants of growth over the first 2 years of life. The CBGS serves as a sound medium for the aims of this thesis because i) it provides a collection of repeated measures and confounding variables of growth, thus enabling time-dependent and well-adjusted evaluations, ii) it has an extensive range of anthropometric indices, iii) it is well-powered within the limits of prospective studies, iv) it is ethnically homogeneous, thus apt for genetic analyses and v) it follows subjects during infancy. Critically, infancy is a period when postnatal growth is at its peak so most susceptible to adverse conditions[185] and the immune system is immature[134], thus bridging the two biological elements that are critical to the pathogenesis of T1D when at their highest vulnerability.

Chapter 2

SYSTEMATIC LITERATURE REVIEW

2.1 CONTEXT

A plethora of research investigations published in the last five decades have grappled with the conundrum of the link between early growth and T1D in childhood, yet the literature has been beset with controversy and claims have pointed research to conflicting tracks. In a turning point, recent studies that tap into a wealth of data are starting to shed light on compelling and replicated associations. Delving into the literature of a topic that is highly-debated and contradictory is imperatively the starting point in the quest for clarity and cues for refining questions. Here, I review the existing literature on birth and infancy anthropometry in relation to risk of T1D in childhood as a prelude to the original findings of this thesis.

2.2 METHODS

Studies published until May 2018 were identified from a literature search conducted in PubMed of the US National Library of Medicine (www.ncbi.nlm.nih.gov) by using the following search terms: T1D ('T1D' OR 'type 1 diabetes' OR 'insulin dependent diabetes mellitus' OR 'IDDM' or 'juvenile onset diabetes' OR 'juvenile diabetes' OR 'diabetes in childhood' OR 'islet autoimmunity'), AND ('head size' OR 'head circumference') OR ('birth weight' OR 'birthweight' OR 'size at birth' or 'birth size') OR ('height' OR 'body length') OR ('weight gain' OR 'growth') OR ('BMI' OR 'fat mass') OR ('anthropometry' OR 'anthropometric'). In ensuring coverage of the latest research developments, additional searches were conducted for the major prospective studies in T1D, namely 'BABYDIAB', 'DAISY', 'DIABIMMUNE', 'DIPP', 'TEDDY' and 'TRIGR'. Filters were not applied in order to minimise inadvertent omissions. Titles and Abstracts returned from the search counted >5,000 and were screened. Reference lists of articles retrieved were hand searched for additional publications to include. Studies on a wide age range, e.g. spanning from birth to disease onset, were considered solely for their findings pertaining to infancy.

2.3 STUDY DETAILS

2.3.1 Variables

Studies exploring perinatal influences on T1D use 'perinatal growth' as a marker of foetal and postnatal environments, and by inference of β-cell mass. Measurable parameters reflecting prenatal growth are *birth weight* and *birth length*, along with their combined measure of *ponderal index*. Parameters that reflect postnatal growth are *weight gain* and *linear growth*, along with their combined index of *BMI gains*. Most studies reported associations between infancy growth parameters as the explanatory variable and T1D risk as the outcome variable. Data on islet autoimmunity as the outcome instead is evolving[355] especially as it offers the benefit of increased sample size. For instance, the TEDDY study reported that greater weight in the first year of life was associated with islet autoimmunity (n=575) but not T1D (n=169)[356], highlighting the requirement for adequate sample sizes when the magnitude of effect is small.

2.3.2 Design

Prospective cohort studies of T1D have been hampered by the rarity of the disease[357]. The most common approach has relied on population registry data, which is limited by the lack of standardised measurement protocols and the use of retrospective data that introduces information bias. Recently, several single-centred prospective studies have followed infants at increased risk of T1D based on family history and/or HLA type, with the objective of identifying determinants of islet autoimmunity and T1D (Appendix I). Associations of polygenic traits identified in small cohorts necessitate replication in larger populations[358], since there is the tendency for the initial report to overestimate the genetic effect on the disease, possibly owing to bias and genuine population diversity[359]. This limitation has been addressed by multi-centre studies, exemplified by TEDDY, that increase statistical power (Appendix I).

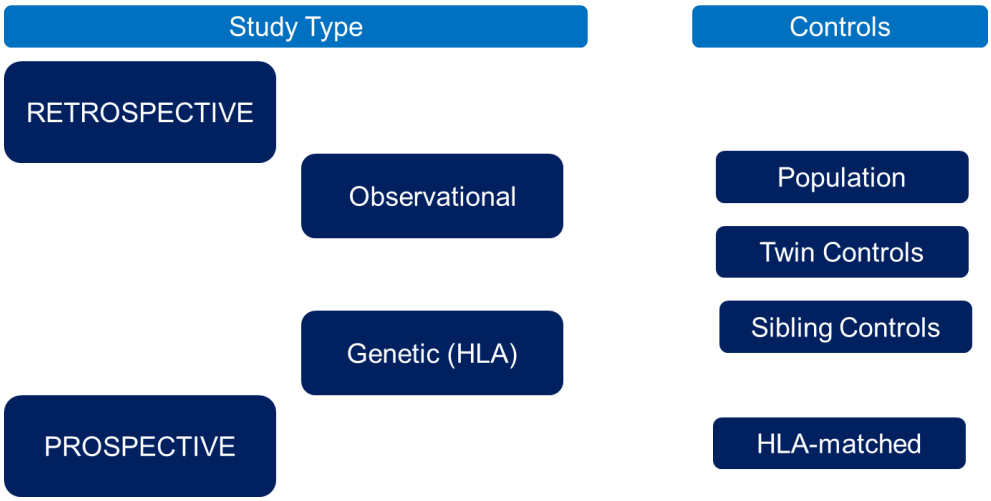
2.3.3 Population

Observational studies used an age cut-off of 15 years for the study population to reduce the possibility of misdiagnosis with T2D. The selection of age limit is supported by population-based evidence dating back to the mid-1980s, which established that nearly all diabetes diagnosed at <19 years of age was T1D and only 6% of subjects in the age groups of 15 to 19 years classified as T2D[360]. Further reinforcement of the designated age cut-off is fuelled by data indicating that birth weight persists as a predictive factor for T1D only for the first decade of life, with minimal, if any, bearing on the risk of late-onset T1D[361].

To ensure homogeneity of study population and preclude subsets with extremes of growth trajectories, clearly-defined exclusion criteria were often used. With the exception of twin-control studies, most investigations were restricted to singleton pregnancies as twins and triplets result in lower birth weight than standard. The criterion of singleton pregnancies additionally lent itself to linkage studies because it circumvents logistical difficulties in distinguishing between details of neonates abstracted from population registries[362]. Excluded were often children born to mothers with preexisting or gestational diabetes since it affects foetal growth and results in higher birth weight. Analysis was usually restricted to term pregnancies, predominantly defined as 37 to 42 gestational weeks, as preterm babies are known to have a lower birth weight.

Comparisons were largely made with growth data from the background infant population regarded as the 'controls'. Twin-control studies, albeit infrequent, have provided a powerful way of investigation as they control for the intrauterine environment and maternal influences of growth (e.g. maternal weight), since the twins share the womb at the same time, and allow for differences between dizygotic twins discordant for birth weight to be attributed to genetic factors[363]. As of recent, some studies have practised the use of healthy siblings as the control population in order to correct for family aspects shared by the siblings, namely genetics as well as environmental factors such as feeding habits, physical exercise and socioeconomic status[205,364]. The novel use of HLA-matched controls, made by a study in Sweden[244], is a new addition to the range of 'control populations', which teases out the effect of the HLA type, thus controlling for the dominant genetic factor. Figure 2-1 outlines the key methodological features of studies discussed herein.

Figure 2-1 | Methodological aspects of observational studies of early growth and T1D.

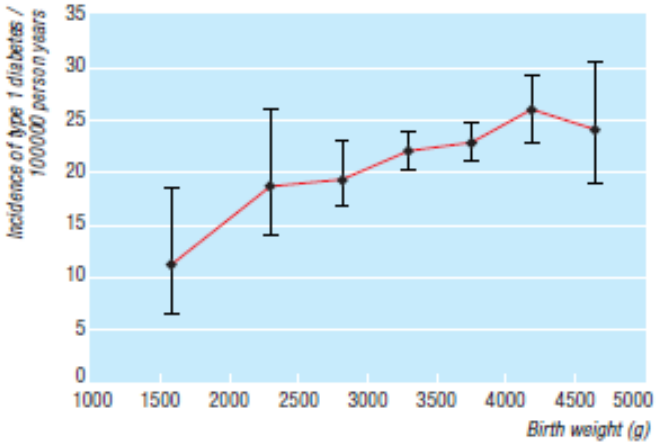


2.4 T1D RISK AND SIZE AT BIRTH: OBSERVATIONAL STUDIES

2.4.1 The direction and magnitude of the association

Numerous studies supported findings that high birth weight is a risk factor for T1D and other studies found that low birth weight displays a protective effect against T1D in non-diabetic pregnancies. A cohort study in Norway conclusively merged the two outcomes by establishing an almost linear relationship (Figure 2-2) between birth weight and risk of T1D, predicting a 1.7% increase in incidence rate per 100 g increase in birth weight[365].

Figure 2-2 | Birth weight and incidence rate of T1D in a Norwegian cohort. From Stene et al.[365].



It has been estimated that an effect size of this magnitude would require a large sample size (2,500 cases and 2,500 controls) to detect significant associations at a power of 80%[365], which was not attained by most published case-control studies[366]. In general, the inconsistency of complex trait association studies stems from the interpretation of findings from inadequately sized samples[367]. Thus, insufficient sample size could account for the inconsistent findings concerned with the existence or not of an association between risk of T1D and birth weight.

The next level of conflict concerns the direction of effect. Amongst studies which detected significant associations, the majority promulgated that high birth weight conferred risk of T1D, with two meta-analyses attempting to distil the magnitude of its effect. The first meta-analysis on 12 studies reported a confounder-adjusted summary OR of 1.43 [95% CI 1.11 to 1.85] for newborns weighing >4 kg, with each increase of 1,000 g in birth weight conferring a 7% increase of T1D risk[368]. Subsequently, Cardwell et al.[366] consolidated data from 29 studies of non-family-based study design (to tease out influence from higher genetic susceptibility) and concluded that children with birth weight of >4 kg had an increased risk of 10% in comparison with children weighing between 3.0 to 3.5 kg at birth (pooled OR=1.10 [95% CI 1.04 to 1.19]), which corresponds to a linear increase in diabetes risk of 3% per 500 g increase in birth weight.

Neither meta-analysis identified a significant effect for low birth weight. Recently, a study conducted on 3.6 million singleton births in Sweden, using prospectively collected data, reported that SGA children were at lower risk and LGA children were at higher risk of T1D[205]. Similar to a retrospective investigation conducted in Sweden 20 years ago[369], the recent study established significant associations with relative birth weight but found no evidence of increased risk for absolute birth weight >4 kg[205], which contrasts the outcomes of both meta-analyses. Nevertheless, I note that the findings of this study resemble an almost linear relationship between *relative* birth weight and risk of T1D, which echoes the correlation identified by Stene et al.[365] for absolute birth weight. Reinforcement of this proposition is fuelled by a more recent study that used registry data of hundreds of thousands of children in England and corroborated that high birth weight for gestational age was associated with subsequent T1D whereas lower birth weight (<2,500 g) was associated with lower disease incidence[370]. The estimates made by the two studies in Sweden and England respectively were almost identical, and in agreement with the adjusted pooled estimate of the latest meta-analysis discussed earlier. In addition, both studies concurred on low gestational age (<38 weeks) conferring risk of T1D[205,370].

The reporting of an inverted U-shaped association between birth weight and risk of T1D for pregnancies with maternal diabetes, meaning that low but also high birth weight display protection in children of mothers with T1D[45], underscores the sensitivity of growth associations, not only in terms of size of effect but also direction of effect, to intrauterine glycaemic conditions. This distinction was recently highlighted in a German prospective study following babies at familial risk, which identified significant associations between islet autoimmunity and growth patterns in children of non-diabetic mothers, but not in children of diabetic mothers[371].

This literature search has an exclusive focus on non-diabetic pregnancies, and as discussed, most observational studies excluded diabetic pregnancies from their analysis. It is noteworthy that no study, to my knowledge, assigned a protective effect to high birth weight in normal pregnancies, with the implication that the long-standing debate in the literature on the direction of the effect on risk of T1D revolves solely around low birth weight (Figure 2-3).

Figure 2-3 | Summary of associations between birth weight and T1D risk in non-diabetic pregnancies.

***P* value of association**	**Effect on T1D risk**	**Exposure: birth weight**	
		High	**Low**
>0.05	Increased risk	√	√
	Decreased risk		
≤0.05	Increased risk	√	√
	Decreased risk		√

2.4.2 The dichotomy of the effect of low birth weight

Low birth weight was initially found to be protective against T1D by case-control studies in Sweden[369] as well as seven study centres of the EURODIAB study group[372]. In contrast, high risk of T1D displayed by low birth weight was found in Taiwan[373,374] and Spain[375] — populations of relatively low T1D prevalence. Of note, a bigger investigation from Spain, which used data that was included in the EURODIAB study, showed that normal and low birth weight presented a lower risk of T1D compared with LGA children[376]. Studies in the Czech Republic and Austria detected no significant association between low birth weight and T1D[377,378]. The conflicting findings amongst low-to-medium T1D prevalence countries in Europe reinforce the assumption that the dichotomy of the effect of low birth weight on risk of T1D is attributable to limitations of study design or sample size, rather than disease prevalence or ethnic differences.

The protective effect of low birth weight on T1D was corroborated by the two largest studies on the subject to-date; a lower incidence of the disease was associated with very low birth weight (<1,500 kg) and SGA in Sweden[205], and with low birth weight (<2,500 kg) in England[370]. This relationship was further reinforced by an independent statistical study conducted as a vehicle for statistically differentiating between a conditional effect size (i.e. conditioning on confounders), which is widely used in clinical research, from its marginal counterpart (i.e. marginalising over confounders), using data on birth weight and T1D[379]. The details of the statistical argument are out of scope, but of relevance to this thesis, the results identified that the conditional effect of low birth weight on the onset of T1D in full-term children was significantly <1, ascribing a protective effect to low birth weight, whereas the marginal estimate presented no significant relationship[379].

Beyond statistical arguments, there is an accruing list of observations that reinforce the plausibility of a protective effect displayed by low birth weight. Support is lent by the finding that maternal smoking during pregnancy, known to result in smaller birth weight, decreased risk of T1D[362,380]. The observation that twins, who have lower birth weight on average than singletons, are not at higher risk for developing T1D provides further testimony[363].

The fallacy of ascribing risk to low birth weight is possibly pivoted on studies which determined that infants of lower birth weight showed a significantly earlier age of onset of T1D than late-onset children. I argue that the extrapolation of conclusions from studies where the outcome is timing of disease onset (early vs. late) onto studies concerned with the development of disease or not lacks robustness, especially when considering that the early- and late-onset diseases might relate to different pathogenic mechanisms.

Experimental evidence supports the differential autoimmune and histological profiles of the early-onset vs. late-onset T1D. Cases presenting at age less than 5 years of age were found to have higher IAA levels but not GADA in contrast to cases presenting at an older age, concluding that IAA positivity was associated with younger age[381]. Children with an early-onset often have a more aggressive clinical presentation of T1D as they are less likely to have a honeymoon period and are burdened by severe complications later in life due to prolonged disease duration[381]. A recent study probed into the reasons why T1D in children diagnosed at an early age is more aggressive by analysing pancreatic tissue samples after death from people with the disease[382]. The researchers identified two distinct patterns of insulitis which differ in the proportion of infiltrating CD20+; children diagnosed by the age of 7 had high CD20+, thus losing β-cells more rapidly, compared with the late-onset children who had low CD20+ and retained ~40% of their insulin-containing islets, thus differentiating loss of β-cells mass as causal for the early onset vs. functional deficit as causal for the late onset of T1D[382].

2.4.3 The predictive power of birth weight on age of T1D onset

The distinction between early- vs. late-onset T1D justifies the search for the predictive power of birth weight on the age of clinical presentation, which was addressed by a few studies that reported inconsistent findings. A national survey in the British Isles[383] and two studies in Australia[10,357] found no differences in birth weight between early-onset (<5 years) vs. late-onset diabetic children inferring that birth weight has no effect by age of disease presentation. In contrast, the same count of studies conducted in Sweden[384], Australia[385] and Japan[386] found that children with early-onset T1D (<5 years) weighed significantly less at birth compared with late-onset diabetic children. Of note, small size at birth accelerated growth in T1D children with younger onset or across the board[384,386]. It is then possible that the observed association between birth weight and risk of T1D is mediated via postnatal growth[366].

2.4.4 Biological mechanisms

The biological mechanism of the association between birth weight and T1D remains unknown. Yet, the almost-linear correlation between the two variables echoes the supposition that poor intrauterine growth decreases, whereas excess growth increases, the risk of diabetes[369]. This has been interpreted upon the mechanistic assumption that a high growth rate increases the demand for insulin production, which acts as a stress factor on the insulin secreting cells, coined as 'β-cell overload', and accelerates the process of β-cell destruction[361,387]. The view was supported by the experimental finding that hyper-functioning β-cells were more prone to cytokine-induced damage than resting β-cells[104,388]. Since foetal insulin is an important growth factor[389], with the number of pancreatic β-cells dramatically increasing >130th fold between the

12th intrauterine week and the 5th postnatal month[144], children with higher intrauterine growth and eventual higher birth weight have more active pancreatic β-cells that secrete more insulin.

At the opposite end of the spectrum, lower birth weight was theorised to be in favour of reducing the risk of T1D but only at values within the physiological range, whereas severe weight deficit at birth could precipitate developmental disturbances that take the preponderance[390]. I argue that this thesis was recently refuted by the very large study from Sweden which found that very preterm (22 to 32 weeks) and very low birth weight (<1,500 g) babies — exposures which had been previously understudied due to their rarity — are at reduced risk of T1D[205]. These authors remarked that the persistence of the associations in the sibling control arm of the study precludes familial factors as a possible explanation and raises the possibility of a biological mechanism of foetal programming that protects vital organs[205].

Another attempt at explaining the association between birth weight and T1D risk is anchored on the notion that during foetal and perinatal life the immune system is immature and vulnerable to immunological triggers[380]. The independent findings that low birth weight was characterised by a significant reduction in the number of T lymphocytes and cell mediated immunity up to 5 years of age[391], and increased mortality from infectious diseases from 1 to 15 years of age[187], underpinned the proposition that low birth weight is linked to impaired cellular immune response that could impart a reduced risk of immune-mediated destruction of β-cells[365]. Insights into the molecular signatures of families with T1D, which are characterised by an innate inflammatory state irrespective of disease state, found temporal increases in the frequency of peripheral activated Treg cells in healthy siblings vs. probands[125]. It follows that the innate inflammatory state becomes more actively regulated in the healthy siblings in an age-dependent manner, which might also explain the decline in T1D susceptibility with age[125]. Perhaps, the suggestion that the sex differential programming *in utero* results in differences in the development of the immune system between male and female neonates[392] could provide a justification for sex-dimorphic observations in this domain, such as the association between increased T1D risk and increased birth weight in Finnish boys but not girls[390].

2.4.5 Length at birth

Population-based studies point to high birth weight conferring risk of T1D and low birth weight displaying protection against T1D. Length has received less attention but, even when studied, there was a high rate of failure to establish significant associations with T1D risk[364,377,378,380,384,393-397]. Studies which established significant associations reported that short birth length was protective[372] and longer newborns were marginally at increased risk of T1D[398,399], which resemble the associations with weight. However, inadequate sample size and inaccurate measurements of length at birth obscure the identification of clear associations.

2.4.6 Body composition at birth

After establishing a link between size at birth and risk of T1D, albeit of a small magnitude of effect, the question honed on the specific contributions of body composition to the development of the disease. One study proposed that birth length exerts a stronger influence than birth weight on the risk of T1D[390], but replication of the data would be of the essence. When considering the relative contributions of lean vs. adipose mass, a small study based on 26 T1D cases put forward that adiposity at birth has no influence on the development of T1D, hinged on the observation that lean, but not total, mid-upper arm circumference at birth significantly associated with risk of T1D[357]. Interpretation of findings pertaining to body composition should be made with caution. Fat and muscle areas, calculated from upper arm circumference and triceps skinfold thickness, are based on estimations which limit their accuracy[400]. Infancy is a time of rapid transition characterised by pronounced changes in compartment, tissue, and chemical composition, with noticeable sex-specific differences and possibly bias of estimates[206].

2.5 T1D RISK AND SIZE AT BIRTH: GENETIC STUDIES

2.5.1 Overview

Over 30 years ago, Court et al.[401] reported that diabetic children with HLA-B8 were of smaller stature than those without the antigen, suggesting that 'within the diabetic spectrum of genotypes the HLA system is relevant to growth and development'. The subsequent finding of an association between HLA genes and foetal growth rates[402] reinforces the pleiotropic effects of the MHC region on susceptibility to T1D and early growth (Figure 2-4).

Figure 2-4 | Possible relationship between HLA genes, size at birth and T1D risk.

Such findings have catalysed the shift towards population genetic studies as a means of i) exploring the role of the genetic constitution in early development in the normal population and ii) assessing the development or not of islet autoimmunity/T1D based on HLA risk stratification. A benefit of population genetic studies in infancy is that they overcome the limitation of the rare incidence of T1D by taking advantage of cord blood banks to stratify the general population by HLA risk type. 'High-risk' HLA haplotypes predispose to T1D and are

more frequently encountered in diabetics than non-diabetics; in contrast, 'protective' haplotypes are less frequent in diabetics that non-diabetics, and 'neutral' haplotypes are distributed with no significant variation between the two groups. The main approach used for interrogating the genetic associations between birth weight and T1D susceptibility variants has been the comparison of birth weight across different HLA risk segments of the population, or the calculation of the frequency of HLA alleles across the birth weight distribution.

2.5.2 HLA studies

The hypothesis that HLA genes associate with size at birth could shed light on the mechanisms by which higher birth weight operates to affect risk of T1D. On the grounds of HLA class II molecules conferring susceptibility or protection for T1D, it would be expected that an association between HLA class II, or a factor linked to it (e.g. cytokine), would be observable with birth weight[203]. Genetic association studies discussed in this review are summarised in Appendix I.

2.5.3 High-risk HLA

Quite a few studies from Scandinavia found a positive association between T1D high-risk HLA alleles and size at birth. A large population-based prospective cohort in Sweden promulgated associations of high relative birth weight with the HLA-risk genotypes *DQ8/06:04*, *DQ8/X*[14] and *DQ2/8*[89]. Length SDS at birth was also shown to correlate with high-risk HLA, confirming that intrauterine growth is under the influence of HLA genotypes[244]. Children with a high-risk HLA haplotype specific to the Finnish population were also shown to be significantly heavier at birth, whether they developed T1D or not[161].

Other studies failed to confirm these associations. Neither a case-control Norwegian study[403] nor analyses of the prospective DIABIMMUNE cohort[245,404] generated evidence of an association between T1D susceptibility HLA and birth weight. Similarly, a prospective Italian study reported no association between T1D risk HLA categories and weight or length at birth[405].

In contrast, a retrospective study in Norway identified an inverse association, i.e. lowest birth weight observed with T1D high-risk DQ2/8 carriers, suggesting that non-genetic factors are likely explanations for the relationship between birth weight and risk of T1D[15]. This conclusion is presumably predicated on the supposition that commonly observed phenotypes in diabetic children would be linked to dominant disease susceptibility loci, led by the *HLA* and *INS* VNTR. As much validity as there might be in this statement, the observation that progressively fewer diagnoses present the classical high-risk HLA genotypes[40] is an indicator that genes of smaller effect sizes might be at play and worth investigating. The TEDDY study

raised the possibility that the relative contribution of genetic influences on birth size varies by country; Swedish babies carrying the T1D high-risk genotypes DQ2/8 and DQ8/8 were longer at birth independent of parental height, whereas in the US the high-risk HLA carriers were longer and heavier but dependent on paternal physical characteristics[230].

2.5.4 Protective HLA

Intriguingly, population studies reported that children with the protective haplotype DQB1*06:03 were or tended to be heavier at birth, albeit of a small magnitude of effect[14,403]. This association was confirmed by studies in Swedish full-term infants[203] and Finnish children[161] for the DR13 haplotype, which is in strong LD with the DQB1*06:03. Further reinforcement is provided by a report that Norwegian children homozygous for the very protective DQB1*06:02 had the highest birth weight across all T1D HLA groups[15]. The DIABIMMUNE study identified a significant relationship between increased relative birth weight and protective alleles but only in cross combination with the strong risk allele DQB1*03:02[404].

2.5.5 Further insights

The findings presented here stemmed from prospective cohorts following T1D children. The inconsistent results may be explained by differences in T1D incidence in the background population, sizes of study cohorts, methodologies applied[404], or the highly diverse HLA class II background. In the aggregate, the findings point to HLA genes, whether predisposing or protective for T1D, exerting effects on birth weight. On the grounds of the protective T1D alleles at the *HLA-DQ* and *INS* loci associating with larger size at birth, Stene et al.[403] hypothesised that correcting for allelic variation at these loci would increase the OR between birth weight and risk of T1D. However, the authors found no significant interaction between birth weight and the protective *HLA-DQ* or *INS*, underscoring that the genetic effects on foetal growth are too small to be detected unless sample sizes are very large[403].

Alternatively, the genetic effects on growth might be modulated by another factor. This supposition is given momentum by the finding that prediabetic children showed increased growth rate in the first 18 months of life compared with their HLA-matched controls, but not with their non-HLA matched controls, indicating that the effect on growth in progressors is mediated by risk factors other than HLA[244]. As discussed in Chapter 1, viral infections constitute a candidate trigger of autoimmunity and increased relative birth weight. The positive association found between high-risk HLA and birth weight was aggravated by recurrent infections in pregnancy[89]. Infections appear to also influence the effect of protective HLA alleles on growth, directly or indirectly:

i. the HLA DRB1*13 allele, known to protect against infections (e.g. malarial anaemia, HPV), was associated with higher birth weight[203];
ii. the DQB1*06:03 had an increased risk of higher birth weight only when several gestational infections were reported[89].

Intriguingly, infections appear to switch on and off the association between autoantibodies and birth weight: a negative association between birth weight and GADA or IAA was observed only in the background of an infection[89]. A link between perinatal growth and islet autoantibodies has been suggested by other studies: i) GADA in the cord blood, but not IA-2A, reduced the risk of higher relative birth weight[14], ii) in contrast, IA-2A, but not GADA, measured at 1 month of age, associated with early growth, i.e. increased BMI, in prediabetic children[406]. These findings collectively underscore that environmental factors and disease traits might influence anthropometric measures, thus highlighting the need to tease out the possible intervening effects of the disease, as well as its prodrome of islet autoimmunity, when investigating associations between genetic variants and growth.

2.6 TID RISK AND POSTNATAL GROWTH

2.6.1 Context

Widdowson[248] described that 'growth in humans is depicted by a sigmoidal curve with its inflection point marking birth'. It is then conceivable that foetal growth and postnatal growth lie on a single continuum under the influence of insulin supply. Just as enhanced foetal growth is closely connected with hyper-insulinaemia as evidenced by infants with nesidioblastosis or the Beckwith-Wiedemann syndrome[232], rapid early growth associates with higher insulin secretion[406]. Whereas birth weight is a crude marker of intrauterine growth rate, detailed patterns of post-delivery growth are richly informative of the postnatal environment, believed to influence risk of islet autoimmunity and T1D as suggested by population studies.

The spectrum of growth parameters investigated by these studies includes *weight gains*, *linear growth* and *BMI changes*. The greater emphasis on weight vs. length possibly arises from their differential rate of change in early life. Birth weight triples during the first year of life, whereas height in infancy takes four years to double[364]. In addition, length is notorious for the reduced accuracy of its measurement in neonates. Studies on the association between postnatal growth and risk of autoimmunity and T1D discussed herein are summarised in Appendix I.

2.6.2 Critical window of development

The identification of the critical window of development that affects risk of T1D was made possible by numerous studies with converging results. A study on Swedish longitudinal data showed that children who developed T1D before 5 years of age gained significantly more weight than controls between 3 months and 3 years of age, whereas at 5 years of age the two groups weighed about the same[396]. Further honing of this critical window in infancy was achieved by a Dutch family-controlled study which showed that beyond the first year of life there was no continued tendency for weight gain in prediabetic children vs. their healthy siblings[406]. The results were replicated in the sequel, a slightly larger study, which found significant differences in weight SDS gain and height SDS gain between cases and family controls in the first year of life but not after, with early onset of T1D relating to growth gains in just the first six months of life[364]. The importance of weight gains within the first six months of life, vs. subsequent weight gains, for T1D risk was recently corroborated in a large population-based study abstracting data from the Norwegian and Danish birth cohorts[407].

The pivotal importance of the early post-delivery window is maintained even when islet autoimmunity is used as the outcome variable instead. The TEDDY study recently reported that weight z score at 12 months, but not at 24 or 36 months, was a risk factor for development of islet autoimmunity[356]. A prospective Australian cohort of children at increased risk of T1D showed that prediction of islet autoimmunity from weight SDS and BMI SDS was significant between birth and 2 years, but not 4 years[408]. The prospective DAISY cohort examined these relationships beyond the first 2 years of life and showed that increased height growth velocity consistently associated with the development of islet autoimmunity (0.18 cm/year), and also associated with progression from islet autoimmunity to T1D (0.54 cm/year)[207].

The last observation attests to the proposition that height gains endure beyond the post-delivery window, as initially suggested by the finding that ‘children who later develop diabetes, especially boys, tend to be slightly but consistently taller, although not heavier, than matched referent children from early childhood up to the year of the clinical onset of T1D’[409]. These results might imply that weight gains are pronounced in infancy of prediabetics, whereas height gains overtake later in life. Considering that growth gains in prediabetic infants are higher but not abnormal[212], it is valid to assume that rapid growth is most discernible in the first 1 to 2 years of life, beyond which the effect attritions in magnitude. Important to my study, the genetic component of growth variations is expected to be larger postnatally when lifestyle variation is limited[410].

2.6.3 Postnatal weight gains

Initial evidence that altered growth is a prelude to T1D dates back to 1975 when a study in the UK reported that increased weight gain in the first year of life preceded onset of clinical diabetes in both boys and girls[4]. Similarly, a subsequent retrospective case-control study in Sweden showed that weight gain from birth to 30 months of age was significantly greater in diabetic children aged <15 years than in referent children (matched in a ratio of 1 to 3 for age, sex and geographical region), even amongst exclusively breastfed children who later developed T1D[384]. Along the same lines, a nationwide case-control study of Finnish diabetic children matched to population-based control children by birthday and sex identified an association between greater weight gain in the first year of life, albeit not abnormal, and 1.5-fold risk of T1D, even when adjusting for early introduction of formula feeding[212]. The finding was replicated in the follow-up analysis which quantified that a 10% unit increment in relative weight predicted 50%-60% increase in T1D risk in children under the age of 3 years[393].

Three multi-centre studies across populations with a range of incidence rates gave credence to the association between early weight gains and risk of T1D. The EURODIAB study across five participating centres found that weight significantly increased among patients from 1 month after birth to a maximum of 0.41 SDS between 1 to 2 years of age[399]. The prospective TRIGR study mapped Northern Europe, which reportedly has the highest incidence of T1D, to the highest weight z-score between birth and 12 months of age, in contrast to Southern Europe and US which had the lowest weight z-scores at most time points[411]. Ecological studies provide an alternative means of hypothesis testing by replacing cases and controls with countries of high and low incidence respectively, but interpretation should be made with caution as they are hampered by the 'ecological fallacy', meaning that associations on the aggregate level may exaggerate, or reverse, true associations in individuals[412]. Recently, the prospective TEDDY study has found relative weight z-scores at 12 months to be associated with development of islet autoimmunity, albeit not with T1D, in boys and girls[356].

To my knowledge, only two studies failed to identify an association between weight gain in infancy and risk of T1D. Both were retrospective, register-based, case-control investigations of anthropometric observations collected, via questionnaires sent to families and/or health records, from birth to age of T1D onset in Swedish children[409] and for the first year of life in Danish children[398]. The lack of associations with weight is possibly attributed to the use of registry data in both studies, which hampers detailed growth patterns, in addition to inadequate power in the latter study.

2.6.4 Postnatal height gains

In 1992 Blom et al.[409] in Sweden reported that rapid linear growth is a risk factor for T1D in childhood, which might either be a promoter of the disease or a marker of a physiological mechanism that affects both growth and the pathogenesis of autoimmunity. The contribution of accelerated height gains in infancy to risk of T1D was further replicated by European retrospective population-based case-control studies[393,399,406], as well as a retrospective family case-control study in the Netherlands that matched each diabetic child to at least one unaffected healthy sibling, as opposed to population controls, as a way of correcting for confounders relating to family characteristics (e.g. socioeconomic status, feeding habits)[364]. Conversely, a study on large prospective cohorts of infants at familial risk of T1D in Germany has recently made the novel observation that rapid normalisation of very high length SDS at birth to average values up to 3 years of age was protective against islet autoimmunity[371].

2.6.5 Weight gains vs. height gains

Whereas Blom et al.[409] in 1992 found an association for height but not weight, a subsequent longitudinal study, also in Sweden[396] and spanning from birth until disease onset, identified that both weight gains and height gains were more pronounced in prediabetics than controls. Studies identifying associations for both parameters found a slightly stronger contribution of weight vs. height to T1D[384,393,399], which was speculated to be attributed to nutritional factors[399].

A sex-specific nature was identified in the association between altered growth and T1D, exemplified by studies which claimed that differences in weight gain between probands and controls were more pronounced in girls than in boys[212,384,396]. In contrast, differences in height gain between probands and controls were more marked for boys than girls[393,409].

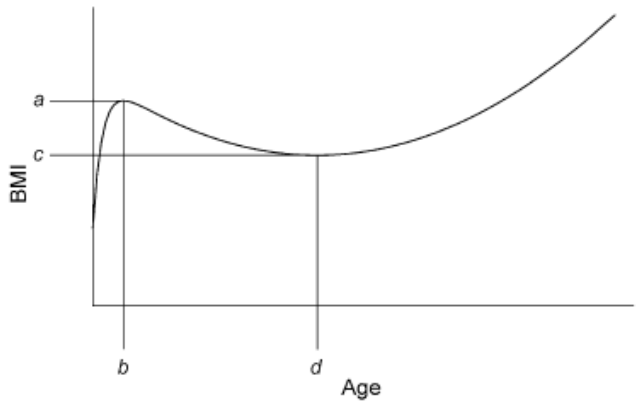
The gamut of studies discussed here includes case-control studies, which are prone to selection and information bias, and cohort studies in susceptible individuals, which are hampered by limited external validity since associations in high-risk populations might not be generalisable. These fallacies were recently addressed by a prospective population-based study of circa 100k babies — the outcome of merging the birth cohorts of Norway and Denmark — which elucidated that T1D positively associated with weight increase in the first year of life without sex-specific differences, but did not associate with linear growth[407]. The finding of this population-based study, in which clinical T1D was the endpoint, resonates with the conclusion reached by a prospective study in at-risk individuals which showed that weight z score, but not length or height z score, were continuous predictors of islet autoimmunity even after adjusting for HLA type[408].

2.6.6 Postnatal BMI

The BMI offers the advantage of adjusting weight for height. However, Weber et al.[206] warned that in the paediatric population the assumption that BMI is independent of height is violated. In addition, the variability in FFM vs. FM in children suggests that different combinations of the two types of body mass could yield the same BMI value, thus rendering the accuracy of BMI dependent on degree of fatness, i.e. BMI is not accurate among thinner children[413].

The trajectory of infancy BMI is not regular or consistent across individuals but is marked by two turning points (Figure 2-5). BMI peaks near the end of infancy, known as the infancy 'BMI maximum', after which it slightly decreases reaching a nadir at around age 5 to 6 years, known as the 'BMI rebound', before it starts to increase again between 3 to 11 years of age[206,414]. The timing of the BMI rebound is under genetic influence[415]. This renewed rise in adiposity is coined 'adiposity rebound', with an early rebound (<5.5 years) followed by a significantly higher adiposity level than a later rebound (>7 years)[416].

Figure 2-5 | BMI trajectory from birth. From Lammi et al.[414].



a: BMI peak
b: age at which BMI peak is reached
c: BMI rebound
d: age at which BMI rebound is reached

The irregularity of the BMI trajectory across infants might account for the contradictory findings reported for BMI in relation to T1D risk. A longitudinal study from Sweden spanning from birth until onset of disease reported that prediabetics had similar BMI with controls in the first three years of life, whilst deviations were detected between 5 to 13 years of age[396]. The finding that T1D risk associates with later vs. early accelerated BMI gains is in accord with two other studies: i) a Dutch family-based study did not detect BMI SDS gains in the first 4 years of life[364], ii) analysis of the 1970 British Birth Cohort found that higher BMI z-score at 10 years of age was a predictor of subsequent T1D[417].

In contrast, a Dutch study found that prediabetic children had a tendency for an increased rate of BMI growth in the first year of life which did not continue beyond[406]. Similarly, the EURODIAB study[399] reported significant excesses of BMI from 6 months of age, with the largest divergence observed at 1 to 2 years of age. A hint of BMI gains as a prelude to T1D

was also provided by a population-based study in Denmark, which found an association between BMI at age 1 year and risk of T1D, albeit of a very small sample size of 13 cases and 6 controls[398]. BMI gains over the first 3 years of life were recently found to be predictive of islet autoimmunity in a prospective study in Germany[371].

The caveat of these studies is that they don't allow for interindividual variation of BMI curves. This limitation has been recently addressed by complex modelling that factors in the considerable differences in the early growth patterns between subjects. A Finnish study reported that higher infancy BMI max before the age of 3 years increased the risk of T1D in young adults whereas the age of BMI max had no influence[414]. In contrast, a prospective study in Germany reported that i) early age of infant BMI peak (mean=9 months of age), or ii) high difference between BMI peak and BMI rebound (mean=1.6 kg/m^2) associated with increased risk of islet autoimmunity[355].

Both studies underscore the importance of making comparisons of growth data at the same stage of development to account for interindividual differences in growth trajectories, as opposed to *a priori* time points. The pooled findings suggest that an accelerated BMI trajectory in infancy acts as a risk factor for T1D, whether it results from an upward shift of the y-coordinate, i.e. BMI max, or a right-to-left shift of the x-coordinate, i.e. age, of the BMI curve.

2.6.7 Postnatal growth gains and insulin resistance

Contrary to expectations, the prospective DIABIMMUNE study showed that children carrying the T1D high-risk HLA gained significantly less weight and less height during the first 24 months vs. children in other HLA groups[245]. A possible explanation was anchored on the assumption that rapid weight gain associates with insulin resistance[418], with the implication that neutral and protective HLA reflect higher insulin resistance that accelerates weight gain[245]. The involvement of insulin resistance as a suspect pathogenic actor in the early growth of prediabetics was hinted in a retrospective Italian study which explored the influence of birth weight on the clinical phenotype of patients with T1D[419]. The authors inferred that diabetics born SGA who maintained a normal BMI had increased insulin requirement not associated with poor metabolic control — an indirect index of insulin resistance — suggesting that SGA patients have reduced insulin sensitivity[419]. This convoluted line of evidence conjures on the hypothesis that small size at birth and eventual rapid growth are predicated on insulin resistance. The speculation that insulin resistance influences development of T1D was recently refuted by prospective studies: the BABYDIAB study found that islet-autoantibody children had similar insulin resistance and BMI at the time of seroconversion with autoantibody negative children[420], and the DIPP study showed that whereas insulin secretion was perturbed in the long presymptomatic stage of T1D, insulin sensitivity remained unaffected[148].

2.6.8 Biological determinants of postnatal growth in prediabetics

Observational studies have shown that children who develop T1D, or islet autoimmunity, are heavier[4,212,356,393,399,407], taller[393,399,409], gain weight faster[384], have excesses of BMI[398,399], or gain BMI faster[371,406], as a result of higher[414] or earlier BMI peak[355], during infancy. Conversely, normalisation of very high relative height at birth to average values in infancy is protective against islet autoimmunity[371]. Of note, studies that established associations between risk of T1D and early growth failed to detect an association with birth weight, casting light on the magnitude of influence exerted by foetal vs. postnatal growth on the risk of T1D.

Having established an association between rapid growth in the first 1 to 2 years of life and risk of T1D, the question hones on the biological pathways. Key biological candidates are insulin and IGF-1, which play important roles in perinatal growth and glucose homeostasis[346]. A Dutch family-controlled study showed that non-carriers of the protective *INS* VNTR class III allele showed increased growth in controls only, with the absence of an effect in patients suggesting that growth operates via a different mechanism in disease vs. health[346]. The study also found that the IGF-1*194 allele was less frequent in cases than controls but its effect on rapid growth was independent of disease[346]. In the German BABYDIAB and BABYDIET studies, the IGF-1 levels were similar between children with vs. without islet autoimmunity, but lower at 9 months of age vs. birth in children at familial risk of T1D, suggesting that *decreases* in IGF-1 levels in infancy signals risk of pathogenesis[355]. A follow-up study of the DIABIMMUNE cohort found that IGF-1 levels and the IGF-1/IGFBP-3 ratio were lower in the autoantibody-positive children at 12 and 24 months respectively but the increase in IGFBP-3 levels was higher in the autoantibody-positive children before seroconversion vs. controls[421]. The puzzling isolated increment of IGFBP-3 was attributed to the disintegration of the GH/IGF-1 axis, alluding to normal levels of IGF-1 but higher GH as observed in T1D and obesity[422]. The study also found that IGF-1 and IGFBP-3 levels at 24 months of age were lower in high-risk HLA carriers, characterised by a slow-down in weight and linear gains by the end of infancy[245], vs. low-risk HLA carriers, suggesting that high IGF-1 levels protect against T1D[421].

2.7 CONCLUSION

Collectively, the findings concur on early growth being mediated via different mechanisms in health vs. disease, thus necessitating the study of genetic associations with physiological processes in a disease-free population as disease might influence gene expression. They further underscore the importance of longitudinal growth data via frequent time points, adequate sample size, and standardised measurement protocols to minimise information bias.

Chapter 3

COLLECTING AND EXPLORING THE PHENOTYPES

3.1 INTENT

The research project herein tapped into the wealth of infancy growth data and biological specimens collected by the completed 'Cambridge Baby Growth Study I' (referred to as CBGS), which monitored infants from the general population in Cambridgeshire over the first 24 months of life. The study was conceived and coordinated by a team of clinical investigators at Addenbrooke's Hospital. Recruitment and assessment were performed by four trained paediatric research nurses at perinatal and postnatal clinic visits according to the study protocol. I familiarised myself with data collection by shadowing the nurses at the Medical Research Council (MRC) Epidemiology Clinical Research Facility during clinic visits of the ongoing extension to the study, the CBGS II, which uses the same protocol. My research concerned the investigation of genotype-phenotype associations in early life that are pertinent to clinical traits pervasive of the prodrome of T1D. I implemented this by a biology-guided screening of genetic variants to genotype from DNA of infants, coupled with analyses of growth, hormonal and genetic data by constructing and testing statistical models based on conventional (cross-sectional) and advanced (longitudinal) statistical techniques.

This chapter explores the measured anthropometric and hormonal traits — denoted as 'phenotypes' — in the context of quantifiable covariates in the CBGS, for a two-fold purpose:

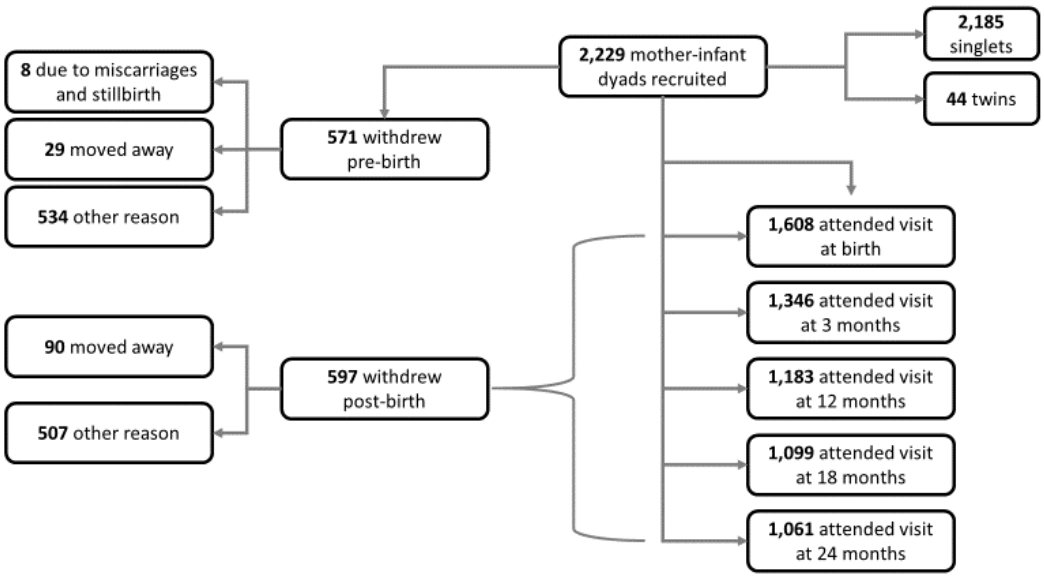
1. Exploration of cohort data
    i. gain an overview of cohort descriptive characteristics;
    ii. understand postnatal growth patterns and underlying endocrine factors in the background population of infants, which serves as a springboard for defining and explaining 'normal growth', and further determines exclusion criteria for analyses;
    iii. evaluate the robustness and limitation of measurements.
2. Definition of statistical framework
    i. check assumptions required for statistical tests;
    ii. identify confounders to adjust for in statistical models, whereby phenotypes are compared amongst genotypic groups of T1D susceptibility loci.

3.2 CAMBRIDGE BABY GROWTH STUDY

3.2.1 Cohort

The CBGS is a prospective observational birth cohort study conducted at a single centre, the Rosie Maternity Hospital in Cambridge, UK, whose aim was the investigation of the antenatal and postnatal determinants of infancy growth and reproductive development, including environmental, genetic, hormonal and nutritional exposures[423]. Recruitment commenced in April 2001 and ended in August 2009. Approximately 2,200 expecting mothers aged ≥16 years were approached and recruited by trained paediatric research nurses at their first routine antenatal ultrasound clinic appointment at 12 weeks of gestation and followed up at 28 weeks of gestation for a 75-g OGTT. Participating offspring were assessed by the research nurses at a minimum of one study visit scheduled within one week from birth, and at ages 3, 12, 18 and 24 months. Weight, height, head circumference and skinfold thickness were routinely measured at each visit. Genomic DNA was extracted from cord blood samples collected at birth upon consent by eligible families, and/or infancy capillary blood samples or buccal swabs. Dried blood spot (DBS) samples were routinely collected from 3 to 24 months of age for assaying hormones. All mothers gave written informed consent to participate in the CBGS. The study was approved by the Cambridge local research ethics committee. Out of 2,229 mother-infant dyads approached, a total of 1,658 enrolled in the study corresponding to a 74% participation rate (Figure 3-1). Measures of birth weight were recorded for two additional infants that were forced to withdraw from the study prior to birth, bringing the total count of subjects with anthropometric data to 1,660.

Figure 3-1 | Flowchart of the study population. Adapted from Prentice et al.[423].



3.2.2 Endpoints and methods

3.2.2.1 Anthropometric measurements

Infancy anthropometric endpoints were collected in triplicates (except for birth weight which was retrieved from maternity hospital records) and used to derive composite anthropometric indices and infancy gains (Table 3-1). Weight was measured to the nearest 1 g using a SECA 757 electronic digital scale (Chasmors Ltd). Supine length was measured to the nearest 0.1 cm using a Kiddimeter (Holtain Ltd, Crosswell, Pembs, UK) and standing length with a stadiometer. Head circumference was measured to the nearest 0.1 cm using a tape measure. Skinfold thickness was measured at four sites (triceps, subscapular, flank and quadriceps) on the left side of the body using a Holtain Tanner/Whitehouse Skinfold Caliper (Holtain Ltd): the triceps skinfold was measured halfway between the acromial process and the olecranon; the subscapular skinfold was measured at the oblique angle below the scapula; the flank skinfold was measured in the posterior axillary line immediately posterior to the iliac crest; the quadriceps skinfold was taken from a vertical line in the middle of the thigh and half way between the top of the patella and the inguinal crease.

Table 3-1 | Anthropometric endpoints by infant's age in months.

	Age (mo)					Gains (mo)	
	0	**3**	**12**	**18**	**24**	**0-12**	**0-24**
Measured anthropometric endpoint							
Weight	•	•	•	•	•	•	•
Height	•	•	•	•	•	•	•
Head circumference	•		•	•	•		
Skinfold thickness: triceps	•	•	•	•	•		
Skinfold thickness: quadriceps	•	•	•	•	•		
Skinfold thickness: flank	•	•	•	•	•		
Skinfold thickness: subscapular	•	•	•	•	•		
Derived anthropometric endpoint							
Skinfold thickness mean	•	•	•	•	•	•	•
Ponderal Index	•						
BMI	•		•	•	•		

Replicates (≤3) of raw measurements were assessed, averaged and appropriately standardised by Dr Philippa Prentice, one of the study investigators[424]. Weight, height, head circumference and BMI had been converted to age- and sex-appropriate SDS using the computerised 1990 British standard that simplifies growth assessment[201]. The score

represents the number of standard deviations that the value differs from the mean of the distribution of the 1990s British babies of the same age and sex. The British 1990 cohort was the reference of choice because it was relatively contemporaneous with the CBGS, it was a Caucasian cohort, and allowed for newborn measurements to be adjusted for gestational age[424]. Anthropometric indices derived were i) BMI, calculated as weight in kg divided by height in meters squared, ii) ponderal index, calculated as weight in kg divided by length in meters cubed at birth. Measurements of skinfold thickness at each anatomical site were converted to internal SDS adjusted for age (or gestational age for newborn measurements): SDS = individual measurement minus cohort mean, divided by the cohort SD; an overall skinfold thickness index was calculated in each infant, as an indicator of subcutaneous adiposity, by taking the mean of the SDS at the four sites[424]. Postnatal gains were calculated for weight and height by taking the difference in SDS between birth and 12 months or birth and 24 months[424]. A gain greater than 0.67 SDS, representing the width of each centile band on standard growth charts (e.g. 2nd to 9th, 9th to 25th, 25th to 50th) reflected significant catch-up growth; a reduction greater than 0.67 SDS was designated as catch-down category[276].

3.2.2.2 Biological sample collection

Biological specimen collected from infants of relevance to this thesis included:

i. buccal swab samples collected using mouthswabs, immersed in 2.5-ml transfer buffer inside a 13-ml Sarstedt tube, and stored at -20 °C until DNA extraction;
ii. capillary blood obtained by heel prick (Tenderfoot; Elitech UK, Barkhamstead, UK), which was partly transferred to a vial and stored at -20 °C until DNA extraction. The remaining blood sampled was blotted onto Whatman™ 903® untreated marked filter cards (Ahlstrom 226, ID Biological Systems) until the spot had spread about 1.5 cm in diameter, air-dried at room temperature overnight and stored in plastic bags at -20 °C until assaying. Prior to biochemical assaying, blood-spot disks of 3.2-mm in diameter were punched out from the filter card into 2-ml polypropylene screw cap microtubes.

Methods for DNA extraction and biochemical analyses are detailed in Appendix II.

3.2.2.3 Questionnaires

Contextual cohort data — sociodemographic status, self-reported maternal anthropometry, maternal general health and diet, and environmental exposures during pregnancy — were gathered through a perinatal questionnaire administered in late pregnancy. Mode of infant feeding at 3 months of age (breast milk and/or formula milk) was assessed by a postnatal questionnaire completed at the designated study visit.

3.3 PHENOTYPE STATISTICS

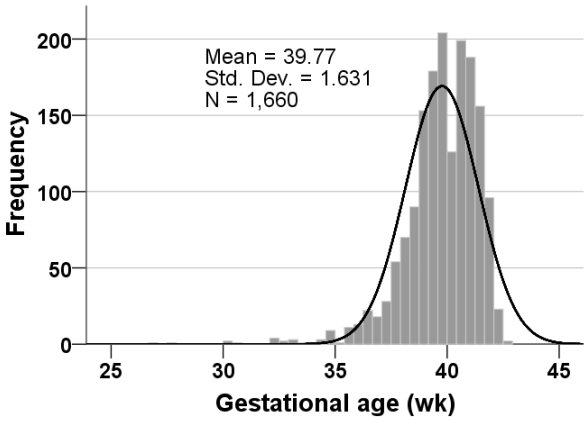
3.3.1 Statistical analysis

Statistical analyses were performed using the IBM Statistical Package for Social Sciences (SPSS 23.0, Chicago, IL, USA) for Windows. Categorical variables were summarised as percentages and continuous variables as mean ± SD. Descriptive trajectories of phenotypes were plotted by averaging data of all available subjects to make most use of measurements. Tests for normality for all response variables are shown in Appendix II. Comparisons of means were made with the independent t-test and one-way ANOVA between 2 and >2 groups respectively where the data approximated a normal distribution. Non-parametric tests were used where distributions deviated from normality. On the grounds of parametric tests being robust to some degree of non-normality for large samples and in the absence of influential outliers[425], parametric tests were used for all standardised anthropometric variables. Correlations between two continuous variables were calculated with the Pearson's correlation coefficient *r* for normal distributions, or Spearman's *rho* for distributions that did not conform to normality. Comparisons of percentages across groups of categorical variables were made with the Pearson's chi square test. A $P \leq 0.05$ was considered statistically significant.

3.3.2 Cohort characteristics

Of the 1,660 children, there were 859 boys and 801 girls. Amongst these, there were 22 pairs of twins. Almost 95% of children were born at term (gestation ≥37.0 weeks). Gestation ranged from as low as 27.0 up to 42.7 weeks, evident of a left-skewed distribution (Figure 3-2).

Figure 3-2 | Distribution of gestational age.



SGA was defined as weighing ≤-1.5 SDS at birth and comprised 4.7% of infants, whereas LGA was defined as weighing ≥1.5 SDS at birth and represented 7.2% of the cohort. This definition

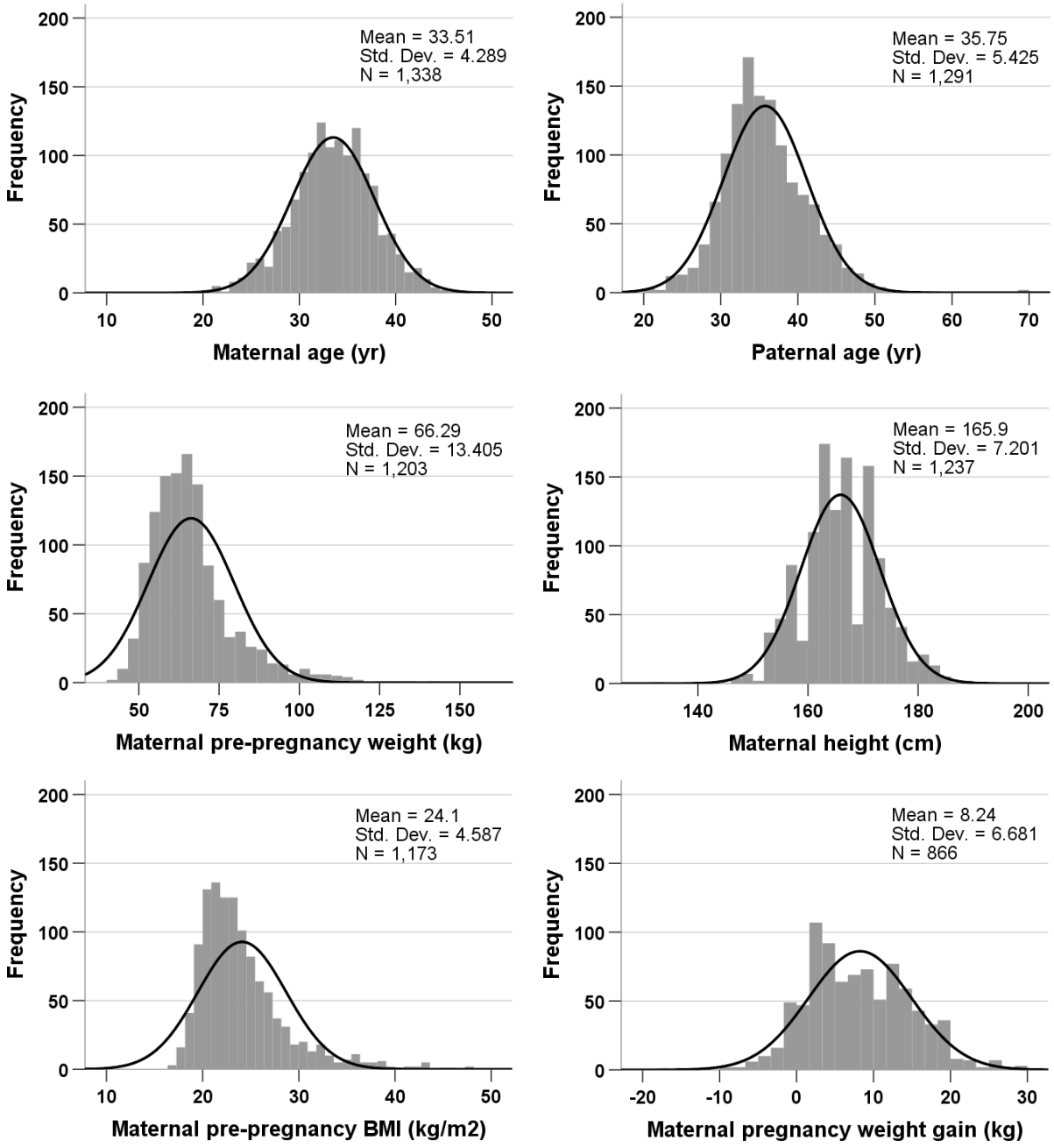
was chosen over the more stringent one of ±2.0 SDS of mean birth weight, corresponding to 1.1% and 2.7% of SGA and LGA infants respectively, to increase the count of subjects in these extreme groups and allow for meaningful subgroup analyses (see Chapter 5). In addition to data collected on the infant, contextual information was gathered on the parents and pregnancy for profiling the cohort, and more importantly, defining the framework for statistical analysis. The characteristics of the CBGS cohort are summarised in Table 3-2.

Table 3-2 | Cohort characteristics.

	N	Mean ± SD or n (%) for specified outcome	Range
Newborn index			
Gestation (wk)	1,660	39.8 ± 1.6	27.0, 42.7
Twins	1,660	44 (2.7)	
Infant sex: male	1,660	859 (51.7)	
Premature birth: < 36 weeks of gestation	1,660	39 (2.3)	
Relative birth weight	1,657		
appropriate for gestational age		1,460 (88.1)	
large for gestational age		119 (7.2)	
small for gestational age		78 (4.7)	
Pregnancy influences			
Maternal T1D	1,638	5 (0.3)	
Primiparous pregnancy	1,629	707 (43.4)	
Maternal smoking during pregnancy	1,575	86 (5.5)	
Maternal height (cm)	1,237	165.9 ± 7.2	132.0, 193.0
Maternal pre-pregnancy weight (kg)	1,203	66.3 ± 13.4	40.0, 143.2
Maternal pre-pregnancy BMI (kg/m^2)	1,173	24.1 ± 4.6	16.6, 48.1
Maternal pregnancy weight gain (kg)	866	8.2 ± 6.7	-17.3, 29.9
Maternal age (yr)	1,338	33.5 ± 4.3	18.1, 47.5
Paternal age (yr)	1,291	35.7 ± 5.4	20.0, 69.7
Post-pregnancy influences			
Birth delivery	1,630		
normal		956 (58.7)	
vacuum		101 (6.2)	
forceps		82 (5.0)	
elective caesarean		252 (15.5)	
acute caesarean		239 (14.7)	
3-month feeding: exclusively breastfed	1,319	557 (42.2)	
Sociodemographic influences			
Ethnicity: White Caucasian	1,098	1,045 (95.2)	
Index of multiple deprivation	920	8.9 ± 4.2	3.0, 32.1
Maternal marital status	1,291		
married		1,070 (82.9)	
co-habiting		193 (14.9)	
single		28 (2.2)	
Maternal education	780		
GCSE		110 (14.1)	
A-Level		181 (23.2)	
University		489 (62.7)	

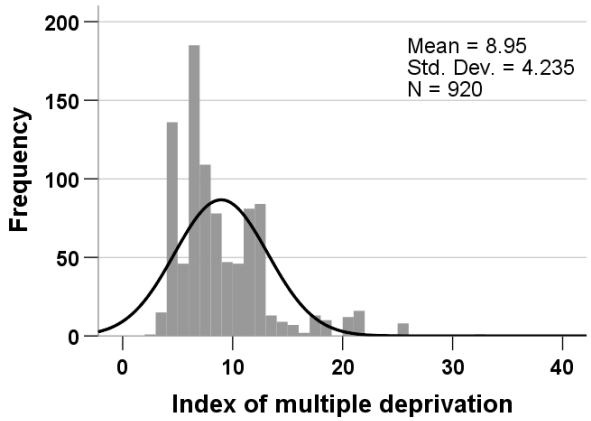
Distributions of parental traits are depicted in Figure 3-3. Only maternal age at birth was normally distributed ($P_{Kolmogorov\text{-}Smirnov}$=0.200) with a mean value of 33.5 years. Paternal age at birth was slightly higher at a mean value of 35.7 years, but its distribution deviated from normality ($P_{Kolmogorov\text{-}Smirnov}$<0.0001) with influential outliers as high as 69 and 70 years of age. Exceptions to normality were markedly shown by maternal pre-pregnancy weight and BMI ($P_{Kolmogorov\text{-}Smirnov}$<0.0001), whose distributions were right-skewed, indicative of a cohort with overweight women. The mean maternal weight before pregnancy was 66.3 ± 13.4 kg at a mean height of 165.9 ± 7.2 cm, corresponding to a BMI of 24.1 ± 4.6 kg/m^2, which sits just below the overweight cut-off for women. An average of 8.2 ± 6.7 kg was gained during pregnancy, with 53 women incurring weight loss during pregnancy.

Figure 3-3 | Distributions of parental characteristics.



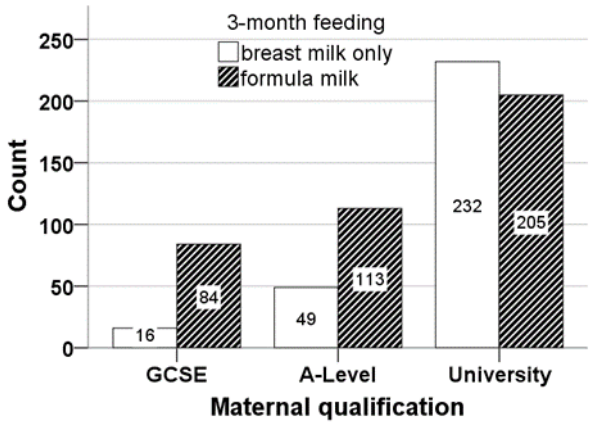
Ethnically, the CBGS cohort was of Caucasian ethnicity in the majority (95.2%). The Index of multiple deprivation (IMD) had a mean value at the lower end of the spectrum, i.e. less socio-economically deprived, but deviated from a normal distribution ($P_{Kolmogorov\text{-}Smirnov}$<0.0001) with a marked skewness to the right indicative of social clusters of higher deprivation (Figure 3-4).

Figure 3-4 | Distribution of parental Index of multiple deprivation.



Most participating mothers were married (82.9%) and educated at university degree-level (62.7%). Exclusive breastfeeding vs. formula feeding at 3 months of age was the preferred choice amongst mothers who had university degrees than not ($P_{chi\text{-}square}$<0.0001) (Figure 3-5), underscoring the influence of maternal education on infant feeding practice.

Figure 3-5 | Occurrence of infants by mode of feeding at 3-months and maternal education.



Overall, the recruited term and late-preterm (gestation ≥36.0 weeks) mother-infant dyads were representative of the overall population of births at the Rosie Maternity Hospital, except that CBGS mothers were slightly older at delivery and pregnancies were more likely to be primiparous[423]. The 1,061 (63.9%) mother-infant dyads who completed the infancy study protocol up to 2 years of age were similar to those lost to follow-up, although both parents were slightly older, and mothers were up to three times less likely to smoke during pregnancy (Table 3-3). The difference in birth weight was of small magnitude and nominal statistical significance.

Table 3-3 | Comparison of mother-infant dyads who were followed up vs. those lost to follow-up.

	Followed-up		Lost to follow-up		
	N	Mean ± SD or n (%) for specified outcome	N	Mean ± SD or n (%) for specified outcome	*P* value
Newborn index					
Gestation (wk)	1,061	39.8 ± 1.6	599	39.7 ± 1.7	0.197
Twins	1,061	30 (2.8)	599	14 (2.3)	0.550
Infant sex: male	1,061	555 (52.3)	599	304 (50.8)	0.542
Premature birth: < 36 weeks of gestation	1,061	24 (2.3)	599	15 (2.5)	0.754
Relative birth weight	1,058		599		0.607
appropriate for gestational age		927 (87.6)		533 (89.0)	
large for gestational age		81 (7.7)		38 (6.3)	
small for gestational age		50 (4.7)		28 (4.7)	
Birth weight (kg)	1,058	3.50 ± 0.54	599	3.44 ± 0.55	0.026
Birth length (cm)	1,024	51.4 ± 2.7	575	51.3 ± 2.5	0.358
Newborn head circumference (cm)	1,025	35.3 ± 1.7	577	35.3 ± 1.7	0.904
Newborn BMI (kg/m^2)	1,021	13.2 ± 1.6	575	13.0 ± 1.7	0.058
Newborn mean skinfold thickness (mm)	1,022	6.3 ± 1.6	577	6.3 ± 1.6	0.372
Pregnancy influences					
Maternal T1D	1,046	3 (0.3)	592	2 (0.3)	0.857
Primiparous pregnancy	1,042	452 (43.4)	587	255 (43.4)	0.980
Maternal smoking during pregnancy	1,002	34 (3.4)	573	52 (9.1)	<0.0001
Maternal height (cm)	917	166.1 ± 7.0	320	165.4 ± 7.7	0.268
Maternal pre-pregnancy weight (kg)	895	66.2 ± 13.1	308	66.5 ± 14.3	0.629
Maternal pre-pregnancy BMI (kg/m^2)	871	24.0 ± 4.5	302	24.3 ± 4.9	0.518
Maternal pregnancy weight gain (kg)	642	8.3 ± 6.7	224	7.8 ± 6.5	0.267
Maternal age (yr)	1,034	33.7 ± 4.2	304	32.8 ± 4.5	0.001
Paternal age (yr)	963	36.1 ± 5.3	328	34.7 ± 5.6	<0.0001
Post-pregnancy influences					
Birth delivery	1,039		591		0.043
normal		599 (57.7)		357 (60.4)	
vacuum		78 (7.5)		23 (3.9)	
forceps		55 (5.3)		27 (4.6)	
elective caesarean		162 (15.6)		90 (15.2)	
acute caesarean		145 (14.0)		94 (15.9)	
3-month feeding: exclusively breastfed	1,022	439 (43.0)	297	118 (39.7)	0.322
Sociodemographic influences					
Ethnicity: White Caucasian	738	703 (95.3)	360	342 (95.0)	0.852
Index of multiple deprivation	779	8.9 ± 4.2	141	9.0 ± 4.5	0.963
Maternal marital status	966		325		0.959
married		799 (82.7)		271 (83.4)	
co-habiting		146 (15.1)		47 (14.5)	
single		21 (2.2)		7 (2.2)	
Maternal education	568		212		0.050
GCSE		82 (14.4)		28 (13.2)	
A-Level		119 (21.0)		62 (29.3)	
University		367 (64.6)		122 (57.5)	

3.3.3 Infancy anthropometry

The exploratory analyses of infant anthropometry and growth were conducted on the entire cohort for elucidating subsets of infants that showed marked deviations from the average growth trajectories. Unless otherwise stated, anthropometric variables are in SDS.

3.3.3.1 Anthropometric variables

Weight, height, head circumference and BMI had been standardised against the British 1990s reference. The effectiveness of this conversion was confirmed by i) plotting the SDS distributions and showing that they conform to standard curves, i.e. centred at zero (histograms not shown), and ii) performing an independent t-test to show that the statistically significant differences of the pre-standardised measurements between males and females largely disappear with the standardisation. Table 3-4 summarises these anthropometric variables by age and sex of infant.

Measures of skinfold thickness, before and after standardisation against an internal reference, are summarised in Table 3-5. Consistently at every age measured, the quadriceps skinfold made the biggest contribution to overall skinfold thickness in mm (32% to 41%), followed by the flank (23% to 25%), the triceps (19% to 23%) and the subscapular (16% throughout infancy down from 21% at birth). Girls presented with, or showed a trend for, higher skinfold thickness than boys. The highest sex-specific difference was observed for quadriceps with the gap progressively increasing in the second year of life; in contrast, adiposity in the flank did not differ between boys and girls beyond birth. Paradoxically, the dimorphic nature of adiposity disappeared at 3 months of age across all anatomical sites, which could not be attributed to statistical reasons as the sample size at this age is comparable with those at other ages measured.

Table 3-4 | Cohort growth characteristics by age and sex.

	Boys		Girls		
	N	**Mean ± SD**	**N**	**Mean ± SD**	***P* value**
Gestation (wk)	859	39.7 ± 1.7	801	39.8 ± 1.6	0.387
Birth					
Weight (kg)	858	3.54 ± 0.56	799	3.40 ± 0.52	<0.0001
Weight SDS	858	0.09 ± 0.97	799	0.07 ± 0.98	0.653
Height (cm)	828	51.8 ± 2.6	771	50.9 ± 2.6	<0.0001
Height SDS	828	-0.05 ± 0.92	771	-0.10 ± 1.02	0.316
HC (cm)	832	35.6 ± 1.7	770	34.9 ± 1.6	<0.0001
HC SDS	832	-0.14 ± 0.98	770	-0.14 ± 0.98	0.914
BMI (kg/m^2)	827	13.2 ± 1.6	769	13.1 ± 1.6	0.285
BMI SDS	822	-0.07 ± 1.24	764	0.03 ± 1.24	0.110
Ponderal Index	827	25.5 ± 3.4	769	25.8 ± 3.4	0.088
3 months					
Weight (kg)	703	6.43 ± 0.83	637	5.83 ± 0.72	<0.0001
Weight SDS	703	0.07 ± 1.01	637	-0.15 ± 1.04	<0.0001
Height (cm)	704	61.9 ± 2.5	631	60.3 ± 2.5	<0.0001
Height SDS	704	0.22 ± 0.94	631	0.07 ± 1.06	0.006
12 months					
Weight (kg)	623	10.32 ± 1.17	559	9.58 ± 1.09	<0.0001
Weight SDS	623	0.08 ± 1.08	558	-0.02 ± 1.09	0.131
Height (cm)	618	76.8 ± 2.8	557	75.0 ± 2.7	<0.0001
Height SDS	618	0.36 ± 1.07	556	0.28 ± 1.09	0.212
HC (cm)	621	47.0 ± 1.4	558	45.8 ± 1.2	<0.0001
HC SDS	621	-0.65 ± 1.14	557	-0.65 ± 1.05	0.949
BMI (kg/m^2)	618	17.5 ± 1.4	557	17.0 ± 1.4	<0.0001
BMI SDS	618	-0.16 ± 1.00	556	-0.24 ± 1.03	0.216
18 months					
Weight (kg)	569	11.70 ± 1.29	525	10.99 ± 1.25	<0.0001
Weight SDS	568	0.11 ± 1.03	525	0.04 ± 1.06	0.263
Height (cm)	571	83.0 ± 3.1	525	81.3 ± 3.1	<0.0001
Height SDS	570	0.37 ± 1.10	525	0.20 ± 1.09	0.011
HC (cm)	566	48.4 ± 1.4	522	47.2 ± 1.4	<0.0001
HC SDS	565	-0.74 ± 1.07	522	-0.73 ± 1.14	0.935
BMI (kg/m^2)	568	16.9 ± 1.2	524	16.6 ± 1.3	<0.0001
BMI SDS	567	-0.19 ± 0.93	524	-0.18 ± 0.98	0.765
24 months					
Weight (kg)	547	12.91 ± 1.45	504	12.27 ± 1.42	<0.0001
Weight SDS	546	0.19 ± 1.02	504	0.11 ± 1.04	0.199
Height (cm)	547	88.4 ± 3.4	496	87.0 ± 3.4	<0.0001
Height SDS	546	0.46 ± 1.07	496	0.31 ± 1.05	0.018
HC (cm)	548	49.3 ± 1.4	503	48.1 ± 1.3	<0.0001
HC SDS	547	-0.73 ± 1.06	503	-0.78 ± 1.07	0.382
BMI (kg/m^2)	543	16.5 ± 1.3	495	16.2 ± 1.2	<0.0001
BMI SDS	542	-0.17 ± 0.95	495	-0.18 ± 0.94	0.895

Table 3-5 | Skinfold thickness measurements by age and sex.

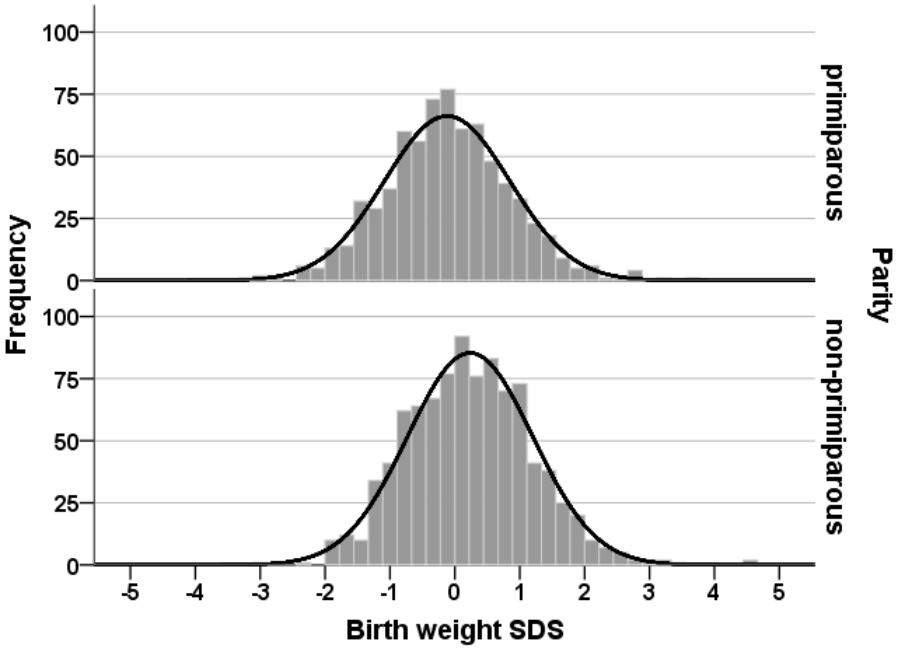
	Boys		Girls		
	N	**Mean ± SD**	**N**	**Mean ± SD**	***P* value**
Birth					
Triceps (mm)	830	5.5 ± 1.5	766	5.6 ± 1.4	0.295
Triceps SDS	826	0.0 ± 1.0	763	0.0 ± 1.0	0.369
Subscapular (mm)	830	5.3 ± 1.4	768	5.5 ± 1.3	0.005
Subscapular SDS	826	-0.1 ± 1.0	765	0.1 ± 1.0	0.007
Flank (mm)	828	6.0 ± 1.8	769	6.3 ± 1.8	0.001
Flank SDS	824	-0.1 ± 1.0	766	0.1 ± 1.0	0.001
Quadriceps (mm)	829	7.9 ± 2.6	768	8.3 ± 2.5	0.002
Quadriceps SDS	825	-0.1 ± 1.0	765	0.1 ± 1.0	0.002
Mean SDS	826	-0.1 ± 0.9	766	0.1 ± 0.9	0.005
3 months					
Triceps (mm)	706	8.4 ± 2.2	635	8.3 ± 2.1	0.447
Triceps SDS	706	0.0 ± 1.0	635	0.0 ± 1.0	0.451
Subscapular (mm)	706	7.0 ± 1.5	635	7.0 ± 1.6	0.698
Subscapular SDS	706	0.0 ± 1.0	635	0.0 ± 1.0	0.694
Flank (mm)	706	10.2 ± 2.7	636	10.2 ± 3	0.731
Flank SDS	706	0.0 ± 0.9	636	0.0 ± 1.1	0.734
Quadriceps (mm)	706	17.6 ± 3.6	636	17.8 ± 3.4	0.246
Quadriceps SDS	706	0.0 ± 1.0	636	0.0 ± 1.0	0.236
Mean SDS	706	0.0 ± 0.8	636	0.0 ± 0.8	0.914
12 months					
Triceps (mm)	621	9.6 ± 2.5	557	9.8 ± 2.7	0.127
Triceps SDS	621	0.0 ± 1.0	556	0.0 ± 1.0	0.140
Subscapular (mm)	620	7.2 ± 1.6	557	7.4 ± 1.7	0.131
Subscapular SDS	620	0.0 ± 1.0	556	0.0 ± 1.0	0.148
Flank (mm)	620	10.4 ± 2.8	557	10.6 ± 3.2	0.283
Flank SDS	620	0.0 ± 0.9	556	0.0 ± 1.1	0.285
Quadriceps (mm)	621	17.4 ± 3.6	556	17.9 ± 3.4	0.028
Quadriceps SDS	621	-0.1 ± 1.0	555	0.1 ± 1.0	0.026
Mean SDS	621	0.0 ± 0.8	556	0.0 ± 0.8	0.042
18 months					
Triceps (mm)	572	9.5 ± 2.3	521	9.8 ± 2.4	0.039
Triceps SDS	571	-0.1 ± 1.0	521	0.1 ± 1.0	0.034
Subscapular (mm)	571	6.7 ± 1.4	524	7.0 ± 1.7	0.009
Subscapular SDS	570	-0.1 ± 0.9	524	0.1 ± 1.1	0.007
Flank (mm)	570	10.5 ± 2.9	520	10.8 ± 3.4	0.212
Flank SDS	569	0.0 ± 0.9	520	0.0 ± 1.1	0.237
Quadriceps (mm)	571	15.8 ± 3.6	521	16.7 ± 3.6	<0.0001
Quadriceps SDS	570	-0.1 ± 1.0	521	0.1 ± 1.0	<0.0001
Mean SDS	571	-0.1 ± 0.7	524	0.1 ± 0.8	0.001
24 months					
Triceps (mm)	545	9.1 ± 2.3	503	9.6 ± 2.5	0.001
Triceps SDS	545	-0.1 ± 1.0	503	0.1 ± 1.0	0.001
Subscapular (mm)	550	6.5 ± 1.5	504	6.8 ± 1.8	0.012
Subscapular SDS	550	-0.1 ± 0.9	504	0.1 ± 1.1	0.012
Flank (mm)	549	10.2 ± 2.7	501	10.5 ± 3.4	0.115
Flank SDS	549	0.0 ± 0.9	501	0.1 ± 1.1	0.116
Quadriceps (mm)	546	14.5 ± 3.5	499	15.5 ± 3.6	<0.0001
Quadriceps SDS	546	-0.1 ± 1.0	499	0.1 ± 1.0	<0.0001
Mean SDS	552	-0.1 ± 0.8	505	0.1 ± 0.8	<0.0001

3.3.3.2 Distributions of size at birth

Size at birth was explored through the lens of key categorical covariates. With only five mothers with T1D, the average birth weight of their children was still many-fold significantly higher than the birth weight of children born to non-T1D mothers (2.85 ± 1.56 vs. 0.07 ± 0.97, *P*<0.0001). Such magnitude of difference underscores the effect of elevated maternal blood glucose levels, even when controlled with exogenous insulin, on foetal growth. Babies born to T1D mothers were on average longer (1.52 ± 0.08 vs. -0.08 ± 0.97, *P*<0.0001), but the impact of the high glycaemic maternal-uterine environment on birth length was reduced compared with birth weight.

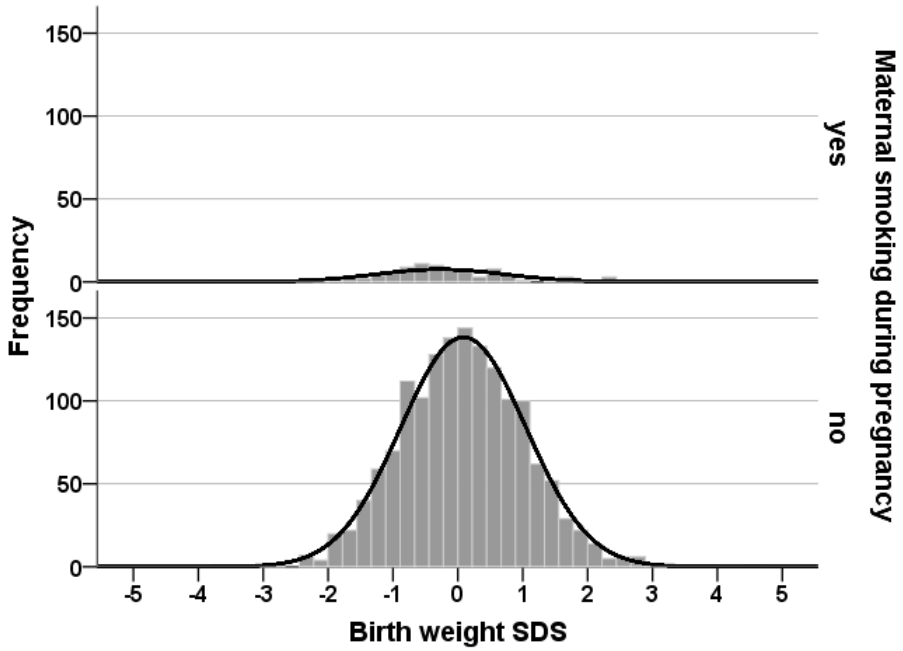
In terms of parity, the average birth weight of first-born infants was significantly less than that of the second or later child (-0.12 ± 0.96 vs. 0.24 ± 0.96, *P*<0.0001). The distribution of birth weight of non-primiparous pregnancies had a longer tail to the right (Figure 3-6), testifying to the thesis that the unprepared maternal-uterine environment restrains the growth of first-born infants[217].

Figure 3-6 | Distribution of birth weight SDS by parity.



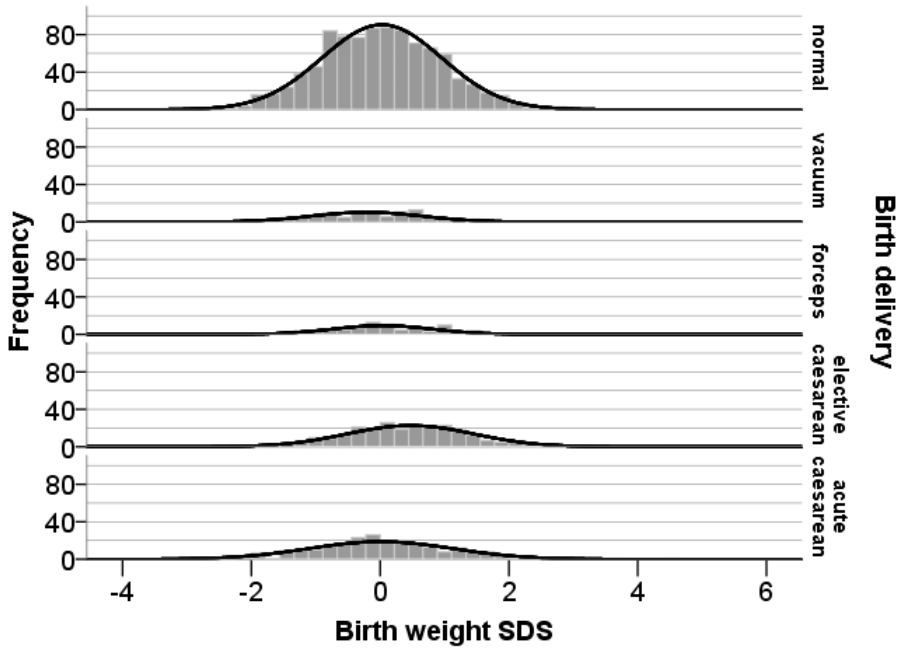
Mothers who smoked during pregnancy gave birth to infants with a lower birth weight compared with non-smoking mothers (-0.28 ± 1.00 vs. 0.08 ± 0.97, *P*=0.01), despite the uneven distribution of smoking vs. non-smoking subpopulations in the CBGS (Figure 3-7).

Figure 3-7 | Distribution of birth weight SDS by maternal smoking during pregnancy.



Infants born by elective caesarean were heavier than those born by all other modes of delivery put together (0.47 ± 0.98 vs. 0.01 ± 0.96, *P*<0.0001) or separately (Figure 3-8).

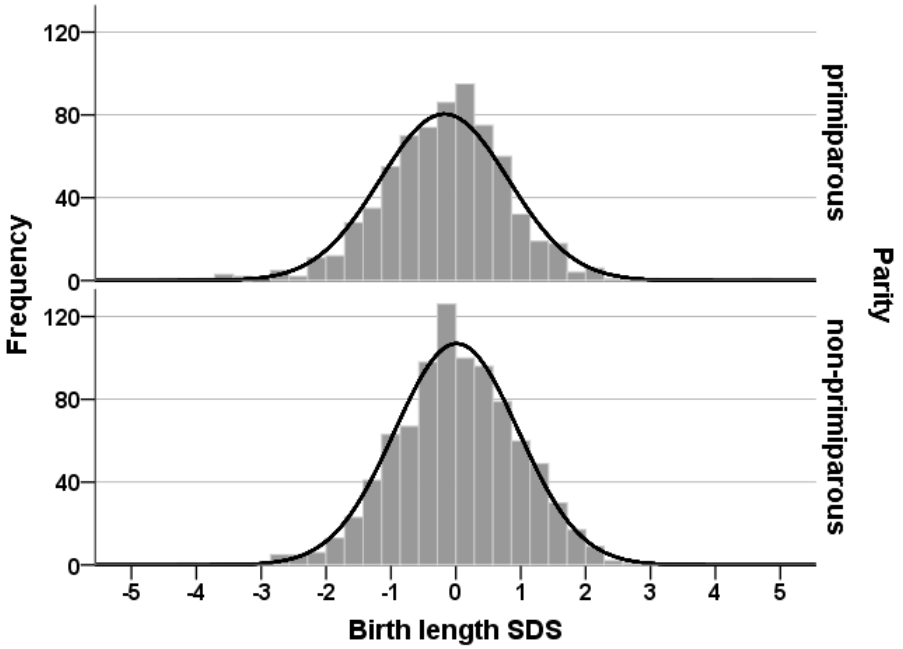
Figure 3-8 | Distribution of birth weight SDS by mode of birth delivery.



(0.03 ± 0.93 vs. -0.21 ± 0.87 vs. 0.03 ± 0.77 vs. 0.47 ± 0.98 vs. 0.01 ± 1.12, P_{ANOVA}<0.0001)

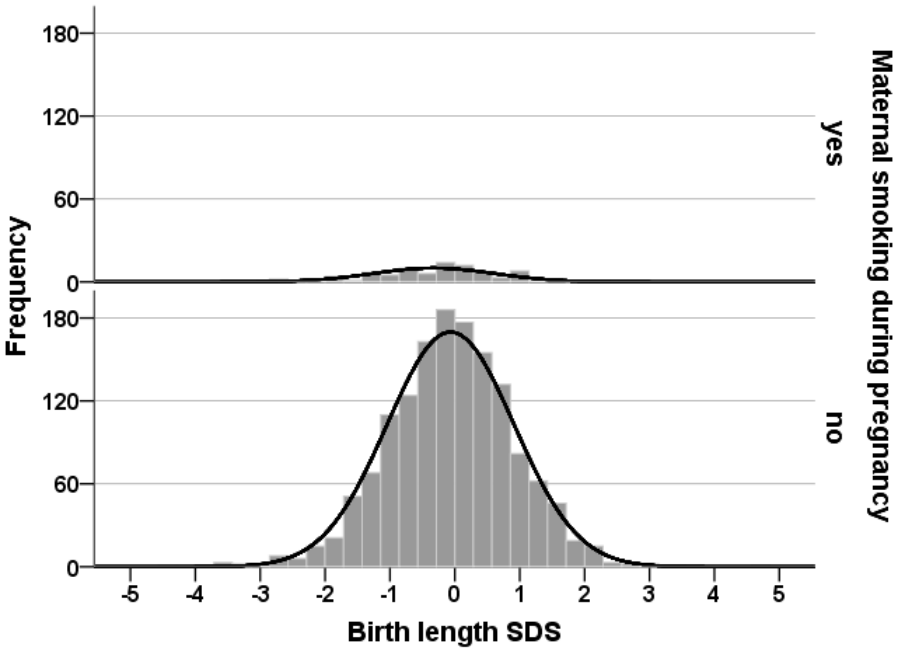
The respective distributions for birth length are similar with those for birth weight (Figures 3-9 to 3-10).

Figure 3-9 | Distribution of birth length SDS by parity.



(-0.18 ± 0.99 vs. 0.01 ± 0.95, *P*<0.0001)

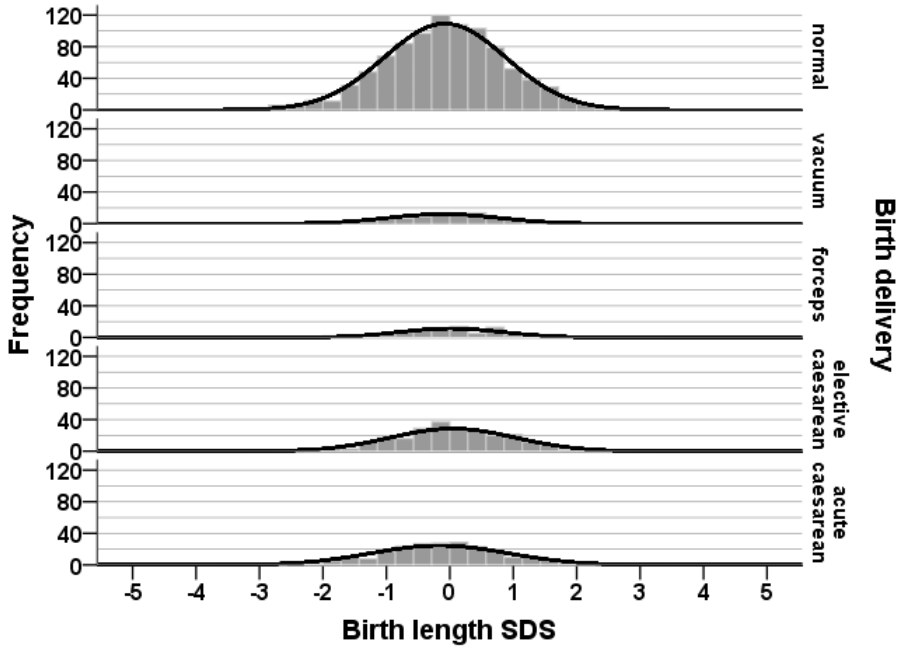
Figure 3-10 | Distribution of birth length SDS by maternal smoking during pregnancy.



(-0.33 ± 0.94 vs. 0.07 ± 0.98, *P*=0.018)

Infants born by elective caesarean were longer than those born by other modes collectively (0.04 ± 0.98 vs. -0.09 ± 0.97, *P*=0.042) or separately (Figure 3-11), but the difference was attenuated compared with weight.

Figure 3-11 | Distribution of birth length SDS by mode of birth delivery.



(-0.08 ± 0.96 vs. -0.10 ± 0.87 vs. 0.01 ± 0.81 vs. 0.04 ± 0.98 vs. -0.16 ± 1.07, P_{ANOVA}=0.180)

3.3.4 Infancy growth

3.3.4.1 Overall trends

The mean birth weight of 3.47 kg (n=1,657) tripled in a year and almost quadrupled (3.6-fold) two years after birth. The mean length at birth of 51.4 cm (n=1,599) increased by only 50% within a year and 70% within two years. Head circumference gains were even more modest, with the mean birth measure of 35.3 cm (n=1,602) going up by only 30% in the first year and less than 40% two years after birth. Weight is the most rapidly changing anthropometric index postnatally, which, combined with its accuracy of measurement, makes for a robust proxy of postnatal growth. Evaluation of mean SDS figures shows that CBGS infants had a weight in infancy comparable with the reference sample albeit marginally growing heavier by 24 months of age (0.15 SDS). Birth length was comparable with the reference, but beyond birth, the CBGS infants were slightly taller than the British 1990 referents with their height SDS increasing from 0.15 at 3 months to 0.39 by 24 months. Mean head circumference SDS values were consistently negative, suggesting that CBGS infants had on average a smaller head than British babies in the 1990s throughout infancy. This trend is unlikely explained by the shift

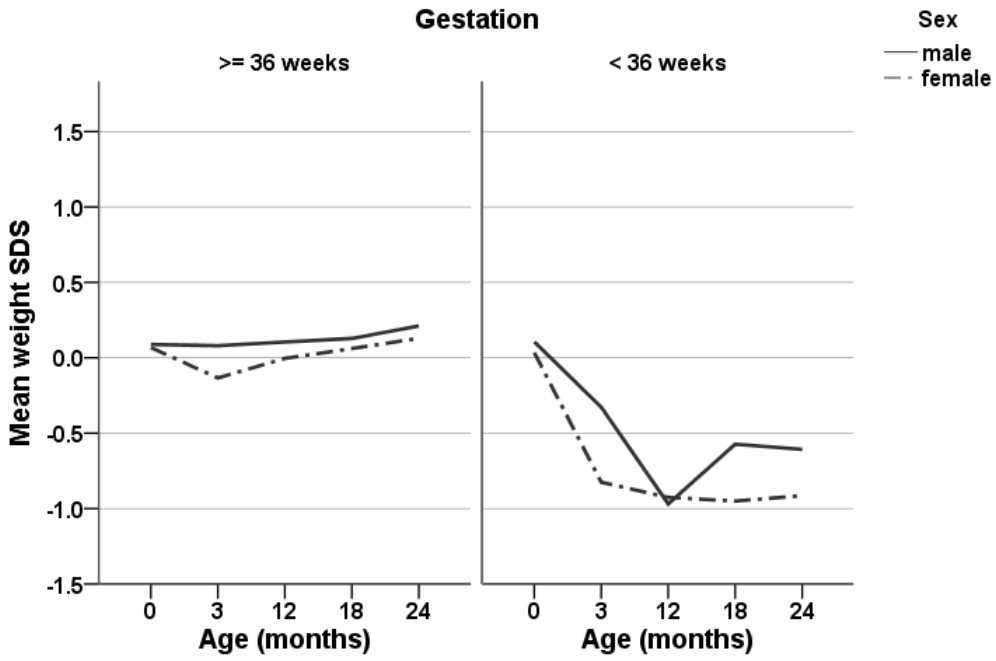
towards breastfeeding as neither weight nor length followed a similar trajectory. Changes in food composition (e.g. hormones) with an effect on neurodevelopment might be accountable.

3.3.4.2 Growth trajectories by primary pregnancy covariates

The maternal T1D environment is arguably the strongest influence on foetal growth, evidenced by the observation that the average size of CBGS infants born to T1D mothers is many-fold higher than the size of infants of non-diabetic mothers (Section 3.3.3.2). However, the very small size of this subset in the CBGS defies further investigation.

Gestational length is another important influence on foetal growth. The weight SDS trajectory of babies born at less than 36 weeks of gestation was declining and nearing a mean of -1.0 SDS at 12 months (Figure 3-12). The below-standard growth of premature babies surfaces that incomplete organogenesis at the point of delivery has persistent adverse effects on physiological processes despite intensive neonatal care. Term birth followed a steady growth trajectory showing a dip at 3 months in girls only.

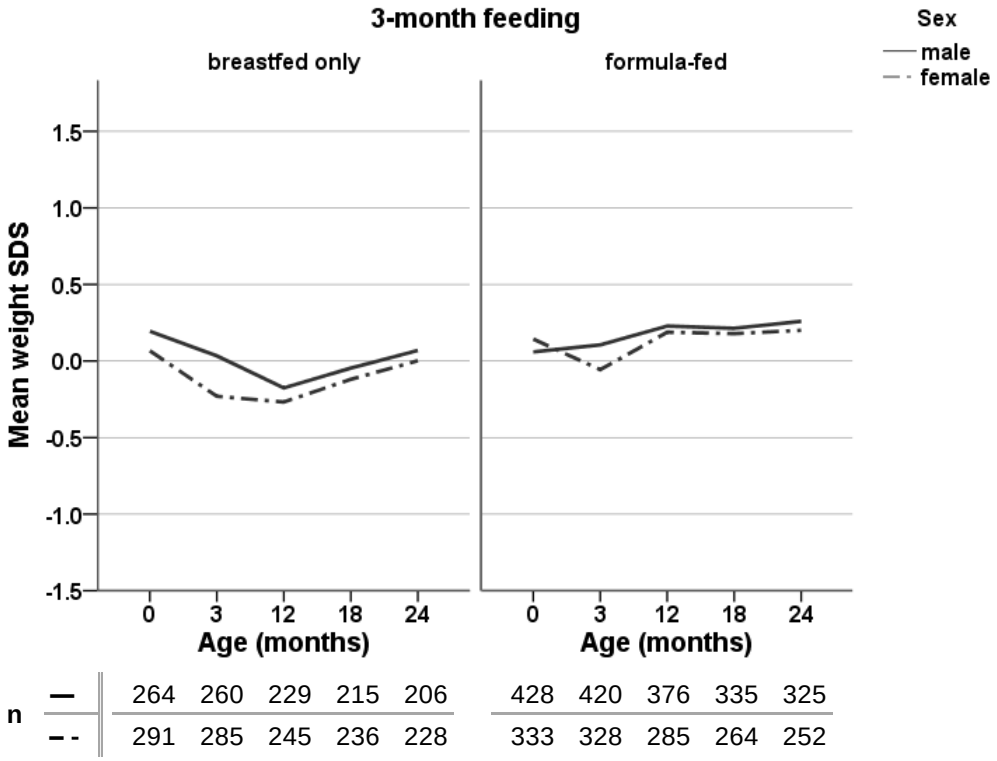
Figure 3-12 | Weight trajectory by sex and gestational category.
Frequencies by age and subgroup are listed below the graphs.



The nadir at 3 months is given context through the lens of feeding (Figure 3-13). The apparent differences in birth weight between boys and girls in the exclusively breastfed subset were not

statistically significant. Boys exclusively breastfed in the first 3 months showed a catch-down trajectory for the first year of life vs. boys on formula milk, possibly explained by the fact that the reference sample was dated prior to the move towards breastfeeding that took place in the UK. Girls were on a declining weight trajectory for the first 3 months irrespective of type of feeding, which accounts for the compounded dip in Figure 3-12. It is possible that breastfeeding constituted a higher component of the mixed nutrition of girls vs. their male counterparts, judging by the steeper weight increase in girls after the first 3 months of life when consumption of complementary foods goes up. Beyond the first year of life, weight across the board caught up (breastfed only) or reached a plateau (formula-fed).

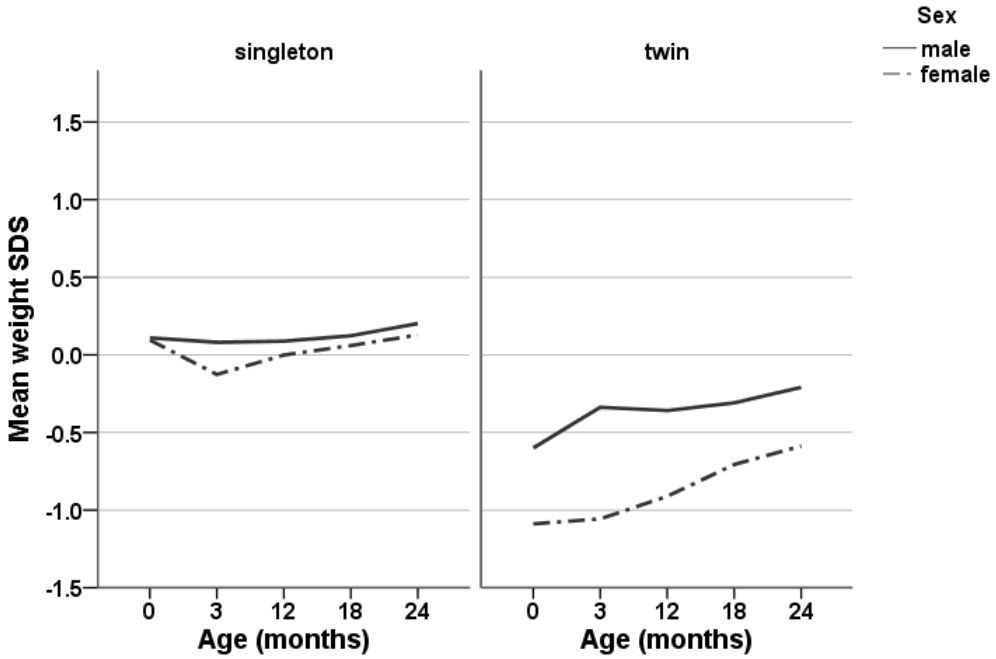
Figure 3-13 | Weight trajectory by sex and type of feeding.
Frequencies by age and subgroup are listed below the graphs.



The third pregnancy factor with a strong restraining effect on birth weight is the presence of twins (Figure 3-14), who compete for resources and space in the womb. Despite catching up beyond birth, twins also fell short of attaining average weight in infancy, highlighting the slow rate of the biological processes underlying growth. The trend in the CBGS concurs with the finding made by a Dutch study in >4,000 twins who caught up in body size after birth but failed to achieve the same weight and height as singletons until 2.5 years of age, justifying the use of distinct reference charts for twins[426].

Figure 3-14 | Weight trajectory by sex and twinship.

Frequencies by age and subgroup are listed below the graphs.

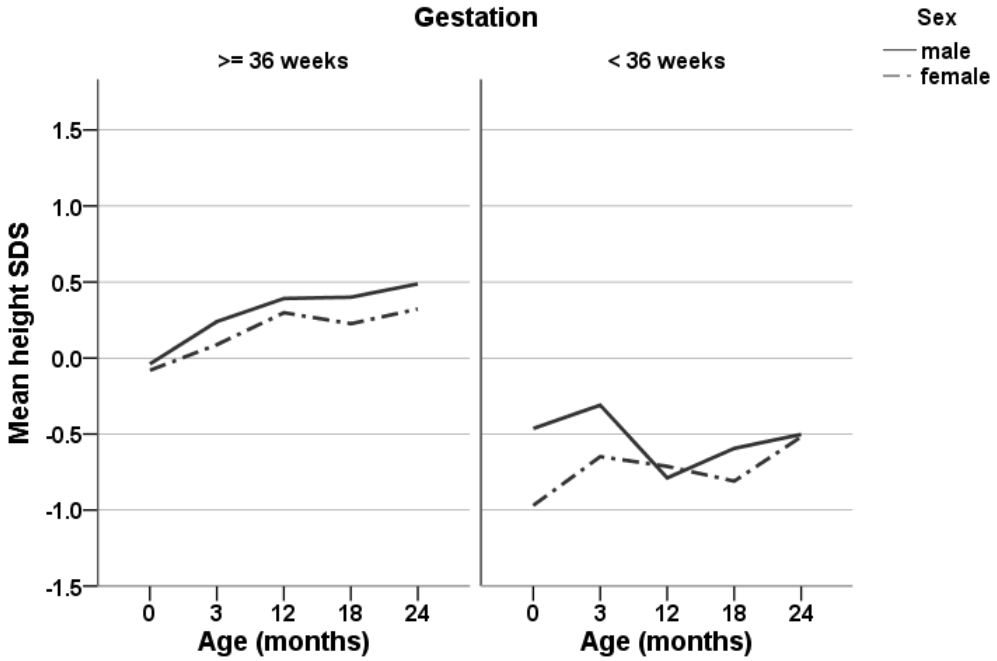


n										
—	833	682	605	552	530	25	21	18	16	16
– -	780	622	546	511	490	19	15	12	14	14

Height appears to be influenced by the same major pregnancy factors that influence weight. Across term pregnancies, the constant linear increase in height SDS (Figure 3-15), in contrast to the weight SDS trajectory which featured a dip at 3 months, might be suggestive of differential contributions of 'type of feeding' and genetics to weight vs. height in early life. A study in Dutch singletons and twins found that whereas heritability for weight remained almost constant up to 4 months postnatally, heritability for height increased strongly in the first month postnatally, with heritability estimates from the 2nd trimester to 36 months changing from 20% to 42% for weight and 13% to 63% for height[427]. The erratic trajectory of height amongst premature babies could be attributed to the inadequate size of the subset (20 boys and 15 girls at birth), as well as the drastically varying effects of the gestational age of preterm babies which ranges from 27 to 36 weeks. Conclusively, the comparison between weight and height trajectories suggests that breastfeeding most likely accounts for the decline in weight SDS observed in the first 3 months of life, yet the discrepancies are of a small magnitude (<0.20) to be of concern.

Figure 3-15 | Height trajectory by sex and gestational category.

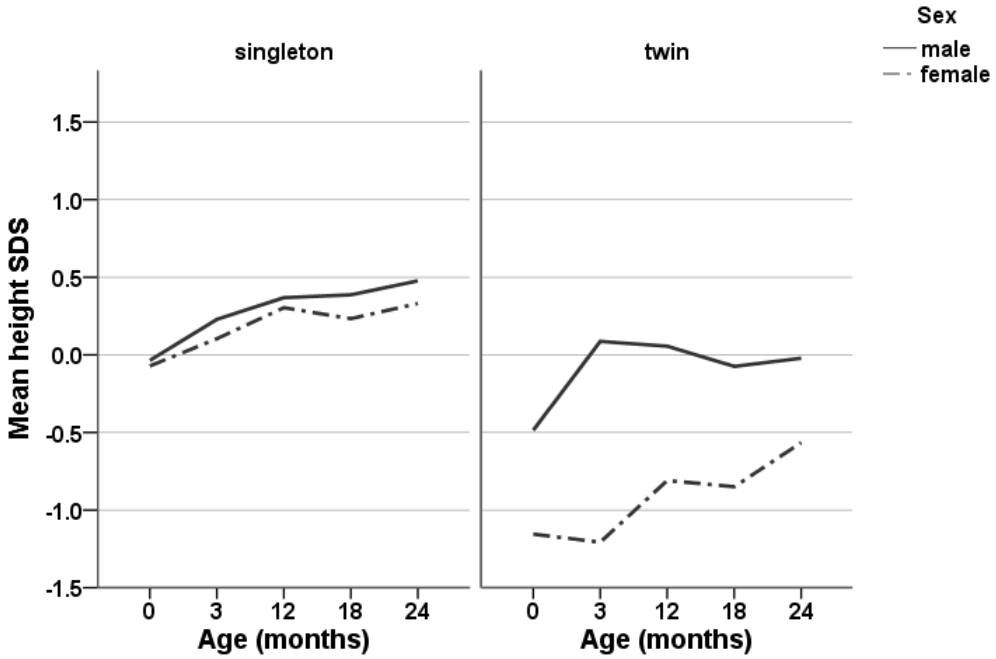
Frequencies by age and subgroup are listed below the graphs.



Twins had a lower height than average during infancy (Figure 3-16). The effect of restrained growth on length at birth was aggravated in girls (n=19) vs. boys (n=25) but the mean difference of 0.67 SDS was not statistically significant (*P*=0.140). The sex-specific difference in height attained statistical significance from 3 until 18 months of age.

Figure 3-16 | Height trajectory by sex and twinship.

Frequencies by age and subgroup are listed below the graphs.



n										
—	803	681	600	554	530	25	23	18	16	16
– -	752	616	544	511	482	19	15	12	14	14

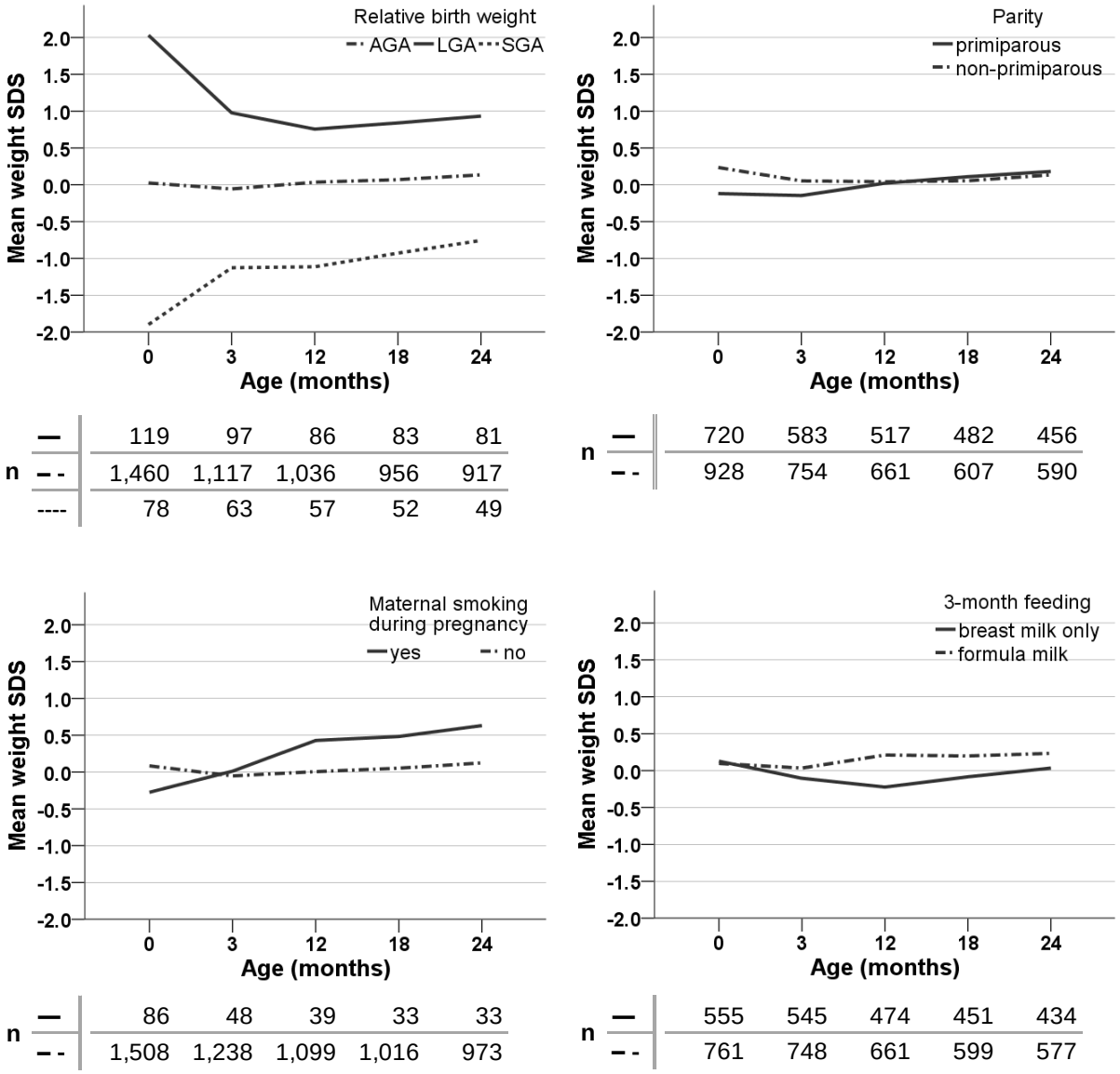
3.3.4.3 Growth trajectories by other covariates

The trajectories of weight throughout infancy were also explored by other influences: relative birth weight category, parity, smoking during pregnancy and type of feeding. Relative birth weight groups were defined as AGA with birth weight SDS lying between -1.5 and 1.5 (3.46 ± 0.46 kg), LGA with birth weight SDS ≥1.5 (4.26 ± 0.46 kg), and SGA with birth weight SDS ≤-1.5 (2.55 ± 0.41 kg). The CBGS cohort contained a higher percentage of LGA (7.2) vs. SGA (4.7) babies (Table 3-1), pointing to possible dominating growth-accelerating intrauterine exposures. Factors that might account for the enriched LGA group include maternal pre-pregnancy BMI and nutrition during pregnancy.

Figure 3-17 depicts the weight trajectories by selected pregnancy and post-pregnancy factors. Infants with appropriate intrauterine growth stay on the same continuum of weight beyond birth. In contrast, babies born SGA or LGA catch up or catch down respectively, yet without bridging the gap with their AGA counterparts even 2 years after birth. In terms of parity, first-born children had a lower weight SDS than second or later children up to 3 months (P<0.0001), but their trajectories coincided thereafter (P>0.1). Smoking reversed its effects

beyond birth, with offspring of mothers who smoked during pregnancy catching up and weighing significantly more than offspring of non-smokers at ages 12, 18 and 24 months (P<0.02). The weight gap between exclusively breastfed vs. formula-fed infants was largest at 12 months of age (-0.22 ± 1.02 vs. 0.21 ± 1.09, P<0.0001), with a mean difference of -0.43 SDS vs. -0.20 SDS at 24 months but remaining significant.

Figure 3-17 | Weight trajectory by perinatal influences.
Frequencies by age and subgroup are listed below the graphs.



n	0	3	12	18	24
—	119	97	86	83	81
– -	1,460	1,117	1,036	956	917
----	78	63	57	52	49

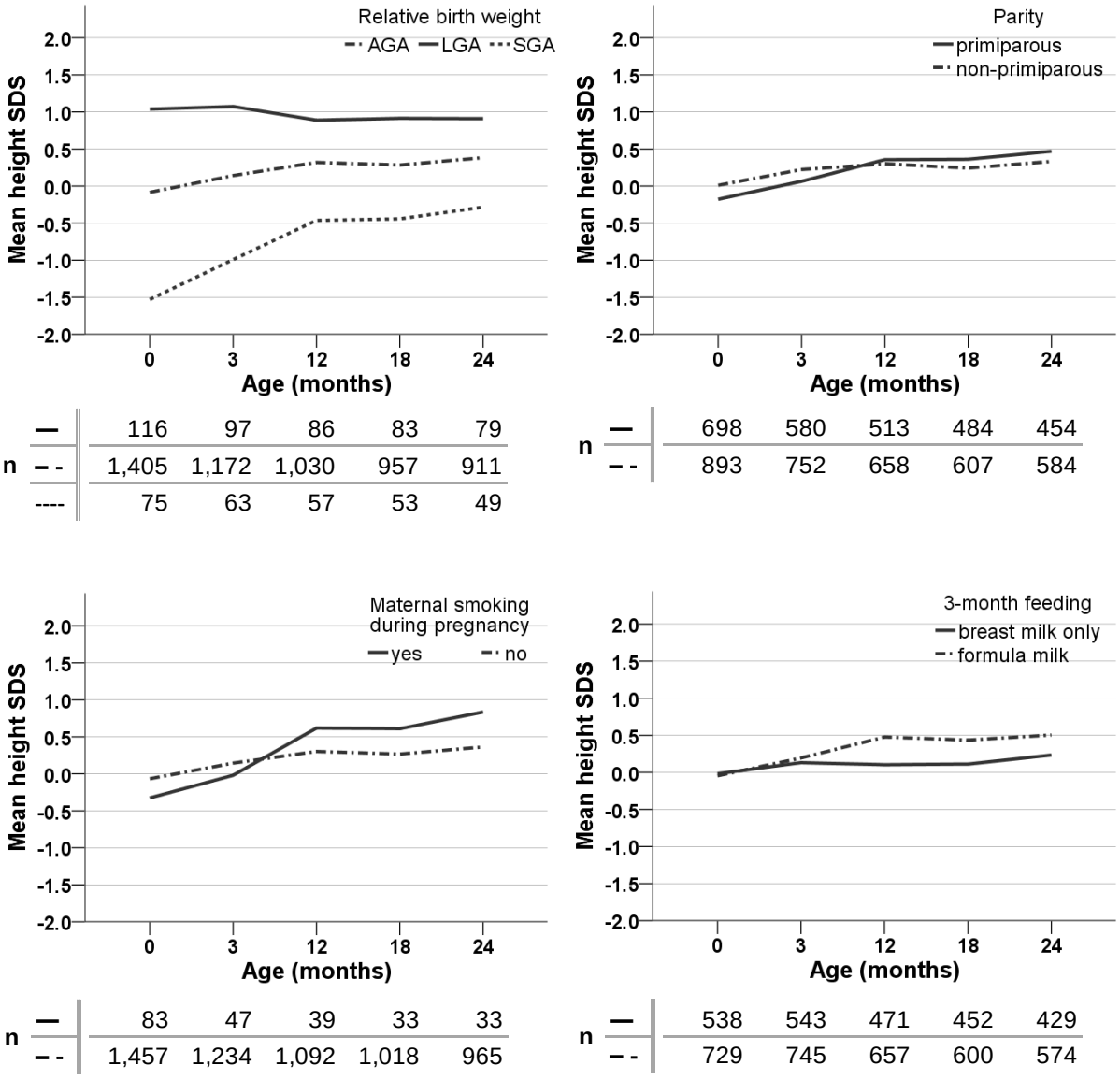
n	0	3	12	18	24
—	720	583	517	482	456
– -	928	754	661	607	590

n	0	3	12	18	24
—	86	48	39	33	33
– -	1,508	1,238	1,099	1,016	973

n	0	3	12	18	24
—	555	545	474	451	434
– -	761	748	661	599	577

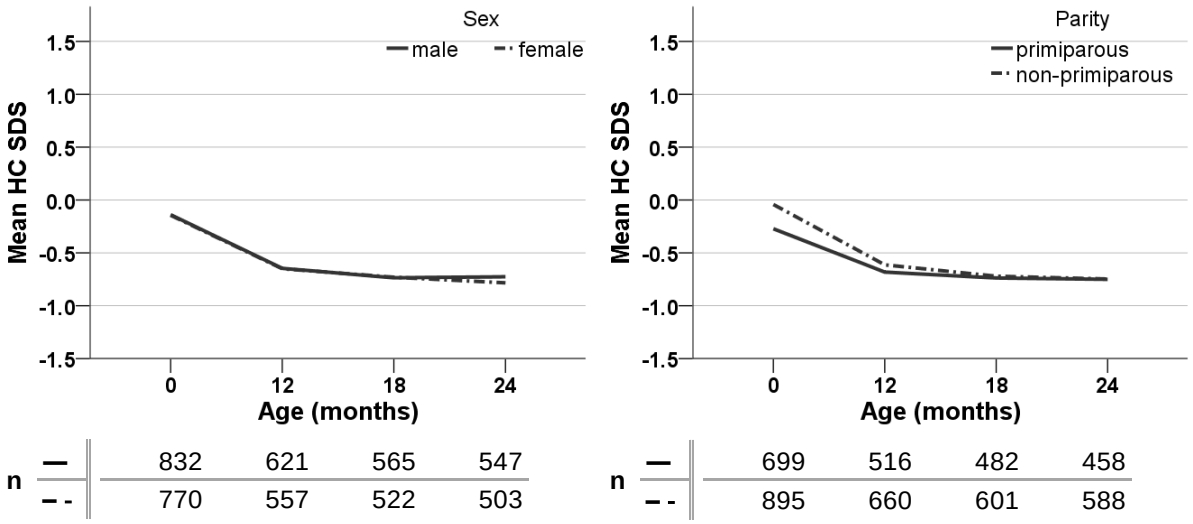
The respective trajectories for height are depicted in Figure 3-18. Mirroring weight gains, standardised linear gains of SGA and LGA babies had pronounced sustained differences vs. babies born with appropriate weight. Similar with weight SDS, height SDS was not influenced by parity beyond the first 3 months of life. In contrast to weight SDS, height SDS differences beyond birth, and until 18 months of age, due to maternal smoking during pregnancy were not of significance; this might be due to lack of power, i.e. scarcity of smokers in the CBGS cohort, in conjunction with the possibility of a reduced effect of smoking on height vs. weight. Breastfeeding had a persistent effect on height SDS from 12 months of age onwards, with the maximum difference of -0.37 at 1 year of age (P<0.0001) and remaining significant thereafter.

Figure 3-18 | Height trajectory by perinatal influences.
Frequencies by age and subgroup are listed below the graphs.



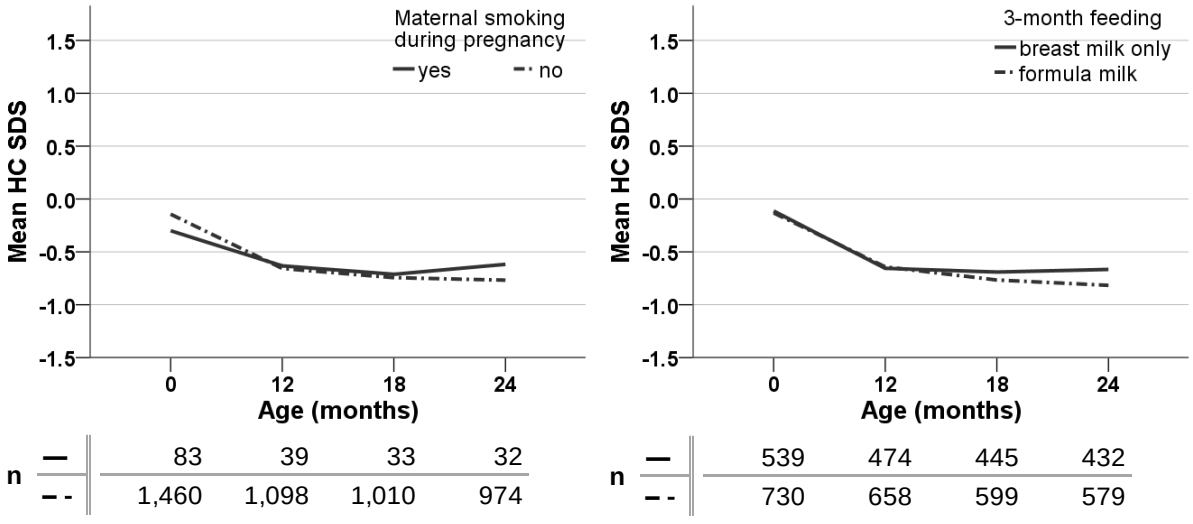
Secondary anthropometric measures were explored by infant sex, parity, maternal smoking during pregnancy and type feeding at 3 months of age. The trajectories of head circumference are shown in Figure 3-19. Head circumference SDS was similar between boys and girls throughout infancy (P>0.3). Primiparous infants had on average 0.2 SDS lower head circumference at birth than non-primiparous infants (P<0.0001), which was reflected in their reduced weight. Neither smoking during pregnancy nor type of feeding at 3 months exerted an effect on head circumference (P>0.1), except for a spurious association of the latter at 24 months of age.

Figure 3-19 | Head circumference (HC) trajectory by perinatal influences.
Frequencies by age and subgroup are listed below the graphs



n					
male	832	621	565	547
female	770	557	522	503

n					
primiparous	699	516	482	458
non-primiparous	895	660	601	588



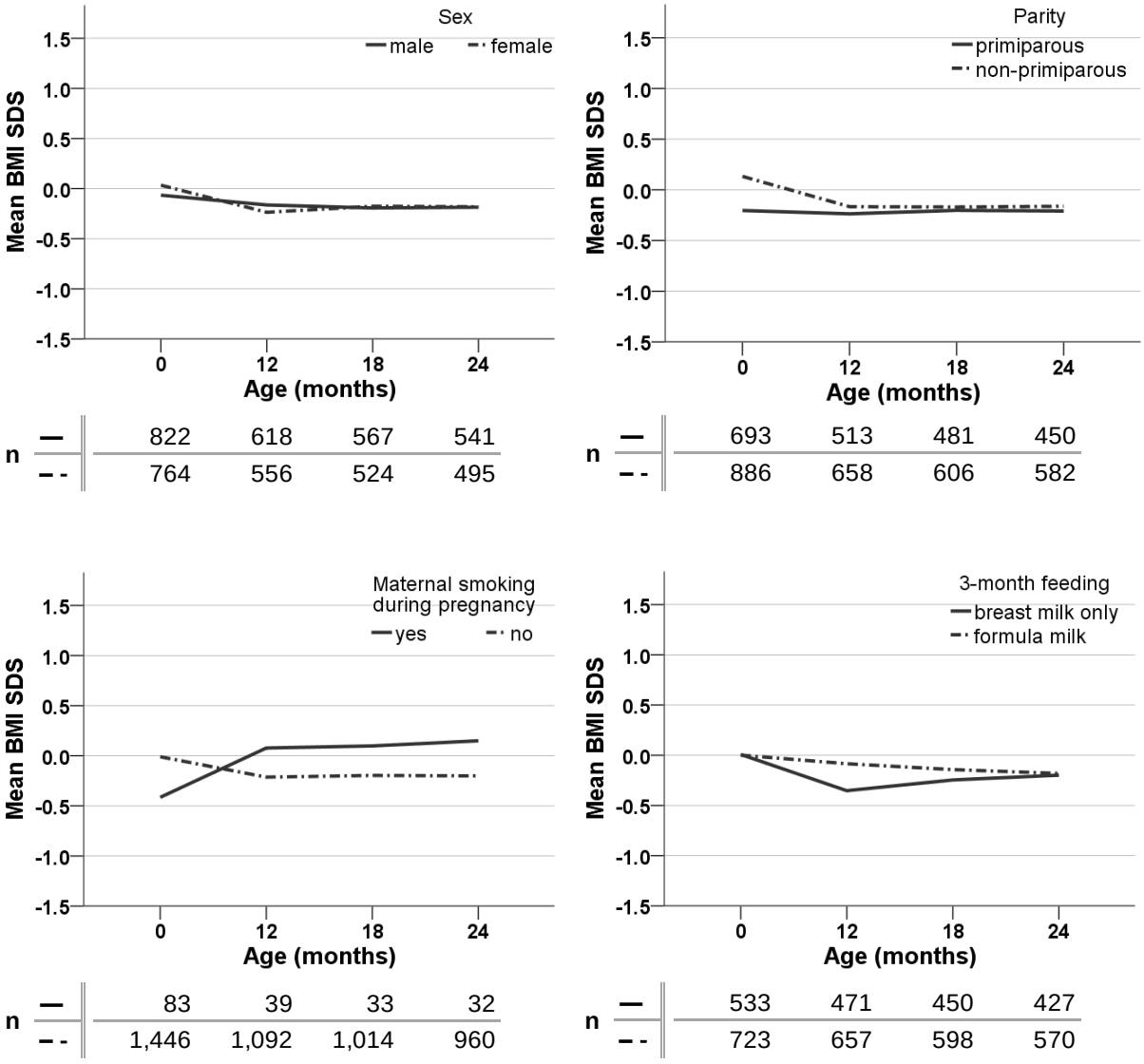
n					
yes	83	39	33	32
no	1,460	1,098	1,010	974

n					
breast milk only	539	474	445	432
formula milk	730	658	599	579

BMI trajectories are shown in Figure 3-20. The BMI SDS was similar between boys and girls throughout infancy. Primiparous babies had on average a lower BMI SDS at birth than non-primiparous babies, attributed to the increased restraining conditions of the first pregnancy (P<0.0001). In line with the influence on weight and height at birth, maternal smoking during pregnancy had a negative effect on BMI SDS at birth (P=0.004), which was not sustainable beyond birth. Infants exclusively breastfed at 3 months of age had a sizably reduced BMI by 0.3 SDS at 12 months of age (P<0.0001), underscoring the potential effect of formula milk on obesity. The differences were not sustained beyond 1 year of age when presumably most babies were on mixed nutrition.

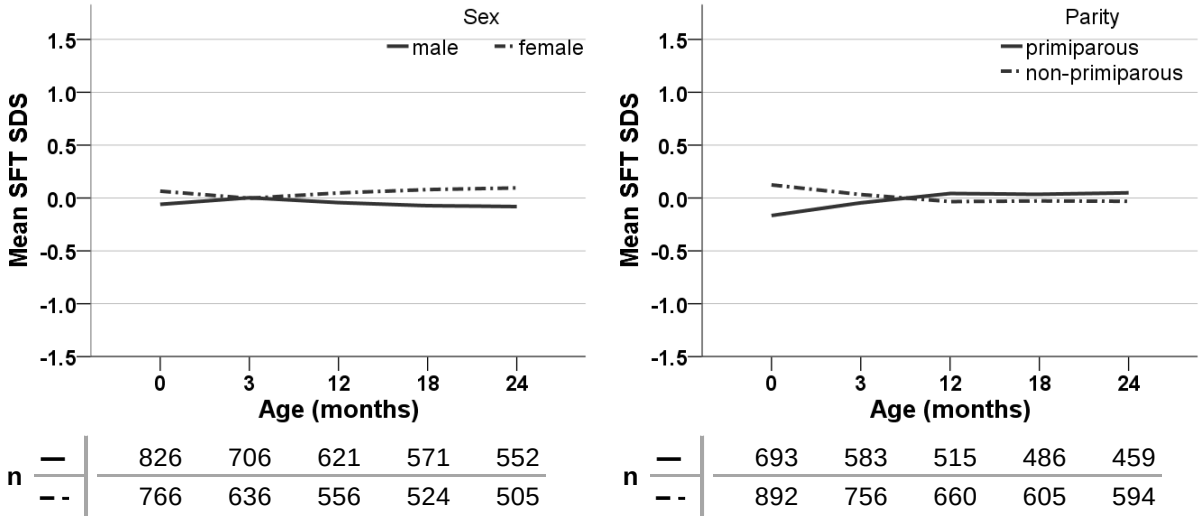
Figure 3-20 | BMI trajectory by perinatal influences.

Frequencies by age and subgroup are listed below the graphs.



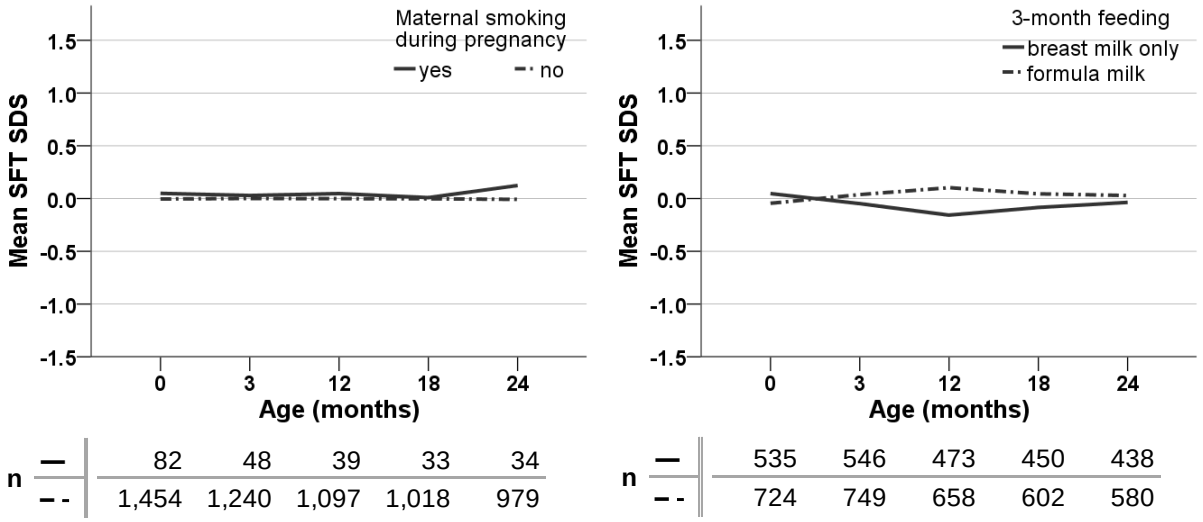
The trajectories of skinfold thickness mean SDS are shown in Figure 3-21. Adiposity in girls was higher at birth by 0.1 SDS than boys (*P*=0.005) but converged at 3 months before it significantly deviated again from 12 months onwards. Primiparous babies had 0.3 SDS lower adiposity at birth than their non-primiparous counterparts (*P*<0.0001) but the difference didn't persist thereafter. Smoking during pregnancy had no influence on infant's adiposity. Exclusively breastfed infants were characterised by 0.3 SDS lower overall adiposity at 12 months of age (*P*<0.0001) with the gap remaining statistically significant until 18 months. It is likely that babies exclusively on breast milk at 3 months of age continued so thereafter, unless type of feeding in the critical first three months of life has a persistent effect on adiposity.

Figure 3-21 | Skinfold thickness (SFT) trajectory by perinatal influences.
Frequencies by age and subgroup are listed below the graphs.



n						
male	—	826	706	621	571	552
female	– -	766	636	556	524	505

n						
primiparous	—	693	583	515	486	459
non-primiparous	– -	892	756	660	605	594

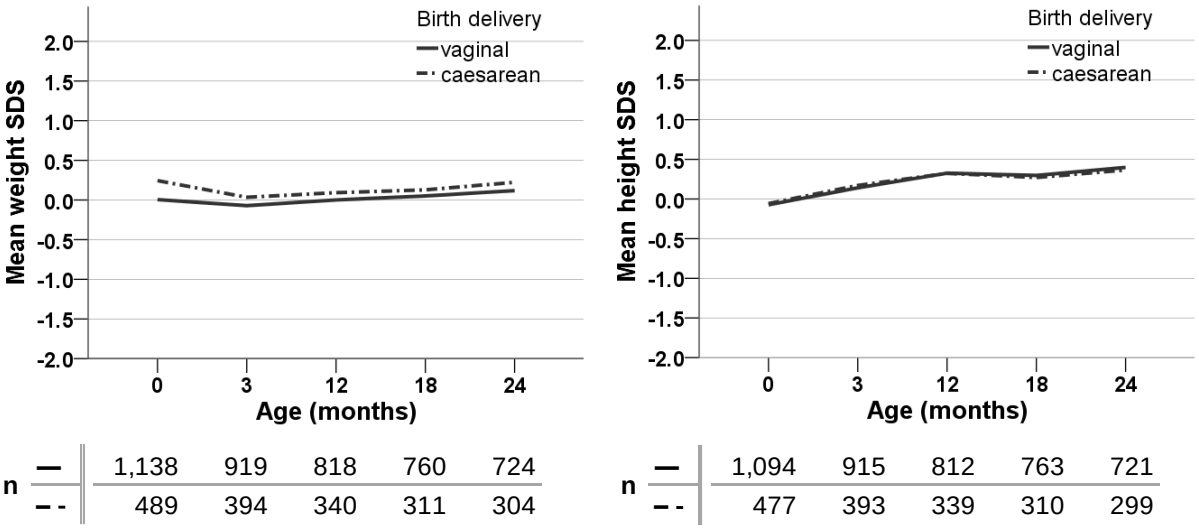


n						
yes	—	82	48	39	33	34
no	– -	1,454	1,240	1,097	1,018	979

n						
breast milk only	—	535	546	473	450	438
formula milk	– -	724	749	658	602	580

On the grounds of birth delivery influencing the offspring's microbiome — whereby vaginal delivery offers contact with the maternal vaginal and intestinal flora, thus contributing to infants' intestinal colonisation and composition of human microbiota[428] — the growth trajectories by mode of delivery were compared (Figure 3-22). Babies born by vaginal delivery (n=1,138) had higher weight SDS (P<0.0001) compared with babies delivered by caesarean (n=489), yet the difference was not significant beyond birth (P_{3mo}=0.087, P_{12mo}=0.190, P_{18mo}=0.274, P_{24mo}=0.135). Height was consistently identical between the two modes of delivery from birth onwards.

Figure 3-22 | Weight and height trajectories by mode of birth delivery.
Frequencies by age and subgroup are listed below the graphs.



n						
—	1,138	919	818	760	724
– -	489	394	340	311	304

n						
—	1,094	915	812	763	721
– -	477	393	339	310	299

3.3.5 Endocrine factors

Having established associations between growth and covariates, the question hones on the underlying biological pathways. Key candidate endocrine factors include IGF-1 and insulin, which play an important role in early growth and glucose homeostasis, as well as leptin which is secreted in adipose tissue. Levels of IGF-1 and IGFBP-3 were measured in DBS samples at ages 3, 12, 18 and 24 months of age. Both C-peptide, the product of insulin biosynthesis with a longer half-life than insulin, and leptin were measured at 3 and 12 months (Table 3-6).

Table 3-6 | Characteristics of endocrine factors measured by age and sex.

	Boys			Girls			
	N	**Mean ± SD**	**Range**	**N**	**Mean ± SD**	**Range**	***P* value**
IGF-1 (ng/ml)							
3-month	303	52.1 ± 20.1	8.0, 124.0	266	48.2 ± 19.8	8.0, 119.0	0.017
12-month	200	46.0 ± 20.6	10.0, 138.0	187	57.1 ± 24.6	14.0, 153.0	<0.0001
18-month	146	52.9 ± 22.6	12.0, 120.0	154	64.1 ± 24.5	13.0, 133.0	<0.0001
24-month	87	58.1 ± 26.8	15.0, 145.0	77	66.8 ± 26.9	21.0, 125.0	0.026
IGFBP-3 (ng/ml)							
3-month	301	1,680.3 ± 381.6	732.4, 2,919.1	265	1,728.2 ± 383.5	910.0, 2,963.0	0.104
12-month	191	1,854.2 ± 453.8	980.0, 3,823.0	186	2,073.7 ± 520.3	878.7, 3,930.4	<0.0001
18-month	141	1,866.0 ± 384.7	1,049.0, 3,017.0	148	2,131.3 ± 512.0	953.0, 5,370.0	<0.0001
24-month	85	1,948.3 ± 436.4	1,153.0, 2,968.7	70	2,190.6 ± 574.4	1,003.9, 3,896.2	0.006
C-peptide (pmol/L)							
3-month	147	710.8 ± 378.0	139.0, 2,847.3	170	684.8 ± 386.7	173.0, 3,059.0	0.363
12-month	78	591.7 ± 300.5	140.3, 1,500.8	66	582.9 ± 342.2	108.3, 1,867.9	0.705
Leptin (ng/ml)							
3-month	145	2.5 ± 2.0	0.6, 15.3	166	3.3 ± 2.3	0.6, 20.0	<0.0001
12-month	123	1.2 ± 0.8	0.6, 5.8	115	1.5 ± 0.9	0.6, 5.6	0.002

The endogenous production of IGF-1 starts at 32 weeks of postmenstrual age[429] but decreases rapidly at birth[430]. IGF-1 levels are known to positively correlate with postnatal age, gestational age and nutritional intake[431]. The CBGS confirmed that the concentrations of IGF-1, as well as IGFBP-3, go up with postnatal age but their trajectories don't quite mirror each other. Nevertheless, there exists a significant correlation between levels of IGF-1 and levels of IGFBP-3, with the degree of correlation increasingly strengthening over age, 3mo: *rho*=0.497 (*P*<0.0001); 12mo: *rho*=0.542 (*P*<0.0001); 18mo: *rho*=0.521 (*P*<0.0001); and 24mo: *rho*=0.638 (*P*<0.0001).

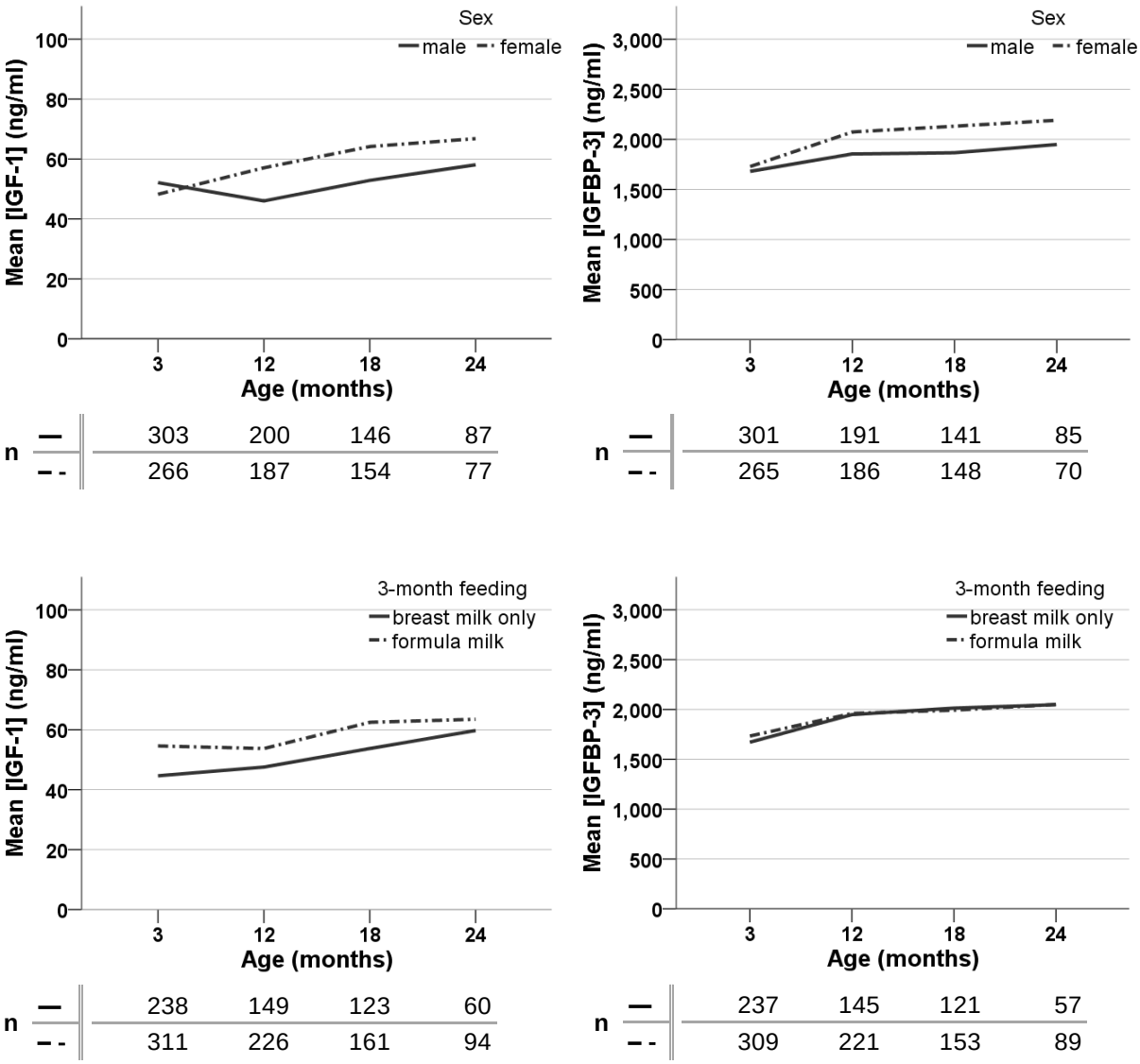
Hormone blood levels would be expected to correlate with growth (Table 3-7). C-peptide at 3 months was the only hormone that significantly correlated with gestational age, pointing to insulin production slowing down after a long gestation. Concentrations of IGF-1 and IGFBP-3 correlated with weight and height at corresponding ages. Contrary to published studies (Chapter 1), no discernible correlation was detected between head circumference and IGF-1 levels questioning the accuracy of the measurement or the statistical power herein.

Table 3-7 | Correlations between endocrine factors and growth parameters by age.

	IGF-1		IGFBP-3		C-peptide		Leptin	
	rho	***P* value**	***rho***	***P* value**	***rho***	***P* value**	***rho***	***P* value**
Gestation (wk)	-0.057	0.175	-0.075	0.074	-0.148	0.008	-0.022	0.702
3 months								
Weight SDS	0.312	<0.0001	0.169	<0.0001	0.239	<0.0001	0.349	<0.0001
Height SDS	0.158	<0.0001	0.105	0.013	0.141	0.013	0.115	0.044
SFmean SDS	0.278	<0.0001	0.155	<0.0001	0.2	<0.0001	0.428	<0.0001
12 months								
Weight SDS	0.224	<0.0001	0.195	<0.0001	0.185	0.144	0.198	0.002
Height SDS	0.138	0.007	0.153	0.003	0.253	0.002	0.029	0.654
HC SDS	-0.046	0.370	-0.074	0.152	0.051	0.545	-0.071	0.277
BMI SDS	0.180	<0.0001	0.131	0.011	0.071	0.398	0.242	<0.0001
SFmean SDS	0.132	0.010	0.036	0.488	0.014	0.869	0.248	<0.0001
18 months								
Weight SDS	0.241	<0.0001	0.052	0.379				
Height SDS	0.211	<0.0001	0.021	0.727				
HC SDS	0.094	0.103	-0.027	0.644				
BMI SDS	0.170	0.003	0.082	0.166				
SFmean SDS	0.093	0.106	0.045	0.450				
24 months								
Weight SDS	0.282	<0.0001	0.186	0.021				
Height SDS	0.163	0.038	0.121	0.137				
HC SDS	0.062	0.436	-0.013	0.876				
BMI SDS	0.316	<0.0001	0.206	0.011				
SFmean SDS	0.108	0.172	0.027	0.741				

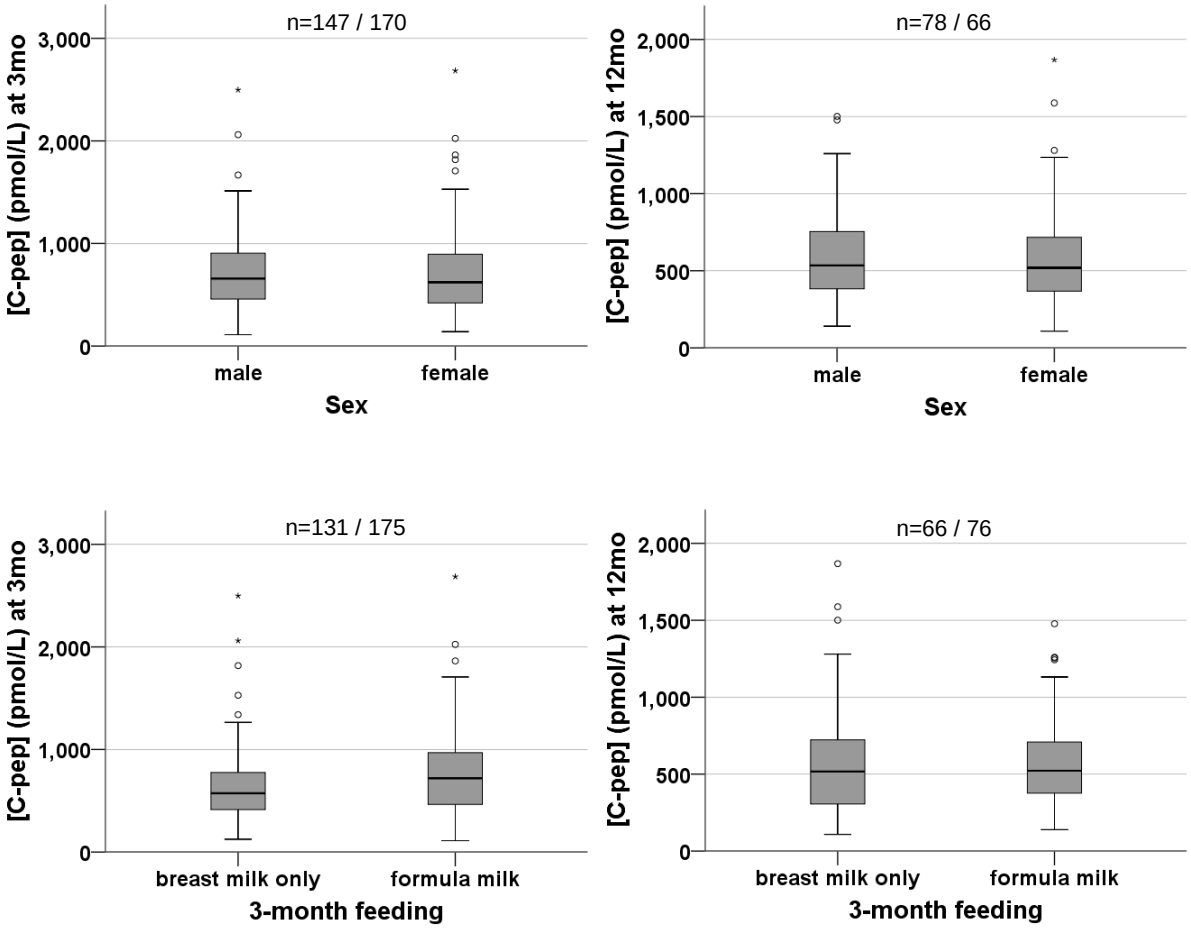
Figure 3-23 shows the corresponding levels of IGF-1 and IGFBP-3 by sex and type of feeding. IGF-1 levels differed significantly between boys and girls at all ages measured (3mo: 52.12 ± 20.12 vs. 48.22 ± 19.78, *P*=0.017; 12mo: 46.02 ± 20.59 vs. 57.10 ± 24.60, *P*<0.0001; 18mo: 52.87 ± 22.62, 64.15 ± 24.48, *P*<0.0001; 24mo: 58.08 ± 26.78 vs. 66.81 ± 26.89 *P*=0.026). Interestingly, beyond the early months of life, IGF-1 levels were consistently higher in girls; such direction of difference does not reflect the respective sex-specific weight or height trajectories. Furthermore, the IGF-1 concentration in boys showed a dip at 12 months, whereas in girls the levels consistently went up with age. IGFBP-3 levels differed significantly by sex from 12 months onwards in a direction almost parallel to that of IGF-1 (3mo: 1,680 ± 382 vs. 1,728 ± 383, *P*=0.104; 12mo: 1,854 ± 453 vs. 2,073 ± 520 *P*<0.0001; 18mo: 1,866 ± 385 vs. 2,131 ± 512 *P*<0.0001; 24mo: 1,948 ± 436 vs. 2,191 ± 574 *P*=0.006). Babies exclusively breastfed had consistently lower IGF-1 levels than babies on formula milk, with the difference being statistically significant up until 18 months of age (3mo: 44.65 ± 16.85 vs. 54.60 ± 21.53, *P*<0.0001; 12mo: 47.54 ± 22.48 vs. 53.74 ± 23.56 *P*=0.006; 18mo: 53.73 ± 21.47 vs. 62.49 ± 25.69, *P*=0.010; 24mo: 59.75 ± 28.36 vs. 63.50 ± 25.51, *P*=0.246).

Figure 3-23 | IGF-1 and IGFBP-3 trajectories by sex and type of feeding.
Frequencies by age and subgroup are listed below the graphs.



C-peptide measurements followed distributions that were skewed to the right both at ages 3 months (697 ± 382 pmol/L) and 12 months (588 ± 319 pmol/L). Figure 3-24 depicts C-peptide levels at each age measured by sex and type of feeding. The concentration of C-peptide was higher at 3 months than 12 months, which follows the decreasing postnatal growth rate beyond the first 6 weeks of life. However, the concentrations of C-peptide did not differ between boys vs. girls (3mo: 710 ± 378 vs. 685 ± 387, *P*=0.363; 12mo: 592 ±301 vs. 583 ± 342, *P*=0.705). However, C-peptide levels were significantly lower in exclusively breastfed vs. formula-fed babies at 3 months (3mo: 636 ± 375 vs. 748 ± 390, *P*=0.003; 12mo: 583 ± 364 vs. 582 ± 271, *P*=0.656), possibly underscoring the sensitivity of insulin secretion to protein.

Figure 3-24 | C-peptide levels by sex and type of feeding.
Frequencies are listed in graphs (male / female; breast milk only / formula milk).



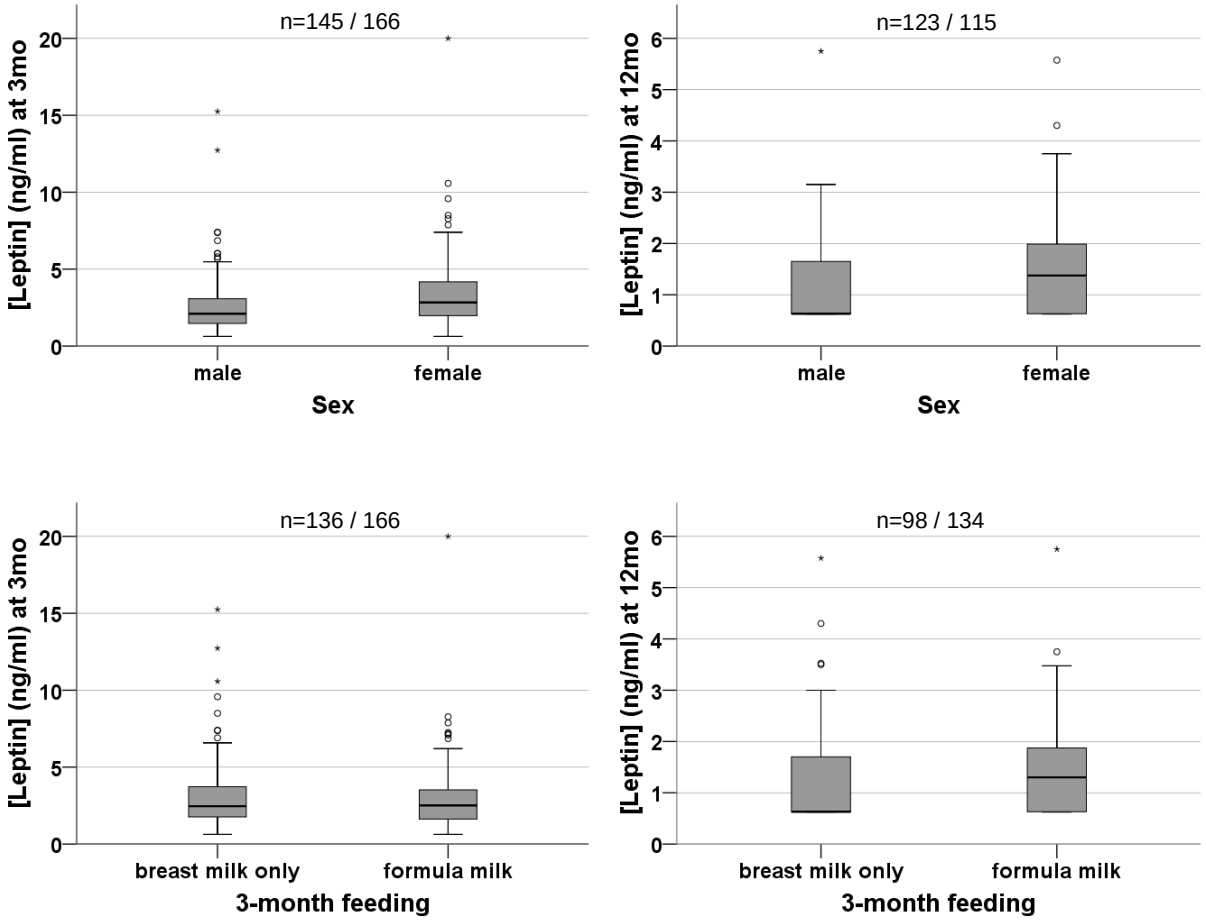
Levels of adipose-secreted leptin correlated with overall adiposity at both 3 and 12 months, and the correlations were strongest with subscapular skinfold thickness amongst all four anatomical sites considered (Table 3-8).

Table 3-8 | Correlations between leptin levels and skinfold thickness SDS by age.

	N	*rho*	*P* value
3 months			
Triceps SDS	311	0.182	0.001
Subscapular SDS	311	0.441	<0.0001
Flank SDS	311	0.369	<0.0001
Quadriceps SDS	311	0.336	<0.0001
Skinfold mean SDS	311	0.428	<0.0001
12 months			
Triceps SDS	237	0.206	0.001
Subscapular SDS	237	0.274	<0.0001
Flank SDS	237	0.151	0.020
Quadriceps SDS	237	0.190	0.003
Skinfold mean SDS	237	0.248	<0.0001

As expected, leptin levels were significantly higher in girls than boys but showed no differences by type of feeding at 3 months (Figure 3-25).

Figure 3-25 | Leptin levels by sex and type of feeding.
Frequencies are listed in graphs (male / female; breast milk only / formula milk).



3.4 GROWTH AND HORMONE COVARIATES

3.4.1 Approach to identifying covariates

In identifying predictors of anthropometric and endocrine phenotypes to adjust for in regression analyses, I considered biological factors (Chapter 1), in conjunction with the following statistical workings:

1. Pairwise associations give a useful indication of the effect that might be expected and inform of significant variables in regressions[432]. Thus, I considered bivariate associations between each outcome and candidate covariate at birth, the midpoint and end of infancy.

2. Regressions of anthropometric measures on phenotypes prior to introducing the genetic predictor.
3. Model selection based on the Akaike information criterion (AIC), which penalises for fitness of good and number of covariates.

Standardised growth variables approximated a normal distribution; however, skinfold thickness SDS and hormone levels did not conform to normality. Nevertheless, the response variables were not transformed as regression models are robust to moderate deviations from normality provided a large sample size and few multivariate outliers[425].

3.4.2 Bivariate associations with growth

Bivariate associations for standardised weight and height are summarised in Tables 3-9 and 3-10, and for other growth indices in Appendix II. In brief, gestational age and sex had a persistent effect on *unadjusted* infant size, justifying the requirement for SD scores. Following standardisation, maternal T1D had the largest effect on size at birth, particularly weight, but not thereafter. Parity affected size until the third month after birth. Maternal smoking during pregnancy resulted in a smaller newborn but its effect reversed after birth, reflecting catch-up growth following the oxygen-deprived maternal-uterine environment. Mode of birth delivery associated with weight at birth and 12 months of age; these correlations represent reverse causality since larger babies are likely delivered by elective caesarean and remain large.

Maternal height, weight and BMI each positively correlated with infant's size throughout infancy. Amongst them, the composite BMI had the smallest magnitude of effect. The limitation of using weight and height, rather than BMI, is collinearity; however, tolerance values in models using maternal weight and height were acceptably ~0.87. Whereas maternal weight and height made comparable contributions to infant's weight, mother's height dominated the effect on infant's length. Maternal weight gain had a weak correlation with size at birth. Parental ages showed a positive yet weak association with size at birth, with the direction of effect reversing from 3 months onwards. Parental ages lost their effect in regression models, possibly due to their close-to-zero correlation coefficients. Their tolerance values (<0.7) suggested a degree of collinearity between the designated variables. For these reasons, models were not adjusted for parental ages. Amongst sociodemographic factors, the IMD showed a negative trend with size at birth, which reversed thereafter. Marital status and maternal education had no influence on birth weight (P>0.05), testifying to the findings of a prior study from the UK according to which social and psychological factors had no direct effect on birth weight[221]. Breastfeeding at 3 months associated with a lower infant size throughout infancy. The cohort lacked information on feeding beyond 3 months, but it is fair to speculate

that these mothers adhered to the recommended guidelines of breastfeeding up to at least 4 to 6 months of age. As previously documented elsewhere[433], size at birth had a strong effect on postnatal growth in the CBGS.

Gains of weight and height from birth to 12 months or birth to 24 months consistently correlated with parity, smoking and type of feeding. Maternal height had a strong effect on both infant's height gains and weight gains; in contrast, maternal weight and maternal BMI lost their effects on gains. These observations warn about the simultaneous input of maternal BMI and weight into models, a practice which i) leads to redundancy and ii) annuls the true effect of height on the response variable. As a justification, 'an important underlying assumption of BMI is that weight scales to height^2', and therefore BMI is independent of height in adults[206].

Table 3-9 | Pairwise associations with Y=Weight SDS (*rho* ≡ Spearman's coefficient; Δ ≡ mean difference for t-test; F ≡ F-statistic for ANOVA).

Variable		Birth			12 months			24 months		
Name	**Type**	***rho***	**Δ / F**	***P* value**	***rho***	**Δ / F**	***P* value**	***rho***	**Δ / F**	***P* value**
Newborn index										
Gestation	Continuous									
Infant Sex (baseline: males)	Binary									
Pregnancy influences										
Maternal T1D (baseline: non-T1D)	Binary		-2.777	<0.0001		-0.928	0.056		-1.024	0.085
Parity (baseline: primiparous)	Binary		-0.355	<0.0001		-0.021	0.741		0.046	0.471
Smoking (baseline: non-smoking)	Binary		0.359	0.001		-0.423	0.017		-0.507	0.006
Maternal height	Continuous	0.220		<0.0001	0.283		<0.0001	0.257		<0.0001
Maternal pre-pregnancy weight	Continuous	0.232		<0.0001	0.251		<0.0001	0.277		<0.0001
Maternal pre-pregnancy BMI	Continuous	0.160		<0.0001	0.143		<0.0001	0.177		<0.0001
Maternal pregnancy weight gain	Continuous	0.086		0.011	0.049		0.194	0.033		0.409
Maternal age	Continuous	0.075		0.006	-0.088		0.003	-0.039		0.206
Paternal age	Continuous	0.068		0.014	-0.065		0.032	-0.046		0.150
Post-Pregnancy influences										
Birth delivery	Categorical		13.756	<0.0001		5.370	<0.0001		1.956	0.099
3-month feeding (baseline: breast milk only)	Binary					-0.434	<0.0001		-0.200	0.002
Birth weight SDS	Continuous				0.407		<0.0001	0.409		<0.0001
Birth length SDS	Continuous				0.394		<0.0001	0.421		<0.0001
Sociodemographic influences										
Ethnicity (baseline: White Caucasian)	Binary		0.046	0.734		-0.111	0.532		-0.066	0.710
Index of multiple deprivation	Continuous	-0.065		0.047	0.064		0.070	0.065		0.071
Marital status	Categorical		1.556	0.211		0.509	0.601		0.428	0.652
Maternal education	Categorical		1.270	0.282		5.363	0.005		0.748	0.474

Table 3-10 | Pairwise associations with Y=Height SDS (*rho* ≡ Spearman's coefficient; Δ ≡ mean difference for t-test; F ≡ F-statistic for ANOVA).

Variable		Birth			12 months			24 months		
Name	**Type**	***rho***	**Δ / F**	***P* value**	***rho***	**Δ / F**	***P* value**	***rho***	**Δ / F**	***P* value**
Newborn index										
Gestation	Continuous									
Infant Sex (baseline: males)	Binary									
Pregnancy influences										
Maternal T1D (baseline: non-T1D)	Binary		-1.594	<0.0001		-0.879	0.070		-0.554	0.463
Parity (baseline: primiparous)	Binary		-0.189	<0.0001		0.056	0.387		0.137	0.043
Smoking (baseline: non-smoking)	Binary		0.260	0.018		-0.316	0.075		-0.473	0.013
Maternal height	Continuous	0.260		<0.0001	0.360		<0.0001	0.337		<0.0001
Maternal pre-pregnancy weight	Continuous	0.197		<0.0001	0.234		<0.0001	0.225		<0.0001
Maternal pre-pregnancy BMI	Continuous	0.131		<0.0001	0.095		0.003	0.087		0.011
Maternal pregnancy weight gain	Continuous	0.101		0.003	0.052		0.170	0.051		0.200
Maternal age	Continuous	0.068		0.014	-0.074		0.012	-0.061		0.049
Paternal age	Continuous	0.049		0.082	-0.062		0.041	-0.046		0.155
Post-Pregnancy influences										
Birth delivery	Categorical		1.569	0.180		1.243	0.291		1.895	0.109
3-month feeding (baseline: breast milk only)	Binary					-0.373	<0.0001		-0.270	<0.0001
Birth weight SDS	Continuous				0.331		<0.0001	0.318		<0.0001
Birth length SDS	Continuous				0.481		<0.0001	0.458		<0.0001
Sociodemographic influences										
Ethnicity (baseline: White Caucasian)	Binary		-0.001	0.993		0.043	0.802		-0.077	0.675
Index of multiple deprivation	Continuous	-0.058		0.083	0.031		0.384	0.087		0.015
Marital status	Categorical		1.633	0.196		0.373	0.689		0.468	0.627
Maternal education	Categorical		1.654	0.192		1.878	0.154		0.021	0.980

3.4.3 Regressions on phenotypes

Each anthropometric index was regressed solely on phenotypes by constructing various models that pressure-tested covariates chosen based on replaceability, measurement error or collinearity. Tables 3-11 to 3-14 present the Beta and *P* value for regression models built as described below. Highlighted are *P* values <0.05, blank entries denote that the variable was not entered in the model, and 'N/A' indicates that the parameter could not be calculated.

- Model Set A: Adjusted for mutually-exclusive, non-collinear covariates.
- Model Set B: Replaced 'maternal height' and 'maternal weight' with 'maternal BMI'.
- Model Set C: Incorporated 'maternal weight gain', which made a meaningful contribution to size at birth and weight at 12 months, but illustrated that low coverage of a single measurement, which results in increased exclusions of subjects with missing data, distort the consistency of predictors.
- Model Set D: Incorporated 'birth size'. Simultaneous addition of weight SDS and length SDS at birth yielded low tolerance values (~0.5) suggesting high collinearity. My analysis concurs with a prior study that adjusted weight models for birth weight SDS and height models for birth length SDS[433]. The advantages of adjusting postnatal growth for birth size are i) consistency of predictors across outcomes and ii) increased model fitness, i.e. R^2.

I further tested my selection of predictors for weight, height and respective gains by regressing using the 'stepwise forward' method, which is sometimes used in statistical analysis to identify covariates (Table 3-15). The results showed very good agreement with the output of the 'forced entry' method on just the selected covariates (Table 3-16), albeit a few exceptions (maternal pre-pregnancy weight was missing from weight models and paternal age was included for birth length). Importantly, the final models that emerged are parsimonious with a few parameters: models for size at birth or size at 3 months are adjusted for 4 covariates, subsequent 'weight' or 'height' models are adjusted for 3 covariates. Regressions were repeated for secondary anthropometric measures (Tables 3-17) and endocrine factors (Table 3-18).

Comparing my statistical framework with the literature, a TEDDY study that examined genetic associations with size at birth adjusted for maternal smoking, maternal alcohol consumption, delivery complications, maternal age, height and weight at end of pregnancy, and paternal height[230]. These authors adjusted for maternal weight at *end* pregnancy, which captures both the pre-pregnancy weight and weight gain during pregnancy, but similar with my analysis, they avoided the composite BMI. I advocate that delivery complications represent reverse causality and the effect of alcohol during pregnancy is restricted to length and is minor in the absence of smoking.

Table 3-11 | Model Set A — Multivariate regressions of primary anthropometric measures: base predictors.

	Maternal T1D		Parity		Smoking		Maternal height		Maternal weight		Maternal wt gain		Maternal age		Paternal age		IMD imputed		3-month feeding		
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	R^2 adj
Weight SDS at birth	0.202	<0.0001	0.147	<0.0001	0.044	0.130	0.162	<0.0001	0.189	<0.0001			0.002	0.954	0.034	0.327	-0.031	0.275			0.141
Weight SDS at 12 months	0.057	0.064	0.030	0.350	-0.039	0.209	0.256	<0.0001	0.144	<0.0001			-0.053	0.151	-0.031	0.399	0.070	0.023	0.184	<0.0001	0.160
Weight SDS at 24 months	0.056	0.094	0.009	0.795	-0.054	0.109	0.193	<0.0001	0.207	<0.0001			-0.030	0.453	-0.026	0.509	0.052	0.116	0.074	0.031	0.123
Length SDS at birth	0.098	0.001	0.065	0.034	0.024	0.424	0.201	<0.0001	0.172	<0.0001			0.039	0.277	0.011	0.768	-0.035	0.236			0.104
Height SDS at 12 months	0.051	0.097	-0.006	0.844	-0.046	0.136	0.326	<0.0001	0.102	0.002			-0.034	0.356	-0.033	0.359	0.029	0.341	0.175	<0.0001	0.176
Height SDS at 24 months	0.005	0.874	-0.017	0.616	-0.044	0.189	0.303	<0.0001	0.119	0.001			-0.056	0.150	-0.004	0.918	0.063	0.055	0.119	<0.0001	0.151
Δweight SDS 0-12 months	-0.138	<0.0001	-0.115	<0.0001	-0.058	0.070	0.109	0.001	-0.033	0.336			-0.037	0.338	-0.049	0.199	0.116	<0.0001	0.206	<0.0001	0.105
Δweight SDS 0-24 months	-0.101	0.004	-0.133	<0.0001	-0.070	0.046	0.058	0.121	-0.002	0.961			-0.008	0.848	-0.037	0.371	0.098	0.005	0.097	0.006	0.052
Δheight SDS 0-12 months	-0.073	0.025	-0.060	0.072	-0.063	0.057	0.152	<0.0001	-0.026	0.456			-0.048	0.222	-0.021	0.595	0.079	0.016	0.213	<0.0001	0.083
Δheight SDS 0-24 months	-0.062	0.078	-0.069	0.054	-0.070	0.049	0.135	<0.0001	-0.058	0.127			-0.054	0.201	0.007	0.873	0.106	0.003	0.168	<0.0001	0.063

Table 3-12 | Model Set B — Multivariate regressions of primary anthropometric measures: maternal BMI replacing maternal weight and height.

	Maternal T1D		Parity		Smoking		Maternal BMI		Maternal weight		Maternal wt gain		Maternal age		Paternal age		IMD imputed		3-month feeding		
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	R^2 adj
Weight SDS at birth	0.204	<0.0001	0.139	<0.0001	0.041	0.170	0.162	<0.0001					-0.002	0.955	0.044	0.217	-0.029	0.326			0.084
Weight SDS at 12 months	0.061	0.060	0.025	0.453	-0.049	0.134	0.120	<0.0001					-0.058	0.135	-0.020	0.612	0.069	0.033	0.165	<0.0001	0.062
Weight SDS at 24 months	0.066	0.058	0.006	0.874	-0.072	0.039	0.180	<0.0001					-0.032	0.439	-0.019	0.649	0.048	0.168	0.054	0.128	0.047
Length SDS at birth	0.103	0.001	0.057	0.072	0.022	0.470	0.145	<0.0001					0.035	0.350	0.021	0.569	-0.034	0.278			0.032
Height SDS at 12 months	0.056	0.090	-0.012	0.724	-0.058	0.081	0.078	0.019					-0.040	0.314	-0.019	0.622	0.028	0.401	0.152	<0.0001	0.041
Height SDS at 24 months	0.025	0.482	-0.022	0.545	-0.067	0.057	0.091	0.010					-0.059	0.160	0.004	0.920	0.058	0.099	0.096	0.007	0.029
Δweight SDS 0-12 months	-0.137	<0.0001	-0.116	<0.0001	-0.061	0.058	-0.035	0.273					-0.038	0.327	-0.045	0.232	0.116	<0.0001	0.201	<0.0001	0.096
Δweight SDS 0-24 months	-0.099	0.004	-0.134	<0.0001	-0.074	0.035	-0.005	0.885					-0.008	0.846	-0.035	0.389	0.097	0.005	0.093	0.009	0.050
Δheight SDS 0-12 months	-0.070	0.033	-0.062	0.068	-0.067	0.043	-0.033	0.323					-0.050	0.209	-0.017	0.674	0.078	0.018	0.206	<0.0001	0.065
Δheight SDS 0-24 months	-0.056	0.115	-0.070	0.052	-0.077	0.031	-0.063	0.079					-0.054	0.202	0.009	0.837	0.104	0.003	0.162	<0.0001	0.052

Table 3-13 | Model Set C — Multivariate regressions of primary anthropometric measures: maternal weight gain incorporated.

	Maternal T1D		Parity		Smoking		Maternal height		Maternal weight		Maternal wt gain		Maternal age		Paternal age		IMD imputed		3-month feeding		
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	R^2 adj
Weight SDS at birth	0.160	<0.0001	0.156	<0.0001	0.034	0.315	0.130	<0.0001	0.180	<0.0001	0.114	0.001	-0.012	0.775	0.060	0.145	-0.043	0.207			0.120
Weight SDS at 12 months	0.073	0.047	0.032	0.401	-0.035	0.341	0.214	<0.0001	0.144	<0.0001	0.094	0.012	-0.046	0.285	-0.001	0.984	0.078	0.035	0.194	<0.0001	0.147
Weight SDS at 24 months	0.071	0.075	0.040	0.331	-0.053	0.190	0.168	<0.0001	0.185	<0.0001	0.060	0.139	-0.020	0.672	-0.003	0.949	0.055	0.172	0.101	0.014	0.105
Length SDS at birth	0.081	0.019	0.070	0.051	0.029	0.409	0.185	<0.0001	0.199	<0.0001	0.117	0.001	-0.016	0.704	0.080	0.053	-0.026	0.455			0.119
Height SDS at 12 months	0.004	0.911	-0.010	0.787	-0.042	0.241	0.362	<0.0001	0.077	0.053	0.052	0.158	-0.030	0.484	0.000	0.991	0.053	0.139	0.166	<0.0001	0.187
Height SDS at 24 months	-0.004	0.918	-0.002	0.961	-0.041	0.297	0.338	<0.0001	0.074	0.086	0.030	0.450	-0.034	0.455	0.023	0.610	0.083	0.034	0.130	0.001	0.153
Δweight SDS 0-12 months	-0.087	0.021	-0.123	0.002	-0.053	0.163	0.104	0.012	-0.034	0.424	-0.013	0.744	-0.015	0.744	-0.045	0.311	0.141	<0.0001	0.210	<0.0001	0.093
Δweight SDS 0-24 months	-0.056	0.174	-0.126	0.003	-0.072	0.083	0.056	0.217	-0.007	0.877	-0.031	0.456	0.006	0.905	-0.048	0.318	0.126	0.002	0.105	0.013	0.048
Δheight SDS 0-12 months	-0.082	0.034	-0.073	0.065	-0.060	0.125	0.207	<0.0001	-0.088	0.040	-0.063	0.107	0.004	0.932	-0.037	0.409	0.093	0.016	0.212	<0.0001	0.087
Δheight SDS 0-24 months	-0.059	0.154	-0.079	0.063	-0.074	0.076	0.194	<0.0001	-0.125	0.007	-0.081	0.054	0.003	0.957	-0.013	0.785	0.123	0.003	0.180	<0.0001	0.073

Table 3-14 | Model Set D — Multivariate regressions of primary anthropometric measures: birth size incorporated.

	Maternal T1D		Parity		Smoking		Maternal height		Maternal weight		Maternal age		Paternal age		IMD imputed		3-month feeding		Birth size*		
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	R^2 adj
Weight SDS at birth																					
Weight SDS at 12 months	-0.023	0.442	-0.029	0.336	-0.048	0.100	0.201	<0.0001	0.074	0.019	-0.047	0.177	-0.039	0.252	0.092	0.001	0.196	<0.0001	0.361	<0.0001	0.270
Weight SDS at 24 months	-0.004	0.891	-0.044	0.175	-0.061	0.052	0.148	<0.0001	0.129	<0.0001	-0.022	0.559	-0.029	0.437	0.073	0.020	0.087	0.007	0.354	<0.0001	0.231
Length SDS at birth																					
Height SDS at 12 months	-0.020	0.487	-0.022	0.436	-0.052	0.064	0.238	<0.0001	0.042	0.170	-0.036	0.290	-0.028	0.407	0.055	0.050	0.197	<0.0001	0.415	<0.0001	0.330
Height SDS at 24 months	-0.025	0.407	-0.035	0.252	-0.061	0.046	0.222	<0.0001	0.037	0.262	-0.051	0.157	-0.001	0.976	0.084	0.006	0.148	<0.0001	0.419	<0.0001	0.309
Δweight SDS 0-12 months	-0.021	0.442	-0.027	0.336	-0.045	0.100	0.191	<0.0001	0.070	0.019	-0.044	0.177	-0.037	0.252	0.087	0.001	0.186	<0.0001	-0.530	<0.0001	0.342
Δweight SDS 0-24 months	-0.004	0.891	-0.041	0.175	-0.057	0.052	0.138	<0.0001	0.119	<0.0001	-0.020	0.559	-0.027	0.437	0.067	0.020	0.080	0.007	-0.574	<0.0001	0.337
Δheight SDS 0-12 months	-0.020	0.487	-0.023	0.436	-0.054	0.064	0.246	<0.0001	0.043	0.170	-0.037	0.290	-0.029	0.407	0.057	0.050	0.203	<0.0001	-0.472	<0.0001	0.285
Δheight SDS 0-24 months	-0.025	0.407	-0.036	0.252	-0.062	0.046	0.225	<0.0001	0.037	0.262	-0.052	0.157	-0.001	0.976	0.085	0.006	0.150	<0.0001	-0.505	<0.0001	0.294

Table 3-15 | Model Set E — Multivariate regression of primary anthropometric measures using the Stepwise Forward method: validation of predictors.

	Maternal T1D		Parity		Smoking		Maternal height		Maternal weight		Maternal wt gain		Maternal age		Paternal age		IMD imputed		3-month feeding		Birth size*		
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	R^2 adj
Weight SDS at birth	0.161	<0.0001	0.170	<0.0001	N/A	N/A	0.134	<0.0001	0.175	<0.0001	0.114	0.001	N/A	N/A	N/A	N/A	N/A	N/A					0.119
Weight SDS at 12 months	N/A	N/A	N/A	N/A	N/A	N/A	0.205	<0.0001	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.116	0.001	0.232	<0.0001	0.378	<0.0001	0.262
Weight SDS at 24 months	N/A	N/A	N/A	N/A	N/A	N/A	0.135	0.001	0.117	0.004	N/A	N/A	N/A	N/A	N/A	N/A	0.094	0.012	0.120	0.001	0.358	<0.0001	0.220
Length SDS at birth	0.081	0.019	0.070	0.049	N/A	N/A	0.188	<0.0001	0.195	<0.0001	0.116	0.001	N/A	N/A	0.074	0.037	N/A	N/A					0.121
Height SDS at 12 months	N/A	N/A	N/A	N/A	N/A	N/A	0.289	<0.0001	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.077	0.020	0.201	<0.0001	0.394	<0.0001	0.331
Height SDS at 24 months	N/A	N/A	N/A	N/A	N/A	N/A	0.268	<0.0001	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.107	0.003	0.168	<0.0001	0.397	<0.0001	0.306
Δweight SDS 0-12 months	N/A	N/A	N/A	N/A	N/A	N/A	0.195	<0.0001	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.110	0.001	0.220	<0.0001	-0.527	<0.0001	0.336
Δweight SDS 0-24 months	N/A	N/A	N/A	N/A	N/A	N/A	0.125	0.001	0.107	0.004	N/A	N/A	N/A	N/A	N/A	N/A	0.086	0.012	0.110	0.001	-0.577	<0.0001	0.338
Δheight SDS 0-12 months	N/A	N/A	N/A	N/A	N/A	N/A	0.296	<0.0001	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.079	0.020	0.207	<0.0001	-0.491	<0.0001	0.295
Δheight SDS 0-24 months	N/A	N/A	N/A	N/A	N/A	N/A	0.269	<0.0001	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.107	0.003	0.169	<0.0001	-0.511	<0.0001	0.299

Table 3-16 | Model Set F — Multivariate regression of primary anthropometric measures using the Forced Entry Method: on selected predictors only.

	Maternal T1D		Parity		Smoking		Maternal height		Maternal weight		Maternal wt gain		Maternal age		Paternal age		IMD_imputed		3-month feeding		Birth size*		
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	R^2 adj
Weight SDS at birth	0.195	<0.0001	0.152	<0.0001	0.054	0.049	0.162	<0.0001	0.192	<0.0001													0.140
Weight SDS at 12 months							0.197	<0.0001	0.083	0.007							0.096	0.001	0.218	<0.0001	0.339	<0.0001	0.257
Weight SDS at 24 months							0.152	<0.0001	0.133	<0.0001							0.075	0.014	0.114	<0.0001	0.335	<0.0001	0.221
Length SDS at birth	0.095	0.001	0.075	0.009	0.035	0.218	0.209	<0.0001	0.160	<0.0001													0.104
Height SDS at 12 months							0.248	<0.0001									0.073	0.007	0.223	<0.0001	0.416	<0.0001	0.324
Height SDS at 24 months							0.225	<0.0001									0.107	<0.0001	0.181	<0.0001	0.422	<0.0001	0.301
Δweight SDS 0-12 months							0.184	<0.0001	0.078	0.007							0.090	0.001	0.204	<0.0001	- 0.555	<0.0001	0.351
Δweight SDS 0-24 months							0.139	<0.0001	0.122	<0.0001							0.068	0.014	0.103	<0.0001	- 0.597	<0.0001	0.352
Δheight SDS 0-12 months							0.254	<0.0001									0.075	0.007	0.228	<0.0001	- 0.477	<0.0001	0.290
Δheight SDS 0-24 months							0.226	<0.0001									0.107	<0.0001	0.181	<0.0001	- 0.503	<0.0001	0.296

* weight models adjusted for birth weight; height models adjusted for birth length

Table 3-17 | Model Set G — Multivariate regression on secondary anthropometric variables and adiposity.

	Maternal T1D		Parity		Smoking		Maternal height		Maternal weight		Maternal age		Paternal age		IMD imputed		Maternal qualif.		3-month feeding		Birth size*		
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	R^2 adj
HC SDS at birth	0.070	0.021	0.129	<0.0001	-0.008	0.787	0.135	<0.0001	0.149	<0.0001	-0.026	0.487	0.058	0.115	-0.068	0.026	0.080	0.008					0.081
HC SDS at 12 months	-0.046	0.104	-0.043	0.148	-0.009	0.748	0.080	0.009	0.034	0.270	-0.032	0.350	0.026	0.452	0.053	0.065	0.035	0.225	0.036	0.221	0.533	<0.0001	0.304
HC SDS at 24 months	-0.038	0.213	-0.056	0.071	-0.024	0.440	0.074	0.022	0.039	0.229	-0.025	0.491	0.023	0.519	0.054	0.074	0.041	0.172	-0.034	0.268	0.538	<0.0001	0.308
BMI SDS at birth	0.155	<0.0001	0.106	0.001	0.065	0.033	0.048	0.142	0.153	<0.0001	-0.017	0.639	0.062	0.092	-0.004	0.906							0.065
BMI SDS at 12 months	0.003	0.926	0.016	0.633	-0.027	0.419	0.060	0.085	0.082	0.021	-0.027	0.491	-0.028	0.475	0.074	0.024			0.117	<0.0001	0.259	<0.0001	0.104
BMI SDS at 24 months	0.015	0.676	0.012	0.732	-0.055	0.117	-0.016	0.672	0.127	0.001	0.040	0.331	-0.049	0.236	-0.001	0.969			0.008	0.820	0.259	<0.0001	0.087
SFT SDS at birth	0.102	0.001	0.159	<0.0001	-0.060	0.054	0.011	0.747	0.087	0.009	-0.040	0.284	-0.015	0.685	-0.040	0.199							0.040
SFT SDS at 12 months	-0.017	0.609	-0.051	0.139	0.028	0.412	-0.011	0.754	0.066	0.068	-0.014	0.730	0.007	0.855	0.092	0.006			0.179	<0.0001	0.129	<0.0001	0.052
SFT SDS at 24 months	-0.045	0.205	-0.079	0.033	-0.007	0.839	-0.022	0.566	0.072	0.061	0.014	0.751	-0.036	0.400	0.030	0.396			0.040	0.270	0.134	<0.0001	0.021

* HC models adjusted for head circumference at birth; BMI models adjusted for BMI at birth; skinfold thickness (SFT) models adjusted for skinfold thickness at birth

Table 3-18 | Model Set H — Multivariate regression on levels of endocrine factors.

	Gestational Age		Sex		Maternal T1D		Parity		Smoking		Maternal height		Maternal weight		Maternal age		Paternal age		IMD imputed		3-month feeding		
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	R^2 adj
3-mo IGF-1	-0.053	0.256	-0.078	0.095	0.025	0.588	-0.031	0.516	0.010	0.826	0.009	0.854	0.003	0.945	0.034	0.554	0.020	0.731	0.113	0.016	0.248	<0.0001	0.063
12-mo IGF-1	-0.111	0.059	0.224	<0.0001	0.039	0.498	-0.167	0.003	-0.094	0.101	-0.028	0.637	-0.028	0.639	0.135	0.057	-0.124	0.080	0.008	0.887	0.107	0.068	0.094
18-mo IGF-1	-0.049	0.459	0.183	0.006	0.062	0.336	-0.163	0.014	-0.126	0.052	-0.020	0.764	-0.020	0.768	0.121	0.112	0.027	0.725	0.062	0.332	0.207	0.003	0.114
24-mo IGF-1	-0.123	0.232	0.159	0.103	0.067	0.491	-0.072	0.459	0.104	0.299	0.148	0.179	-0.126	0.262	0.052	0.629	-0.227	0.047	0.058	0.533	0.092	0.367	0.048
3-mo IGFBP-3	-0.043	0.376	0.087	0.069	0.038	0.433	0.084	0.086	-0.046	0.343	-0.024	0.636	0.019	0.707	0.046	0.438	-0.013	0.825	0.043	0.379	0.085	0.084	0.007
12-mo IGFBP-3	-0.052	0.403	0.173	0.005	0.093	0.128	0.003	0.957	0.020	0.737	0.064	0.314	-0.037	0.567	0.025	0.734	-0.029	0.702	-0.001	0.981	0.007	0.913	0.011
18-mo IGFBP-3	-0.034	0.631	0.248	<0.0001	0.055	0.421	0.024	0.730	-0.100	0.144	-0.044	0.541	0.067	0.345	0.072	0.380	-0.042	0.614	-0.057	0.397	0.003	0.966	0.041
24-mo IGFBP-3	0.011	0.917	0.241	0.016	0.219	0.029	-0.003	0.971	0.206	0.044	-0.003	0.977	0.006	0.960	0.180	0.107	-0.207	0.072	-0.060	0.526	-0.014	0.894	0.070
3-mo C-peptide	-0.037	0.562	-0.020	0.760	0.075	0.238	-0.120	0.066	0.073	0.256	0.092	0.187	0.047	0.497	-0.021	0.799	0.005	0.954	0.036	0.574	0.145	0.031	0.025
12-mo C-peptide	-0.149	0.099	-0.167	0.073	N/A	N/A	-0.230	0.013	-0.051	0.572	0.253	0.015	0.035	0.728	0.137	0.201	-0.220	0.037	-0.027	0.766	0.056	0.564	0.120
3-mo Leptin	-0.005	0.940	0.186	0.008	-0.037	0.585	0.068	0.323	-0.235	0.001	-0.072	0.333	-0.095	0.206	0.021	0.800	0.058	0.477	0.049	0.481	-0.046	0.518	0.062
12-mo Leptin	0.025	0.819	0.012	0.915	-0.033	0.758	-0.181	0.092	-0.040	0.713	-0.019	0.863	0.143	0.231	0.054	0.669	-0.127	0.321	0.064	0.554	-0.104	0.365	-0.025

The associations of head circumference with sociodemographic indices merited exploration. It transpired that these associations are due to confounding by smoking. Specifically, the proportion of women who smoked during pregnancy was significantly higher amongst single vs. co-habiting vs. married women ($P_{\text{chi-square}}$=0.01, Table 3-19).

Table 3-19 | Distribution of mothers by smoking during pregnancy and marital status.

N (%)				
		Maternal marital status		
		married	**co-habiting**	**single**
Maternal smoking during pregnancy	**yes**	30 (2.9)	11 (5.9)	4 (14.8)
	no	995 (97.1)	176 (94.1)	23 (85.2)

In a similar pattern, there was a significantly higher proportion of women who did not smoke during pregnancy with increasing academic qualification ($P_{\text{chi-square}}$<0.0001, Table 3-20). The IMD (unimputed) was lower amongst women who didn't smoke (8.9, n=849) compared with smokers (10.1, n=33) but the difference was not significant (P=0.208). Nevertheless, the abovementioned findings collectively warn against the use of sociodemographic variables in models due to possible confounding.

Table 3-20 | Distribution of mothers by smoking during pregnancy and maternal education.

N (%)				
		Maternal education		
		GCSE	**A-Level**	**University**
Maternal smoking during pregnancy	**yes**	9 (8.8)	8 (4.7)	7 (1.5)
	no	93 (91.2)	163 (95.3)	459 (98.5)

Assumptions for linear regression require that residuals are i) independent, ii) follow a normal distribution and iii) have a constant variance[432]. The main regressions herein were shown to meet these requirements by plotting the distributions of standardised residuals and showing they are in a random scatter plot with no change in variance across predicted values (distributions not shown).

As a final method of validating predictors, the fitness of each model was assessed by the AIC, which additionally penalises for the number of parameters included in an iterative process of excluding variables one at a time and favouring the model with the lowest AIC value. This method surfaced that BMI models are optimised by using maternal BMI, in lieu of maternal height and weight. The statistically-identified covariates were justified by biological reasoning and the literature. Table 3-21 lists the final predictors by outcome and age.

Table 3-21 | Predictors to adjust for in regression analyses by response variable (growth parameter or levels of endocrine factor) and age (where appropriate).

	W_0H_0	W_3,H_3	$W_{\geq 12}$	$H_{\geq 12}$	ΔW	ΔH	HC_0	$HC_{\geq 12}$	BMI_0	$BMI_{\geq 12}$	SFT_0	$SFT_{\geq 3}$	IGF-1	C-pep	Leptin
Newborn index															
Gestation	SDS values used														
Sex	SDS values used												**X**		**X**
Pregnancy influences															
Maternal T1D	Excluded from analysis														
Parity	**X**	**X**					**X**		**X**		**X**				**X**
Maternal smoking in pregnancy	**X**								**X**						
Maternal height	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**							
Maternal pre-pregnancy weight	**X**	**X**	**X**	**X**	**X**		**X**	**X**							
Maternal pre-pregnancy BMI									**X**	**X**					
Maternal pregnancy weight gain	Non-robust measurements and low coverage in the cohort (<60%)														
Maternal age															
Paternal age															
Post-Pregnancy influences															
Birth delivery	Reverse causality with size at birth														
3-month feeding		**X**	**X**	**X**	**X**	**X**				**X**		**X**	**X**	**X**	
Birth weight SDS					**X**										
Birth length SDS						**X**									
Sociodemographic influences															
Ethnicity	Caucasian cohort in majority														
Index of multiple deprivation	Possible confounding														
Maternal marital status	Confounding with smoking														
Maternal education	Confounding with smoking														

3.5 EXCLUSION CRITERIA

The comparative growth trajectories described in Section 3.3 showed that twins, offspring of mothers with T1D, and infants born prematurely (<36 weeks of gestation) were characterised by extremes of size at birth, which intrinsically affects postnatal growth, thus representing non-tolerable sources of noise to omit from analyses. Late preterm infants (36 weeks of gestation) were included to increase the number of subjects. The question whether these exclusion criteria are applicable to hormones was investigated by correlational analysis (Appendix II). Despite lack of statistical power, clinically significant differences were observed for hormone levels, justifying the application of the abovementioned exclusion criteria across the board.

3.6 COMPARING GENOTYPED VS. NON-GENOTYPED SUBJECTS

A total of 612 infants of the CBGS had normalised DNA samples, which were readily used for genotyping. Of these, 597 subjects also had phenotypes of varied extent across the course of the study depending on stage of withdrawal (Figure 3-26).

Figure 3-26 | Count of CBGS genotyped infants by stage of withdrawal from the study.

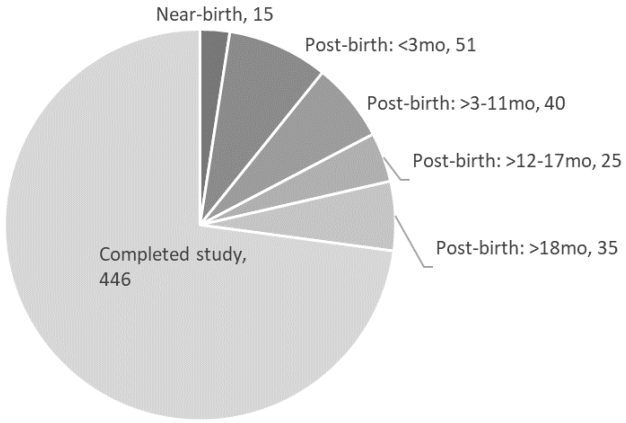


Table 3-22 compares the genotyped against the non-genotyped subjects. Genotyped infants were on average of longer gestation and born with higher weight, higher height and higher head circumference than babies who were not genotyped, yet the magnitude of differences was not clinically significant. The groups were similar with respect to cohort characteristics, except that the genotyped subgroup had a circa 10% higher percentage of normal deliveries (P<0.0001), lower frequency of twins (P=0.001) and lower frequency of premature births (P=0.007). The difference in maternal weight gain during pregnancy was of nominal statistical significance (P=0.033).

Table 3-22 | Comparison of cohort characteristics between non-genotyped vs. genotyped infants.

	Non-genotyped		Genotyped		
	N	**Mean ± SD or n (%) for specified outcome**	**N**	**Mean ± SD or n (%) for specified outcome**	***P* value**
Newborn index					
Gestation (wk)	1,063	39.7 ± 1.7	597	40.0 ± 1.4	0.002
Twins	1,063	39 (3.7)	597	5 (0.8)	0.001
Infant sex: male	1,063	538 (50.6)	597	321 (53.8)	0.217
Premature birth: < 36 weeks of gestation	1,063	33 (3.1)	597	6 (1.0)	0.007
Relative birth weight	1,063		594		0.104
appropriate for gestational age		950 (89.4)		510 (85.9)	
large for gestational age		69 (6.5)		50 (8.4)	
small for gestational age		44 (4.1)		34 (5.7)	
Birth weight (kg)	1,063	3.44 ± 0.55	594	3.53 ± 0.53	0.022
Birth length (cm)	1,021	51.3 ± 2.6	578	51.6 ± 2.5	0.030
Newborn head circumference (cm)	1,022	35.2 ± 1.7	580	35.4 ± 1.6	0.027
Newborn BMI (kg/m^2)	1,021	13.1 ± 1.7	575	13.2 ± 1.6	0.094
Newborn mean skinfold thickness (mm)	1,021	6.3 ± 1.6	578	6.3 ± 1.6	0.862
Pregnancy influences					
Maternal T1D	1,044	3 (0.3)	594	2 (0.3)	0.862
Primiparous pregnancy	1,040	451 (43.4)	589	256 (43.5)	0.969
Maternal smoking during pregnancy	1,014	61 (6.0)	561	25 (4.5)	0.192
Maternal height (cm)	740	165.7 ± 7.3	497	166.3 ± 7.0	0.140
Maternal pre-pregnancy weight (kg)	722	66.5 ± 14.1	481	65.9 ± 12.1	0.591
Maternal pre-pregnancy BMI (kg/m^2)	702	24.3 ± 4.9	471	23.8 ± 4.0	0.596
Maternal pregnancy weight gain (kg)	514	7.8 ± 6.7	352	8.8 ± 6.6	0.031
Maternal age (yr)	811	33.4 ± 4.4	527	33.6 ± 4.1	0.504
Paternal age (yr)	776	35.7 ± 5.4	515	35.7 ± 5.4	0.697
Post-pregnancy influences					
Birth delivery	1,042		588		<0.0001
normal		576 (55.3)		380 (64.6)	
vacuum		57 (5.5)		44 (7.5)	
forceps		60 (5.8)		22 (3.7)	
elective caesarean		175 (16.8)		77 (13.1)	
acute caesarean		174 (16.7)		65 (11.1)	
3-month feeding: exclusively breastfed	793	324 (40.9)	526	233 (44.3)	0.216
Sociodemographic influences					
Ethnicity: White Caucasian	753	718 (95.4)	345	327 (94.8)	0.683
Index of multiple deprivation	542	9.1 ± 4.3	378	8.8 ± 4.2	0.451
Maternal marital status	777		514		0.519
married		640 (82.4)		430 (83.7)	
co-habiting		122 (15.7)		71 (13.8)	
single		15 (1.9)		13 (2.5)	
Maternal education	439		341		0.357
GCSE		61 (13.9)		49 (14.4)	
A-Level		94 (21.4)		87 (25.5)	
University		284 (64.7)		205 (60.1)	

3.7 COVERAGE OF MEASUREMENTS

The coverage of variables measured was comparable between the entire cohort and the genotyped subset (Table 3-23).

Table 3-23 | Coverage (%) of measured variables in the entire study population vs. genotyped subset.

	Entire CBGS (n=1660)			Genotyped CBGS (n=597)		
	Valid	**Missing**	**%**	**Valid**	**Missing**	**%**
Newborn index						
Gestation	1,660	0	100.0	597	0	100.0
Twin	1,660	0	100.0	597	0	100.0
Infant sex	1,660	0	100.0	597	0	100.0
Anthropometric measures (SDS)						
Birth weight	1,657	3	100.0	594	3	99.5
Birth length	1,599	61	96.3	578	19	96.8
Newborn head circumference	1,602	58	100.0	580	17	97.2
Newborn skinfold thicknesses	1,592	68	100.0	575	22	96.3
3-month weight	1,340	320	80.7	528	69	88.4
3-month height	1,335	325	80.4	522	75	87.4
3-month skinfold thicknesses	1,342	318	80.8	528	69	88.4
12-month weight	1,181	479	71.1	478	119	80.1
12-month height	1,174	486	70.7	474	123	79.4
12-month head circumference	1,178	482	71.0	476	121	79.7
12-month skinfold thicknesses	1,177	483	70.0	476	121	79.7
18-month weight	1,093	567	65.8	458	139	76.7
18-month height	1,095	565	66.0	457	140	76.5
18-month head circumference	1,087	573	65.5	455	142	76.2
18-month skinfold thicknesses	1,095	565	66.0	458	139	76.7
24-month weight	1,050	610	63.3	441	156	73.9
24-month height	1,042	618	62.8	435	162	72.9
24-month head circumference	1,050	610	63.3	438	159	73.4
24-month skinfold thicknesses	1,057	603	63.7	442	155	74.0
Pregnancy influences						
Maternal T1D	1,638	0	100.0	594	0	100.0
Parity	1,629	9	99.5	589	5	99.2
Maternal smoking during pregnancy	1,575	63	96.2	561	33	94.4
Maternal height	1,237	401	75.5	497	97	83.7
Maternal pre-pregnancy weight	1,203	435	73.4	481	113	81.0
Maternal pre-pregnancy BMI	1,173	465	71.6	471	123	79.3
Maternal pregnancy weight gain	866	772	52.9	352	242	59.3
Maternal age	1,338	300	81.7	527	67	88.7
Paternal age	1,291	347	78.8	515	79	86.7
Post-pregnancy influences						
Birth delivery	1,630	32	98.2	588	9	98.5
3-month feeding	1,319	343	79.5	526	71	88.1
Sociodemographic influences						
Ethnicity	1,098	564	66.1	345	252	57.8
Index of multiple deprivation	920	718	56.2	378	216	63.6
Index of multiple deprivation: imputed	1,638	0	100.0	594	0	100.0
Maternal marital status	1,291	347	78.8	514	80	86.5
Maternal education	780	858	47.6	341	253	57.4
Endocrine factors						
3-month IGF-1	569	1,091	34.3	213	384	35.7
12-month IGF-1	387	1,273	23.3	170	427	28.5
18-month IGF-1	300	1,360	18.1	138	459	23.1
24-month IGF-1	164	1,496	9.9	90	507	15.1
3-month C-peptide	317	1,343	19.1	123	474	20.6
12-month C-peptide	144	1,516	8.7	56	541	9.4
3-month leptin	311	1,349	18.7	122	475	20.4
12-month leptin	238	1,422	14.3	96	501	16.1

3.8 CONCLUSION

The cohort is representative of an almost ethnically homogeneous population of infants in Cambridgeshire and possesses robust measures of early growth and relevant endocrine regulators measured in whole blood, as well as a range of covariates to adjust for, with the limitation that paternal height was not collected. The participation rate was high at 74% and the rate of complete follow-up to 2 years of age was respectably 64%. Infants who completed the follow-up were very similar to those who were lost to follow-up, but their parents were slightly older in age and their mothers less likely to smoke during pregnancy. Coverage of anthropometric variables is sufficient, but attritions as the study progresses and is lowest at 24 months of age.

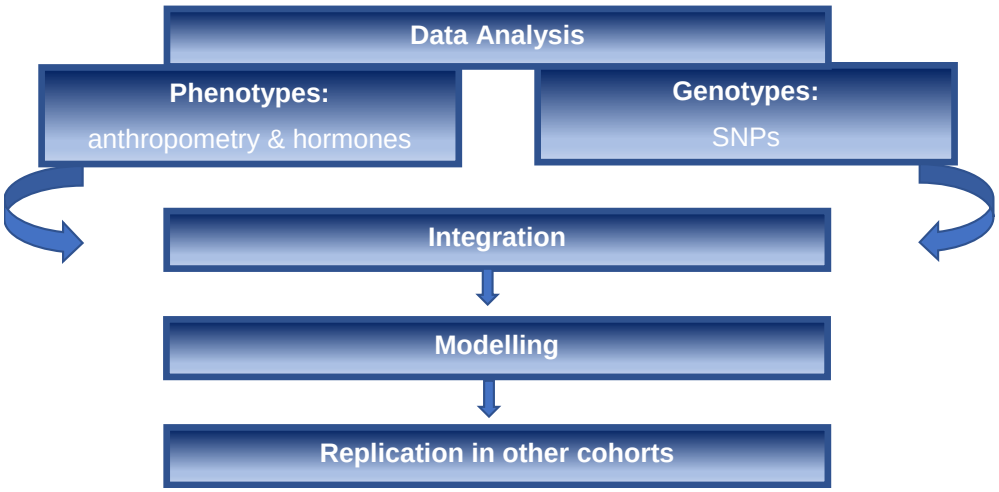
Chapter 4

COLLECTING AND EXPLORING THE GENOTYPES

4.1 INTENT

The research question tackled in this thesis is a hypothesis-driven probe for the genetic basis of phenotypes commonly observed in the prodrome of childhood diabetes during infancy. The elucidation of such genotype-phenotype correlates requires focussed collection and processing of anthropometric and biological observations, in tandem with genetic data (Figure 4-1). In fulfilment of the second half of the objective, selected genetic loci implicated in T1D were determined in a subset of the CBGS cohort using DNA extracted from blood or buccal samples. The screening of T1D loci that might play a role in the biological events of the preclinical stage of the disease was the initial step in teasing out an unbiased association.

Figure 4-1 | A flowchart for the integration of genotype and phenotype data.



This chapter describes and explores the collected genetic data in three stages:

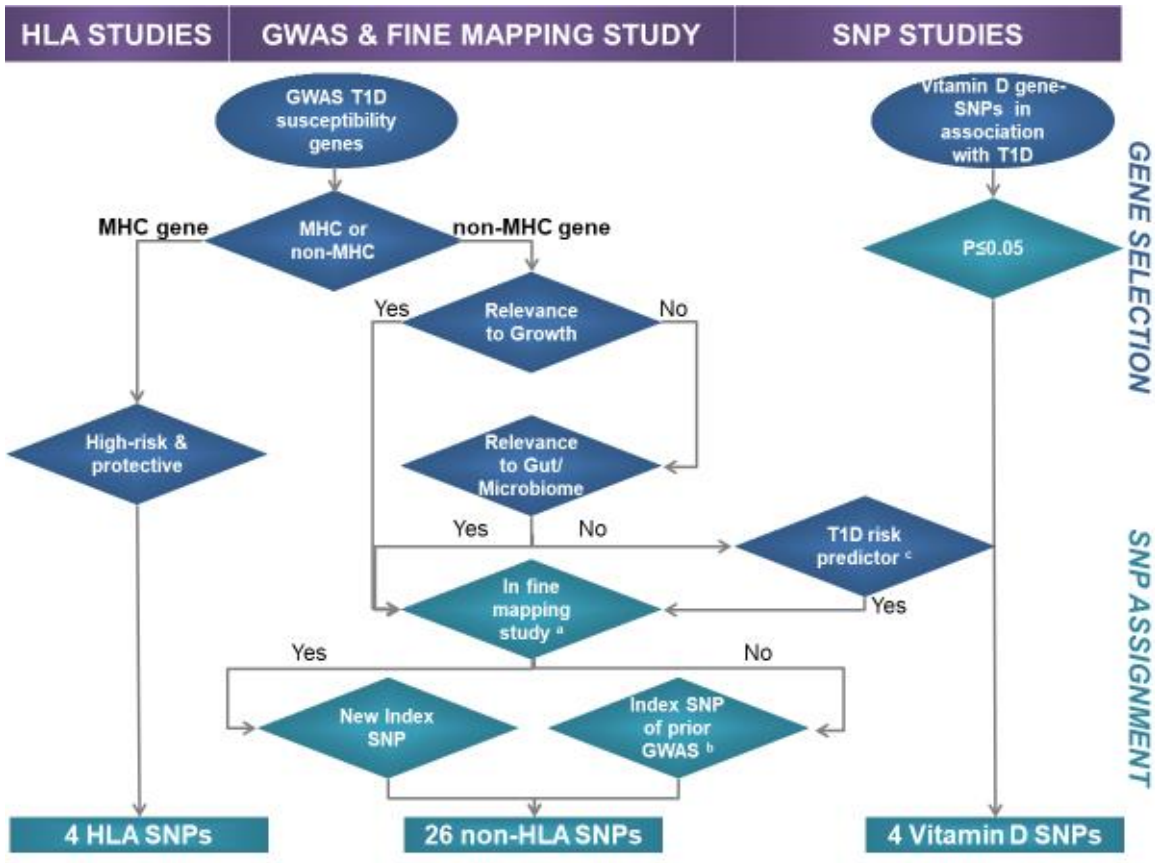
i. description of methods for selecting and genotyping candidate SNPs;
ii. presentation of descriptive statistics and quality checks done in accord with genetic studies;
iii. preparation of data to be used in analyses as individual SNPs or genetic risk scores.

4.2 SELECTING T1D GENETIC VARIANTS

4.2.1 Rationale and approach

Chapter 1 highlighted the putative link between T1D in childhood and a trio of factors: growth, gut/microbiome and vitamin D, which served as the selection criteria for genetic variants to feed into the hypothesis-driven testing. The focussed search for biologically meaningful variants, as opposed to scanning the net wide and capturing all SNPs implicated in T1D, meant that hypotheses were underpinned by science and potential findings justified *a priori*, in contrast to serendipitous discoveries made by pure statistical workings. Targeted identification of SNPs also offered the benefit of minimising the use of DNA, which is a precious resource for a birth cohort. Here, I present the rationale and approach for generating a list of SNPs to take forward as reference markers. The decision flow is outlined in Figure 4-2.

Figure 4-2 | Decision diagram for selecting T1D genetic variants for hypothesis-driven testing.



JA Todd group, 2016

[a] *Onengut-Gumuscu et al, 2015*

[b] *ImmunoBase, accessed 2015*

[c] *Winkler et al, 2014*

Cooper et al, 2011

In principle, I rigorously searched through existing biological facts on T1D candidate genes for clues that supported a conjecture about their involvement in somatic growth. Next, I identified T1D susceptibility variants in each gene and selected the most appropriate SNP (e.g. index). The assignment of SNPs to genes in a one-to-one relationship served to i) isolate the effect a single SNP from each region and ii) maximise coverage of regions across the human genome. The premise of my selection is the assumption that SNPs exert some regulatory effect on their pinpointed genes. Of note, this approach is inverse to the one employed by GWAS, which scan for SNPs that are in strong association with the outcome and subsequently map them to genes.

The starting point of the screening process was the curated list of 59 T1D susceptibility chromosomal regions sourced from ImmunoBase (www.immunobase.org, accessed 2015), a web-based portal built on the same platform as its sister T1Dbase[434], which is an excellent resource for up-to-date information on the genetics of autoimmune diseases. ImmunoBase was created in parallel with the fine mapping study of T1D susceptibility loci by Onengut-Gumuscu et al.[102], whose publication coincided with the outset of this project. In general, fine mapping studies make use of the Immunochip — an Illumina microarray chip designed to interrogate 195,806 SNPs and 718 small insertions-deletions — which is intended for i) deep replication of meta-GWAS by narrowing the content to 2,000 independent association signals and ii) fine mapping of GWAS loci by using tens of thousands of samples (as opposed to the few hundred samples used in HapMap analyses and genotyping all identified SNPs within a region), which results in improved resolution[435]. In principle, fine mapping studies define a 'credible set of polymorphisms' that are likely to contain the causal variants[436]. The fine mapping study of T1D served as the source of non-MHC genetic variants used in this thesis. Oliver Burren, one of the first authors of the study, at the Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory (JDRF/WTDIL) of the University of Cambridge provided insight via personal communication. The fine mapping study, conducted on case-control and family populations of European ancestry, transcended prior genetic investigations on the subject as it refined the genetics of T1D and expanded the disease scope to allow for comparison with 15 other autoimmune diseases[102] — an outcome which lent itself to the qualification of short-listed loci as described next.

4.2.2 Selection of genes

Genes were selected by functional category of pertinence to the project context (search steps I to III). The implication of selecting genes based on their established biology is the exclusion of genes with unknown or as-yet unidentified function, which I bypassed with search step (IV):

I. MHC genes associated with susceptibility to T1D (*HLA-DR* and *HLA-DQ*);
II. non-MHC genes associated with T1D at GWAS-significance level and having a putative biological relevance to 'growth' or 'gut/microbiome';
III. vitamin D genes associated with T1D but below GWAS-significance level, and thus not captured in the previous step or included in ImmunoBase;
IV. genes identified, via pure statistical reasoning and not biological knowledge, to predict progression to T1D in the German BABYDIAB and BABYDIET studies of infants at familial risk of T1D[437].

Whereas selection of MHC and vitamin D genes was straightforward, screening for non-MHC genes required a thorough search. It is known that an associated region may contain a single gene or a few genes or found in a 'gene desert'[117]. As such, I systematically applied the following sub-steps for the selection of non-MHC genes (step II) with relevance to 'growth' or 'gut/microbiome':

1. *Selection*: For each of 59 chromosomal regions, I selected the causal gene pinpointed.
2. *Enrichment*: I enriched the list of causal genes by searching through ImmunoBase and publications for 'alternative protein-coding genes'. This search lent itself to regions which had not been assigned a causal gene by ImmunoBase. A total of 65 genes were listed.
3. *Screening I*: Genes were screened for evidence of association with *'Growth'* on the grounds of fulfilling at least one of the following requirements based on information sourced from GeneCards (www.genecards.org, accessed 2015):
    i. Molecular function of gene products as defined by
        a. functional classification (e.g. growth factors),
        b. gene ontology (e.g. growth hormone receptor),
        c. functional interaction with growth factor receptors;
    ii. Phenotype in
        a. human (e.g. associations with growth-related disease),
        b. mouse, where the phenotype includes relevant terms (e.g. 'growth', 'size');
    iii. Functional interactions with genes identified in steps (3i) or (3ii), e.g. *CTLA4* was selected based on its interaction with *PTPN22* which qualified from (3i).
4. *Screening II*: Genes were screened for evidence of an association with the 'gut' or 'microbiome' based on literature or information sourced from GeneCards.

The rationale for selecting each gene used herein based on the above algorithm is briefed in Appendix II. The final panel of selected candidate genes was cross-referenced with the fine mapping study[102] and GWAS[438,439]. Some genes that were previously in focus were absent from ImmunoBase (e.g. *RGS1*), or not pinpointed in the fine mapping study on the grounds of having no known role in immunity or functional data linking it to T1D (e.g. *RNLS*)[102].

4.2.3 Selection of SNPs

4.2.3.1 MHC SNPs

The HLA class II region is strongly implicated in the pathogenesis of T1D and encompasses both high-risk and protective alleles. Genotyping 2 or 3 SNPs for capturing T1D risk in the HLA region has emerged as a simple, rapid and cost-effective alternative to the gold-standard method of HLA typing[440,441]. Barker et al.[440] showed that two SNPs alone tag the high-risk DR3/DR4-DQ8 with 98.6% sensitivity and 99.7% specificity.

Information on the high-risk HLA class II variants and genotype risk groups was provided by Neil Walker at the JDRF/WTDIL of the University of Cambridge (N. Walker, personal communication, 2016). Walker and colleagues identified a 3-SNP tag model that makes it possible to identify the majority of known HLA risk in a population of infants not preselected by family history. The tags are rs2187668 for *HLA-DRB1*03* (DR3), rs17426593 for *HLA-DRB1*04* (DR4) and rs7454108 for *HLA-DQB1*03:02* (DQ8). Walker's 3-SNP tag model replaces the previous 2-SNP tag model — that keeps the DQ8 tag but omits the DR4 tag — as it captured more risk than was supposed to, yet the gains are marginal. Walker's model offers an advantage over the alternative 3-SNP tag model proposed by Nguyen et al.[441] which cannot be replicated using TaqMan.

Amongst HLA haplotypes with a protective effect against T1D, the most known is DR2, which is the serological nomenclature that refers to the Caucasian haplotype DRB1*15:01-DQA1*01:02-DQB1*06:02[168]. Other protective haplotypes in decreasing order of magnitude of effect include (DR7)-DQA1*02:01-DQB1*03:03, (DR14)-DQB1*05:03, DRB1*04:03-DQB1*03:02, (DR13)-DQB1*06:03, (DR11/12/13)-DQA1*05-DQB1*03:01, (DR1)-DQB1*05:01[354]. Nevertheless, the HLA-DQB1*06:02 associated with dominant protection against T1D[172], and in the UK case-control cohort it occurred in a ratio of 1 to 26 (N. Walker, personal communication, 2016). In population studies in Scandinavia, this haplotype significantly associated[442] with higher birth weight[15,403]. On these grounds, the *DQB1*06:02* was my selected T1D protective HLA reference marker to genotype in the CBGS.

Published methods employed for capturing the protective effect of HLA against T1D in study subjects were based on oligotyping procedures[14,15,161,203]. As there was no precedence, to my knowledge, of genotyping SNPs tagging protective HLA-II alleles at the outset of this project, I sourced the tag SNP for *HLA-DQB1*06:02*, which also tags the *HLA-DRB1*15:01,* from the study of De Bakker et al.[156].

4.2.3.2 Non-MHC SNPs

The fine mapping study by Onengut-Gumuscu et al.[102] served as the source of SNPs for the selected non-MHC genes on the grounds that it is:

i. the most recent genetic study in T1D in the largest population to-date of 6,670 cases, 9,716 controls and 2,601 families;
ii. done on Immunochip, thus increasing the likelihood that causal variants are among the SNPs genotyped (SNPs showing the strongest association with the disease are not necessarily causal owing to sampling variation and LD);
iii. conducted in a UK population.

Five genes harboured more than just the causal variant, i.e. the one with the strongest association in the primary analysis, namely (number of associations is shown in square brackets): *IFIH1* [3], *IL2RA* [3], *INS* [2], *PTPN2* [2], *TYK2* [2][102]. As explained previously, only the index SNP was selected.

4.2.3.3 Vitamin D SNPs

SNPs mapped to genes of the Vitamin D milieu were sourced from Cooper et al.[77], who tested for associations between genetic variants in seven vitamin D metabolism genes and both circulating 25(OH)D levels and T1D status in British cases (paediatric and adult diabetes clinics at 150 National Health Service hospitals in the UK), controls (British 1958 Birth Cohort and UK Blood Services Common Control Collection), and families of white European ancestry. A total of four SNPs significantly associated ($P \leq 0.05$) with T1D in the case-control analysis.

4.2.4 Expanding the search

The screening process anchored on existing properties of T1D candidate genes that are conjectured to influence growth. For the search to be comprehensive, the flipside of the association under test was considered. Thus, I conducted a blanket search for genetic variants that have been associated with phenotypes of relevance: anthropometric traits or growth regulators. Towards this end, I used the GWAS catalogue — founded in 2008 and provided jointly by the National Human Genome Research Institute (NHGRI) and the European Bioinformatics Institute (EBI) — which is a quality controlled, curated, literature-derived collection of all published GWAS assaying at least 100,000 SNPs and *P* values $<1.0 \times 10^{-5}$ for all SNP-trait association[442]. Amongst a list of >1,400 GWAS traits reported, I identified 13 searchable traits of relevance, including ones pertaining to infancy, across 76 studies (Table 4-1).

Table 4-1 | GWAS traits and counts of respective studies searched for genetic associations.

GWAS trait	Counts of	
	studies	**associations**
Adiponectin	11	58
Adiposity (newborn)	0	0
Anthropometric (newborn)	1	5
Birth weight	3	13
BMI	25	183
Bone mineral density paediatric (upper limb, total body less head)	2	26
Head circumference (infant)	1	3
Height	22	525
Insulin resistance/response	2	14
Insulin growth factors	1	8
Vitamin D levels	4	15
Visceral fat	1	40
Weight	3	33

Out of >900 GWAS associations yielded, I identified 5 T1D candidate genes that had been 'reported' by the author or 'mapped' by the NCBI from a total of 6 SNPs (Table 4-2).

Table 4-2 | GWAS traits that associate with SNPs mapped to genes of relevance to T1D.

GWAS trait	**SNP**	**Gene**	**Mapped**	**Reported**	***P* value**
Bone mineral density	rs1262476	*CENPW*	X	X	4×10^{-8}
BMI	rs1561288	*EFR3B*		X	5×10^{-8}
Height	rs13428823	*EFR3B*		X	6×10^{-12}
BMI	rs12229654	*CUX2*	X		5×10^{-9}
Height	rs1570106	*RAD51B*	X		8×10^{-9}
Insulin resistance/response	rs2407103	*SMARCE1*	X		2×10^{-6}

The *P* value of the SNPs associating with their respective traits ranged from 10^{-6} to 10^{-12}. The SNP pinpointed to *CUX2*, a protein-coding gene, is mapped to the same region as *SH2B3,* which I already selected. On the grounds of the effect of ethnicity on genetic associations, *CUX2* and *SMARCE1* were not taken forward because their reported associations were established in East-Asian ancestry and African American populations respectively. The rest of the genes were questionably of little interest due to minimal existing information on their function (*EFR3B*) or not highly compelling *P* values for large scale genetic studies (*RAD51B, CENPW).*

4.2.5 Screening output

A total of 34 genetic variants were selected. Table 4-3 lists the selected genes by category, the SNPs, alleles, minor allele frequency (MAF), functional consequence and the source of the reference sequence (rs) by last name of author or database. The MAF was sourced from the same source as the rs, except for SNPs in the MHC region and vitamin D metabolism milieu which were sourced from Ensembl (www.ensembl.org, accessed 2016). Some SNPs were mapped to more than one gene: rs75793288 in *IL2* and *IL21*, rs12927355 in *CLEC16A* and *DEXI*, rs12453507 in *IKZF3*, *ORMDL3*, *GSDMB*. For these SNPs, the tables herein list only the biology-guided screened candidate gene.

4.2.6 Comments on the gene panel

On the grounds of T1D being a disease of the pancreas as much as of the immune system[26], and loci unique to T1D having the potential to illuminate disease-specific pathways[176], it was deemed important to include genes in this investigation with the following attributes: non-immune, expressed in the pancreas, and not related to other autoimmune diseases. A post-hoc assessment of the panel of genes confirmed that the output comprised largely immune genes — classified based on the definition provided by the Immunogenetic Related Information Source (IRIS), whereby an immune gene produces a functional transcript and demonstrates a defence characteristic[443] — but, within the boundaries of immune diseases, also included non-immune genes: *INS*, *DLK1*, *GLIS3*, *RNLS*. Approximately half of the selected genes are expressed in the pancreas, e.g. *IFIH1*, *TNFAIP3*, *GLIS3*, *INS*, *CYP27B1*, *ERBB3*, *SH2B3*, *DLK1*, *CTSH*, *CLEC16A*, *PTPN2*, *TYK2*, *FUT2, RNLS*, *ORMDL3.* In fulfilment of the final attribute, the panel contained genes specific to T1D and not related to other autoimmune or autoinflammatory diseases: *GLIS3*, *PRKCQ*, *INS*, *ITGB7*, *DLK1, RNLS* (www.immunobase.org, accessed 2016).

The panel includes commonly-occurring polymorphisms. The MAF exceeds 0.4 for SNPs mapped to *IFIH1, CTLA4, GLIS3, SH2B3, DLK1, FUT2, ORMDL3*. Two genetic variants with reported MAF <0.1 are mapped to *TYK2* (MAF=0.04) and *ITGB7* (MAF=0.05), which might be sensitive to false positives (type I errors).

Finally, on the grounds of including SNPs that overlap potential enhancers near *CTSH*, *TYK2*, *UBASH3A*[102] and causal genes which contain high-confidence missense variants (*PTPN22*, *TYK2*) the gene panel was deemed compelling.

Table 4-3 | Selected T1D genetic variants to test for associations with growth phenotypes.

	Region	Locus	SNP	Alleles	MAF	Func.	SNP Source
HLA	6p21	*HLA-II*	rs17426593	DRB1*04	0.16		Walker
	6p21	*HLA-II*	rs2187668	DRB1*03	0.11		Walker
	6p21	*HLA-II*	rs7454108	DQB1*03:02	0.18		Walker
	6p21	*HLA-II*	rs3135388	DQB1*06:02	0.12		De Bakker
Relevance to growth or gut/microbiome	1p13.2	*PTPN22*	rs2476601	G>A	0.09	missense	Onengut
	1q32.1	*IL10*	rs3024505	G>A	0.16		Onengut
	2q24.2	*IFIH1*	rs2111485	G>A	0.39		Onengut
	2q33.2	*CTLA4*	rs3087243	G>A	0.45		Onengut
	4q27	*IL2*	rs75793288	C>G	0.36		Onengut
	6q23.3	*TNFAIP3*	rs6920220	G>A	0.22		ImmunoBase
	9p24.2	*GLIS3*	rs6476839	A>T	0.40		Onengut
	10p15.1	*IL2RA*	rs61839660	C>T	0.10		Onengut
	10p15.1	*PRKCQ*	rs11258747	G>T	0.26	synonymous	ImmunoBase
	11p15.5	*INS*	rs689	T>A	0.3		Onengut
	11q13.1	*BAD*	rs694739	T>C	0.37		ImmunoBase
	12q13.13	*ITGB7*	rs11170466	C>T	0.05		ImmunoBase
	12q13.2	*ERBB3*	rs705705	G>C	0.34		Immunobase
	12q14.1	*CYP27B1*	rs10877012	G>T	0.33		ImmunoBase
	12q24.12	*SH2B3*	rs653178	T>C	0.48		Onengut
	14q32.2	*DLK1*‡	rs56994090	T>C	0.41		Onengut
	15q25.1	*CTSH*	rs34593439	G>A	0.10		Onengut
	16p13.13	*CLEC16A*	rs12927355	C>T	0.32		Onengut
	18p11.21	*PTPN2*	rs1893217	A>G	0.16		Onengut
	19p13.2	*TYK2*	rs34536443	G>C	0.04	missense	Onengut
	19q13.33	*FUT2*	rs516246	T>C	0.49		Onengut
	21q22.3	*UBASH3A*	rs11203202	C>G	0.33		Onengut
T1D predictor	6q15	*BACH2*	rs72928038	G>A	0.17		Onengut
	10q23.31	*RNLS*‡	rs12416116	C>A	0.28		Onengut
	16p11.2	*IL27*	rs151234	G>C	0.12		Onengut
	17q12	*ORMDL3*	rs12453507	G>C	0.49		Onengut
Vitamin D	11p15.2	*CYP2R1*	rs10741657	G>A	0.38		Cooper
	11p15.2	*CYP2R1*	rs12794714	C>T	0.35	synonymous	Copper
	11q13.4	*DHCR7*	rs12785878	T>G	0.30		Cooper
	4q12-q13	*GC*	rs4588	C>A	0.25	missense	Cooper

‡ Gene not named in the fine mapping study[102] due to lack of evidence of an established role in the immune system.

4.3 DETERMINING THE GENOTYPES

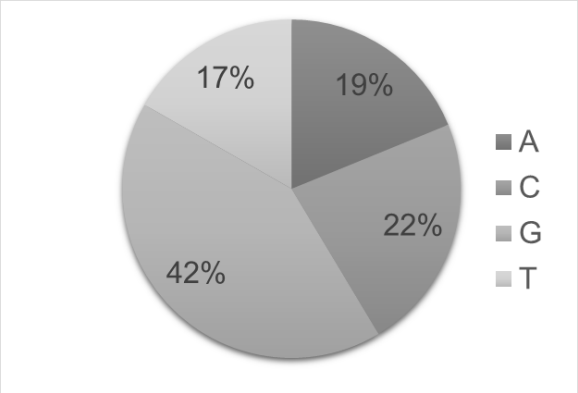
4.3.1 Genotyping

A total of 34 T1D susceptibility SNPs were determined in 612 CBGS infants using the 'Kompetitive Allele Specific PCR' (KASP) genotyping assay by LGC Genomics (Hoddesdon, UK). Wet human DNA samples (n=612 + 2 duplicates) in 16-µl aliquots at a concentration of 50 ng/µl were supplied by Professor Ken Ong's group at Addenbrooke's Hospital, Cambridge. Towards designing each assay, I sourced flanking sequences of 50 bases on each side of the variant of reference (www.ensembl.org, accessed 2016). A genotype was confirmed based on at least two reads and depicted as clusters of data points on Cartesian plots of the FAM (X axis) and HEX (Y axis) fluorescence values, accessible via the LGC SNPviewer software[444].

4.3.2 Yield

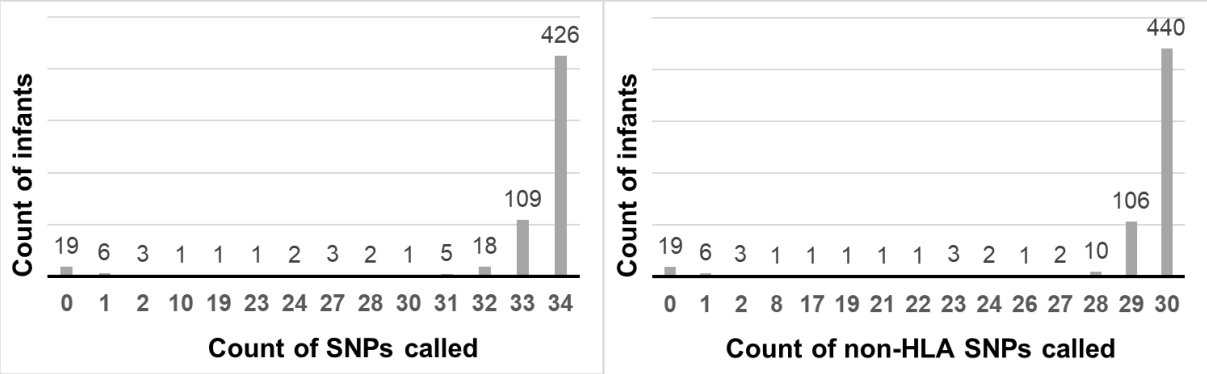
A total of 39,202 alleles were called in biallelic combinations for 34 SNPs in 612 samples. The distribution of calls by nucleotide showed approximately equal proportions for Adenine (A), Cytosine (C) and Thymine (T), whereas Guanine (G) was double the content (Figure 4-3).

Figure 4-3 | Distribution of nucleotides called for 34 SNPs (n=612).



Among the 597 subjects with phenotypes, the yield of successfully genotyping all 34 SNPs per subject was 71% (n=426), and the yield of calling all 30 non-HLA SNPs was 74% (n=440). A total of 19 samples had no SNP called (Figure 4-4). Missing data rate per sample serves as an indicator of low DNA quality[445]. Hence, the 19 samples with nil genotyping success were excluded from data quality controls on the grounds of poor DNA quality.

Figure 4-4 | Distribution of counts of SNPs called by infant with phenotype data (n=597).



Interpretation of the genotyping output generated is given in Table 4-4.

Table 4-4 | Key to genotyping outputs. Adapted from LGC SNPviewer User guide[444].

Title	Description
NTC	No template controls. Wells that do not contain any template DNA and will not generate any fluorescent signal.
Bad	Samples which consistently provide unreliable genotyping results across several assays indicating an inherent problem with the DNA.
?	Unidentified calls. Samples that have not been assigned a genotyping call because they did not generate consistent signal or failed to amplify.
Uncallable	Unable to get 2 confirming reads.
DUPE	Duplication error. Duplicated samples have no matching calls.

4.4 GENOTYPE DESCRIPTIVE STATISTICS

4.4.1 Overview of quality control

Comprehensive information on the SNPs genotyped in the CBGS, and juxtaposed against findings from prior studies, is summarized in Table 4-5. Out of 30 non-HLA SNPs, 17 are protective against T1D (OR: 0.42 to 0.96) and 13 confer susceptibility to the disease (OR: 1.04 to 1.89). Beyond the HLA, the *INS* VNTR exerts the highest magnitude of effect with an OR of 0.42, followed by the *PTPN22* which is predisposing and has an OR of 1.89. The location of the SNPs in the region indicates that variants are non-coding.

Quality control is essential for genetic studies. Erroneous SNP genotype data can inflate the number of false positives (type I error), which should be avoided with higher priority than the increase in the number of false negatives (type II error); false positive results misguide research and expends resources, whereas missed genetic signals would emerge in subsequent larger studies[446]. Thus, poor quality SNPs ought to be removed to obviate inflation of standard errors. The power to determine genetic effects is also contingent on the MAF[447]. A study found that SNPs with MAF of 5% exhibited significantly more false positives than MAF at 25% and 50% at $\alpha=10^{-4}$, yet rare SNPs did not show more positive results than expected by chance, suggesting that removal of low MAF SNPs from analysis due to concerns for inflated type I error are inappropriate[447]. Quality control was conducted across all unique genotyped samples (n=612) and comprised i) removing samples deemed to have poor DNA quality (n=19), ii) including SNPs with call rates >95%, consistent with the recommended GWAS cut-off[445], iii) excluding SNPs with MAF <1% and iv) testing for Hardy-Weinberg equilibrium (HWE).

4.4.2 Call rates

Call rates, a measure of the quality of the assay, were calculated as the percentage of samples successfully genotyped, i.e. excluding calls designated 'Uncallable' 'Unidentified' and 'DUPE'. Call rates in the CBGS ranged from 95% to 98%. Thus, no SNPs were excluded at this stage.

4.4.3 Counts and frequencies

Each infant was assigned a genotype (AA, Aa, aa) per SNP. *Observed genotype counts* were drawn for each SNP (Table 4-6) and used to calculate allele counts and allele frequencies (p, q) by considering that homozygotes carry two identical alleles and heterozygotes one of each. *Genotype frequencies* were calculated as below and presented in Table 4-7:

- *Observed genotype frequency*: genotype counts divided by sample size;
- *Expected genotype frequency*: calculated from allele frequencies (p, q) as

  f (AA) = p^2 f (Aa) = 2pq f (aa) = q^2

In testing for the HWE assumption, *expected genotype counts* were calculated by multiplying the expected genotype frequency by the sample size.

Comparison of observed against expected genotype frequencies showed agreement. The yardstick for crudely assessing population genetic statistics is the MAF, which showed deviation from published data solely for *FUT2* — whose alleles occur in equal ratios — and *HLA-DQB1**06:02, whose alleles in the CBGS were reversed to those published. As expected, the lowest MAF were for *ITGB7* (0.06) and *TYK2* (0.05), neither of which fell below the cut-off.

Table 4-5 | Cohort genotype information on 34 T1D genetic variants typed in the CBGS.

			CBGS							Published			
Chromosome	**SNP**	**Gene(s)**	**Sample N**	**Call rate**	**Alleles**	**MAF**	**HWE χ^2**	**Context Sequence**	**Orientation**	**Alleles**	**MAF**	**Source**	**Functional consequence**
1p13.2	rs2476601	*PTPN22*	579	98%	G>A	0.09	0.60	ATGATTCAGGTGTCC[A/G]TACAGGAAGTGGAG	Forward	G>A	0.09	Onengut	nonsynonymous
1q32.1	rs3024505	*IL10*	579	98%	G>A	0.13	0.09	GGCAGAGCGTGAGGG[A/G]GACTAGTGTTTACT	Forward	G>A	0.16	Onengut	downstream
2q24.2	rs2111485	*IFIH1*	583	98%	G>A	0.40	0.37	ATAAATATAAAGCCT[A/G]GAAGGGTGGAATTT	Forward	G>A	0.39	Onengut	intergenic
2q33.2	rs3087243	*CTLA4*	570	96%	G>A	0.44	0.00	TATTTGGGATATAAC[A/G]TGGGTTAACACAGA	Forward	G>A	0.45	Onengut	downstream
4q12-q13	rs4588	*GC*	581	98%	C>A	0.30	0.07	CTGATGCCACACCCA[A/C]GGAACTGGCAAAGC	Reverse	C>A	0.25	Ensemble	nonsynonymous
4q27	rs75793288	*IL2*	576	97%	C>G	0.38	0.11	TGCTGTCCACTTTAT[G/C]TTAGAGCCCTTGGC	Forward	C>G	0.36	Onengut	intron
6p21	rs17426593	*HLA-DR4*	569	96%	T>C	0.19	3.30	AGACCATGCCTGAT[C/T]GGTGTTTTACACATC	Forward	T>C	0.16	Onengut	intron
6p21	rs2187668	*HLA-DR3*	584	98%	G>A	0.13	0.75	GGCAGCTGAGAGTAA[A/G]TGAGGACCATGTGG	Reverse	C>T	0.11	Ensemble	intron
6p21	rs7454108	*HLA-DQ8*	577	97%	T>C	0.10	0.02	CTTCCCTCCCTAATA[C/T]TCATGCTATTTTGT	Forward	T>C	0.18	Ensembl	upstream
6p21	rs3135388	*HLA-DQB1*06:02*	571	96%	G>A	0.14	0.00	AAAACCTAAAGTGGG[A/G]TTGGTTTGTTGGGA	Forward	T>C	0.12	Ensembl	downstream
6q15	rs72928038	*BACH2*	576	97%	G>A	0.16	0.25	AGGGACGGATTTCCT[A/G]TAAGCTGATCTTGA	Forward	G>A	0.17	Onengut	intron
6q23.3	rs6920220	*TNFAIP3*	578	97%	G>A	0.22	0.02	ACTTCTCCACTAAAA[A/G]GATATGGTTCTGTA	Forward	G>A	0.22	Onengut	intron
9p24.2	rs6476839	*GLIS3*	577	97%	A>T	0.43	0.65	AATTTTAACCAGAGG[A/T]AAGTGCAAACATGT	Forward	A>T	0.40	Onengut	intron
10p15.1	rs61839660	*IL2RA*	574	97%	C>T	0.11	0.15	TTCTGAAGGAGGTAT[C/T]TATTTTGGTCCCAA	Forward	C>T	0.10	Onengut	intron
10p15.1	rs11258747	*PRKCQ*	578	97%	G>T	0.22	0.01	TCTTCTCAGGTTCTC[G/T]CACGAAGAGCTGAA	Forward	G>T	0.13	Barrett	synonymous
10q23.31	rs12416116	*RNLS*	576	97%	C>A	0.28	0.54	TCATTAGAACAACAA[C/A]AAAAAAATTAGCCT	Forward	C>A	0.28	Onengut	intron
11p15.2	rs10741657	*CYP2R1*	572	96%	C>T	0.41	0.29	TTCCTTGACAGCCCT[C/T]GCCTGCTAAAGTAT	Reverse	G>A	0.38	Ensembl	upstream
11p15.2	rs12794714	*CYP2R1*	581	98%	G>A	0.40	0.03	GACATGGGGAAGCTC[A/G]GATGAGGCTGCCAG	Forward	C>T	0.35	Ensembl	synonymous
11p15.5	rs689	*INS*	566	95%	A>T	0.29	1.20	TCAGCCCTGCCTGTC[A/T]CCCAGATCACTGTC	Reverse	T>A	0.30	Onengut	intron
11q13.1	rs694739	*BAD*	572	96%	A>G	0.38	4.02	AGGGAGCCAGGTGTG[A/G]GGGCTGTTGCCTCT	Forward	T>C	0.37	Evangelou	upstream
11q13.4	rs12785878	*DHCR7*	580	98%	T>G	0.24	0.61	ATATCACAAAGCTTC[G/T]ATCCTCTCCTGGCC	Forward	T>G	0.30	Ensembl	intron
12q13.13	rs11170466	*ITGB7*	578	97%	C>T	0.06	0.01	ATAGGGACAGATCAT[C/T]AGTTGCTCACAGTG	Forward	C>T	0.05	Evangelou	intron
12q13.2	rs705704	*ERBB3*	568	96%	G>A	0.34	0.33	TCCTACACCTACCTA[G/A]TAGAAGGACTGTCA	Forward	G>A	0.33	Ensembl	upstream
12q14.1	rs10877012	*CYP27B1*	579	98%	G>T	0.32	0.04	AAAATTAAAATAAAA[G/T]AATCTCCCACAGTT	Forward	C>A	0.33	Ensembl	intron
12q24.12	rs653178	*SH2B3*	576	97%	A>G	0.45	1.94	CTGCAATATTGGACA[A/G]CATGACATAGGACA	Reverse	T>C	0.48	Onengut	intron
14q32.2	rs56994090	*DLK1*	577	97%	T>C	0.43	0.34	CCCAACATGAGGGAC[T/C]GTGTGTGAGAGAGG	Forward	T>C	0.41	Onengut	intron
15q25.1	rs34593439	*CTSH*	583	98%	G>A	0.11	5.26	ATGCCAGATATGAAC[G/A]ATTTATTACCGTTT	Forward	G>A	0.10	Onengut	intron
16p13.13	rs12927355	*CLEC16A*	580	98%	C>T	0.32	0.40	TACAGTTTATACTCA[C/T]AGATTTCTTTCAGT	Forward	C>T	0.32	Onengut	intron
16p11.2	rs151234	*IL27*	578	97%	G>C	0.14	0.73	AAATGGCTGAGGTGC[G/C]GATACGAGACCTGA	Forward	G>C	0.12	Onengut	upstream
17q12	rs12453507	*ORMDL3*	578	97%	G>C	0.48	0.08	ACACACAATGGAAGG[C/G]GAAGGCCTCTGGCT	Forward	G>C	0.49	Onengut	intergenic
18p11.21	rs1893217	*PTPN2*	583	98%	A>G	0.18	0.01	CCATTCCTAGGGACA[A/G]AGGTAGAGGAAGAA	Forward	A>G	0.16	Onengut	intron
19p13.2	rs34536443	*TYK2*	573	97%	G>C	0.05	1.57	CTCTCACCGTGGGGG[C/G]GCTCTGGCTGGAGT	Forward	G>C	0.04	Onengut	missense
19q13.33	rs516246	*FUT2*	576	97%	C>T	0.49	0.25	AGCGCCCCGGGCCTC[C/T]ATCTCCCAGCTAAC	Forward	T>C	0.49	Onengut	intron
21q22.3	rs11203202	*UBASH3A*	573	97%	C>G	0.36	1.80	GTGCAGGGCGGCCAC[C/G]GGGCACCGTCCAGG	Forward	C>G	0.33	Onengut	intron

Table 4-6 | Observed genotype counts.

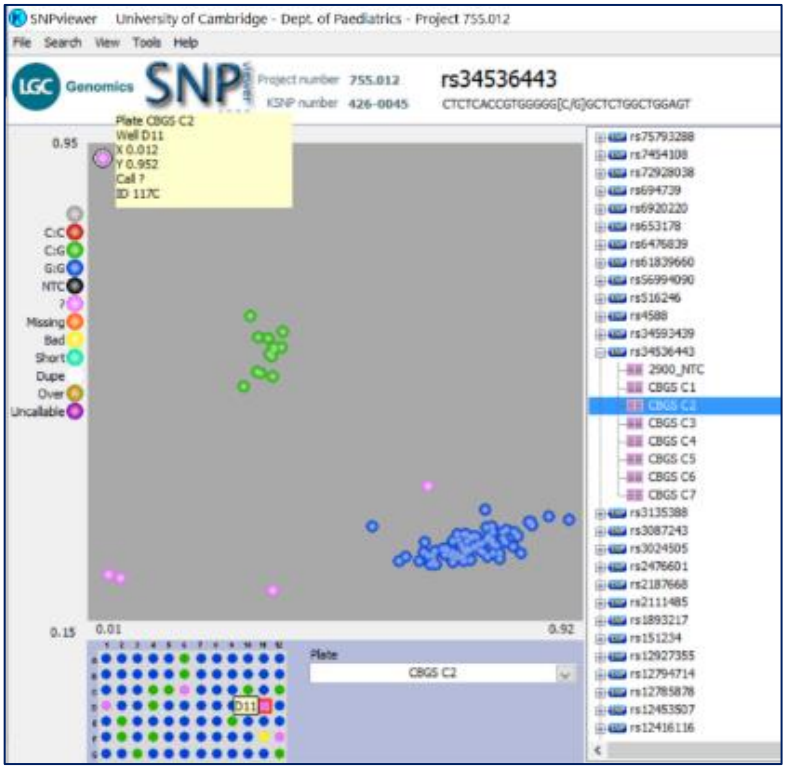
SNP	Gene(s)	A/A	C/C	G/G	T/T	Hetero
rs2476601	*PTPN22*	3	0	480	0	96
rs3024505	*IL10*	10	0	443	0	126
rs2111485	*IFIH1*	90	0	206	0	287
rs3087243	*CTLA4*	108	0	182	0	280
rs4588	*GC*	52	281	0	0	248
rs75793288	*IL2*	0	220	81	0	275
rs17426593	*HLA-DR4*	0	14	0	366	189
rs2187668	*HLA-DR3*	8	0	436	0	140
rs7454108	*HLA-DQ8*	0	6	0	468	103
rs3135388	*HLA-DQB1*06:02*	11	0	424	0	136
rs72928038	*BACH2*	14	0	400	0	162
rs6920220	*TNFAIP3*	28	0	349	0	201
rs6476839	*GLIS3*	180	0	0	104	293
rs61839660	*IL2RA*	0	454	0	6	114
rs11258747	*PRKCQ*	0	0	350	29	199
rs12416116	*RNLS*	50	299	0	0	227
rs10741657	*CYP2R1*	0	196	0	93	283
rs12794714	*CYP2R1*	96	0	207	0	278
rs689	*INS*	278	0	0	43	245
rs694739	*BAD*	231	0	94	0	247
rs12785878	*DHCR7*	0	0	37	338	205
rs11170466	*ITGB7*	0	514	0	2	62
rs705704	*ERBB3*	69	0	250	0	249
rs10877012	*CYP27B1*	0	0	263	60	256
rs653178	*SH2B3*	180	0	127	0	269
rs56994090	*DLK1*	0	110	0	191	276
rs34593439	*CTSH*	12	0	470	0	101
rs12927355	*CLEC16A*	0	273	0	62	245
rs151234	*IL27*	0	9	424	0	145
rs12453507	*ORMDL3*	0	133	153	0	292
rs1893217	*PTPN2*	390	0	19	0	174
rs34536443	*TYK2*	0	0	516	0	57
rs516246	*FUT2*	0	144	0	138	294
rs11203202	*UBASH3A*	0	245	80	0	248

Table 4-7 | Observed and expected genotype frequencies.

		Observed genotype frequencies					Expected genotype frequencies				
SNP	**Gene(s)**	**A/A**	**C/C**	**G/G**	**T/T**	**Hetero**	**A/A**	**C/C**	**G/G**	**T/T**	**Hetero**
rs2476601	*PTPN22*	0.01		0.83		0.17	0.01		0.83		0.16
rs3024505	*IL10*	0.02		0.77		0.22	0.02		0.76		0.22
rs2111485	*IFIH1*	0.15		0.35		0.49	0.16		0.36		0.48
rs3087243	*CTLA4*	0.19		0.32		0.49	0.19		0.32		0.49
rs4588	*GC*	0.09	0.48			0.43	0.09	0.49			0.42
rs75793288	*IL2*		0.38	0.14		0.48		0.39	0.14		0.47
rs17426593	*HLA-DR4*		0.02		0.64	0.33		0.04		0.65	0.31
rs2187668	*HLA-DR3*	0.01		0.75		0.24	0.02		0.75		0.23
rs7454108	*HLA-DQ8*		0.01		0.81	0.18		0.01		0.81	0.18
rs3135388	*HLA-DQB1*06:02*	0.02		0.74		0.24	0.02		0.74		0.24
rs72928038	*BACH2*	0.02		0.69		0.28	0.03		0.70		0.28
rs6920220	*TNFAIP3*	0.05		0.60		0.35	0.05		0.60		0.35
rs6476839	*GLIS3*	0.31			0.18	0.51	0.32			0.19	0.49
rs61839660	*IL2RA*		0.79		0.01	0.20		0.79		0.01	0.20
rs11258747	*PRKCQ*			0.61	0.05	0.34			0.60	0.05	0.35
rs12416116	*RNLS*	0.09	0.52			0.39	0.08	0.51			0.41
rs10741657	*CYP2R1*		0.34		0.16	0.49		0.35		0.17	0.48
rs12794714	*CYP2R1*	0.17		0.36		0.48	0.16		0.35		0.48
rs689	*INS*	0.49			0.08	0.43	0.50			0.09	0.41
rs694739	*BAD*	0.40		0.16		0.43	0.38		0.14		0.47
rs12785878	*DHCR7*			0.06	0.58	0.35			0.06	0.58	0.37
rs11170466	*ITGB7*		0.89		0.00	0.11		0.89		0.00	0.11
rs705704	*ERBB3*	0.12		0.44		0.44	0.12		0.43		0.45
rs10877012	*CYP27B1*			0.45	0.10	0.44			0.46	0.11	0.44
rs653178	*SH2B3*	0.31		0.22		0.47	0.30		0.21		0.50
rs56994090	*DLK1*		0.19		0.33	0.48		0.18		0.33	0.49
rs34593439	*CTSH*	0.02		0.81		0.17	0.01		0.80		0.19
rs12927355	*CLEC16A*		0.47		0.11	0.42		0.46		0.10	0.43
rs151234	*IL27*		0.02	0.73		0.25		0.02	0.74		0.24
rs12453507	*ORMDL3*		0.23	0.26		0.51		0.23	0.27		0.50
rs1893217	*PTPN2*	0.67		0.03		0.30	0.67		0.03		0.30
rs34536443	*TYK2*			0.90		0.10		0.00	0.90		0.09
rs516246	*FUT2*		0.25		0.24	0.51		0.26		0.24	0.50
rs11203202	*UBASH3A*		0.43	0.14		0.43		0.41	0.13		0.46

The rs34536443 mapped to *TYK2* featured only two genotypes raising questions about the viability of minor homozygous individuals. Mining through the genotyping cluster plots surfaced one infant that was homozygous recessive (Figure 4-5), as exactly anticipated from the expected frequencies, but not called out on the grounds of it being a single data point (LGC Genomics, personal communication, 2016). Baby #117 was female, born at 40 weeks of gestation, weighing 2.7 kg at birth, and with a height of 47 cm and head circumference of 36 cm at the first study visit. Measurements at 3 months were weight=4.9 kg, height=57.7 cm and head circumference=39.9 cm. The baby was subsequently withdrawn from the study.

Figure 4-5 | Genotyping cluster plot for the rs34536443 shows presence of one minor homozygous.



4.4.4 Frequencies by sex

Genotype frequencies by sex showed similarities between boys and girls except for the *HLA-DQ8*, for which homozygosity for the minor (high-risk T1D) allele was 5-fold more frequent in boys than in girls (Table 4-8). This discrepancy might be random due to there being only 5 boys and 1 girl with this genotype.

Table 4-8 | Male to female ratios by genotypic group (n=597).

SNP	Gene(s)	A/A	C/C	G/G	T/T	Hetero
rs2476601	*PTPN22*	2.0		1.1		1.3
rs3024505	*IL10*	0.3		1.2		1.4
rs2111485	*IFIH1*	1.4		1.0		1.3
rs3087243	*CTLA4*	1.3		0.9		1.3
rs4588	*GC*	1.0	1.2			1.2
rs75793288	*IL2*		1.2	1.0		1.3
rs17426593	*HLA-DR4*		0.8		1.2	1.2
rs2187668	*HLA-DR3*	1.0		1.1		1.3
rs7454108	*HLA-DQ8*		5.0		1.2	1.0
rs3135388	*HLA-DQB1*06:02*	1.8		1.3		0.9
rs72928038	*BACH2*	0.4		1.1		1.5
rs6920220	*TNFAIP3*	1.2		1.2		1.3
rs6476839	*GLIS3*	1.2			1.1	1.2
rs61839660	*IL2RA*		1.2		1.0	1.2
rs11258747	*PRKCQ*			1.2	1.0	1.1
rs12416116	*RNLS*	0.5	1.3			1.2
rs10741657	*CYP2R1*		1.4		1.2	1.0
rs12794714	*CYP2R1*	1.4		1.1		1.2
rs689	*INS*	1.1			1.1	1.2
rs694739	*BAD*	1.3		1.2		1.1
rs12785878	*DHCR7*			1.1	1.2	1.1
rs11170466	*ITGB7*		1.2		0.0	0.9
rs705704	*ERBB3*	1.4		1.3		1.0
rs10877012	*CYP27B1*			1.4	0.7	1.1
rs653178	*SH2B3*	1.1		1.3		1.2
rs56994090	*DLK1*		1.0		1.4	1.2
rs34593439	*CTSH*	0.7		1.1		1.5
rs12927355	*CLEC16A*		1.2		0.7	1.3
rs151234	*IL27*		0.8	1.1		1.5
rs12453507	*ORMDL3*		1.0	1.2		1.3
rs1893217	*PTPN2*	1.2		0.9		1.1
rs34536443	*TYK2*			1.2		1.0
rs516246	*FUT2*		1.1		1.3	1.2
rs11203202	*UBASH3A*		1.1	1.7		1.2

4.4.5 Hardy-Weinberg equilibrium

Prior to conducting association analyses, each SNP needs to be tested for HWE, which is a state in which the maternally and paternally inherited alleles at a locus are statistically independent[448]. The Hardy-Weinberg law states that the genotype frequencies and allele frequencies remain constant, i.e. no evolution is occurring, in each succeeding generation of sexually reproducing diploids provided the following conditions: i) a large population size so that genetic drift (random chance) does not cause a random change in allele frequencies; ii) random mating (as inbreeding causes an increase in homozygosity); iii) no mutation (i.e. no new alleles enter the population), no migration (i.e. gene flow), no selection (i.e. specific alleles are not selected for or against); iv) equal allele frequencies between the sexes. A significant departure from HWE indicates non-random mating, non-random genotyping error, or missing genotype data[448]. Of note, genotypes in the MHC region may not segregate according to HWE as this region is known to be undergoing selection[449].

Each SNP was tested for HWE (Table 4-5) using the goodness-of-fit chi-square test as shown in Figure 4-6, which was deemed more appropriate — albeit linked to inflated type 1 error rates — than the Exact test, which is conservative and preferred for smaller sample sizes and/or multi-allelic loci[450]. The chi-square test for HWE assumes one degree of freedom (df) as the number of quantities free to vary is number of genotypes minus number of alleles. For 1 df, a chi-square value of ≥3.841 corresponds to a *P* value <0.05, which provides statistical support to reject the null model and conclude that the population is not in HWE.

Figure 4-6 | The HWE test.

Genotype	Observed Counts	Expected Counts	HWE test
AA (major homozygote)	O_{AA}	E_{AA}	$(O_{AA}-E_{AA})^2 / E_{AA}$
Aa (heterozygote)	O_{Aa}	E_{Aa}	$(O_{Aa}-E_{Aa})^2 / E_{Aa}$
aa (minor homozygote)	O_{aa}	E_{aa}	$(O_{aa}-E_{aa})^2 / E_{aa}$
			$\chi^2 = \sum$ (above)

The criterion used to decide whether to omit a SNP from association analyses or not depends on various factors, including the number of SNPs tested and the call rate, but often SNPs with HWE test values of $P<0.01$ or 0.001 are omitted[448]. In the CBGS, the rs694739 in *BAD* and rs34593439 in *CTSH* marginally deviated from HWE at the 0.05 significance level, with *P* values of 0.0450 and 0.0218 respectively. However, both SNPs were included in the analyses as the HWE test *P* values were >0.01, their call rates were respectable and genotype frequencies by sex were similar.

4.5 STATISTICAL FRAMEWORK

4.5.1 Rationale and approach

Multivariate linear regression was used for the hypothesis testing of associations between each phenotype, i.e. anthropometric index or whole blood hormone levels (explained variable), and genotype (explanatory variable) after adjusting for covariates. Regression offers two advantages over the alternative technique of Analysis of Covariance (ANCOVA): i) it provides an estimate of the size of the difference in the outcome, ii) it is more flexible as some of its assumptions are not as restrictive, e.g. ANCOVA requires that the cell size ratio should not be larger than 1:4[425]. This assumption would be problematic for my study as there are genotypic groups which differ immensely in size due to the rarity of the minor allele, e.g. *PTPN22*, *ITGB7*, *TYK2*. Regressions were performed with the 'forced entry' method as it is agnostic to the order of variables entered. Hence, estimates are not dependent on certain combinations of variables, which is a limitation of the alternative stepwise method[451]. Another flaw of the stepwise method is that it fits best on the dataset used than on a new dataset due to sample variance[451]. Herein, predictors of growth were determined in the entire CBGS cohort (n=1,660) and subsequently adjusted for in regressions performed on the genotyped subset (36%).

Genetic variants were tested individually or aggregated in a genetic risk score. Genotype-phenotype association analyses were carried out under an additive genetic model, which is considered pragmatic for testing biological associations[452] and consistent with the general notion, predicated on empirical data and theoretical models, that genetic variance across a range of complex traits is additive[453]. Specific to T1D, an additive model was considered the best fit based on a study from the Finnish Twin Cohort reporting that 88% of phenotypic variance was due to additive genetic effects and 12% due to environmental factors[454]. In an additive genetic model, genotypes are coded depending on the number of copies of T1D risk-increasing alleles. The overarching null hypothesis stated that there is no difference in the outcome by the infant's T1D susceptibility loci after adjusting for covariates. As the hypothesis tested was the presence of a correlation of unspecified direction, a two-way test was used. So long as SNPs are independent, i.e. not associated with each other, an effect found by a SNP on the outcome is reflective of a relationship not influenced by other SNPs.

4.5.2 SNP independence

The issue of association redundancy of SNPs was assessed by the extent of LD, which refers to the non-random association of alleles at two or more different loci in the genome. Physical proximity on the same chromosome is the main reason that variants are inherited together;

variants are in LD if their distributions are correlated[455]. The undesirable effect of LD lies in the possibility that genetic variants that correlate with the variant used in the analysis influences competing risk factors, raising doubts about the inferences made[455]. LD is judged by the measures of r^2 (square of correlation coefficient) and D' (a correlation coefficient that teases out the allele frequency, which could obscure the presence of LD if the allele of one polymorphism is rare). The extent of LD between pairs of the 34 SNPs in the panel was assessed by three online sources: SNAP (https://data.broadinstitute.org, accessed 2016), rAaggr (http://raggr.usc.edu/, accessed 2016) and ImmunoBase (www.immunobase.org, accessed 2016). In line with the assumptions of large genetic association studies, SNPs sourced from GWAS were not in disequilibrium with each other. However, the two SNPs mapped to *CYP2R1* were in high disequilibrium based on the D' (1.0) but not so based on the r^2 (0.5). Judging by their high and similar MAFs of 0.4, it is possible that the r^2 value bears more credibility. However, to mitigate the possibility of LD between these two SNPs, models that used the collective panel of SNPs excluded the rs10741657 and retained the rs12794714 due to the latter's higher genotyping success and stronger association with T1D.

4.5.3 SNP codes

Genotypes were coded as 0, 1, or 2 depending on the number of copies of T1D risk-increasing allele carried by each infant. Hence, a genetic code of '2' is interpreted as carrying 2 copies of a T1D predisposing allele (OR>1) or nil copy of a protective allele (OR<1). Table 4-9 lists the genotype codes assigned to each SNP necessary for the analysis of individual SNPs or creating genetic risk scores. For non-HLA SNPs, the genetic codes were assigned based on the published OR, i.e. <1 or >1 in consideration of the strand that the assay was designed for. For instance, for the SNP rs689, T is listed as the major and T1D risk-increasing allele in the reference fine mapping study[102], whereas in the CBGS A is the major allele as the assay was designed on the reverse strand. For HLA SNPs, the assignment of genetic codes was guided by the allele frequencies in the CBGS, i.e. the minor allele was considered the risk allele except for the *HLA-DQB1*06:02* which is protective. The codes were validated by calculating their frequencies in the CBGS (Table 4-10) and ensuring they adhered to the following pattern:

- Risk allele (OR>1): count of genotype code '0' > count of genotype code '2';
- Protective allele (OR<1): count of genotype code '0' < count of genotype code '2'.

Put otherwise, homozygosity for the risk allele or protective allele is less frequent than the alternative, i.e. most individuals possess average risk of T1D. This validation step surfaced that the rs516246 in *FUT2* did not conform to the expected pattern, most likely owing to the equal distribution of its alleles. For the analysis herein, the risk allele of *FUT2* is considered to be the one claimed by the reference fine mapping study[102].

Table 4-9 | SNP codes per allele and weights, calculated from OR, for generating genetic risk scores.

SNP	Gene(s)	Weight	Effect size	OR	Alleles	Risk codes			
						A/A	C/C	G/G	T/T
rs2187668 rs17426593 rs7454108	*DR3/DR4-DQ8*	3.87	48.18	48.18					
	DR4-DQ8/DR4-DQ8	3.09	21.98	21.98					
	DR3/DR3	3.05	21.12	21.12					
	DR4-DQ8/DRx	1.95	7.03	7.03					
	DR3/DRx	1.51	4.53	4.53					
rs2476601	*PTPN22*	0.64	1.89	1.89	G>A	2		0	
rs3024505	*IL10*	0.15	1.16	0.86	G>A	0		2	
rs2111485	*IFIH1*	0.16	1.18	0.85	G>A	0		2	
rs3087243	*CTLA4*	0.17	1.19	0.84	G>A	0		2	
rs4588	*GC*	0.05	1.05	0.95	C>A	0	2		
rs75793288	*IL2*	0.14	1.15	1.15	C>G		0	2	
rs3135388	*HLA-DQB1*06:02*	2.70	14.88		G>A	0		2	
rs72928038	*BACH2*	0.18	1.20	1.20	G>A	2		0	
rs6920220	*TNFAIP3*	0.11	1.12	1.12	G>A	2		0	
rs6476839	*GLIS3*	0.11	1.12	1.12	A>T	0			2
rs61839660	*IL2RA*	0.48	1.61	0.62	C>T		2		0
rs11258747	*PRKCQ*	0.37	1.45	0.69	G>T			2	0
rs12416116	*RNLS*	0.16	1.18	0.85	C>A	0	2		
rs10741657	*CYP2R1*	0.04	1.04	0.96	C>T		2		0
rs12794714	*CYP2R1*	0.04	1.04	1.04	G>A	2		0	
rs689	*INS*	0.87	2.38	0.42	A>T	2			0
rs694739	*BAD*	0.05	1.05	0.95	A>G	2		0	
rs12785878	*DHCR7*	0.07	1.07	1.07	T>G			2	0
rs11170466	*ITGB7*	0.17	1.19	1.19	C>T		0		2
rs705704	*ERBB3*	0.30	1.35	1.35	G>A	2		0	
rs10877012	*CYP27B1*	0.20	1.22	0.82	G>T			2	0
rs653178	*SH2B3*	0.26	1.30	1.30	A>G	0		2	
rs56994090	*DLK1*	0.13	1.14	0.88	T>C		0		2
rs34593439	*CTSH*	0.25	1.28	0.78	G>A	0		2	
rs12927355	*CLEC16A*	0.20	1.22	0.82	C>T		2		0
rs151234	*IL27*	0.17	1.19	1.19	G>C		2	0	
rs12453507	*ORMDL3*	0.11	1.11	0.90	G>C		0	2	
rs1893217	*PTPN2*	0.19	1.21	1.21	A>G	0		2	
rs34536443	*TYK2*	0.40	1.49	0.67	G>C		0	2	
rs516246	*FUT2*	0.14	1.15	0.87	T>C		0		2
rs11203202	*UBASH3A*	0.15	1.16	1.16	C>G		0	2	

Table 4-10 | Counts (N) of infants by genotype code (0, 1, or 2) assigned.

SNP	Gene(s)	Effect on T1D	N[0]	N[1]	N[2]	N[0] > N[2]
rs2476601	*PTPN22*	Risk	480	96	3	yes
rs3024505	*IL10*	Protective	10	126	443	no
rs2111485	*IFIH1*	Protective	90	287	206	no
rs3087243	*CTLA4*	Protective	108	280	182	no
rs4588	*GC*	Protective	52	248	281	no
rs75793288	*IL2*	Risk	220	275	81	yes
rs17426593	*HLA-DR4*	Risk	366	189	14	yes
rs2187668	*HLA-DR3*	Risk	436	140	8	yes
rs7454108	*HLA-DQ8*	Risk	468	103	6	yes
rs3135388	*HLA-DQB1*06:02*	Protective	11	136	424	no
rs72928038	*BACH2*	Risk	400	162	14	yes
rs6920220	*TNFAIP3*	Risk	349	201	28	yes
rs6476839	*GLIS3*	Risk	180	293	104	yes
rs61839660	*IL2RA*	Protective	6	114	454	no
rs11258747	*PRKCQ*	Protective	29	199	350	no
rs12416116	*RNLS*	Protective	50	227	299	no
rs10741657	*CYP2R1*	Protective	93	283	196	no
rs12794714	*CYP2R1*	Risk	207	278	96	yes
rs689	*INS*	Protective	43	245	278	no
rs694739	*BAD*	Protective	94	247	231	no
rs12785878	*DHCR7*	Risk	338	205	37	yes
rs11170466	*ITGB7*	Risk	514	62	2	yes
rs705704	*ERBB3*	Risk	250	249	69	yes
rs10877012	*CYP27B1*	Protective	60	256	263	no
rs653178	*SH2B3*	Risk	180	269	127	yes
rs56994090	*DLK1*	Protective	110	276	191	no
rs34593439	*CTSH*	Protective	12	101	470	no
rs12927355	*CLEC16A*	Protective	62	245	273	no
rs151234	*IL27*	Risk	424	145	9	yes
rs12453507	*ORMDL3*	Protective	133	292	153	no
rs1893217	*PTPN2*	Risk	390	174	19	yes
rs34536443	*TYK2*	Protective	0	57	516	no
rs516246	*FUT2*	Protective	144	294	138	yes
rs11203202	*UBASH3A*	Risk	245	248	80	yes

4.6 GENETIC RISK SCORES

4.6.1 Overview

Genetic risk scores were generated as indicators of the infant's T1D cumulative genetic risk by considering the HLA regions and non-HLA loci separately, owing to the disparate nature of their effect on T1D risk. Specifically, the *HLA-DRB1* and *HLA-DQB1* do not behave multiplicatively in conferring risk of T1D, hence genotypes, rather than alleles, were used to model the HLA risk effect[449]. For non-HLA loci, GWAS summary statistics were sourced from appropriate study samples (Table 4-3). Regression analyses on individual SNPs was done using the genotyping output as delivered. However, this approach would imply a complete case analysis in models that use multiple SNPs in tandem (i.e. genetic risk scores) as they exclude subjects with missing genotypes. The reduction in sample size would compromise statistical power and bias estimates. Towards maximising use of subjects for this analysis, missing genotypes were tackled by a two-stage approach: 'enhancement' and imputation.

4.6.2 SNP enhancement

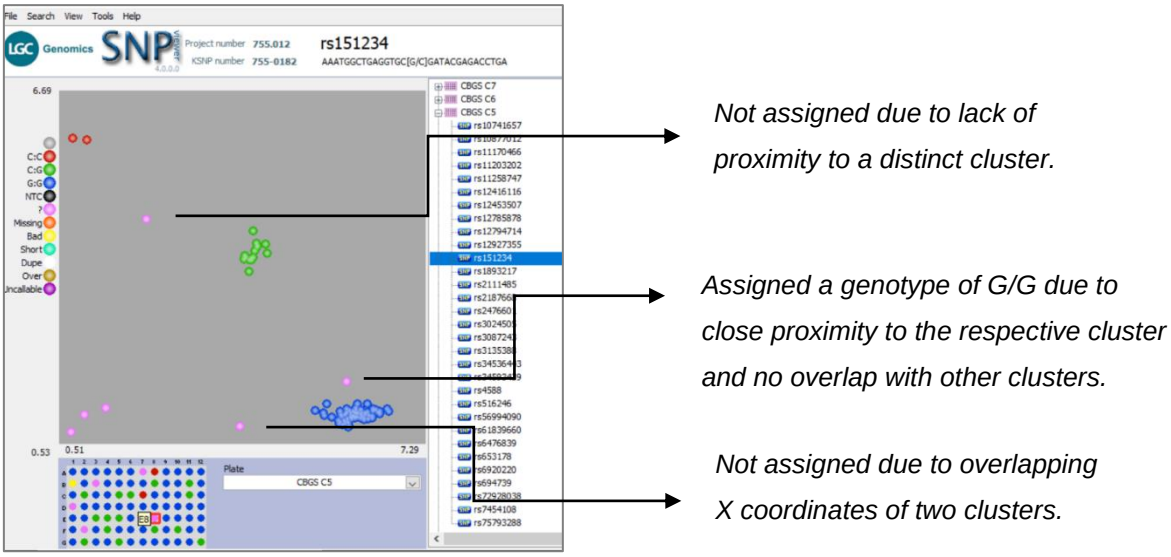
SNP enhancement refers to the self-devised method of assigning unreported genotypes by mining through raw data files and genotype cluster plots. This was carried out via three steps of decreasing degree of confidence and resulted in three additional data outputs (Table 4-11):

1. Supplementation of rs689 genotypes with data obtained with KASP-genotyping in a bigger subset of the CBGS, provided by Dr Clive Petry at the Department of Paediatrics.
2. LGC raw data files were reviewed to identify calls not assigned due to non-confirmatory but non-conflicting reads of the same or duplicated sample(s), e.g the first read was assigned a call but the second was labelled as 'Unidentified', thus annulling the call.
3. Genotyping cluster plots were inspected for unassigned calls that could qualify based on Cartesian coordinates by applying the following criteria: proximity of the data point to a distinct genotype cluster and absence of overlap with another cluster (Figure 4-7).

Table 4-11 | Outputs of SNP enhancement.

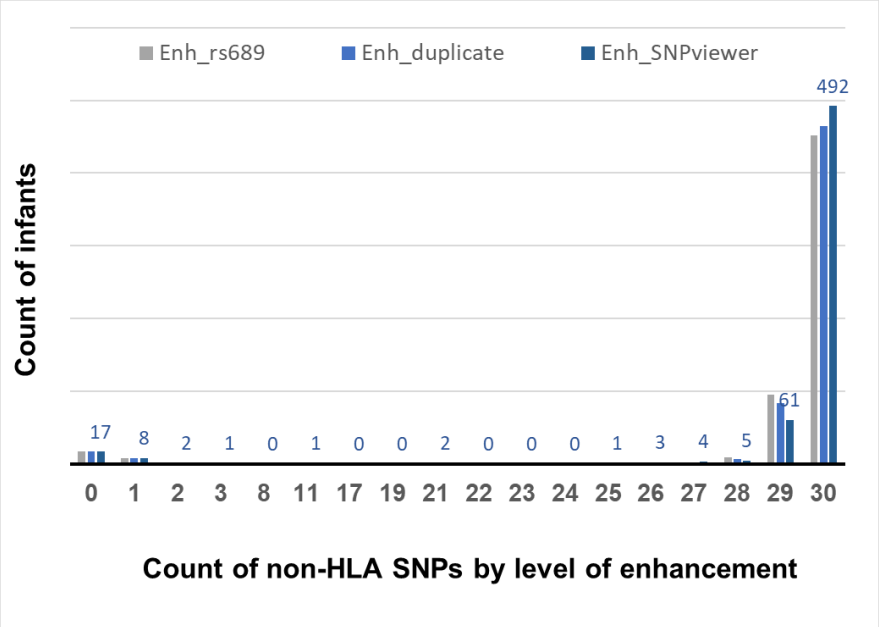
Output	Source of data	Counts of SNPs enhanced	
		HLA	non-HLA
Enh _rs689	LGC output (Petry, C)	-	+17
Enh _duplicate	LGC raw data	+2	+16
Enh _SNPviewer	LGC SNPviewer	+6	+65

Figure 4-7 | Application of SNP enhancement rules for reviewing missing data from LGC SNPviewer.



Despite marginal gains (relatively low count of data points enhanced), enhancement increased the counts of infants with the full set of non-HLA data (Figure 4-8) without affecting mean SNP codes. Thus, analyses were done on the Enh_SNPviewer output albeit subject to judgement.

Figure 4-8 | Effect of SNP enhancement on counts of infants for non-HLA genotypes.



4.6.3 SNP imputation

Missing data ranged between 28 to 43 across HLA SNPs and between 29 to 46 across non-HLA SNPs, which were reduced after SNP enhancement. Imputation was not conducted on SNPs tagging the HLA region owing to its large and non-additive genetic effect on T1D risk. Non-HLA missing data were imputed by the *mean* SNP risk code (Table 4-12).

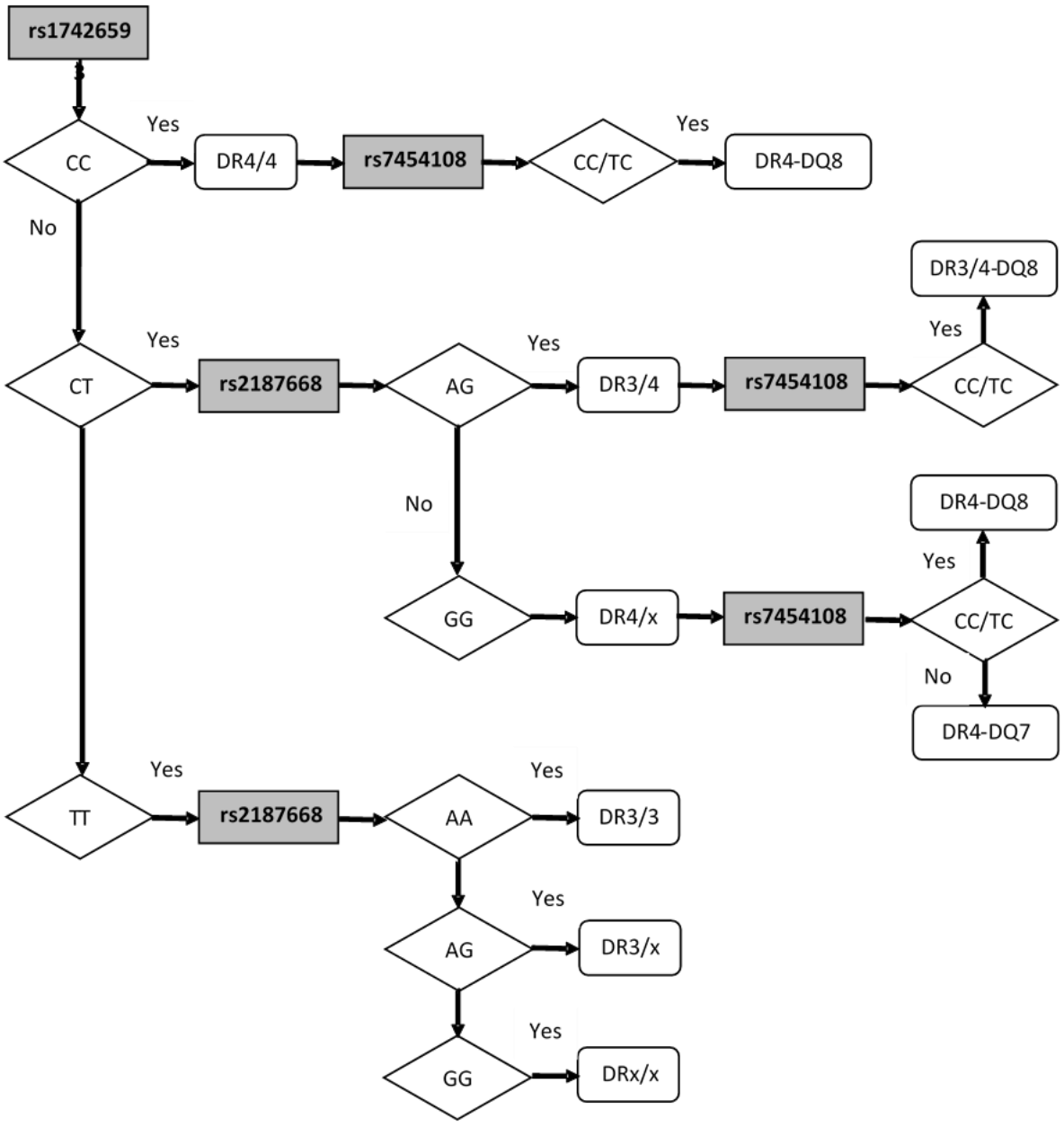
Table 4-12 | Statistical mean and counts (N[ ]) of non-HLA risk codes (0, 1, 2, mean) after imputation.

SNP	Gene(s)	Effect on T1D	OR	Mean	N[0]	N[1]	N[2]	N[Mean]
rs2476601	*PTPN22*	Risk	1.89	0.19	466	97	4	30
rs3024505	*IL10*	Protective	0.86	1.75	10	122	433	32
rs2111485	*IFIH1*	Protective	0.85	1.19	89	281	200	27
rs3087243	*CTLA4*	Protective	0.84	1.13	105	276	176	40
rs4588	*GC*	Protective	0.95	1.39	52	242	272	31
rs75793288	*IL2*	Risk	1.15	0.76	213	271	80	33
rs72928038	*BACH2*	Risk	1.20	0.34	387	163	14	33
rs6920220	*TNFAIP3*	Risk	1.12	0.44	342	194	28	33
rs6476839	*GLIS3*	Risk	1.12	0.87	174	290	100	33
rs61839660	*IL2RA*	Protective	0.62	1.78	6	111	445	35
rs11258747	*PRKCQ*	Protective	0.69	1.56	28	194	343	32
rs12416116	*RNLS*	Protective	0.85	1.44	47	222	295	33
rs10741657	*CYP2R1*	Protective	0.96	1.18	92	273	195	37
rs12794714	*CYP2R1*	Risk	1.04	0.81	202	270	95	30
rs689	*INS*	Protective	0.42	1.41	44	250	275	28
rs694739	*BAD*	Protective	0.95	1.23	95	241	224	37
rs12785878	*DHCR7*	Risk	1.07	0.48	329	200	36	32
rs11170466	*ITGB7*	Risk	1.19	0.11	507	60	2	28
rs705704	*ERBB3*	Risk	1.35	0.68	248	248	69	32
rs10877012	*CYP27B1*	Protective	0.82	1.36	58	252	261	26
rs653178	*SH2B3*	Risk	1.30	0.91	176	263	124	34
rs56994090	*DLK1*	Protective	0.88	1.14	106	273	184	34
rs34593439	*CTSH*	Protective	0.78	1.79	12	98	459	28
rs12927355	*CLEC16A*	Protective	0.82	1.36	62	241	263	31
rs151234	*IL27*	Risk	1.19	0.28	415	142	9	31
rs12453507	*ORMDL3*	Protective	0.90	1.03	131	286	147	33
rs1893217	*PTPN2*	Risk	1.21	0.37	379	170	19	29
rs34536443	*TYK2*	Protective	0.67	1.90	1	55	507	34
rs516246	*FUT2*	Protective	0.87	0.99	140	287	137	33
rs11203202	*UBASH3A*	Risk	1.16	0.72	237	248	79	33

4.6.4 HLA genetic risk score

The tags for the high-risk HLA are rs17426593 for *HLA-DR4* [T>C], rs2187668 for *HLA-DR3* [G>A], and rs7454108 for *HLA-DQ8* [T>C]. The combinatorial rules for deriving the genotypes based on the 3-tag SNP model are depicted in Figure 4-9.

Figure 4-9 | Algorithm for determining HLA genotypic risk groups. Adapted from Nguyen et al.[441].



Two principles guiding the algorithm are:

i. DR3 and DR4 are in biallelic combinations,
ii. DQ8 is in haplotypic association with DR4.

According to Walker's 3-SNP model (unpublished data), the allelic permutations of DR3, DR4 and DQ8 yield ten distinct HLA-DR genotypes which fall under six T1D HLA risk levels, with the highest risk level assigned a score of 6 and the lowest assigned a score of 1 (Table 4-13). A binary HLA risk score was generated by coding high-risk genotypes (DR3/DR4-DQ8, DR4-DQ8/DR4-DQ8, DR3/DR3) as 1, and the rest as 0. A 3-level HLA risk score was also constructed — in reflection of the non-HLA risk codes — whereby codes 0,1 and 2 were respectively assigned to HLA risk levels 1 and 2, 3 and 4, 5 and 6. In parallel, the HLA-DR genotypes were grouped according to the six risk levels defined by Winkler *et al*[437] based on the 2-SNP model, as they underpin the method for creating the combined genetic risk score.

Table 4-13 | HLA-DR classification by genotype, risk level and risk score.

***HLA-DR* genotypes**		**HLA risk level**		**HLA risk score**	
Winkler	Walker	Winkler	Walker	binary	3-level
DR3/DR4-DQ8		***6***	***6***	*1*	*2*
DR4-DQ8/DR4-DQ8		***5***	***5***	*1*	*2*
DR3/DR3		***4***	***5***	*1*	*2*
DR4-DQ8/DRx	*DR4-DQ8/DR4*	***3***	***4***	*0*	*1*
	DR4-DQ8/DRx	***3***	***3***	*0*	*1*
DR3/DRx	*DR3/DR4*	***2***	***3***	*0*	*1*
	DR3/DRx	***2***	***3***	*0*	*1*
DRx/DRx	*DR4/DR4*	***1***	***3***	*0*	*1*
	DR4/DRx	***1***	***2***	*0*	*0*
	DRx/DRx	***1***	***1***	*0*	*0*

The main distinction between Winkler's broad and Walker's detailed classification of genotypes into risk levels is the effect of DR4 on T1D risk; the former classification (2-SNP model) is agnostic to DR4, whereas the latter (3-SNP model) makes a fine distinction.

Each CBGS infant was assigned an *HLA-DR* type based on the count of minor alleles of DR3, DR4 and DQ8. A total of 556 infants (93%) were indisputably assigned an HLA-DR/DQ genotype based on 3 or 2 HLA-tag SNPs, and 13 additional infants were made possible to be classified as 'high-risk' (1) or not (0) even in the absence of complete SNP data (Table 4-14). Excluded were 28 samples for which none of the high-risk HLA tag SNPs was genotyped.

Table 4-14 | Allelic permutations of DR3-DR4-DQ8 based on their count of minor alleles in the CBGS.

DR3/DR4	DR3	DR4	DQ8	*HLA-DR* type	Risk level	Risk score	CBGS count
1/1	1	1	2	DR3/DR4-DQ8	***6***	***1***	-
	1	1	1	DR3/DR4-DQ8	***3***	***1***	22
	1	1	0	DR3/DR4	***3***	***0***	19
2/0	2	0	2	DR3/DR3	***5***	***1***	-
	2	0	1	DR3/DR3	***5***	***1***	-
	2	0	0	DR3/DR3	***5***	***1***	8
1/0	1	0	2	DR3/DRx	***3***	***0***	-
	1	0	1	DR3/DRx	***3***	***0***	1
	1	0	0	DR3/DRx	***3***	***0***	93
0/2	0	2	2	DR4-DQ8/DR4-DQ8	***5***	***1***	6
	0	2	1	DR4-DQ8/DR4	***4***	***0***	7
	0	2	0	DR4/DR4	***3***	***0***	3
0/1	0	1	2	DR4-DQ8/DRx	***3***	***0***	-
	0	1	1	DR4-DQ8/DRx	***3***	***0***	73
	0	1	0	DR4/DRx	***2***	***0***	68
0/0	0	0	2	DRx/DRx	***1***	***0***	-
	0	0	1	DRx/DRx	***1***	***0***	-
	0	0	0	DRx/DRx	***1***	***0***	251
?	1	0	?	DR3/DRx	***3***	***0***	2
	1	?	0	DR3/DR4 or DR3/DRx	***3***	***0***	2
	0	1	?	DR4-DQ8/DRx or DR4/DRx	***3 or 2***	***0***	2
	0	?	0	DR4/DRx or DRx/DRx	***2 or 1***	***0***	9
	0	0	?	DRx/DRx	***1***	***0***	3
	?	?	?	NA	***NA***	***NA***	28
Total							**597**

Table 4-15 summarises the aggregate counts of infants by definitive *HLA-DR* types and risk levels. A total of 36 infants carried the high-risk T1D HLA-DR genotypes. In contrast, almost half of the children carried none of the DR3, DR4 or DQ8 risk alleles.

Table 4-15 | HLA-DR genotypes in the CBGS (n=597).

	HLA risk level		**CBGS**	
HLA-DR genotypes	Winkler	Walker	Counts	Valid %
DR3/DR4-DQ8	***6***	***6***	22	3.7%
DR4-DQ8/DR4-DQ8	***5***	***5***	6	1.0%
DR3/DR3	***4***	***5***	8	1.3%
DR4-DQ8/DR4	***3***	***4***	7	1.2%
DR4-DQ8/DRx	***3***	***3***	73	12.2%
DR3/DR4	***2***	***3***	19	3.2%
DR3/DRx	***2***	***3***	96	16.1%
DR3/DR4 or DR3/DRx	***2***	***3***	2	0.3%
DR4/DR4	***1***	***3***	3	0.5%
DR4/DRx	***1***	***2***	68	11.4%
DR4/DRx or DRx/DRx	***1***	***2 or 1***	9	1.5%
DRx/DRx	***1***	***1***	254	42.5%
	Total		**567**	**94.9%**
		Total	**558**	**93.5%**

In comparison, the combinatorial rules proposed by Nguyen et al.[441], based on a 3-SNP model, require two copies of DQ8 minor alleles for the very high risk DR3/DR4-DQ8 genotype; this haplotypic association does not feature in the CBGS cohort of ~600 HLA-DR genotyped infants and was thus discounted as extremely rare or biologically implausible. In support of my claim, Barker et al.[440], who correlated 2 HLA-tag SNPs with HLA-DR genotypes (the rs2040410 A allele associating with *DRB1*03:02* and the rs7454108 C allele associating with DQB1*03:02) in children of the DAISY cohort and Type 1 Diabetes Genetics Consortium (T1DGC), found two genotypes as markers of the DR3/DR4-DQ8: AG/CT and AA/CT. Neither of these contains two copies of the DQ8 minor allele, thus confirming the finding in the CBGS.

On the grounds of the 2-SNP tag model retaining the SNPs that tag the DR3 and DQ8, I compared the effect of the 3-tag and 2-tag models on genotype distributions in the CBGS.

This comparison surfaced that the 3-tag model annuls the genotype AA/CT of Barker et al.[440] as a marker for the DR3/DR4-DQ8, i.e. there were no infants in the CBGS carrying two DR3 minor alleles and one DQ8 minor allele. Further, the comparison identified two gains of the 3-SNP tag model vs. the 2-SNP model, predicated on the latter being agnostic to the DR4 allele:

i. The 2-SNP model captures more risk than expected, e.g. it allows infants with DR3-DR4-DQ8 associations of minor alleles 1/0/1 or 2/0/1 to be misclassified as carriers of the high-risk HLA-DR3/DR4-DQ8. However, the discrepancies are small; in the CBGS the 2-tag model over-captured only one infant as DR3/DR4-DQ8;
ii. the 3-tag model differentiates between DR4 on DQ8 vs. DR4 not on DQ8, allowing for a finer classification, i.e. DR4/DR4 maps to Risk Level 3 vs. being bundled with DRx/DRx.

Finally, the mutual exclusivity of two DR3 minor alleles and two DR4 minor alleles, i.e. DR3 is biallelic with DR4, is further testimony to the genotyping success of the HLA-DR tag SNPs in the CBGS.

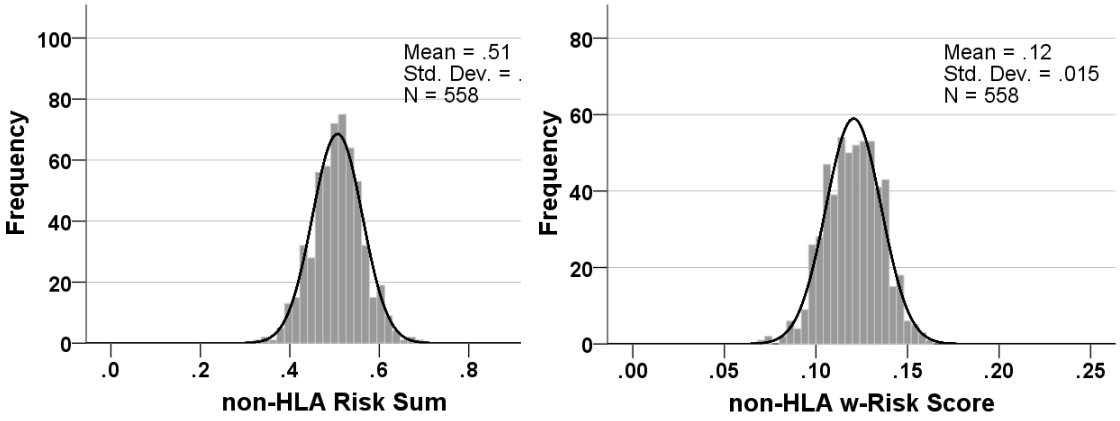
4.6.5 Non-HLA genetic risk score

Two non-HLA genetic risk scores were created per infant, designated and calculated as below:

i. Genetic risk sum score ('Risk Sum'): arithmetic sum across SNPs of the number of copies of risk-increasing alleles (0, 1, or 2) at each SNP, converted to an integer where missing genotypes were imputed by the mean;
ii. Genetic weighted risk score ('w-Risk Score'): sum across SNPs of the number of copies of risk-increasing alleles (0, 1, or 2) at each SNP multiplied by the natural logarithm of the respective effect size (Table 4-9). The premise of this score is that each risk allele has a log-additive effect on T1D risk[456].

The genetic risk scores were scaled to unity by dividing them by the total number of alleles. It was confirmed that using the scaled-to-unity risk score had no bearing on parameter estimates of regression models vs. using the score prior to scaling. Quality control was conducted by excluding subjects where genotyping failed for >2 SNPs. As a result, 39 infants were excluded. As the vitamin D genetic risk score was generated by collating a very small number of SNPs, subjects with one or more missing SNPs were excluded. The distributions of the Risk Sum ($P_{\text{Kolmogorov-Smirnov}}<0.001$) and weighted genetic Risk Scores ($P_{\text{Kolmogorov-Smirnov}}=0.200$) are shown in Figure 4-10.

Figure 4-10 | Distribution of non-HLA genetic risk scores.



4.6.6 Combined genetic risk score

An overall T1D genetic risk score was generated by combining the weighted contributions of variants in the HLA region and non-HLA loci based on the approach described by Oram et al.[456] and replicated in other cohorts[457,458]. The weights used for the combined risk score are shown in Table 4-9. Weights for the DR3/DR4 haplotype combinations and the protective HLA-DRB1*15 were sourced from Oram et al. (Supplementary Table)[456], which had been based on prior publications[171,437]. Whereas the abovementioned studies made use of the rs3129889 to tag the *DRB1*15* region on the grounds of it being predisposing to multiple sclerosis[457], I selected the rs3135388 because it specifically tags the *DQB1*06:02*, as well as *DRB1*15*, and also associates with multiple sclerosis[459]. The rationale lies in that the *DRB1*15:01* without the *DQA1*01:02-DQB1*06:02* did not associate with protection against T1D in family studies, which confirms that protection from the disease associates more strongly to the DQ molecule than the DR2 of this haplotype[460].

As the protective HLA allele fits the log-additive model for T1D risk, it was used along with the 29 non-HLA independent SNPs to create a non-DR weighted genetic risk score, by taking the sum across SNPs of the number of copies of risk-increasing alleles multiplied by the natural log of the effect size, which was added to the weight carried by the HLA-DR (Equation 4-1). Excluded were subjects with missing genotypes for alleles of the greatest weight, namely *DR3/DR4-DQ8* or *HLA-DQB1*06:02*[456,457], and >2 missing genotypes for the non-HLA SNPs.

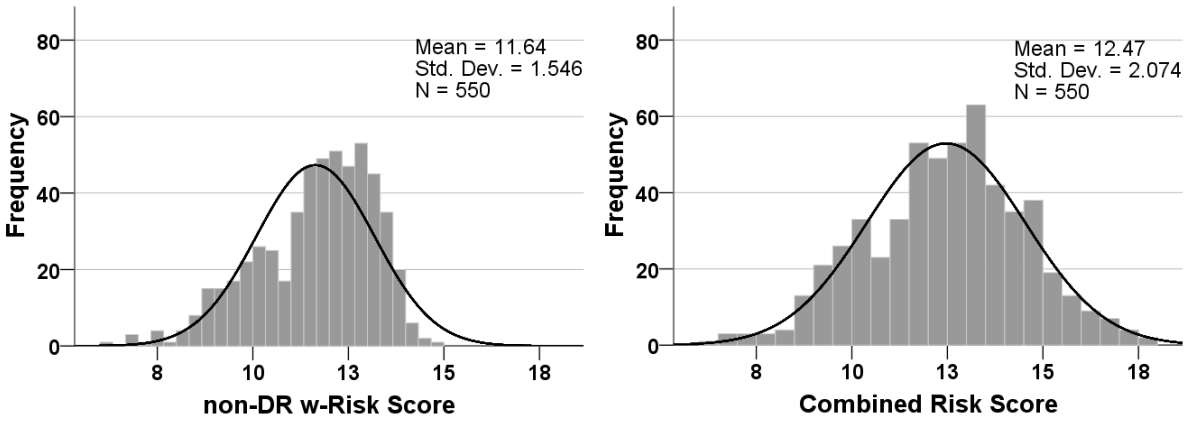
Equation 4-1 | Construction of Combined Genetic Risk Score.

$$\text{Combined Genetic Risk Score} = \text{DR}_\text{j} + \sum_{i=1}^{30} \ln(\beta\text{i}) \times \text{SNPcodeij}$$

for SNP i=1, 2,…30, jth study participant.

The distribution of the non-DR genetic risk score is right-handed, compared with the distribution of the non-HLA genetic risk score (Figure 4-11), because it incorporates alleles of the T1D protective *DQB1*06:02*, which exert a relatively large effect.

Figure 4-11 | Distributions of non-DR and Combined Genetic Risk Scores.



4.6.7 Summary of genetic risk scores

Tables 4-16 and 4-17 respectively summarise and describe the genetic risk scores generated. The non-HLA and vitamin D scores were scaled to unity by dividing by the number of total alleles, which allows for a comparison between their weighted and unweighted aggregates.

Table 4-16 | Summary of genetic risk scores constructed.

	Categorical		Continuous	
Gene cluster	**2-level**	**3-level**	**unweighted**	**weighted**
HLA	√	√		
non-HLA			√	√
vitamin D			√	√
non-DR				√
combined				√

Table 4-17 | Description of genetic risk scores constructed.

Genetic risk score	Definition
HLA Risk Score	Assignment based on DR3/DR4-DQ8 genotypes
HLA Risk_2level	Binary HLA-DR risk score (1='high risk', 0='no risk')
HLA Risk_3level	Three-level HLA-DR risk score (2='high risk',1='medium risk', 0='no risk')
Genetic Risk Sum Score	Arithmetic sum of risk-increasing alleles
nonHLA RiskSum	29 SNPs: all non-HLA excluding rs10741657 in *CYP2R1*
VitD RiskSum	4 SNPs: in vitamin D metabolism genes excluding rs10741657 in *CYP2R1*
Genetic Weighted Risk Score	Weighted sum of risk-increasing alleles multiplied by the respective natural log of the effect size for T1D risk
non-DR w-RiskScore	30 SNPs: *HLA-DQB1*06:02 + 29 non-HLA* excluding rs10741657 in *CYP2R1*
non-HLA w-RiskScore	29 SNPs: all non-HLA excluding rs10741657 in *CYP2R1*
VitD w-RiskScore	4 SNPs: in vitamin D metabolism genes excluding rs10741657 in *CYP2R1*
Combined Risk Score	Sum of HLA-DR haplotype score + non-DR w-RiskScore

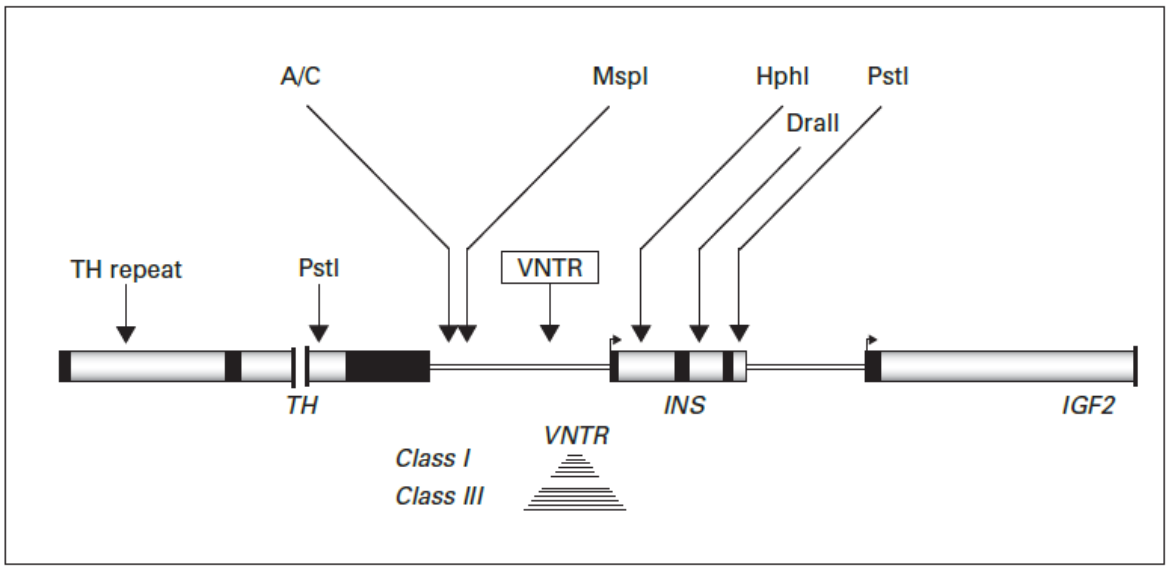
Chapter 5

INS VNTR AND EARLY GROWTH

5.1 CONTEXT

Observational studies in the past few decades have concurred on high birth weight, an established indicator of foetal growth, correlating with increased risk of developing T1D (Chapter 2). However, prospective studies have not provided causative proof for any exogenous agent, such as gestational virus infection, to account for this association. It is thus plausible that the child's genetic background could explain the link between birth weight and T1D risk. Insulin is tasked with the dual role of regulating foetal growth[232] and glucose metabolism. Plasma insulin levels positively correlate with birth weight[293]. In this context, variation at the insulin gene VNTR minisatellite on chromosome 11p15.5, which following its discovery[461] was shown to be in association with T1D[462], is the obvious suspect. This hypothesis is fuelled by the critical location of the *INS* VNTR polymorphism; it is 365 bp upstream of the *INS* promoter in the 5' flanking region of the insulin gene and 5 kb upstream of *IGF-2* (Figure 5-1)[461], both of which encode the major regulators of foetal growth.

Figure 5-1 | *INS* VNTR upstream of the *INS* promoter and *IGF-2* and 10kb downstream of the tyrosine hydroxylase gene (*TH*). From Kim and Polychronakos[175].



The polymorphic nature of *INS* VNTR minisatellite constitutes heterogeneity in length and, to a lesser extent, variation in the nucleotide sequence within individual repeats[463]. VNTRs are dispersed throughout the human genome and located both upstream and downstream of genes, as well as within introns[464]. They are tandem sequences of short DNA motifs, which are moderately repeated and found clustered near telomeres, i.e. end of chromosomes[247]. The most known structure of the *INS* VNTR minisatellite is composed of a variable number of tandem repetition of the 14-15 bp oligonucleotide sequence ACAGGGGTGTGGGG[461]. The number of repeats that are of relevance to T1D corresponds to a biallelic polymorphism: short alleles cluster at 26-63 repeats (class I) and long alleles cluster at 141-209 repeats (class III)[462]. Alleles of intermediate sizes (class II) are very rare[462].

The *INS* VNTR polymorphism is located outside coding regions, and therefore does not alter the sequence of the insulin peptide. Its location upstream of the *INS* promoter suggests that its biological effect is mediated through differences in insulin gene transcription levels[175]. Studies have shown that class III alleles are associated with lower levels of steady-state *INS* mRNA expression in both foetal[465] and adult pancreatic tissue[28]. In non-diabetic adults, *INS* VNTR class III alleles were associated with variations in insulin secretion at the physiological level[466]. On the population level, class III alleles were found to confer dominant protection with a 60% to 70% reduction in T1D risk[28], whereas homozygosity for the short class I allele was predisposing with a moderate effect[27].

On these grounds, several studies explored associations between the *INS* VNTR and size at birth, but the findings have been inconsistent. In the British Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC) cohort of 758 normal full-term newborns, homozygosity for class III was associated with larger head circumference at birth, and the effect was pronounced in a subset of infants who showed no postnatal growth realignment ('non-changers')[16]. The authors found that within the subset of non-changers, the III/III genotype was also associated with higher birth weight and birth length[16]. The finding with head circumference, but not weight or length, at birth was subsequently replicated in a second ALSPAC sub-cohort independent of the first[17]. The association between larger birth weight and the III/III genotype was also observed in a cohort of precocious pubertal girls in Spain[467]. However, the results have not been replicated in other large normal Caucasian cohorts from Britain[468,469], Germany[470], France[471], Finland[472], the Netherlands[410,473] and, more recently, Poland[474]. One of these studies found that the III/III genotype was associated with lower gestational length and lower birth weight, but not birth weight after adjusting for gestational age[473]. Nor an association was found between the *INS* VNTR and birth weight in a Dutch family-control study on early growth and juvenile onset of T1D[346]. A study in Pima Indians showed an association in the opposite direction, i.e. III/III was associated with lower birth weight, which might have been confounded by the high prevalence of obesity and gestational

diabetes in this population[475]. Of note, two studies in non-Caucasian populations, Gambia[476] and Japan[477], generated evidence in support of the influence of the *INS* VNTR on accelerated growth via the involvement of an imprinting mechanism. This seems a plausible scenario considering that *INS* and *IGF2* are both imprinted genes which preferentially express the paternal allele[478,479]. It was originally suggested that the *INS* VNTR was unlikely the susceptibility locus itself, but a marker for linked genes that influence susceptibility to phenotypes[462]. However, *IGF2* has been precluded from being a target for allelic regulation by the *INS* VNTR on the grounds of non-differential *IGF2* mRNA expression between the VNTR class alleles[175,480,481].

Prior to drawing on mechanisms, the replication of associations between genetic variants and polygenic traits is vital to conclusively establish the functional effect of the locus. The aim of this chapter is to verify the effect of the *INS* VNTR on size at birth, gestational age, and postnatal growth in the CBGS population, which is ethnically similar but of larger size than the ALSPAC cohort of the initial positive studies. The rationale for this pursuit is reinforced by the finding that the *INS* VNTR is specific to T1D and not shared by other autoimmune diseases[102]. Thus, the *INS* VNTR stands a good chance of explaining the growth alterations that are specifically observed in prediabetic children.

5.2 METHODS

5.2.1 Study design and subjects

The CBGS is a prospective observational birth cohort that examines infancy growth in children from the general population of Cambridgeshire, UK. The details of the study were described in Chapter 3. The CBGS enrolled 1,660 children, of which >95% were Caucasians. Of these, 915 were genotyped for the *INS* VNTR.

5.2.2 *INS* VNTR genotyping

INS VNTR typing was performed using genomic DNA from blood or buccal samples collected as described in Chapter 3. The SNP rs689, which displays almost complete linkage with *INS* VNTR[27,482], was genotyped by LGC Genomics (Hoddesdon, UK) using the KASP genotyping assay designed on the reverse strand. In Caucasians specifically, both loci display >99.7% LD thus qualifying the rs689 as a precise surrogate for VNTR genotyping[483]. The rs689 A allele corresponds to class I VNTR allele, and the rs689 T allele tags the class III allele.

5.2.3 Outcome measures

Age- and sex-appropriate SD scores of weight, height, head circumference, BMI and mean skinfold thickness were available for the first 2 years of life as detailed in Chapter 3. Gains of weight or height between birth and 12 months or birth and 24 months had been calculated. Whole blood concentrations of IGF-1, C-peptide and leptin were measured in DBS samples as described in Appendix II.

5.2.4 Statistical analysis

After excluding infants born to mothers with T1D, premature newborns (<36 weeks of gestation) and twins, a total of 875 children remained in the study. HWE was assessed using chi-square tests to compare the observed genotype frequencies against the expected genotype frequencies inferred from allele frequencies. Stratification of the cohort by postnatal growth realignment was based on changes in weight SDS between birth and 12 months or birth and 24 months. Assignment of subjects as catch-up, catch-down or non-changers was done using the criteria defined by the ALSPAC studies. A gain in weight SDS >0.67 between birth and 12 months or birth and 24 months indicated clinically significant catch-up gain; a decrease in weight SDS indicated catch-down and in between values were characteristic of non-changers.

Statistical analyses were done by assuming additive genetic effects as explained in Chapter 4. The *INS* VNTR genotypes were coded as 0, 1, or 2 depending on the count of copies of the risk-increasing allele carried by each child. A dominant model was also considered on the grounds of the known dominant protective effect of the class III allele, whereby genotypes were coded as 1 (I/I) or 0 (I/III, III/III). Gestational age, which did not follow a normal distribution, was reflected, i.e. each value was subtracted from a constant, and log-transformed. Differences in gestational age were tested across VNTR genotypic groups using one-way ANOVA (additive effects) and independent samples t-test (dominant effects). Multivariate linear regression models were used to test for associations between each concerned phenotypic trait (explained variable) and the rs689 (explanatory variable) with adjustment for preselected covariates (Table 3-21) by co-entering those variables in the models (Equation 5-1).

Equation 5-1

$$Y_i = \beta_0 + \sum_{j=1}^{m} \beta_j \times X_{ij} \quad + \beta_{SNP} \times SNP_i$$

for study participant i=1, 2,…n; X is one of m covariates.

The null hypothesis tested was H_0: β_{SNP}=0 vs. H_1: $\beta_{SNP}\neq 0$. Statistical analyses were performed using the IBM Statistical Package for Social Sciences (SPSS 23.0, Chicago, IL, USA) for Windows. Tests were two-tailed and a $P\leq 0.05$ was considered statistically significant.

5.3 RESULTS

5.3.1 Growth characteristics

Overall, 875 children (462 boys and 413 girls) were included in the analysis. Table 5-1 displays the anthropometric characteristics of the participants. IGF-1 levels at 3, 12, 18 and 24 months were measured in 335, 255, 203 and 118 children respectively. C-peptide levels at 3 and 12 months were measured in 176 and 84 children respectively. Leptin levels at 3 and 12 months were measured in 173 and 145 children respectively.

Table 5-1 | Growth characteristics of participating infants genotyped for the *INS* VNTR.

	Boys		Girls	
	N	**Mean ± SD**	**N**	**Mean ± SD**
Gestation (wk)	462	40.0 ± 1.3	413	40.0 ± 1.2
Birth				
Weight (kg)	461	3.61 ± 0.50	412	3.45 ± 0.47
Weight SDS	461	0.11 ± 0.97	412	0.10 ± 0.94
Height (cm)	443	52.1 ± 2.3	397	51.1 ± 2.6
Height SDS	443	-0.02 ± 0.87	397	-0.06 ± 1.01
HC (cm)	445	35.7 ± 1.6	397	35.1 ± 1.6
HC SDS	445	-0.14 ± 0.96	397	-0.13 ± 0.96
BMI (kg/m^2)	442	13.3 ± 1.5	396	13.2 ± 1.5
BMI SDS	442	-0.05 ± 1.23	395	0.06 ± 1.24
Ponderal Index	442	25.6 ± 3.2	396	25.9 ± 3.3
Skinfold thickness SDS	444	-0.05 ± 0.88	393	0.11 ± 0.87
3 months				
Weight (kg)	420	6.50 ± 0.78	372	5.87 ± 0.68
Weight SDS	420	0.11 ± 0.97	372	-0.13 ± 1.01
Height (cm)	420	62.2 ± 2.2	367	60.4 ± 2.4
Height SDS	420	0.28 ± 0.86	367	0.08 ± 1.04
Skinfold thickness SDS	421	0.02 ± 0.77	371	0.01 ± 0.77
12 months				
Weight (kg)	395	10.41 ± 1.12	343	9.58 ± 1.09
Weight SDS	395	0.16 ± 1.01	343	-0.03 ± 1.09
Height (cm)	391	77.0 ± 2.5	342	75.1 ± 2.7
Height SDS	391	0.47 ± 0.97	342	0.31 ± 1.09
HC (cm)	393	47.1 ± 1.3	342	45.8 ± 1.2
HC SDS	393	-0.58 ± 1.10	342	-0.66 ± 1.01
BMI (kg/m^2)	391	17.5 ± 1.4	342	17.0 ± 1.3
BMI SDS	391	-0.14 ± 0.99	342	-0.27 ± 1.03
Skinfold thickness SDS	393	-0.03 ± 0.79	342	0.07 ± 0.78
18 months				
Weight (kg)	373	11.74 ± 1.22	330	10.98 ± 1.23
Weight SDS	373	0.16 ± 0.98	330	0.03 ± 1.05
Height (cm)	373	83.3 ± 2.9	331	81.4 ± 3.1
Height SDS	373	0.48 ± 1.1	331	0.22 ± 1.08
HC (cm)	369	48.4 ± 1.3	329	47.1 ± 1.3
HC SDS	369	-0.68 ± 1.05	329	-0.77 ± 1.09
BMI (kg/m^2)	373	16.9 ± 1.2	330	16.6 ± 1.3
BMI SDS	373	-0.22 ± 0.91	330	-0.20 ± 0.99
Skinfold thickness SDS	373	-0.07 ± 0.74	330	0.09 ± 0.84
24 months				
Weight (kg)	362	12.98 ± 1.37	317	12.27 ± 1.42
Weight SDS	361	0.25 ± 0.97	317	0.11 ± 1.04
Height (cm)	362	88.4 ± 3.0	311	87.0 ± 3.4
Height SDS	361	0.60 ± 0.97	311	0.32 ± 1.07
HC (cm)	364	49.3 ± 1.4	316	48.1 ± 1.3
HC SDS	363	-0.66 ± 1.04	316	-0.79 ± 1.05
BMI (kg/m^2)	360	16.4 ± 1.4	311	16.2 ± 1.3
BMI SDS	359	-0.22 ± 0.96	311	-0.19 ± 0.96
Skinfold thickness SDS	364	-0.08 ± 0.78	317	-0.09 ± 0.85

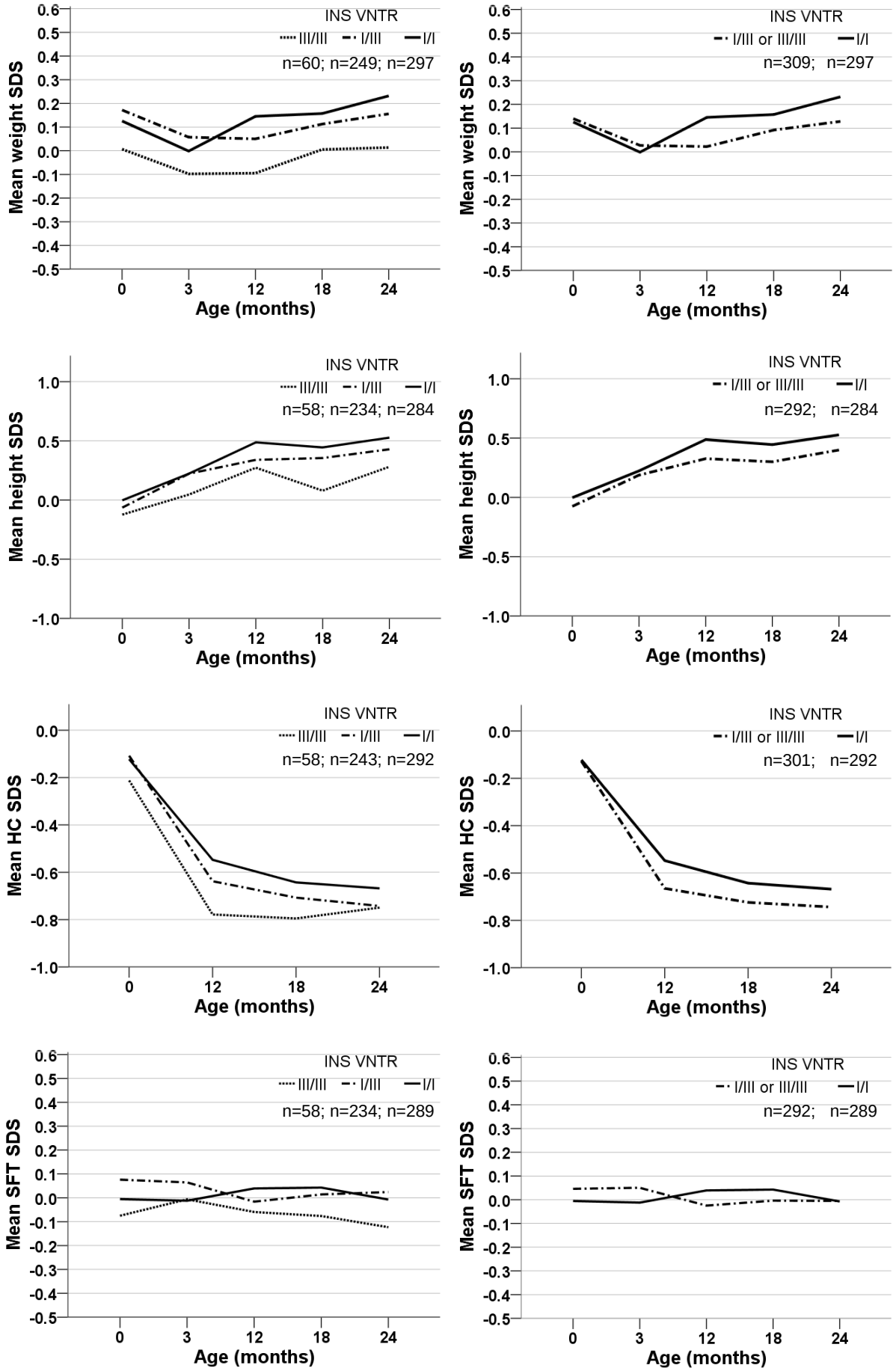
5.3.2 Allele and genotype distributions

The *INS* VNTR distribution of genotypes were I/I 48.0%, I/III 42.5%, III/III 9.5%. The frequencies were comparable with those reported for the ALSPAC cohort as 47.6%, 43.0% and 9.4% respectively[16]. The rs689 SNP was common in the CBGS (MAF=0.3) and in HWE ($P_{chi\text{-}square}$=0.96).

5.3.3 Growth trajectories by *INS* VNTR genotypes

Trajectories of mean SDS for weight, height, head circumference and overall skinfold thickness by VNTR class genotype were plotted for the infancy period by excluding cases with missing values to follow the same subjects (Figure 5-2). The mean weight SDS from birth to 2 years of age appeared consistently lower in homozygotes for class III alleles compared with the other two genotypes, providing ground for proceeding with statistical analyses. Similarly, height SDS and head circumference SDS were consistently but marginally higher for homozygotes I/I. No differences were discerned for overall skinfold thickness SDS across genotypic groups.

Figure 5-2 | Trajectories of SDS for infancy growth parameters by *INS* VNTR class, assuming additive or dominant genetic effects. Genotype frequencies are listed in legends.



5.3.4 Distribution of *INS* VNTR genotypes by size at birth

The occurrence of the *INS* VNTR genotypes I/I, I/III and III/III were similar amongst the SGA, AGA and LGA groups (Table 5-2).

Table 5-2 | Distribution of the *INS* VNTR genotypes by relative birth weight class.

Newborns N (%)	***INS* VNTR genotype**		
Relative birth weight	**I/I**	**I/III**	**III/III**
SGA (n=40)	19 (47.5)	16 (40.0)	5 (12.5)
AGA (n=774)	377 (48.7)	325 (42.0)	72 (9.3)
LGA (n=59)	24 (40.7)	29 (49.2)	6 (10.2)

$P_{chi\text{-}square} = 0.755$

Likewise, the frequency of the genotypes did not differ amongst categories of postnatal growth realignment between birth and 12 months (Table 5-3) or birth and 24 months (Table 5-4).

Table 5-3 | Distribution of the *INS* VNTR genotypes by weight gains between birth and 12 months.

Newborns N (%)	***INS* VNTR genotype**		
Birth-12mo growth	**I/I**	**I/III**	**III/III**
catch-down (n=191)	87 (45.5)	86 (45.0)	18 (9.4)
normal (n=347)	166 (47.8)	146 (42.1)	35 (10.1)
catch-up (n=198)	104 (52.5)	75 (37.9)	19 (9.6)

$P_{chi\text{-}square} = 0.679$

Table 5-4 | Distribution of the *INS* VNTR genotypes by weight gains between birth and 24 months.

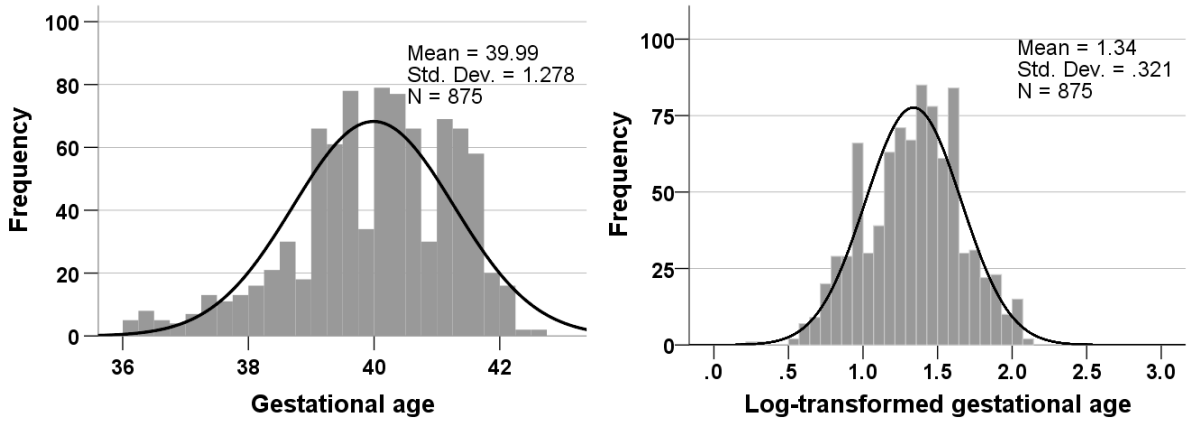
Newborns N (%)	***INS* VNTR genotype**		
Birth-24mo growth	**I/I**	**I/III**	**III/III**
catch-down (n=166)	82 (49.4)	68 (41.0)	16 (9.6)
normal (n=304)	139 (45.7)	134 (44.1)	31 (10.2)
catch-up (n=206)	109 (52.9)	75 (36.4)	22 (10.7)

$P_{chi\text{-}square} = 0.529$

5.3.5 Gestational age by *INS* VNTR genotypes

The natural log-transformed gestational age approximated a normal distribution (Figure 5-3). There was no evidence of a difference in the transformed gestational age between the VNTR class genotypic groups I/I, I/III and III/III (P_{ANOVA}=0.593) or between I/I and III+ (P=0.315). The results remained non-significant after expanding the dataset to include preterm newborns.

Figure 5-3 | Distribution of gestational age before and after logarithmic transformation (n=875).



5.3.6 *INS* VNTR and early growth

Table 5-5 displays the summary statistics (standardised Beta, *P* value) of multivariate regressions of each standardised growth parameter by age on the rs689 by additive or dominant genetic effects after adjusting for covariates as summarised in Table 3-21. With additive genetic effects, the standardised regression coefficient Beta is interpreted as the change in the response for each additional risk allele. With dominant effects, the regression coefficient represents the difference in the response between carrying no copy of the T1D protective allele (I/I) vs. carrying a minimum of one copy of the T1D protective allele (III+). There was no significant effect of the *INS* VNTR genotype on any anthropometric parameter at birth or postnatally, or on growth gains between birth and 12 months or birth and 24 months.

Table 5-5 | Associations between the rs689 and standardised growth parameters by age in infancy, assuming additive or dominant genetic effects†.

	rs689 additive model		**rs689 dominant model**	
	Beta	*P* value	Beta	*P* value
Weight SDS				
Birth	0.011	0.776	-0.007	0.860
3 months	-0.001	0.987	-0.030	0.442
12 months	0.066	0.080	0.067	0.074
18 months	0.041	0.291	0.028	0.472
24 months	0.072	0.077	0.064	0.117
Height SDS				
Birth	0.007	0.847	0.004	0.914
3 months	0.015	0.703	-0.009	0.821
12 months	0.050	0.182	0.049	0.187
18 months	0.067	0.091	0.039	0.329
24 months	0.052	0.197	0.037	0.359
HC SDS				
Birth	-0.026	0.496	-0.020	0.603
12 months	0.018	0.648	0.025	0.531
18 months	0.006	0.876	0.005	0.903
24 months	-0.006	0.886	0.017	0.691
BMI SDS				
Birth	0.032	0.416	0.027	0.501
12 months	0.047	0.251	0.050	0.221
18 months	-0.014	0.742	-0.005	0.903
24 months	0.044	0.308	0.039	0.360
Skinfold thickness SDS				
Birth	-0.020	0.554	-0.036	0.291
3 months	-0.018	0.617	-0.029	0.424
12 months	0.042	0.255	0.051	0.168
18 months	0.057	0.136	0.045	0.245
24 months	0.009	0.820	-0.001	0.985
Deltas				
Δ weight SDS 0-12 months	0.058	0.080	0.062	0.058
Δ weight SDS 0-24 months	0.055	0.109	0.053	0.123
Δ height SDS 0-12 months	0.002	0.957	-0.024	0.508
Δ height SDS 0-24 months	0.051	0.132	0.049	0.141

†Models were adjusted for covariates as outlined in Table 3-21 by growth parameter and infant's age.

5.3.7 *INS* VNTR and endocrine factors

Similarly, the *INS* VNTR genotype did not influence the major regulators of growth in infancy, namely IGF-1 and C-peptide, but showed a weak association with leptin levels at 12 months (Table 5-6). Highlighted are *P* values <0.05

Table 5-6 | Associations between the rs689 and levels of IGF-1, C-peptide and leptin by age in infancy, assuming additive or dominant genetic effects†.

	rs689 additive model		**rs689 dominant model**	
	Beta	*P* value	Beta	*P* value
IGF-1 (ng/ml)				
3 months	-0.042	0.438	-0.024	0.655
12 months	0.045	0.460	0.061	0.310
18 months	0.055	0.423	0.037	0.590
24 months	-0.163	0.093	-0.154	0.115
C-peptide (pmol/L)				
3 months	-0.058	0.445	-0.049	0.518
12 months	0.111	0.324	0.068	0.544
Leptin (ng/ml)				
3 months	-0.113	0.137	-0.019	0.804
12-months	0.182	0.026	0.170	0.037

†Models were adjusted for covariates as outlined in Table 3-21 by endocrine factor.

5.3.8 Stratification by postnatal growth realignment

After allowing for postnatal growth realignment between birth and 12 months (Tables 5-7 and 5-8) or birth and 24 months (Tables 5-9 and 5-10), the T1D predisposing rs689 A allele was associated with higher BMI at 24 months in the group of catch-down growth realignment from birth to 12 months (overall per-A allele increase=0.306 SDS units, *P*=0.0004) and in the group of catch-down growth realignment from birth to 24 months (overall per-A allele increase=0.292 SDS units; *P*=0.001). The associations were corroborated by the dominant model. Associations, albeit not as strong, were also observed for weight at 24 months. There were no significant associations with head circumference or height in any category by postnatal growth realignment, or with weight in the catch-up group.

Table 5-7 | Associations between the rs689 and standardised growth parameters by age in infancy, stratified by birth-12mo weight SDS realignment and assuming additive genetic effects†.

rs689 additive model	**catch-down**		**non-changers**		**catch-up**	
	Beta	*P* value	Beta	*P* value	Beta	*P* value
Weight SDS						
Birth	0.066	0.387	0.077	0.175	-0.052	0.514
3 months	0.058	0.482	0.062	0.289	-0.087	0.275
12 months	0.110	0.155	0.099	0.080	-0.019	0.808
18 months	0.108	0.177	0.081	0.175	-0.058	0.473
24 months	0.214	0.010	0.090	0.149	-0.043	0.598
Height SDS						
Birth	0.024	0.768	0.020	0.738	0.061	0.454
3 months	-0.068	0.397	0.055	0.339	0.071	0.371
12 months	0.017	0.833	0.076	0.176	0.036	0.624
18 months	0.021	0.795	0.086	0.153	0.111	0.171
24 months	0.018	0.827	0.077	0.207	0.079	0.319
HC SDS						
Birth	-0.003	0.967	0.037	0.522	-0.114	0.142
12 months	-0.002	0.980	0.094	0.104	-0.117	0.131
18 months	-0.011	0.896	0.069	0.245	-0.072	0.385
24 months	-0.076	0.380	0.079	0.195	-0.071	0.384
BMI SDS						
Birth	0.081	0.324	0.075	0.221	-0.054	0.516
12 months	0.118	0.153	0.055	0.371	-0.052	0.518
18 months	0.118	0.171	0.011	0.856	-0.182	0.031
24 months	0.306	<0.0001	0.019	0.771	-0.148	0.077
Skinfold thickness SDS						
Birth	0.048	0.516	-0.007	0.893	-0.060	0.398
3 months	-0.007	0.925	0.038	0.501	-0.126	0.085
12 months	0.146	0.047	-0.032	0.565	0.043	0.557
18 months	0.126	0.103	0.027	0.642	0.019	0.803
24 months	0.110	0.161	-0.028	0.632	-0.044	0.566
Deltas						
Δ weight SDS 0-12 months	0.101	0.151	0.060	0.319	0.076	0.315
Δ weight SDS 0-24 months	0.193	0.011	0.031	0.620	0.020	0.794
Δ height SDS 0-12 months	-0.093	0.206	0.063	0.286	0.049	0.516
Δ height SDS 0-24 months	0.009	0.896	0.080	0.157	0.016	0.810

†Models were adjusted for covariates as outlined in Table 3-21 by growth parameter and infant's age.

Table 5-8 | Associations between the rs689 and standardised growth parameters by age in infancy, stratified by birth-12mo weight SDS realignment and assuming dominant genetic effects†.

rs689 dominant model	**catch-down**		**non-changers**		**catch-up**	
	Beta	*P* value	Beta	*P* value	Beta	*P* value
Weight SDS						
Birth	0.072	0.344	0.054	0.337	-0.047	0.561
3 months	0.066	0.426	0.008	0.889	-0.123	0.124
12 months	0.110	0.156	0.074	0.191	-0.022	0.770
18 months	0.117	0.144	0.044	0.467	-0.094	0.244
24 months	0.193	0.020	0.081	0.196	-0.069	0.397
Height SDS						
Birth	0.000	0.997	0.017	0.773	0.064	0.431
3 months	-0.053	0.507	0.020	0.734	0.019	0.816
12 months	0.028	0.719	0.059	0.292	0.026	0.720
18 months	0.013	0.869	0.048	0.425	0.062	0.450
24 months	0.022	0.795	0.030	0.623	0.088	0.269
HC SDS						
Birth	0.038	0.646	0.014	0.805	-0.103	0.185
12 months	0.022	0.793	0.078	0.175	-0.106	0.173
18 months	-0.005	0.953	0.047	0.428	-0.068	0.411
24 months	-0.037	0.672	0.080	0.189	-0.055	0.498
BMI SDS						
Birth	0.117	0.154	0.061	0.320	-0.033	0.689
12 months	0.099	0.229	0.040	0.518	-0.048	0.550
18 months	0.127	0.137	0.001	0.993	-0.176	0.037
24 months	0.252	0.004	0.039	0.550	-0.190	0.023
Skinfold thickness SDS						
Birth	0.022	0.768	-0.037	0.495	-0.075	0.294
3 months	-0.051	0.500	0.034	0.542	-0.152	0.037
12 months	0.156	0.032	-0.028	0.614	0.028	0.698
18 months	0.110	0.154	0.010	0.870	-0.022	0.776
24 months	0.086	0.274	-0.016	0.784	-0.099	0.192
Deltas						
Δ weight SDS 0-12 months	0.100	0.157	0.056	0.349	0.042	0.577
Δ weight SDS 0-24 months	0.172	0.023	0.059	0.347	-0.025	0.746
Δ height SDS 0-12 months	-0.071	0.329	0.004	0.943	-0.028	0.710
Δ height SDS 0-24 months	0.024	0.724	0.059	0.297	-0.010	0.884

†Models were adjusted for covariates as outlined in Table 3-21 by growth parameter and infant's age.

Table 5-9 | Associations between the rs689 and standardised growth parameters by age in infancy, stratified by birth-24mo weight SDS realignment and assuming additive genetic effects†.

rs689 additive model	**catch-down**		**non-changers**		**catch-up**	
	Beta	*P* value	Beta	*P* value	Beta	*P* value
Weight SDS						
Birth	0.113	0.178	0.018	0.761	0.050	0.526
3 months	0.139	0.098	-0.024	0.697	-0.001	0.991
12 months	0.175	0.039	0.072	0.229	0.029	0.697
18 months	0.117	0.160	0.072	0.229	0.011	0.887
24 months	0.207	0.013	0.056	0.353	0.027	0.725
Height SDS						
Birth	0.056	0.543	0.020	0.750	0.081	0.306
3 months	-0.007	0.936	0.031	0.619	0.061	0.434
12 months	0.073	0.383	0.078	0.200	0.054	0.451
18 months	0.066	0.427	0.084	0.174	0.118	0.130
24 months	0.032	0.703	0.062	0.304	0.071	0.346
HC SDS						
Birth	0.018	0.842	0.028	0.649	-0.064	0.410
12 months	-0.019	0.836	0.095	0.132	-0.030	0.691
18 months	-0.048	0.591	0.108	0.083	-0.051	0.515
24 months	-0.124	0.163	0.055	0.376	-0.034	0.652
BMI SDS						
Birth	0.117	0.193	0.020	0.759	0.030	0.718
12 months	0.164	0.066	0.013	0.847	-0.003	0.969
18 months	0.092	0.306	0.004	0.945	-0.116	0.156
24 months	0.292	0.001	-0.012	0.857	-0.049	0.541
Skinfold thickness SDS						
Birth	-0.012	0.877	-0.062	0.289	0.035	0.612
3 months	0.027	0.732	-0.060	0.303	-0.032	0.663
12 months	0.119	0.135	-0.070	0.237	0.099	0.169
18 months	0.035	0.658	0.028	0.636	0.063	0.396
24 months	0.123	0.117	-0.063	0.282	0.029	0.688
Deltas						
Δ weight SDS 0-12 months	0.161	0.043	0.070	0.271	0.002	0.977
Δ weight SDS 0-24 months	0.214	0.005	0.070	0.255	-0.024	0.745
Δ height SDS 0-12 months	-0.072	0.344	0.058	0.334	-0.011	0.881
Δ height SDS 0-24 months	0.036	0.618	0.086	0.154	0.012	0.859

†Models were adjusted for covariates as outlined in Table 3-21 by growth parameter and infant's age.

Table 5-10 | Associations between the rs689 and standardised growth parameters by age in infancy, stratified by birth-24mo weight SDS realignment and assuming dominant genetic effects†.

rs689 dominant model	**catch-down**		**non-changers**		**catch-up**	
	Beta	*P* value	Beta	*P* value	Beta	*P* value
Weight SDS						
Birth	0.099	0.231	0.007	0.913	0.031	0.694
3 months	0.127	0.130	-0.050	0.417	-0.065	0.402
12 months	0.149	0.076	0.074	0.213	-0.014	0.854
18 months	0.113	0.173	0.053	0.381	-0.069	0.384
24 months	0.176	0.032	0.061	0.313	-0.041	0.597
Height SDS						
Birth	0.021	0.817	0.025	0.685	0.076	0.338
3 months	-0.001	0.988	0.006	0.925	0.011	0.887
12 months	0.069	0.404	0.076	0.211	0.029	0.693
18 months	0.039	0.636	0.062	0.322	0.049	0.533
24 months	0.015	0.858	0.032	0.597	0.045	0.547
HC SDS						
Birth	0.058	0.506	0.022	0.713	-0.068	0.381
12 months	-0.002	0.982	0.094	0.139	-0.053	0.489
18 months	-0.065	0.467	0.106	0.092	-0.063	0.422
24 months	-0.073	0.405	0.076	0.221	-0.036	0.636
BMI SDS						
Birth	0.136	0.126	0.018	0.781	0.040	0.629
12 months	0.131	0.137	0.016	0.808	-0.041	0.609
18 months	0.109	0.217	-0.001	0.991	-0.144	0.078
24 months	0.243	0.006	0.008	0.895	-0.116	0.144
Skinfold thickness SDS						
Birth	-0.051	0.522	-0.066	0.256	0.013	0.849
3 months	-0.009	0.914	-0.026	0.657	-0.093	0.200
12 months	0.110	0.166	-0.038	0.521	0.066	0.362
18 months	0.038	0.630	0.016	0.792	0.014	0.855
24 months	0.096	0.224	-0.040	0.498	-0.032	0.659
Deltas						
Δ weight SDS 0-12 months	0.143	0.068	0.088	0.163	-0.043	0.550
Δ weight SDS 0-24 months	0.190	0.012	0.090	0.144	-0.100	0.176
Δ height SDS 0-12 months	-0.047	0.542	0.015	0.800	-0.065	0.376
Δ height SDS 0-24 months	0.051	0.483	0.078	0.196	-0.026	0.699

†Models were adjusted for covariates as outlined in Table 3-21 by growth parameter and infant's age.

5.4 DISCUSSION

5.4.1 Associations with size at birth and during infancy

Reports on the association between size at birth and the VNTR class genotype in the promoter region of the insulin gene locus are equivocal. The class III variant has been reported to have a positive effect, no effect, or a negative effect on anthropometric measures at birth, namely weight, length, head circumference and BMI[16,17,410,467-476]. I addressed the inconclusive findings by taking advantage of the homogeneous Caucasian CBGS population, which is of British origins but larger in size than the ALSPAC study of 758 singletons in which a relationship was found between the *INS* VNTR and head circumference[16]. Nevertheless, no material differences could be evidenced for any measure of size at birth, or at ages 3, 12, 18 and 24 months by VNTR class genotype based on multivariate linear regression assuming either additive or dominant genetic effects. My results on growth beyond birth concur with a Dutch study that examined associations in infancy[410]. In comparison, a Dutch family-controlled study in T1D showed that non-carriers of the protective *INS* VNTR class III showed increased first-year growth in siblings of patients but absence of an effect in patients[346].

On the grounds of familial aggregation of SGA[484] and significant paternal contribution to SGA[485], which point to a genetic contribution to IUGR, studies in other Caucasian populations tested for an association between the *INS* VNTR genotypes and SGA vs. AGA[471,474]. Analogous to these findings, I found that the frequency of the *INS* VNTR genotypes was similar between newborns by relative size at birth (SGA, AGA, LGA). In contrast to a study that found associations between III/III genotypes and lower gestational age[473], I found no evidence of differences in gestation by *INS* VNTR genotype even after including preterm singletons.

The ALSPAC study reported that the association with birth weight was limited to the 348 children that did not show catch-up growth in the first 2 years of life, which was explained by infants following their genetic growth trajectory post-delivery in the absence of maternal-uterine growth constraints, such as nutrition[16]. However, this finding was not replicated in other European cohorts[17,468,472]. Similarly, in my study there was no association with any measure of size at birth and the *INS* VNTR after stratifying by growth realignment between birth and 12 or 24 months of age.

However, the stratified analysis provided evidence of a consistent positive association between BMI SDS at 24 months and the *INS* VNTR in the catch-down group. Specifically, under the additive genetic model, infants whose weight SDS grew by less than -0.67 between birth and 12 months showed an average BMI increase of 0.31 SDS per class I allele (P=0.0004). The findings suggest that the pronounced slow-down in weight gain after birth,

presumably arising from accelerated foetal growth, is countered by a pronounced increase in adiposity in the T1D-predisposing I/I carriers compared with children carrying the III+ genotypes. I speculate that the effect is specific to the catch-down group due to absence of strong environmental factors, such as nutrition, that drive sizeable weight gains which would otherwise mask minuscule genetic effects on adiposity. In the ALSPAC study, class I+ genotypes were associated with higher BMI at age 7 years compared with class III homozygotes but in the group of children that showed postnatal catch-up growth instead[17]. These findings are further perplexed by the subsequent claim that III/III children of the ALSPAC cohort had higher BMI arising from FM at age 9 years[486] and young British adults homozygous for class III had higher BMI than class I homozygotes at a nominal significance level[468]. In contrast, a study investigating the body composition in early life in 738 children[410] and another study in 7,999 middle-aged Europeans found that *INS* VNTR made no contribution to BMI[487].

5.4.2 Associations with hormone levels during infancy

In concordance with my findings on foetal and postnatal growth in the CBGS population, no compelling association was evident between the *INS* VNTR polymorphisms and the major endocrine regulators of infancy growth. The relationship between the *INS* VNTR variants and insulin secretion or obesity, which are established T2D-related metabolic traits, has been studied with equal vigour as infancy growth, predicated on the propositions that *INS* VNTR class III allele associates with increased risk of T2D[469] and paternal transmission of class I allele predisposes to early onset obesity[488]. Notwithstanding an abundance of studies, the relevance of the *INS* VNTR polymorphism with insulin resistance and secretion remains inconclusive. The III/III genotype was claimed to confer higher insulin resistance in British men[469], lower insulin resistance in British young adults[468], higher insulin resistance in young SGA adults[471], lower insulin resistance in prepubertal SGA children[474] and lower insulin resistance in precocious pubarche girls[467]. With respect to insulin supply and β-cell function, the class I allele was associated with higher insulin secretion at the physiological level in healthy adults[466]. In contrast, associations between class I and lower insulin secretion were predominantly observed in disease-enriched populations, namely obesity[489] and polycystic ovary syndrome (PCOS)[490], as well as paediatric populations, such as infants at age 1 year[491] and prepubertal SGA children[474]. Still, other studies found no association with insulin secretion in healthy individuals[468] or obese children[492]. A large-scale study in Danish Caucasians found that the III/III genotype was associated with reduced glucose-induced insulin and C-peptide release in young healthy adults but not in middle-aged subjects[493]. A study in healthy adult Caucasians reported an association between allelic variation at the *INS* VNTR locus and the pulsatility of insulin secretion, but not the rate or amount of insulin secretion[494].

Collectively, studies lack consensus on the role, if any, of the *INS* VNTR in β-cell dysfunction and altered insulin resistance, and in effect, T2D. Credence to the lack of an effect is given by a case-control study which found no differences in the VNTR class genotype distribution amongst T2D patients and controls[493]. The most compelling evidence stems from a recent meta-analysis of GWAS data from subjects of European ancestry which established no association between the SNP rs689 and T2D (OR=1.01 [95% CI 0.98, 1.04])[495].

The lack of an association with IGF-1 reported here does not provide support of the contribution of the *INS* VNTR to infancy growth. The positive association of nominal statistical significance detected with leptin levels at 1 year of age echoes prior associations found between the *INS* VNTR genotypes and fat mass, as well as BMI, in children[486,496]. Similar results for endocrine factors have not been previously reported for infants, but a study in Poland found no evidence of an association between either leptin or IGF-1 and the *INS* VNTR class in SGA-born prepubertal children[474]. Could there be a role of the *INS* VNTR in foetal growth instead, considering its proximity to *INS* and *IGF2*[463,465,497,498], both of which are major foetal growth regulators? Despite non-differential *IGF2* mRNA expression between the VNTR class alleles[175,480,481], the *INS* VNTR genotypes previously associated independently with cord blood IGF-2 levels[17] and foetal growth in mid-pregnancy but not late pregnancy[473], leading me to suggest that *INS* VNTR might exert its effect early *in utero*. Stratification analysis by postnatal growth realignment (data not shown) elucidated that lower IGF-1 levels at 24 months of age significantly associated with the T1D high-risk *INS* VNTR alleles in infants who did not show weight realignment between birth and 24 months of age (Beta=-0.413, *P*=0.006, additive genetic effects). Since 'non-changers' are indicative of genetic growth potential, because they are not subject to significant maternal-uterine enhancement or restraint of foetal growth[17], the finding opens the possibility for a genetic interaction between the *INS* VNTR and IGF-1, albeit not reflected in differences in postnatal growth between genotypic groups of 'non-changers'.

5.4.3 The possible role of the *INS* VNTR

The VNTR class III allele was previously associated with lower levels of *INS* mRNA in both foetal and adult pancreas[28,465], suggesting that the class III allele encodes insulin deficiency[490]. In my study, the association of borderline significance observed between the rs689 and weight at 12 months, supported by the additive (*P*=0.080) and dominant models (*P*=0.074), concur with these findings and feebly suggest that if class III were to associate with infancy growth, it would correlate with lower size. This argument resonates with the main outcome of my systematic literature review (Chapter 2), which cemented that high birth weight is a predisposing factor for T1D whereas low birth weight confers a protective effect.

It has been argued that the preponderant dominant effect of the VNTR class III cannot be explained by the marginal influence of class III on the reduced *INS* expression levels in pancreatic tissue[175]. In fact, thymus appears to be the site where the *INS* VNTR manifests a prominent biological effect, where class III alleles were independently associated with higher steady-state levels of *INS* mRNA expression[499]. This finding frames a plausible explanation for the class III dominant protective effect via central tolerance mechanisms, such as hampering negative selection of autoreactive T-cells, which enhances deletion of insulin-specific autoreactive T-cells and averts development of autoimmunity against insulin[499]. Conceivably, T1D susceptibility and resistance may derive from differential expression of insulin in the thymus but not the pancreas[499] and the absence of insulin in the thymus induces insulin autoimmunity and aggravates the risk of T1D[500]. The role of central tolerance in organ-specific autoimmunity has been reassessed following discovery of the remarkable phenomenon of the promiscuous expression of diverse tissue-specific self-antigens in the thymus[501]. I posit that the role of the *INS* VNTR in immunity, which distances it from the view of its effect on the pancreas, is supported by additional lines of argument. A meta-analysis found a lack of association between the *INS* VNTR and PCOS[502], which hints that the link between this genetic polymorphism and T1D might not be mediated by an endocrine gland. Compellingly, recent breakthrough findings identified that the autoimmune regulator (AIRE) gene on chromosome 21q22.3 modulates, along with the *INS* VNTR, the differential expression of insulin in the thymus[503]. Specifically, mutations in the *AIRE* gene affect human insulin gene expression in epithelial cells in the thymus through *INS* VNTR, which subsequently result in either insulin tolerance or autoimmunity[504]. Hence, it is likely that the *INS* VNTR has a central role in immunity at an early age.

5.4.4 Explaining the discrepant findings

Attempts have been made to explain the discrepant findings on the *INS* VNTR and early growth. Ethnic differences are unlikely to be accountable since LD relationships and nearby modifying variants are quite similar in Caucasians[483]. A possible explanation could be the biological differences between study samples (environmental exposures, antenatal management) and different measures of growth used[472]. Additionally, it has been proposed that the variation in the molecular structure of the VNTR allele, and not its length, accounts for the divergent insulin levels relating with this polymorphism[505]. A methodological consideration is the exploitation of the high concordance of the -23HphI A and T alleles with the adjacent VNTR class I and III alleles, which justifies the use the -23HphI as a tag of the VNTR polymorphism[464]. Nevertheless, the -23Hphl might have a functional effect on its own[506], which could obscure associations. In addition, the complementary nature of the alternative alleles -23HphI A and T, misleadingly denoted in genome browsers as the ‘ancestral’ and ‘minor’ allele

respectively (when A is the minor allele on the forward strand), may obfuscate interpretation of genotyping results. An advantage to my study, in addition to its large sample size and ethnic homogeneity, was the multiple testing of phenotypes, which surfaced associations of nominal significance that might have arisen by chance; this warns about stringent requirements for *P* values at significance levels well below 0.05 in genetic associations of small effects to avoid haphazard results. Nevertheless, the power of this study to detect small genetic effects, particularly those of imprinted genes which express one parental allele, is questioned. The recent GWAS for birth weight detected effect sizes in the range of 0.020 to 0.053 SDS per allele (equivalent to 10 to 26 g) at the $8x10^{-8}$ level of significance in a sample of circa 150 thousand individuals, a substantial subset of which was proved underpowered to conduct parent-of-origin analyses[336].

5.4.5 Conclusion

My results do not support the positive ALSPAC studies that showed associations between *INS* VNTR polymorphisms and size at birth. In fact, if an association were to exist, it would be in the opposite direction to that reported by the ALSPAC studies, judging by the borderline associations found here between the III/III genotype and lower weight, and reconciled with the unequivocal relationship between increased risk of T1D and higher birth weight and vice versa. This discussion underscores the need to interrogate the biological plausibility of statistical associations with molecular- and observational-based arguments. The authors of the initial ALSPAC study proposed that the III/III genotype may confer the survival advantage of increased growth *in utero* in times of low food availability[16]. If the VNTR polymorphism is a relatively recent evolutionary acquisition found only in primates, as suggested after the discovery of this polymorphism[463], I argue that its role is unlikely to be implicated in the ability to survive or thrive on food supply, common to all species since creation. Instead, my view is that we might as well search for features of the immune system of primates, which evolves in response to the changing landscape of threats in the inter-species race of survival, as a possible influence of the *INS* VNTR. Nevertheless, the significant association with BMI leaves open the possibility that the *INS* VNTR could influence obesity at the end of and beyond infancy with a small magnitude of effect. This emerging hypothesis is supported by studies which found associations between T1D risk in childhood and accelerated BMI gains beyond the first 2 years of life[364,396,417]. Replication in larger cohorts is required but it remains to be seen whether a SNP is a representative tag of a VNTR.

Chapter 6

T1D LOCI AND GROWTH: A CROSS-SECTIONAL VIEW

6.1 CONTEXT

The remarkable advances in human genetics, made possible by GWAS that exploit high-throughput genotyping platforms, mapped out the genetic architecture of T1D by identifying approximately 50 susceptibility loci[31,117] which explain up to 80% of heritability[507]. Nevertheless, understanding the precise aetiological disease mechanisms has been hindered by two factors: i) the lack of identification of causal genes, owing to T1D susceptibility loci harbouring numerous genes[508], and ii) the challenge of assigning molecular function to most GWAS hits as they comprise indistinguishable causal and neutral variants in LD and do not alter protein coding sequences, which complicates their interpretation[349]. At its simplest, the genetic factors that contribute to risk of developing T1D influence β-cell autoimmunity (aetiology) and/or affect the progression to clinical onset (pathogenesis)[31]. These candidate genes had been known to exert their effects at the immune system level, specifically on T cells[117]. An interesting twist has been the recent finding that >80% of T1D candidate genes are expressed in pancreatic islets[101,509], indicating that they also act at the level of the β-cell[26]. This notion has been given momentum by a subsequent study that compared SNP locations with chromatin maps for different cell types and identified that the primary signature of T1D SNPs found in T-cell enhancers, is also significantly enriched in pancreatic islet enhancers[349].

Conceivably, T1D is a disease of the pancreas as much as of the immune system[26], but the nature and exact timing of the injurious processes inflicted on the β-cell are not clear. The DIPP study has recently reported that insulin secretion is perturbed as early as 6 years prior to the onset of T1D, whilst insulin sensitivity remains unaffected, 'implying an intrinsic defect in β-cell mass and/or function'[148]. This stretched timeframe from the impairment of insulin secretion to clinical onset suggests that the β-cell pathology emerges very early in life. Genotype-phenotype correlation studies are amongst integrated methods that are employed to probe into the underlying mechanisms of β-cell destruction[508]. This approach enables our understanding of how a susceptible infant's pertinent genetic constitution determines phenotypes early in life, when small genetic effects are most discernible due to the absence

of strong lifestyle influences. However, population studies in search for a genetic justification of the association between somatic growth and risk of T1D have limited their scope to the *HLA* and *INS*. The findings on the former have been contradictory (Chapter 2) and the *INS* VNTR locus failed to provide evidence of a robust association with size at birth or during infancy in the CBGS cohort (Chapter 5). It follows that there is scope beyond the two most predisposing candidate genes to contribute, individually or collectively, to traits that pervade the prodrome of T1D.

The aim of this chapter is to test the hypothesis that the biology-screened T1D variants and candidate genes — exclusively selected upon their functional properties, exemplified by actions on the β-cell — influence somatic growth in infancy, a proxy for β-cell function, at each age measured (cross-sectionally) with the purpose of gaining insights into the mechanisms behind disease susceptibility. Growth in disease operates via a different mechanism than in health[346], hence it is essential to disentangle the confounding effect of the disease (e.g. altering gene expression) on physiological processes. I have accounted for this by taking advantage of the CBGS cohort of infants from the general population to test for unbiased genotype-phenotype associations.

6.2 METHODS

6.2.1 Study design and subjects

The CBGS is a prospective observational birth cohort that examines infancy growth in children from the general population of Cambridgeshire, UK. The details of the study were described in Chapter 3. During 2001 and 2009, the CBGS recruited >2,000 mother-baby dyads antenatally and enrolled 1,660 children of which >95% were Caucasians. Of these, 612 infants were included in the current study as they had genomic DNA available. The number of children with both phenotype and genotype data was 597.

6.2.2 SNP selection

T1D susceptibility loci and candidate genes were screened for biological evidence of relevance to somatic growth as described in Chapter 4. At the end of the screening process, I selected a panel of 34 SNPs (Table 4-3) that have been robustly associated with T1D in large-scale genetic studies and conjectured on having a link with i) growth, ii) gut/microbiome, or iii) vitamin D, based on biological facts which I collected from credible sources.

6.2.3 SNP genotyping

Genotyping was performed on genomic DNA extracted from blood or buccal samples collected as described in Chapter 3. All 34 SNPs were genotyped by LGC Genomics (Hoddesdon, UK) using the KASP genotyping assay. Quality control filters for SNPs and samples were based on a prior publication[445] as described in Chapter 4. In brief, the following exclusion criteria were applied to SNPs: missing data rate >5%, study-wide MAF <5% and HWE *P* value <0.01. Out of 597 samples, 426 had genotype data across all 34 SNPs; for the remaining samples missing data were 'enhanced' and imputed for use in generating genetic risk scores.

6.2.4 Genetic risk scores

I constructed genetic risk scores for T1D using the SNPs genotyped and clustered by i) HLA-DR3/DR4-DQ8 haplotypes, ii) non-HLA loci, iii) vitamin D loci, and iv) combined, as detailed in Chapter 4. Due to LD between the two SNPs in *CYP2R1*, the rs10741657 was excluded on the grounds of a lower study-wide genotyping rate than the alternative rs12794714. In brief, non-HLA loci were collated in unweighted and weighted risk scores under the assumptions that risk alleles have an additive or log-additive effect on risk of T1D respectively. For the DR3/DR4-DQ8 haplotypes, which do not have a log-additive effect on risk of T1D, the weights were obtained from a published study[456]. Table 4-9 summarises the SNPs and respective weights used. Quality control was performed by excluding samples with ≥1 missing genotype(s) for the HLA SNPs or >2 missing genotypes for the non-HLA SNPs. For the vitamin D genetic risk scores, only subjects with complete genotyping data were included.

6.2.5 Outcome measures

Age- and sex-appropriate SD scores of weight, height, head circumference, BMI and mean skinfold thickness were available for the first 2 years of life as detailed in Chapter 3. Gains of weight or height from birth to 12 or 24 months had been calculated.

6.2.6 Statistical analysis

Analyses were restricted to 586 children after excluding extremes of growth trajectories (maternal T1D, <36 weeks of gestation, twins). Genetic correlation with the T1D high-risk HLA SNPs was explored for every SNP by pairwise associations using the chi-square test. Multivariate linear regression was used to test for associations between each phenotype (explained variable) at a single age measured and each individual SNP or genetic risk score

(explanatory variable), with adjustment for preselected covariates (Table 3-21) by co-entering those variables in the model (Equation 6-1). Additive genetic effects were assumed.

Equation 6-1

$$Y_i = \beta_0 + \sum_{j=1}^{m} \beta_j \times X_{ij} \quad + \beta_{SNP} \times SNP_i$$

for study participant i=1, 2,…n; X is one of m covariates.

The null hypothesis tested was H_0: β_{SNP}=0 vs. H_1: β_{SNP}≠0. Statistical analyses were performed using the IBM Statistical Package for Social Sciences (SPSS 23.0 Chicago, IL, USA) for Windows. Tests were two-tailed and a P≤0.05 was considered statistically significant.

6.3 RESULTS

6.3.1 Descriptive statistics

Characteristics of the children genotyped for the study are provided in Chapter 3. Compared with infants in the birth cohort that were not genotyped, there were no material differences in newborn traits or confounders. All SNPs were in HWE at the 0.01 level of significance and their frequencies in the CBGS were in good agreement with those published (Table 4-5).

6.3.2 Genetic correlations with HLA SNPs

Genetic correlations (i.e. associations of population frequencies) were explored with each of the SNPs tagging the T1D HLA which have the greatest effect on risk of T1D. There was a significant co-occurrence of homozygosity for the protective *HLA-DQB1*06:02* minor allele (coded as '0') with homozygosity for each of the major DR3, DR4, DQ8 alleles (Figure 6-1). Conversely, homozygous for the high-risk HLA minor alleles did not carry any protective *HLA-DQB1*06:02* allele. This pattern points to a mutual exclusivity of homozygosity for SNPs tagging the extremes on the spectrum of T1D risk, i.e. very high-risk and very protective HLA-II SNP-tag alleles. These associations might result from the rarity of the HLA alleles in the study population. However, the genotype distributions of the rarest alleles in the study (*TYK2* and *ITGB7*) relative to *HLA-DQB1*06:02* genotypes did not conform to this pattern.

Figure 6-1 | Counts (%) of infants by T1D risk genotypic group between the *HLA-DQB1*06:02* and the HLA- a. D*R3*, b. *DQ8*, c. *DR4*.

a. *HLA-DQB1*06:02 and HLA-DR3*

N (%)				
		rs3135388 *HLA-DQB1*06:02*		
		0	1	2
rs2187668 *HLA-DR3*	0	11 (2.7)	113 (27.6)	285 (69.7)
	1	0 (0.0)	19 (13.7)	120 (86.3)
	2	0 (0.0)	0 (0.0)	8 (100.0)

$P_{chi\text{-}square}$ = 0.001

b. *HLA-DQB1*06:02 and HLA-DQ8*

N (%)				
		rs3135388 *HLA-DQB1*06:02*		
		0	1	2
rs7454108 *HLA-DQ8*	0	11 (2.5)	117 (26.4)	315 (71.1)
	1	0 (0.0)	12 (11.8)	90 (88.2)
	2	0 (0.0)	0 (0.0)	6 (100.0)

$P_{chi\text{-}square}$ = 0.004

c. *HLA-DQB1*06:02 and HLA-DR4*

N (%)				
		rs3135388 *HLA-DQB1*06:02*		
		0	1	2
rs17426593 *HLA-DR4*	0	10 (2.8)	102 (29.1)	239 (68.1)
	1	0 (0.0)	30 (16.9)	148 (83.1)
	2	0 (0.0)	0 (0.0)	14 (100.0)

$P_{chi\text{-}square}$ < 0.0001

6.3.3 T1D SNPs and early growth

Tables 6-1 to 6-6 display the summary statistics (standardised Beta, *P* value) of regressions of standardised growth parameters by age on individual SNPs. Birth weight showed associations with SNPs in *DLK1*, *CYP2R1* and *TYK2*. Height in infancy was associated with SNPs in *BACH2* and *HLA-DR3*. The strongest genetic associations were identified for measures of obesity: *PRKCQ* and BMI at birth and 12 months; *HLA-DR4*, *GC, BAD* and *SH2B3* each strongly associated with overall skinfold thickness postnatally.

Table 6-1 | Associations between T1D SNPs and weight SDS at birth and ages 3, 12, 18, 24 months†.

	SNP	Gene	Birth		3 months		12 months		18 months		24 months	
			Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
1	rs2476601	*PTPN22*	-0.022	0.645	0.002	0.964	0.022	0.640	0.057	0.245	0.029	0.572
2	rs3024505	*IL10*	-0.025	0.592	-0.032	0.502	-0.024	0.624	0.024	0.620	-0.020	0.688
3	rs2111485	*IFIH1*	-0.076	0.099	-0.047	0.333	0.025	0.608	0.009	0.862	0.019	0.708
4	rs3087243	*CTLA4*	0.015	0.750	-0.019	0.696	-0.012	0.800	-0.050	0.321	0.003	0.948
5	rs4588	*GC*	-0.021	0.655	0.107	0.026	0.037	0.449	0.040	0.415	0.028	0.578
6	rs75793288	*IL2*	0.019	0.680	-0.043	0.379	-0.005	0.916	0.009	0.859	0.049	0.331
7	rs17426593	*HLA-DR4*	0.005	0.923	-0.001	0.977	-0.043	0.378	-0.049	0.326	-0.029	0.582
8	rs2187668	*HLA-DR3*	-0.021	0.655	-0.020	0.679	0.047	0.324	0.064	0.194	0.061	0.232
9	rs7454108	*HLA-DQ8*	-0.001	0.976	-0.006	0.901	-0.080	0.094	-0.052	0.297	-0.049	0.342
10	rs3135388	*HLA-DQB1*06:02*	-0.016	0.732	-0.026	0.585	-0.019	0.698	-0.019	0.699	0.004	0.931
11	rs72928038	*BACH2*	0.047	0.321	0.118	0.014	0.089	0.066	0.094	0.058	0.068	0.184
12	rs6920220	*TNFAIP3*	-0.027	0.568	0.041	0.392	-0.015	0.760	-0.027	0.591	0.024	0.636
13	rs6476839	*GLIS3*	0.036	0.445	0.096	0.048	0.022	0.646	0.000	0.993	0.027	0.594
14	rs61839660	*IL2RA*	-0.044	0.353	0.054	0.270	0.020	0.682	-0.020	0.694	0.005	0.929
15	rs11258747	*PRKCQ*	-0.063	0.177	-0.061	0.205	-0.074	0.121	-0.003	0.953	0.006	0.908
16	rs12416116	*RNLS*	0.054	0.248	0.078	0.108	0.056	0.242	0.032	0.524	0.032	0.535
17	rs10741657	*CYP2R1*	-0.108	0.022	0.021	0.674	0.005	0.917	0.002	0.969	-0.024	0.648
18	rs12794714	*CYP2R1*	-0.094	0.043	-0.025	0.602	-0.008	0.866	-0.039	0.428	-0.043	0.396
19	rs689	*INS*	0.026	0.580	-0.054	0.271	0.053	0.280	0.018	0.726	0.054	0.297
20	rs694739	*BAD*	0.015	0.745	0.065	0.179	-0.016	0.735	-0.017	0.734	-0.029	0.573
21	rs12785878	*DHCR7*	-0.010	0.831	-0.014	0.770	-0.096	0.047	-0.087	0.082	-0.065	0.206
22	rs11170466	*ITGB7*	-0.053	0.254	0.013	0.795	0.020	0.677	-0.026	0.602	-0.018	0.729
23	rs705704	*ERBB3*	-0.065	0.164	-0.022	0.652	-0.028	0.562	-0.009	0.859	-0.018	0.731
24	rs10877012	*CYP27B1*	0.040	0.384	-0.012	0.799	-0.009	0.853	-0.022	0.655	-0.035	0.488
25	rs653178	*SH2B3*	-0.025	0.589	0.040	0.409	0.058	0.236	0.027	0.584	0.078	0.130
26	rs56994090	*DLK1*	0.111	0.018	0.061	0.209	0.076	0.114	0.038	0.443	0.040	0.430
27	rs34593439	*CTSH*	0.011	0.817	0.057	0.239	-0.015	0.757	0.001	0.986	-0.065	0.208
28	rs12927355	*CLEC16A*	0.050	0.280	0.068	0.158	0.095	0.047	0.034	0.497	0.015	0.764
29	rs151234	*IL27*	0.023	0.619	-0.004	0.935	-0.036	0.461	-0.035	0.486	-0.058	0.255
30	rs12453507	*ORMDL3*	0.004	0.931	0.050	0.302	0.061	0.212	0.067	0.181	0.045	0.380
31	rs1893217	*PTPN2*	0.035	0.452	0.070	0.147	0.028	0.566	0.015	0.760	0.042	0.414
32	rs34536443	*TYK2*	0.105	0.024	0.057	0.238	0.075	0.116	0.071	0.150	0.110	0.031
33	rs516246	*FUT2*	-0.041	0.379	-0.053	0.271	-0.017	0.721	-0.036	0.473	-0.024	0.639
34	rs11203202	*UBASH3A*	-0.027	0.566	-0.092	0.059	-0.044	0.364	-0.009	0.861	-0.084	0.103

†Models at birth were adjusted for parity, smoking during pregnancy, maternal height and maternal pre-pregnancy weight. Models at 3 months were adjusted for parity, maternal height, maternal pre-pregnancy weight and type of milk feeding at age 3 months. Models at subsequent ages were adjusted for maternal height, maternal pre-pregnancy weight and type of milk feeding at age 3 months.

Table 6-2 | Associations between T1D SNPs and height SDS at birth and ages 3, 12, 18, 24 months †.

	SNP	Gene	Birth		3 months		12 months		18 months		24 months	
			Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
1	rs2476601	*PTPN22*	-0.046	0.339	0.000	0.995	-0.033	0.487	-0.011	0.831	0.006	0.908
2	rs3024505	*IL10*	-0.052	0.281	-0.020	0.671	0.040	0.399	0.054	0.286	-0.017	0.740
3	rs2111485	*IFIH1*	-0.026	0.586	-0.038	0.425	0.066	0.163	0.060	0.237	0.048	0.342
4	rs3087243	*CTLA4*	0.005	0.914	-0.025	0.609	-0.019	0.693	-0.004	0.942	0.026	0.612
5	rs4588	*GC*	-0.027	0.579	-0.031	0.514	-0.016	0.744	-0.013	0.793	0.014	0.779
6	rs75793288	*IL2*	-0.009	0.855	-0.026	0.595	-0.052	0.280	-0.065	0.200	-0.009	0.865
7	rs17426593	*HLA-DR4*	0.046	0.337	0.007	0.881	0.002	0.964	-0.022	0.670	-0.023	0.653
8	rs2187668	*HLA-DR3*	0.018	0.698	-0.005	0.911	0.069	0.143	0.074	0.141	0.116	0.021
9	rs7454108	*HLA-DQ8*	0.056	0.243	-0.012	0.800	-0.053	0.269	-0.046	0.364	-0.058	0.254
10	rs3135388	*HLA-DQB1*06:02*	-0.020	0.674	-0.047	0.324	-0.027	0.571	-0.051	0.321	0.013	0.793
11	rs72928038	*BACH2*	0.067	0.166	0.094	0.050	0.118	0.014	0.130	0.011	0.089	0.083
12	rs6920220	*TNFAIP3*	-0.017	0.715	0.021	0.662	-0.015	0.754	0.007	0.895	0.069	0.173
13	rs6476839	*GLIS3*	-0.029	0.541	0.007	0.880	-0.007	0.890	0.005	0.915	0.011	0.831
14	rs61839660	*IL2RA*	-0.052	0.277	0.004	0.940	-0.033	0.495	-0.008	0.873	-0.016	0.755
15	rs11258747	*PRKCQ*	0.068	0.155	-0.037	0.435	0.031	0.515	0.010	0.844	0.070	0.171
16	rs12416116	*RNLS*	0.012	0.795	0.031	0.524	-0.013	0.789	-0.009	0.858	0.014	0.776
17	rs10741657	*CYP2R1*	-0.090	0.063	-0.039	0.418	-0.037	0.447	-0.065	0.210	-0.023	0.657
18	rs12794714	*CYP2R1*	-0.023	0.636	-0.048	0.320	-0.029	0.543	-0.110	0.031	-0.071	0.160
19	rs689	*INS*	-0.003	0.950	-0.041	0.405	-0.006	0.909	0.046	0.373	0.001	0.982
20	rs694739	*BAD*	0.003	0.943	0.048	0.319	-0.036	0.452	-0.005	0.930	0.038	0.458
21	rs12785878	*DHCR7*	-0.046	0.333	-0.085	0.077	-0.046	0.343	-0.066	0.195	-0.079	0.120
22	rs11170466	*ITGB7*	-0.060	0.207	-0.087	0.070	-0.073	0.126	-0.086	0.091	-0.031	0.539
23	rs705704	*ERBB3*	-0.055	0.254	-0.030	0.526	-0.044	0.354	-0.029	0.575	-0.055	0.285
24	rs10877012	*CYP27B1*	-0.009	0.854	-0.008	0.862	-0.043	0.369	-0.023	0.643	-0.022	0.664
25	rs653178	*SH2B3*	-0.085	0.076	0.009	0.850	0.007	0.880	0.013	0.804	0.030	0.557
26	rs56994090	*DLK1*	0.058	0.223	0.036	0.455	-0.005	0.919	-0.030	0.555	-0.014	0.790
27	rs34593439	*CTSH*	0.094	0.051	0.004	0.941	-0.006	0.901	0.006	0.901	-0.011	0.838
28	rs12927355	*CLEC16A*	0.006	0.908	0.063	0.193	0.014	0.765	-0.012	0.809	-0.066	0.196
29	rs151234	*IL27*	0.051	0.281	0.016	0.735	-0.038	0.427	-0.033	0.521	-0.056	0.273
30	rs12453507	*ORMDL3*	0.026	0.591	0.052	0.281	0.046	0.344	0.078	0.130	0.075	0.144
31	rs1893217	*PTPN2*	0.005	0.918	0.062	0.194	0.038	0.421	0.002	0.970	-0.009	0.856
32	rs34536443	*TYK2*	0.080	0.097	0.095	0.047	0.084	0.079	0.083	0.103	0.044	0.386
33	rs516246	*FUT2*	0.011	0.817	-0.032	0.499	-0.001	0.981	-0.036	0.488	-0.047	0.362
34	rs11203202	*UBASH3A*	0.023	0.634	-0.015	0.756	-0.003	0.955	0.046	0.373	0.008	0.880

†Models at birth were adjusted for parity, smoking during pregnancy, maternal height and maternal pre-pregnancy weight. Models at 3 months were adjusted for parity, maternal height, maternal pre-pregnancy weight and type of milk feeding at age 3 months. Models at subsequent ages were adjusted for maternal height, maternal pre-pregnancy weight and type of milk feeding at age 3 months.

Table 6-3 | Associations between T1D SNPs and HC SDS at birth and ages 12, 18, 24 months†.

	SNP	Gene	Birth		3 months		12 months		18 months		24 months	
			Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
1	rs2476601	*PTPN22*	-0.025	0.586			-0.005	0.918	-0.005	0.924	0.003	0.957
2	rs3024505	*IL10*	0.021	0.653			0.024	0.640	-0.003	0.956	-0.010	0.854
3	rs2111485	*IFIH1*	-0.059	0.202			0.084	0.099	0.057	0.272	0.051	0.339
4	rs3087243	*CTLA4*	0.027	0.565			-0.062	0.232	-0.032	0.539	0.007	0.896
5	rs4588	*GC*	-0.057	0.221			0.020	0.700	0.027	0.606	0.033	0.532
6	rs75793288	*IL2*	-0.019	0.679			0.024	0.639	0.018	0.737	0.039	0.468
7	rs17426593	*HLA-DR4*	0.011	0.822			-0.082	0.116	-0.100	0.059	-0.090	0.096
8	rs2187668	*HLA-DR3*	-0.037	0.423			0.027	0.591	0.041	0.432	0.062	0.242
9	rs7454108	*HLA-DQ8*	-0.015	0.742			-0.051	0.321	-0.078	0.137	-0.067	0.213
10	rs3135388	*HLA-DQB1*06:02*	-0.021	0.657			-0.039	0.455	-0.040	0.450	-0.037	0.492
11	rs72928038	*BACH2*	0.013	0.787			-0.051	0.328	-0.054	0.308	-0.046	0.396
12	rs6920220	*TNFAIP3*	-0.004	0.927			-0.057	0.269	-0.044	0.396	-0.002	0.977
13	rs6476839	*GLIS3*	0.074	0.113			0.004	0.945	-0.012	0.816	-0.003	0.950
14	rs61839660	*IL2RA*	-0.088	0.059			-0.020	0.697	-0.027	0.604	-0.002	0.973
15	rs11258747	*PRKCQ*	-0.010	0.834			-0.048	0.348	-0.003	0.947	0.004	0.933
16	rs12416116	*RNLS*	0.024	0.599			-0.019	0.708	-0.044	0.398	-0.047	0.378
17	rs10741657	*CYP2R1*	-0.044	0.350			-0.048	0.357	0.003	0.961	-0.043	0.423
18	rs12794714	*CYP2R1*	-0.065	0.163			-0.093	0.070	-0.056	0.281	-0.056	0.292
19	rs689	*INS*	-0.039	0.413			-0.001	0.977	-0.009	0.865	-0.025	0.650
20	rs694739	*BAD*	0.002	0.974			0.017	0.751	0.005	0.929	0.005	0.923
21	rs12785878	*DHCR7*	0.020	0.666			-0.109	0.034	-0.086	0.101	-0.084	0.117
22	rs11170466	*ITGB7*	-0.033	0.481			0.008	0.882	0.000	0.995	0.008	0.887
23	rs705704	*ERBB3*	-0.073	0.115			-0.050	0.336	-0.097	0.063	-0.081	0.133
24	rs10877012	*CYP27B1*	0.058	0.213			0.029	0.577	-0.003	0.955	-0.024	0.646
25	rs653178	*SH2B3*	0.011	0.810			-0.021	0.685	-0.010	0.844	-0.003	0.957
26	rs56994090	*DLK1*	0.027	0.563			0.101	0.051	0.055	0.289	0.096	0.073
27	rs34593439	*CTSH*	-0.008	0.857			-0.013	0.799	-0.012	0.814	-0.018	0.737
28	rs12927355	*CLEC16A*	0.040	0.393			0.105	0.041	0.056	0.287	0.049	0.358
29	rs151234	*IL27*	0.001	0.988			0.001	0.990	0.034	0.509	0.000	0.993
30	rs12453507	*ORMDL3*	-0.029	0.537			-0.024	0.648	-0.010	0.841	0.001	0.980
31	rs1893217	*PTPN2*	-0.004	0.933			-0.004	0.940	0.025	0.624	-0.009	0.864
32	rs34536443	*TYK2*	0.088	0.059			0.116	0.024	0.100	0.055	0.084	0.118
33	rs516246	*FUT2*	-0.027	0.555			0.036	0.485	0.003	0.951	0.000	1.000
34	rs11203202	*UBASH3A*	0.001	0.975			-0.111	0.032	-0.068	0.192	-0.078	0.144

†Models at birth were adjusted for parity, maternal height and maternal pre-pregnancy weight. Models at subsequent ages were adjusted for maternal height and maternal pre-pregnancy weight.

Table 6-4 | Associations between T1D SNPs and BMI SDS at birth and ages 12, 18, 24 months†.

	SNP	Gene	Birth		3 months		12 months		18 months		24 months	
			Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
1	rs2476601	*PTPN22*	-0.006	0.909			0.057	0.270	0.081	0.124	0.021	0.696
2	rs3024505	*IL10*	-0.002	0.966			-0.064	0.219	-0.017	0.755	-0.013	0.812
3	rs2111485	*IFIH1*	-0.050	0.312			-0.028	0.590	-0.050	0.345	-0.031	0.570
4	rs3087243	*CTLA4*	0.022	0.658			-0.011	0.839	-0.078	0.148	-0.032	0.572
5	rs4588	*GC*	0.035	0.482			0.073	0.159	0.091	0.087	0.031	0.579
6	rs75793288	*IL2*	0.019	0.708			0.033	0.523	0.066	0.211	0.090	0.101
7	rs17426593	*HLA-DR4*	0.005	0.927			-0.050	0.340	-0.037	0.484	0.000	0.997
8	rs2187668	*HLA-DR3*	-0.033	0.500			0.009	0.857	0.019	0.723	-0.034	0.539
9	rs7454108	*HLA-DQ8*	-0.025	0.613			-0.048	0.355	-0.013	0.803	-0.008	0.881
10	rs3135388	*HLA-DQB1*06:02*	-0.037	0.462			0.004	0.936	0.041	0.440	-0.010	0.858
11	rs72928038	*BACH2*	0.003	0.958			0.015	0.782	0.002	0.971	-0.003	0.955
12	rs6920220	*TNFAIP3*	-0.055	0.263			0.006	0.910	-0.039	0.466	-0.004	0.949
13	rs6476839	*GLIS3*	0.053	0.285			0.033	0.529	-0.004	0.941	0.023	0.673
14	rs61839660	*IL2RA*	-0.037	0.452			0.051	0.325	-0.031	0.568	0.015	0.782
15	rs11258747	*PRKCQ*	-0.127	0.010			-0.127	0.014	-0.018	0.737	-0.056	0.312
16	rs12416116	*RNLS*	0.036	0.466			0.097	0.061	0.061	0.252	0.020	0.721
17	rs10741657	*CYP2R1*	-0.073	0.142			0.030	0.566	0.078	0.151	-0.001	0.991
18	rs12794714	*CYP2R1*	-0.100	0.042			0.006	0.906	0.059	0.267	0.008	0.886
19	rs689	*INS*	0.063	0.212			0.084	0.108	-0.021	0.698	0.064	0.250
20	rs694739	*BAD*	-0.056	0.258			0.000	0.996	-0.034	0.528	-0.105	0.059
21	rs12785878	*DHCR7*	-0.001	0.978			-0.107	0.040	-0.074	0.163	-0.041	0.459
22	rs11170466	*ITGB7*	-0.045	0.360			0.086	0.098	0.050	0.346	-0.012	0.828
23	rs705704	*ERBB3*	-0.073	0.140			-0.020	0.702	0.004	0.942	0.022	0.687
24	rs10877012	*CYP27B1*	0.048	0.331			0.024	0.650	-0.003	0.961	-0.024	0.667
25	rs653178	*SH2B3*	0.049	0.322			0.068	0.191	0.031	0.567	0.072	0.191
26	rs56994090	*DLK1*	0.101	0.041			0.107	0.040	0.076	0.153	0.087	0.115
27	rs34593439	*CTSH*	-0.007	0.894			0.008	0.872	0.021	0.691	-0.056	0.311
28	rs12927355	*CLEC16A*	0.091	0.066			0.117	0.024	0.061	0.256	0.095	0.084
29	rs151234	*IL27*	-0.010	0.845			-0.017	0.749	-0.018	0.741	-0.029	0.600
30	rs12453507	*ORMDL3*	-0.022	0.664			0.058	0.267	0.025	0.639	0.012	0.826
31	rs1893217	*PTPN2*	0.015	0.761			0.013	0.806	0.016	0.756	0.096	0.078
32	rs34536443	*TYK2*	0.067	0.181			0.047	0.367	0.038	0.479	0.118	0.033
33	rs516246	*FUT2*	-0.059	0.231			-0.041	0.433	-0.031	0.567	-0.015	0.789
34	rs11203202	*UBASH3A*	-0.028	0.579			-0.058	0.264	-0.059	0.270	-0.125	0.024

†Models at birth were adjusted for parity, smoking during pregnancy and maternal pre-pregnancy BMI. Models at subsequent ages were adjusted for maternal pre-pregnancy BMI and type of milk feeding at age 3 months.

Table 6-5 | Associations between T1D SNPs and skinfold thickness SDS at birth and ages 3, 12, 18, 24 months†.

	SNP	Gene	Birth		3 months		12 months		18 months		24 months	
			Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
1	rs2476601	*PTPN22*	-0.021	0.616	0.031	0.497	0.103	0.029	0.093	0.056	0.075	0.136
2	rs3024505	*IL10*	-0.075	0.080	-0.056	0.220	-0.040	0.397	0.034	0.485	0.013	0.793
3	rs2111485	*IFIH1*	-0.083	0.051	-0.008	0.854	-0.007	0.885	-0.029	0.549	-0.039	0.442
4	rs3087243	*CTLA4*	0.059	0.172	-0.020	0.668	0.009	0.857	-0.024	0.621	0.043	0.399
5	rs4588	*GC*	-0.018	0.674	0.143	0.002	-0.042	0.374	0.052	0.286	0.065	0.194
6	rs75793288	*IL2*	0.079	0.064	-0.028	0.542	0.044	0.354	0.108	0.026	0.057	0.255
7	rs17426593	*HLA-DR4*	0.029	0.506	0.093	0.043	0.043	0.372	0.082	0.095	0.150	0.003
8	rs2187668	*HLA-DR3*	0.072	0.089	-0.003	0.940	0.017	0.717	-0.054	0.269	-0.026	0.608
9	rs7454108	*HLA-DQ8*	0.004	0.933	0.067	0.142	-0.012	0.809	0.063	0.201	0.120	0.017
10	rs3135388	*HLA-DQB1*06:02*	0.049	0.259	-0.042	0.358	-0.071	0.135	-0.057	0.244	-0.044	0.384
11	rs72928038	*BACH2*	0.026	0.543	0.009	0.840	-0.016	0.739	-0.004	0.936	-0.017	0.736
12	rs6920220	*TNFAIP3*	0.039	0.366	0.009	0.847	-0.104	0.029	-0.047	0.338	-0.010	0.845
13	rs6476839	*GLIS3*	0.040	0.347	0.015	0.744	-0.001	0.976	-0.048	0.329	-0.052	0.302
14	rs61839660	*IL2RA*	-0.009	0.828	0.069	0.132	0.056	0.241	0.074	0.132	0.065	0.200
15	rs11258747	*PRKCQ*	0.023	0.595	-0.042	0.361	-0.070	0.142	0.019	0.705	-0.086	0.086
16	rs12416116	*RNLS*	-0.021	0.617	0.075	0.100	0.041	0.391	0.008	0.864	0.083	0.098
17	rs10741657	*CYP2R1*	0.012	0.788	0.015	0.751	0.027	0.572	0.046	0.351	0.053	0.297
18	rs12794714	*CYP2R1*	0.030	0.477	-0.054	0.238	0.002	0.974	-0.005	0.911	0.035	0.486
19	rs689	*INS*	-0.079	0.069	-0.063	0.172	0.052	0.276	0.008	0.873	-0.023	0.655
20	rs694739	*BAD*	0.059	0.171	0.078	0.089	-0.046	0.333	-0.059	0.231	-0.132	0.009
21	rs12785878	*DHCR7*	0.026	0.534	-0.006	0.888	-0.043	0.371	-0.056	0.253	0.006	0.906
22	rs11170466	*ITGB7*	0.051	0.237	0.037	0.420	0.058	0.222	0.008	0.870	0.083	0.098
23	rs705704	*ERBB3*	-0.049	0.257	-0.023	0.619	0.026	0.594	-0.027	0.582	-0.001	0.992
24	rs10877012	*CYP27B1*	-0.011	0.792	-0.043	0.342	0.002	0.964	-0.038	0.439	-0.060	0.230
25	rs653178	*SH2B3*	-0.065	0.130	0.064	0.165	0.089	0.063	0.141	0.004	0.128	0.012
26	rs56994090	*DLK1*	0.077	0.073	0.032	0.480	0.085	0.074	0.072	0.140	0.078	0.120
27	rs34593439	*CTSH*	-0.017	0.684	-0.040	0.384	-0.095	0.046	-0.034	0.485	-0.024	0.629
28	rs12927355	*CLEC16A*	-0.005	0.908	-0.001	0.975	0.079	0.095	0.027	0.586	-0.005	0.928
29	rs151234	*IL27*	0.004	0.929	0.013	0.778	-0.044	0.360	0.049	0.321	-0.037	0.468
30	rs12453507	*ORMDL3*	0.028	0.515	0.059	0.195	-0.021	0.658	0.032	0.517	0.063	0.211
31	rs1893217	*PTPN2*	0.029	0.493	0.094	0.037	0.059	0.211	0.040	0.409	0.074	0.138
32	rs34536443	*TYK2*	0.081	0.059	-0.014	0.768	0.038	0.429	0.057	0.244	0.029	0.570
33	rs516246	*FUT2*	-0.026	0.546	-0.054	0.238	0.025	0.603	0.045	0.358	0.057	0.256
34	rs11203202	*UBASH3A*	0.002	0.960	-0.068	0.139	-0.016	0.734	-0.039	0.431	-0.052	0.303

†Models at birth were adjusted for parity. Models at subsequent ages were adjusted for type of milk feeding at age 3 months.

Table 6-6 | Associations between T1D SNPs and gains of weight SDS or height SDS between birth and 12 months or birth and 24 months†.

			Δ weight SDS					Δ height SDS			
	SNP	Gene	0-12 months		0-24 months			0-12 months		0-24 months	
			Beta	*P* value	Beta	*P* value		Beta	*P* value	Beta	*P* value
1	rs2476601	*PTPN22*	0.024	0.557	0.030	0.473		-0.018	0.676	0.025	0.566
2	rs3024505	*IL10*	-0.017	0.679	-0.014	0.745		0.070	0.100	0.004	0.928
3	rs2111485	*IFIH1*	0.056	0.182	0.049	0.252		0.061	0.153	0.027	0.534
4	rs3087243	*CTLA4*	-0.021	0.612	-0.005	0.903		-0.031	0.475	0.009	0.842
5	rs4588	*GC*	0.045	0.278	0.029	0.502		-0.003	0.950	0.008	0.855
6	rs75793288	*IL2*	-0.016	0.703	0.030	0.481		-0.033	0.430	-0.011	0.795
7	rs17426593	*HLA-DR4*	-0.036	0.394	-0.025	0.568		-0.008	0.862	-0.017	0.703
8	rs2187668	*HLA-DR3*	0.034	0.405	0.052	0.218		0.048	0.259	0.089	0.041
9	rs7454108	*HLA-DQ8*	-0.068	0.100	-0.039	0.357		-0.067	0.118	-0.070	0.118
10	rs3135388	*HLA-DQB1*06:02*	-0.015	0.711	0.011	0.796		-0.002	0.957	0.026	0.546
11	rs72928038	*BACH2*	0.071	0.089	0.058	0.174		0.096	0.025	0.061	0.167
12	rs6920220	*TNFAIP3*	-0.018	0.658	0.010	0.811		-0.030	0.475	0.044	0.320
13	rs6476839	*GLIS3*	0.010	0.800	0.010	0.823		-0.010	0.808	-0.006	0.889
14	rs61839660	*IL2RA*	0.040	0.332	0.021	0.631		-0.012	0.778	-0.008	0.857
15	rs11258747	*PRKCQ*	-0.041	0.321	0.014	0.747		0.015	0.724	0.037	0.405
16	rs12416116	*RNLS*	0.030	0.463	0.001	0.978		-0.041	0.344	0.011	0.804
17	rs10741657	*CYP2R1*	0.037	0.379	0.004	0.918		-0.030	0.478	-0.011	0.802
18	rs12794714	*CYP2R1*	0.026	0.529	-0.009	0.838		-0.027	0.530	-0.052	0.236
19	rs689	*INS*	0.042	0.327	0.027	0.539		0.021	0.630	0.003	0.953
20	rs694739	*BAD*	-0.002	0.964	-0.020	0.638		-0.055	0.200	0.001	0.975
21	rs12785878	*DHCR7*	-0.089	0.032	-0.061	0.154		-0.022	0.602	-0.048	0.275
22	rs11170466	*ITGB7*	0.028	0.495	-0.005	0.908		-0.047	0.270	-0.013	0.770
23	rs705704	*ERBB3*	-0.016	0.711	0.002	0.969		-0.049	0.249	-0.043	0.330
24	rs10877012	*CYP27B1*	-0.022	0.602	-0.045	0.291		-0.064	0.132	-0.044	0.317
25	rs653178	*SH2B3*	0.058	0.160	0.075	0.079		0.010	0.811	0.034	0.447
26	rs56994090	*DLK1*	0.038	0.363	0.016	0.705		0.001	0.987	0.015	0.728
27	rs34593439	*CTSH*	-0.013	0.756	-0.056	0.192		-0.029	0.505	-0.025	0.570
28	rs12927355	*CLEC16A*	0.055	0.189	-0.028	0.518		0.007	0.870	-0.074	0.092
29	rs151234	*IL27*	-0.035	0.397	-0.056	0.191		-0.050	0.238	-0.065	0.138
30	rs12453507	*ORMDL3*	0.061	0.142	0.045	0.294		0.036	0.406	0.036	0.416
31	rs1893217	*PTPN2*	0.002	0.960	0.018	0.668		0.022	0.599	0.015	0.738
32	rs34536443	*TYK2*	0.044	0.291	0.074	0.086		0.064	0.133	0.020	0.649
33	rs516246	*FUT2*	0.008	0.854	-0.001	0.988		0.007	0.874	-0.045	0.305
34	rs11203202	*UBASH3A*	-0.035	0.399	-0.065	0.130		-0.010	0.814	0.007	0.874

†Weight models were adjusted for maternal height, maternal pre-pregnancy weight, type of milk feeding at age 3 months and birth weight SDS. Height models were adjusted for maternal height, type of milk feeding at age 3 months and birth length SDS.

6.3.4 Genetic risk scores and early growth

Tables 6-7 to 6-12 display the summary statistics (standardised Beta, *P* value) of regressions of standardised growth parameters by age on T1D genetic risk scores. There was no evidence of an association between any genetic risk score and primary growth measures. BMI showed a borderline positive association with the weighted non-HLA risk score at 12 months (*P*=0.082). There was a weak positive association between the HLA binary risk score and skinfold thickness at 24 months, i.e. carrying the high-risk HLA-DR/DQ genotype was associated with 0.115 higher SD score of mean skinfold thickness at the end of infancy (*P*=0.022).

Table 6-7 | Associations between T1D genetic risk scores and weight SDS at birth and ages 3, 12, 18, 24 months†.

Genetic Risk Score	Birth		3 months		12 months		18 months		24 months	
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
HLA										
HLA Risk_2level	-0.029	0.536	-0.022	0.649	-0.002	0.972	0.020	0.687	-0.013	0.807
HLA Risk_3level	-0.021	0.650	-0.022	0.658	-0.009	0.859	0.026	0.600	0.031	0.549
Unweighted										
nonHLA RiskSum	-0.008	0.859	0.076	0.121	0.048	0.320	0.006	0.909	0.021	0.679
VitD RiskSum	-0.043	0.364	0.021	0.673	-0.038	0.438	-0.058	0.247	-0.063	0.223
Weighted										
nonHLA w-RiskScore	-0.016	0.733	0.007	0.882	0.061	0.210	0.028	0.575	0.060	0.249
VitD w-RiskScore	0.006	0.894	-0.012	0.813	-0.043	0.382	-0.061	0.228	-0.071	0.171
Combined RiskScore	-0.023	0.631	-0.025	0.605	0.006	0.901	0.011	0.827	0.038	0.465

†Models were adjusted for covariates as described in the legend of Table 6-1.

Table 6-8 | Associations between T1D genetic risk scores and height SDS at birth and ages 3, 12, 18, 24 months†.

Genetic Risk Score	Birth		3 months		12 months		18 months		24 months	
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
HLA										
HLA Risk_2level	-0.019	0.696	-0.043	0.372	-0.004	0.928	0.016	0.754	0.024	0.643
HLA Risk_3level	0.050	0.290	-0.006	0.905	0.037	0.438	0.040	0.430	0.061	0.229
Unweighted										
nonHLA RiskSum	-0.008	0.869	-0.002	0.960	-0.024	0.626	-0.014	0.791	0.006	0.912
VitD RiskSum	-0.053	0.273	-0.082	0.088	-0.067	0.165	-0.097	0.058	-0.081	0.114
Weighted										
nonHLA w-RiskScore	-0.020	0.679	-0.015	0.756	-0.010	0.839	0.026	0.612	0.015	0.765
VitD w-RiskScore	-0.041	0.401	-0.052	0.278	-0.067	0.168	-0.064	0.212	-0.064	0.217
Combined RiskScore	0.009	0.857	-0.041	0.398	-0.007	0.881	-0.007	0.897	0.042	0.419

†Models were adjusted for covariates as described in the legend of Table 6-2.

Table 6-9 | Associations between T1D genetic risk scores and HC SDS at birth and ages 12, 18, 24 months†.

Genetic Risk Score	Birth		3 months		12 months		18 months		24 months	
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
HLA										
HLA Risk_2level	-0.053	0.248			0.005	0.927	0.013	0.798	0.015	0.782
HLA Risk_3level	-0.037	0.430			-0.014	0.783	-0.011	0.827	0.021	0.696
Unweighted										
nonHLA RiskSum	-0.033	0.475			-0.045	0.382	-0.060	0.257	-0.036	0.501
VitD RiskSum	-0.030	0.522			-0.084	0.104	-0.074	0.160	-0.074	0.169
Weighted										
nonHLA w-RiskScore	-0.058	0.211			-0.033	0.523	-0.045	0.396	-0.043	0.429
VitD w-RiskScore	0.023	0.619			-0.031	0.554	-0.051	0.333	-0.063	0.240
Combined RiskScore	-0.052	0.267			-0.046	0.376	-0.053	0.320	-0.037	0.498

†Models were adjusted for covariates as described in the legend of Table 6-3.

Table 6-10 | Associations between T1D genetic risk scores and BMI SDS at birth and ages 12, 18, 24 months†.

Genetic Risk Score	Birth		3 months		12 months		18 months		24 months	
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
HLA										
HLA Risk_2level	-0.018	0.712			0.010	0.852	0.021	0.694	-0.038	0.487
HLA Risk_3level	-0.051	0.302			-0.031	0.555	0.006	0.915	-0.024	0.665
Unweighted										
nonHLA RiskSum	-0.018	0.724			0.080	0.125	0.014	0.792	0.016	0.772
VitD RiskSum	-0.009	0.864			-0.002	0.973	0.019	0.717	-0.018	0.742
Weighted										
nonHLA w-RiskScore	0.003	0.948			0.091	0.082	0.006	0.906	0.058	0.298
VitD w-RiskScore	0.033	0.504			-0.007	0.900	-0.018	0.731	-0.042	0.455
Combined RiskScore	-0.046	0.360			0.027	0.612	0.037	0.496	0.004	0.937

†Models were adjusted for covariates as described in the legend of Table 6-4.

Table 6-11 | Associations between T1D genetic risk scores and skinfold thickness SDS at birth and ages 3, 12, 18, 24 months†.

Genetic Risk Score	Birth		3 months		12 months		18 months		24 months	
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
HLA										
HLA Risk_2level	0.037	0.389	0.059	0.190	0.044	0.353	0.064	0.190	0.115	0.022
HLA Risk_3level	0.063	0.143	0.044	0.332	0.012	0.801	0.008	0.877	0.074	0.142
Unweighted										
nonHLA RiskSum	0.024	0.572	0.026	0.567	0.021	0.665	0.063	0.200	0.047	0.356
VitD RiskSum	0.012	0.785	0.003	0.947	-0.052	0.278	-0.029	0.551	0.006	0.900
Weighted										
nonHLA w-RiskScore	-0.052	0.222	-0.018	0.689	0.075	0.120	0.088	0.076	0.041	0.424
VitD w-RiskScore	-0.010	0.823	-0.035	0.445	-0.042	0.380	-0.058	0.244	-0.049	0.334
Combined RiskScore	0.044	0.310	-0.009	0.850	-0.010	0.836	0.006	0.904	0.020	0.694

†Models were adjusted for covariates as described in the legend of Table 6-5.

Table 6-12 | Associations between T1D genetic risk and gains of weight SDS or height SDS between birth and 12 months or birth and 24 months†.

Genetic Risk Score	Δ weight SDS					Δ height SDS			
	0-12 months		0-24 months			0-12 months		0-24 months	
	Beta	*P* value	Beta	*P* value		Beta	*P* value	Beta	*P* value
HLA									
HLA Risk_2level	0.007	0.866	0.000	1.000		-0.020	0.645	0.019	0.663
HLA Risk_3level	-0.010	0.818	0.030	0.480		0.011	0.793	0.036	0.416
Unweighted									
nonHLA RiskSum	0.047	0.255	0.012	0.786		-0.038	0.375	-0.022	0.620
VitD RiskSum	-0.019	0.643	-0.048	0.266		-0.057	0.187	-0.066	0.133
Weighted									
nonHLA w-RiskScore	0.060	0.153	0.043	0.316		-0.001	0.980	0.006	0.899
VitD w-RiskScore	-0.041	0.319	-0.068	0.111		-0.073	0.087	-0.069	0.122
Combined RiskScore	0.006	0.883	0.034	0.429		-0.002	0.960	0.033	0.457

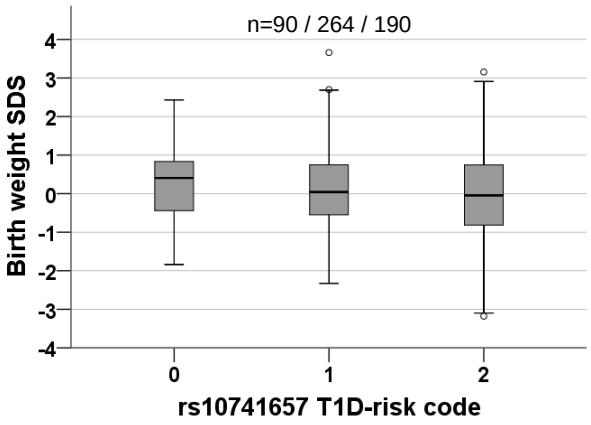
†Models were adjusted for covariates as described in the legend of Table 6-6.

6.3.5 Plots of birth weight by genotypes

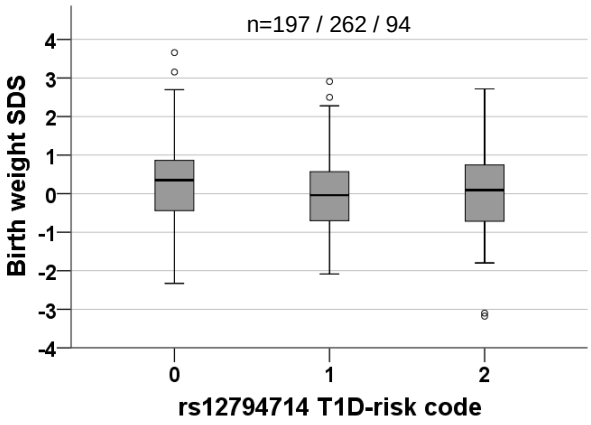
Box plots of birth weight by genotypes were plotted for significant associations (Figure 6-2).

Figure 6-2 | Weight SDS by T1D-risk genotypic group in a. *CYP2R1*, b. *CYP2R1*, c. *DLK1*, d. *TYK2*. Genotype frequencies are listed in graphs (0 / 1 / 2).

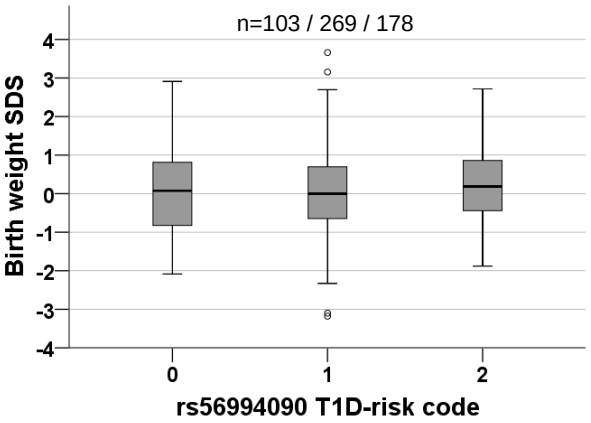
a. *CYP2R1*



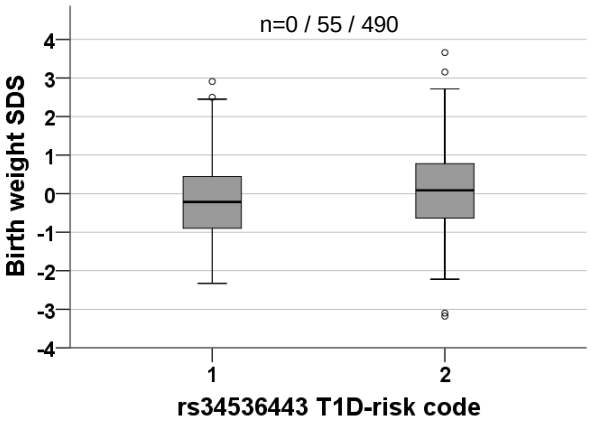
b. *CYP2R1*



c. *DLK1*



d. *TYK2*



6.4 DISCUSSION

6.4.1 Evaluation of the gene panel

In this study, a panel of T1D candidate genes and SNPs were biology-screened and prioritised as potential markers of altered early growth — a preponderant phenotype of the presymptomatic stage of juvenile-onset diabetes — which serves as a proxy for defects in pancreatic β-cells. My unbiased screening process was assessed by comparing the resulting gene panel with the make-up of two published T1D genetic risk scores. First, juxtaposed against a genetic risk score proposed as a predictor of β-cell function[510], seven out of eleven candidate genes selected by those authors were included in my panel: *TNFAIP3*, *IFIH1*, *GSDMB* (*ORMDL3*), *IL10*, *SH2B3*, *CTSH*, *INS*. Second, compared with the referent combined genetic risk score for identifying children at risk of T1D, and very recently applied to other cohorts including the TEDDY study[353,456,457], my 'combined' genetic risk score (vs. published) is composed of a total of 33 (vs.30) SNPs, and comprises 3 (vs. 2) SNPs tagging the high-risk HLA-II, 1 SNP (vs. 1) tagging the protective HLA-II, nil SNPs (vs. 2) tagging the high-risk HLA-I, 25 SNPs (vs. 25) at non-HLA loci, and 4 (vs. 0) SNPs at vitamin D metabolism loci. Within the non-HLA region, there is a high degree of shared candidate genes albeit different SNPs typed; I sourced the rsID from the contemporaneous fine mapping study. The most distinctive discrepancy between my SNP panel and the SNPs included in the published combined genetic risk score is the absence of HLA class I variants from my selection; the effect of HLA class I is small compared with class II (OR of HLA-B vs. DR3/DR4-DQ8 is 2.5 vs. 48[456]). In addition, HLA class I influences rate of progression to clinical disease, as opposed to initiation of autoimmunity which is a class II effect[122] and thus more relevant to the biological events explored in this thesis. Common to both published genetic risk scores but absent from my SNP panel was the *COBL*, a brain-enriched protein known to control neuronal morphology and development[511] but present in pancreatic islets[101]. Notably, numerous of the candidate genes in my panel (*IL27*, *BAD*, *PRKCQ*, *CLEC16A*, *ERBB3*, *CTSH*) were identified, by means of ranking the peak signal SNP on likely functional effect (RegulomeDB[512]), to be causal (or in LD with causal SNPs), allowing for a straightforward interrogation of their effect on T1D risk[31].

6.4.2 Genetic correlations with HLA SNPs

My study generated a wealth of data on genotypes, which passed the quality control assessments performed, and allowed for calculation of frequencies in search for genetic correlations. The incompatibility of homozygosity for the rs3135388 minor allele tagging the *HLA-DQB1*06:02* — which confers dominant protection against T1D[172] and predisposes to multiple sclerosis in Caucasians[459] — and homozygosity for the T1D high-risk *HLA-DR/DQ*

alleles merits further investigation and could indicate an evolutionary advantage. Phenome-wide association studies, whose goal is the identification of disease associations, could lend themselves to deciphering patterns between common infections in carriers of T1D high-risk vs. protective HLA types. The protective effect of the haplotype DRB1*15:01-DQB1*06:02 against T1D is long-range, spanning from the development of islet autoimmunity (e.g. lower likelihood of multiple autoantibodies) through all stages of progression (e.g. no metabolic worsening)[513].

In contrast to the DIPP study[119], I did not observe a higher prevalence of the secretor genotype in *FUT2* amongst carriers of the high-risk HLA genotype, possibly attributed to the lower frequencies of HLA-DR/DQ haplotypes in the CBGS compared with the Finnish cohort.

6.4.3 The sexual dimorphism of T1D genetic risk

The sexual dimorphism of the occurrence of homozygosity for the high-risk *HLA-DQ8* allele is striking (5 boys vs. 1 girl) and not observed for other equally rare variants (*HLA-DR3*, *PTPN22*) or rarer (*TYK2*, *ITGB7*). Of note, all these children but one, who had missing data for the *HLA-DR4*, were homozygous for the *HLA-DR4/DQ8* genotype, i.e. they also carried two *HLA-DR4* minor alleles. In the CBGS, the rs7454108 tagging the *HLA-DQ8* was in HWE and the MAFs were almost identical between boys and girls, which probably disqualifies DQ8 from being the culprit of sexual dimorphism observed in T1D. Nevertheless, it still instigates a discussion.

The non-synonymous variant at position 57 of HLA-DQB1 (DQ8) has long been known as a driver of T1D susceptibility, albeit not sufficient by itself to explain inheritance[514], and was shown by a phenotype-wide association study to have the stronger association with the disease than any other HLA variant of T1D[515]. That study also elucidated that the effects of HLA-DQB1 might stretch beyond pathogenesis; it had a stronger association with T1D complications (e.g. ketoacidosis, renal manifestation) than with the T1D phenotype[515]. I posit that sexual dimorphism, if any, of the occurrence of HLA-DQ8 predisposing genotypes, as seen in the CBGS, would reflect the relative incidence of the disease between men vs. women. In contrast to the female preponderance observed in most autoimmune diseases, prevalence of T1D by sex is near parity with country-specific gender bias[516], and controversy exists about the factors for the male to female excess in T1D in early adulthood[517]. As a comparator, the HLA-DQ8 is also a high-risk locus for Coeliac disease, which is twice as prevalent in women than men, but the DQ8 heterodimer is reportedly found in only 5-10% of these patients[518,519].

The interpretation of my observation is contrasted by prior reports restricting the bias in male incidence of T1D to the HLA-DR3 type[520]. Is there a biological plausibility for the male excess in T1D and islet autoimmunity? A linkage between the X chromosome and diabetes was speculated[520], but GWAS of T1D susceptibility loci identified only one understudied SNP

on the X chromosome with an OR of 1.16[438], which is inadequate to explain the gender bias. Another explanation offered almost two decades ago was based upon the higher BMI SDS and isolated GADA-positivity in male, but not female, first-degree relatives of T1D patients, conjuring on similarities with T2D and suggesting that the extra metabolic burden incurred by puberty and greater BMI precipitated progression to clinical onset in males[517]. A subsequent study in relatives of T1D showed that male sex conveyed increased risk for development of islet cell autoimmunity and multiple autoantibodies, but did not find a sex difference in the prevalence of GADA[521]. Of note, in the CBGS the BMI at birth, which points to genetic or intrauterine differences, was similar between boys and girls. On a molecular level, the HLA-DQ8 shows an enhanced binding of insulin peptides, leading to peripheral activation of autoreactive T cells[347,522]. The TEDDY study has recently shown that homozygosity for the *DR4-DQ8/DR4-DQ8* predominantly associates with the appearance of IAA as the first autoantibody in young children, whereas the *DR3-DQ2/DR3-DQ2* is strongly linked to the appearance of GADA as the first autoantibody in older children[115]. It appears that evidence exists to support a putative link between HLA-DR3 and GADA-positivity as initially suggested.

Whether or not the sex-dimorphic occurrence of HLA, or associated autoantibodies, exists, this discussion surfaces the observation of the sex-dimorphism of birth weight, i.e. males have a higher birth weight vs. girls (a mean of 140 g larger in the CBGS, $P<0.0001$). Could the reason for the association between male sex and larger birth weight provide a seed to our understanding of the pathway that connects risk of T1D and increased birth weight? The 'gender insulin hypothesis' proposes that boys at birth have lower insulin levels than girls, which accounts for intrinsically lower insulin resistance and higher birth weight[523]. However, a lower insulin secretion in male vs. female infants does not reconcile with the theory of the hypersecreting stressed β-cell[361,387], which is more prone to cytokine-induced damage[104,388] and sets off the injurious processes that lead to T1D. In addition, the DIPP study reported that whereas insulin secretion is perturbed in the long presymptomatic stage of the disease, insulin sensitivity remains unaffected[148]. The debate is unresolved, but the differences between male and female immune systems, which translate to sex-dimorphic resilience to infectious and non-infectious diseases[516], calls for the investigation of the effect of male sex on T1D.

6.4.4 T1D SNPs and size at birth

6.4.4.1 HLA

T1D strongly associates with the HLA-DR3 and HLA-DR4/DQ8 haplotypes. In Scandinavia, almost 90% of children diagnosed with T1D have at least one of these haplotypes[524,525]. On these grounds, studies in Scandinavia and elsewhere explored the possibility that the HLA

confers susceptibility to T1D via its association with higher birth weight, but the results have been inconsistent[14,15,89,161,244,245,404,405]. In contrast, studies consistently found that the protective HLA types were associated with higher birth weight[14,161,403,203,368,404], but such direction of effect does not reconcile with the notion that low birth weight is protective against T1D. Notwithstanding these findings, I posit that the familial aggregation found between the conditions of IUGR, which results in small size at birth, and preeclampsia[526], which was previously associated with HLA-DQB1[160], makes a strong case for the involvement of genetic factors, namely HLA, in early growth. Nevertheless, the results in the CBGS did not provide support for this hypothesis. An explanation for the lack of association might rest with the HLA region. Larsen and Alper[36] had long expressed the view that the *DRB1*/*DQB1* alleles are simply markers for T1D susceptibility but these loci are not themselves the susceptibility genes.

6.4.4.2 Summary of associations with birth weight

The significant SNP associations with birth weight found in the CBGS (rs10741657 and rs12794714 in *CYP2R1*, *P*=0.022 and *P*=0.043 respectively; rs56994090 in *DLK1*, *P*=0.018; rs34536443 in *TYK2*, *P*=0.024) are borderline and would fail the correction for multiple testing. Moreover, none of the SNPs here features amongst the 60 loci that were associated with birth weight in the mutli-ancestral GWAS meta-analysis in 153,781 subjects[336].

6.4.4.3 *CYP2R1*

In the CBGS, both SNPs mapped to *CYP2R1*, which encodes the 25-hydroxylase that catalyses the formation of 25(OH)D in the liver, showed a borderline negative association with birth weight, with the rs10741657 exhibiting a SNP dose-dependent effect (Figure 6-2). A prior study established that the minor allele for the rs10741657 is protective against T1D and associates with increased levels of 25(OH)D, whereas the rs12794714 is predisposing and associates with reduced levels of 25(OH)D[77]. Taken together, variants at this locus that associate with predisposition to T1D and lower 25(OH)D levels, also associate with reduced birth weight. Recently, a study in Denmark was the first to measure levels of 25(OH)D from neonatal DBS in 2,868 subjects and report a borderline inverted U-shaped association with birth weight[527]. The findings were in agreement with a prior study that measured cord blood levels of 25(OH)D in Chinese neonates, whose authors interpreted the U-shaped relationship as vitamin D having a dual role or conferring no benefit on foetal growth when in excess[528]. Reports on the association between SGA and blood 25(OH)D levels have been conflicting[527] but could be addressed by the novel finding here which isolates the *CYP2R1*, particularly the rs10741657, as the culprit of the relationship between 25(OH)D levels and foetal growth (variants in *GC*, *DHCR7* or *CYP27B1* did not associate with birth weight in this study). On these grounds, alterations in levels of 25(OH)D arising from the rs12785878 polymorphism in

DHCR7 — shown to associate with serum 25(OH)D concentrations at GWAS significance level[529] — would unlikely show an effect on birth weight. Whether CYP2R1 is on the causal pathway of foetal growth, or the relationship found here is an association arising from pleiotropy in the network involving a growth factor (e.g. IGF-1, IGF-2) in the liver, is a question to consider.

6.4.4.4 *DLK1*

The *DLK1* is the only shared candidate gene, apart from *INS/IGF2*, between my study and the GWAS meta-analysis for birth weight[336]. However, the GWAS strong significant signal rs6575803 ($P<0.5x10^{-8}$) near *DLK1*[336] is not in LD with the rs56994090 ($r^2=0.01$) of my SNP panel. The *DLK1* gene on Chr14 encodes a developmental protein, interchangeably known as delta homolog 1 or delta like non-canonical Notch ligand 1, which is a member of the EGF-like family of homeotic proteins[530]. Maternal circulating DLK1 correlated with embryonic mass in mice and differentiated between healthy SGA vs. pathologically small newborns in a human cohort[531]. Prior to the fine mapping study of T1D susceptibility loci, the rs941575 SNP was the GWAS hit that was associated with risk of T1D and mapped to the DLK1-MEG3 region. Paternal transmission of the T1D protective G allele on the rs941576 (in linkage with the index rs56994090, $r^2=0.91$) was previously associated with lower birth weight ($P=0.01$) and reduced head circumference ($P=0.01$) in a British birth cohort[18]. My study did not show an association with head circumference. Notably, doubts are cast on whether the *DLK1* is a T1D candidate gene due to the absence of functional evidence of its involvement in immune functions[102].

6.4.4.5 *TYK2*

The rs34536443 (G>C), a missense variant (Pro1104Ala) at the *TYK2* locus on Chr19p13 which confers protection against T1D, was associated with lower birth weight in the CBGS, which reconciles with the association between protection against T1D and small size at birth. *TYK2* encodes tyrosine kinase 2 (TYK2), one of four members of the Janus family of intracellular nonreceptor tyrosine kinases (JAKS), which serve essential roles in the intracellular signalling of several cytokines and type I IFNs through the phosphorylation and activation of signal transducers and activators of transcription (STATs)[532]. Experimental studies ascribed a 'self-defense' role to TYK2, explaining its protective effect against autoimmune attack of the islets, by showing that TYK2 inhibition decreases double-stranded RNA- (a non-physiological form of RNA) induced apoptosis in human β-cells via reduction of the type I IFN pathway activation, and downstream decrease in β-cell expression of HLA class I proteins (a mechanism that renders β-cells 'visible' to the immune system[533]) and CXCL10 (a chemokine that recruits T cells to the islet)[534]. The rs34536443 has been identified as the only SNP in *TYK2* with a demonstrable protective impact against autoimmunity, confirming that homozygosity of the rs34536443 minor allele is driven by a loss of TYK2 function[535].

The question that ensues revolves around the possible mechanism underlying the influence of the rs34536443 on birth weight. If inhibition of TYK2, which is protective against T1D, averts the deployment of an immune response against the virally-infected β-cells, insulin secretion would be uncompromised and birth weight would remain at normal values, which contradicts the established view of growth alterations in at-risk children. Another possible explanation is that inhibition of TYK2 allows for replication of viruses in the β-cells, eventually killing the host cells. However, my finding is drawn from the general population of infants and not a cohort of prediabetic children who might have a higher frequency of β-cell virus-residing infections. The remaining explanation for the association between reduced foetal growth and the polymorphism resulting in inhibition of TYK2 is via interactions with growth factors. Almost 20 years ago, an experimental study showed that TYK2 was associated with the GHR in the human liver, possibly modulating GHR signalling[536]. Hence, a likely mechanism behind the influence of TYK2 on birth weight is that TYK2 modulates GHR signalling, which, on the basis of GH regulating transcription of IGF-1[537] and IGF-2[538] in the human liver, influences production of insulin-like growth factors, which in turn regulate intrauterine growth. An alternative scenario is an as-yet unidentified role of TYK2 in insulin receptor signalling. The latter hypothesis arises from the recent finding that the rs34536443 is one of few T1D susceptibility SNPs that reached nominal association (P<0.05) with T2D (OR=1.12, [95% CI 1.04, 1.22]) in a European ancestry meta-analysis of GWAS data in 26,676 T2D patients and 132,532 controls[495].

Speculations about the underlying mechanisms aside, the long-standing correlation between low birth weight and risk of infectious disease mortality[187,188] and hospitalisation[189] resonates with the functional effects of TYK2. Rare mutations that result in TYK2 deficiencies are clinically manifested by primary immunodeficiencies, including the hyperIgE syndrome, which shows increased susceptibility to infections[539]. In the CBGS, there was only one homozygous for the minor rs34536443 allele out of 573 infants genotyped for this variant, a girl born small, implying a selective advantage of *TYK2* and a possible common genetic background to the condition of IUGR and susceptibility to infections in neonates.

6.4.5 T1D SNPs and postnatal size

The association found here between height and *HLA-DR3* at the end of infancy (P=0.021) supports the notion that accelerated linear growth in prediabetics is mediated by the high-risk HLA (Chapter 2). This study is strengthened by analysing SNPs tagging the T1D high-risk HLA haplotypes individually, thus disentangling the HLA effects, which is critical considering the 'heterogeneity between the aetiology of DR4-associated and DR3-associated diabetes'[520].

The negative associations found between BMI at birth and 12 months and the SNP in *PRKCQ* surfaces the long-standing debate of the role, if any, of BMI in T1D development. For

example, a previous study in a Swedish diabetic cohort identified that the association between higher BMI and increased predisposition to T1D was mediated by the low-risk HLA-DQ types, A1*05:01-B1*02:01[540]. A study in first-degree relatives of T1D patients reported an association between GADA and higher BMI SDS in males only[517]. It is thus reasonable to query if *PRCQK* — which encodes protein kinase C theta with important roles in cytokine secretion from cytotoxic T cells[541] — associates with islet autoantibodies. This would be essential to explore considering that *PRKCQ* is deemed a candidate gene based on its functional effects[31].

Strong significant associations were identified between SNPs in putative causal genes with established effects on the immune system, exemplified by the *SH2B3* and *BAD*, and measures of skinfold thickness in infancy. These findings warrant investigation in longitudinal analysis (Chapter 7) and in relation to endocrine factors to probe for mechanisms (Chapter 8). A recent study has detected a significantly greater co-occurrence of T1D and coeliac disease autoimmunities in children who had T1D family history, HLA-DR3/4 haplotypes and the SNP rs3184504 (in perfect linkage with the rs653178) at the SH2B3 locus[542]. In the CBGS, there was no evidence of co-occurrence between the *SH2B3* alleles and the T1D high-risk HLA-DR3/4 genotype, but the SNPs tagging the HLA-DR4 and HLA-DQ8 were also associated with skinfold thickness, indirectly conveying a possible link between *SH2B3* and *DRB1/DQB1*.

6.4.6 Genetic risk scores and early growth

As most T1D SNPs have modest effects on disease risk, a combination of variants would best capture effects of clinical relevance and allow for interactions in a common network[510]. In the current study, I investigated the aggregate impact of SNPs via genetic risk scores, identifying a borderline association between the HLA risk score and skinfold thickness (P=0.022). As discussed in Chapter 2, one study found no role of the adipose tissue in T1D development, but it was conducted in only 26 T1D cases[357]. The association found here is driven by the SNPs tagging the *HLA-DR4* and *DQ8*, evidenced by the weaker association with the genetic risk score vs. individual SNPs, *DR4* (P=0.003) or *DQ8* (P=0.017), suggesting that the cumulative effect is 'diluted' by incorporating a SNP, namely *HLA-DR3*, of no robust influence on adiposity.

6.4.7 Limitations

There is a limitation worthy of mention. Albeit the advantages of a background population of infants, the frequencies of the T1D high-risk HLA loci are too low in the general population to allow for conclusive findings. Nevertheless, the strong significance attained for associations with SNPs that tag the HLA is indicative of robust, yet differentiating, HLA effects on growth, namely *DR3* on height and *DR4-DQ8* on adiposity, which merit replication in larger cohorts.

Chapter 7

T1D LOCI AND GROWTH: A LONGITUDINAL VIEW

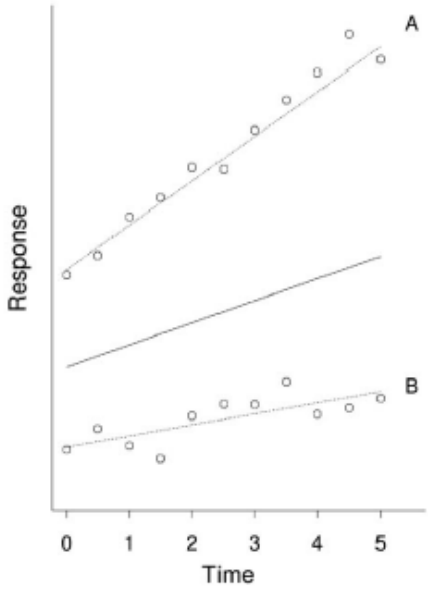
7.1 CONTEXT

Longitudinal study design, comprising consecutive measurements on the same individual over time, has become increasingly popular for examining trends in clinical outcomes as it provides the knowledge base for the natural history and development of the disease, and answers clinically relevant questions with higher precision than a study of simpler design[543]. The main goal of a longitudinal study is the characterisation of changes in a recurrent event over time, in contrast to a cross-sectional study which assesses the response of interest at a single point in time[543-545]. Longitudinal analysis has enabled the investigation of genetic variation influencing phenotype values over time[546] offering clear advantages. Whereas a cross-sectional study based on the baseline value can only capture the mean differences of a trait across genotypic groups of subjects, a longitudinal study additionally estimates the rates of change in the trait in the genotypic groups[547]. Despite the complexity arising from following subjects over time and measuring change during the follow-up, longitudinal analysis has lent itself to medical studies characterised by a long clinical or preclinical development, including the Multicentre AIDS Cohort Study[548], Alzheimer's disease[547] and obesity[549]. As national funding agencies are becoming particularly interested in studies with longitudinal designs, new (longitudinal) statistical techniques that analyse the whole pattern of recurrent events over time have replaced traditional (naive) techniques which use only one observation per subject[543,545].

The distinctive features of longitudinal studies that complicate their analysis include i) correlated observations (knowing the value of the response on one occasion is predictive of the likely value on a future occasion), ii) heterogeneous variability (the variance of the response is not constant over the duration of the study), iii) unbalanced design (due to mistimed measurements), iv) missing data (due to skipped visits or study dropouts), v) existence of multiple covariates[543,544]. The first two features violate the fundamental assumptions of independence and homogeneity of variance required for standard statistical techniques, such as ANOVA[544]. Longitudinal techniques have overcome this limitation by correcting for the dependency of observations within a subject. Specifically, linear mixed effects model (MEM)

is a flexible advanced statistical method that accommodates all the distinct features of longitudinal study designs by assuming natural heterogeneity among subjects, i.e. each subject has its own mean response trajectory in terms of their initial level of response (subject-specific intercept at baseline), and possibly, in terms of the rate of change in the response over time (slope)[544]. In practice, the mixed effects model the mean response based on population characteristics which are assumed to be shared by all individuals ('fixed' effects) and subject-specific effects that are unique to each individual ('random' effects). Figure 7-1 shows a graphic representation of a population-based trajectory and subject-specific trajectories with measurement errors above and below the lines that allow for the response on any occasion to vary randomly[544]. In essence, the introduction of random effects is a means of modelling the correlation and variability among repeated measures[543,544] and is interpreted as the difference between the population-averaged intercept (or slope) and the subject-specific intercept (or slope). In this way, the estimation of the single measure of the variance of intercepts obviates the estimation of a different intercept per subject, which would be practically impossible[545].

Figure 7-1 | Graphic representation of mean response trajectory for each of two individuals (A, B) and mean response trajectory averaged across all individuals. From Fitzmaurice and Ravichandran[544].



Here, I applied a linear MEM to evaluate associations between the longitudinal measures of anthropometric phenotypes in the CBGS and SNPs known to robustly associate with the development of T1D. The premise of this analysis is the assumption of a long preclinical stage of the disease, which is thought to start early in life and inflict injurious processes on β-cells manifested by alterations in growth.

7.2 METHODS

7.2.1 Study design and measures

The details of the CBGS design and subjects, SNP selection and genotyping, genetic scores, quality controls and outcome measures were described in Chapter 6.

7.2.2 Statistical analysis

7.2.2.1 Variables

Longitudinal analyses relating growth rates from birth to 24 months of age (explained variable) and individual T1D susceptibility SNP or genetic risk score (explanatory variable) were performed using multivariate linear regression with adjustment for preselected covariates (Table 3-21) by co-entering those variables in the model. Additive genetic effects were assumed for the analysis of individual SNPs.

7.2.2.2 Observations

Missing data is a distinct characteristic of longitudinal analysis which may bias estimates of change over time when subjects with complete data are not a representative sample from the target population[544]. In contrast to repeated measures ANOVA, which conducts a 'complete case analysis' by excluding subjects with one or more missing measurements from the analysis[544], advanced longitudinal techniques make use of all observations for each subject. Consequently, performing such kind of analysis necessitated a long data structure, whereby each row represented one observation time per case. Analyses were restricted to 586 children after excluding extremes of growth trajectories (maternal T1D, <36 weeks of gestation, twins). Out of the 586 babies included, 381 had complete data for weight SDS at all visits, 79 for four visits, 28 for three visits, 46 for two visits, and 52 for only one visit. For height SDS, 365 babies had complete data for five visits, 87 for four visits, 35 for three visits, 43 for two visits, 55 for one visit, and 1 for none. For head circumference SDS, 374 babies had complete data for four visits, 89 for three visits, 30 for two visits, 89 for one visit and 4 for none. For BMI SDS, 371 babies had complete data for four visits, 87 for three visits, 34 for two visits, 90 for one visit and 4 for none. For overall skinfold thickness SDS, 371 babies had complete data for five visits, 84 for four visits, 31 for three visits, 43 for two visits, 54 for one visit and 1 for none of the visits.

7.2.2.3 Covariance structure

Two common features of longitudinal data are the correlation between repeated measurements on a subject and the heterogeneous variability of the response variable across the population over time. The term 'covariance' captures both characteristics[544]. Advanced statistical techniques allow for a variety of covariances to be modelled. A common structure is the 'compound symmetry' (or exchangeable), which indicates the same correlations between all pairs of measurements (sphericity assumption). Alternatively, we could model the correlation of the repeated measurements for each subject in different ways. For example, the 'unstructured' covariance is the most liberal as it allows for every term to be different, but this comes at the cost of using the most parameters[550]. However, as the sample size in this study is not adequate to fit such a covariance structure, longitudinal analyses were performed without modelling the covariance which makes use of fewer parameters.

7.2.2.4 Mixed effects model

A linear MEM with random intercept and fixed slope (Equation 7-1) was used because it captures the interindividual variation at baseline, i.e. birth (random intercept) and assumes identical growth rates (fixed slope). My choice for the model structure is reinforced by the suggestion that a mean-based statistic is more powerful to detect a genetic association than a slope-based statistic (e.g. random slope)[551].

Equation 7-1

$$Y_{ik} = (\beta_0 + U_{0i}) + \beta_1 \times t_{ik} + \beta_2 \times SNP_i + \beta_3 (SNP_i \times t_{ik}) + \sum_{j=1}^{m} \beta_j \times X_{ij}$$

for study participant i=1, 2,…n; k is timepoint, U_{0i} is random intercept, X is one of m covariates.

The aim of the longitudinal analysis is to test for no interaction effects, i.e. the growth trajectories are parallel between genotypic groups of subjects. The null hypothesis tested was H_0: $\beta_3=0$ vs. H_1: $\beta_3\neq0$. Linear MEMs were generated using the IBM Statistical Package for Social Sciences (SPSS 23.0, Chicago, IL, USA) for Windows and fitted using maximum likelihood with number of maximum iterations of 150. Tests were two-tailed and a $P\leq0.05$ was considered statistically significant.

7.3 RESULTS

7.3.1 Descriptive statistics

Table 7-1 presents the mean and SD of the anthropometric phenotypes used in the analyses.

Table 7-1 | Mean (SD) of standardised growth parameters by age of infants included in the analysis.

	Weight SDS	Height SDS	HC SDS	BMI SDS	SFT SDS
Birth	0.10 (1.02)	-0.05 (0.97)	-0.13 (0.99)	-0.00 (1.26)	0.08 (0.90)
3 months	0.00 (1.04)	0.16 (0.96)			0.08 (0.75)
12 months	0.06 (1.08)	0.36 (1.01)	-0.63 (1.10)	-0.20 (1.04)	0.11 (0.81)
18 months	0.08 (1.01)	0.32 (1.05)	-0.76 (1.05)	-0.20 (0.93)	0.08 (0.76)
24 months	0.17 (1.00)	0.45 (1.01)	-0.75 (1.05)	-0.22 (0.91)	0.11 (0.82)

SFT, skinfold thickness

7.3.2 T1D SNPs and early growth

Table 7-2 displays the summary statistics — estimate of the interaction between the SNP and age of infant, standard error (SE) and *P* value — of linear multivariate regressions of longitudinal standardised growth parameters collected from birth to 3 months on individual SNPs. Similarly, Table 7-3 presents the summary statistics of longitudinal analysis of repeated measures from birth to 24 months. Weight and height models were adjusted for parity, smoking during pregnancy, maternal height, maternal pre-pregnancy weight and type of milk feeding at age 3 months. Head circumference models were adjusted for parity, maternal height and maternal pre-pregnancy weight. BMI models were adjusted for parity, smoking during pregnancy, maternal pre-pregnancy BMI and type of milk feeding at age 3 months. Skinfold thickness models were adjusted for parity and type of milk feeding at age 3 months.

With additive genetic effects, the estimate of the interaction is interpreted as the per-allele change in the response variable per unit time. There was evidence that the rate of change in skinfold thickness SDS over the first 3 months of life is significantly influenced by the polymorphisms in *GC* (*P*=0.008) and *SH2B3* (*P*=0.003). When analysing the whole development over infancy (birth to 24 months), the longitudinal measures of skinfold thickness showed strong significant associations with T1D susceptibility SNPs in *BAD* ($P=2.4\times10^{-4}$) and *SH2B3* ($P=7.0\times10^{-6}$). The rate of development of weight and height in infancy were significantly influenced by variants in *CLEC16A* and *IFIH1*. The SNP tagging the high-risk *HLA-DR3* significantly influenced linear growth over infancy.

Table 7-2 | Summary statistics of T1D SNP-age interactions for longitudinal measures of SDS for weight, height and skinfold thickness (SFT) between birth and 3 months of age†.

	SNP	Gene	Weight SDS			Height SDS			SFT SDS		
			Estimate	SE	*P* value	Estimate	SE	*P* value	Estimate	SE	*P* value
1	rs2476601	*PTPN22*	0.016	0.130	0.905	0.053	0.108	0.625	0.091	0.121	0.449
2	rs3024505	*IL10*	-0.030	0.111	0.788	0.052	0.094	0.581	0.094	0.103	0.358
3	rs2111485	*IFIH1*	0.084	0.075	0.264	-0.007	0.064	0.918	0.095	0.072	0.184
4	rs3087243	*CTLA4*	-0.080	0.075	0.283	-0.068	0.063	0.285	-0.077	0.070	0.273
5	rs4588	*GC*	0.190	0.078	0.016	-0.039	0.066	0.561	0.199	0.075	0.008
6	rs75793288	*IL2*	-0.111	0.078	0.153	-0.007	0.066	0.912	-0.143	0.073	0.052
7	rs17426593	*HLA-DR4*	-0.008	0.096	0.932	-0.086	0.081	0.287	0.062	0.092	0.500
8	rs2187668	*HLA-DR3*	-0.024	0.105	0.821	-0.018	0.089	0.844	-0.129	0.100	0.197
9	rs7454108	*HLA-DQ8*	0.058	0.126	0.646	-0.136	0.100	0.178	0.033	0.119	0.782
10	rs3135388	*HLA-DQB1*06:02*	-0.043	0.106	0.682	0.002	0.091	0.979	-0.154	0.101	0.127
11	rs72928038	*BACH2*	0.142	0.099	0.154	0.013	0.084	0.877	-0.036	0.095	0.707
12	rs6920220	*TNFAIP3*	0.068	0.089	0.444	0.021	0.075	0.779	-0.066	0.083	0.423
13	rs6476839	*GLIS3*	0.078	0.074	0.295	0.063	0.063	0.318	-0.046	0.071	0.518
14	rs61839660	*IL2RA*	0.243	0.117	0.038	0.146	0.099	0.139	0.158	0.112	0.158
15	rs11258747	*PRKCQ*	0.013	0.085	0.881	-0.150	0.071	0.035	-0.093	0.081	0.252
16	rs12416116	*RNLS*	-0.016	0.078	0.836	-0.019	0.067	0.781	0.096	0.076	0.210
17	rs10741657	*CYP2R1*	0.184	0.074	0.013	0.023	0.063	0.718	0.006	0.072	0.930
18	rs12794714	*CYP2R1*	0.087	0.072	0.229	-0.047	0.061	0.442	-0.103	0.070	0.140
19	rs689	*INS*	-0.104	0.081	0.202	-0.049	0.069	0.474	0.048	0.078	0.538
20	rs694739	*BAD*	0.117	0.073	0.112	0.057	0.062	0.362	0.024	0.069	0.729
21	rs12785878	*DHCR7*	-0.028	0.083	0.733	-0.103	0.070	0.143	-0.050	0.079	0.524
22	rs11170466	*ITGB7*	0.138	0.168	0.409	-0.207	0.141	0.143	-0.111	0.155	0.475
23	rs705704	*ERBB3*	0.054	0.075	0.470	0.011	0.063	0.865	0.021	0.073	0.772
24	rs10877012	*CYP27B1*	-0.073	0.077	0.343	-0.005	0.065	0.944	-0.055	0.074	0.457
25	rs653178	*SH2B3*	0.103	0.070	0.143	0.117	0.059	0.049	0.198	0.067	0.003
26	rs56994090	*DLK1*	-0.067	0.075	0.366	-0.014	0.063	0.825	-0.070	0.070	0.316
27	rs34593439	*CTSH*	0.134	0.111	0.227	-0.134	0.094	0.153	0.031	0.105	0.768
28	rs12927355	*CLEC16A*	-0.019	0.076	0.805	0.049	0.064	0.451	-0.028	0.073	0.703
29	rs151234	*IL27*	-0.055	0.109	0.612	-0.041	0.093	0.663	0.027	0.102	0.789
30	rs12453507	*ORMDL3*	0.103	0.075	0.169	0.040	0.063	0.527	0.039	0.072	0.581
31	rs1893217	*PTPN2*	0.050	0.094	0.596	0.104	0.079	0.186	0.015	0.091	0.869
32	rs34536443	*TYK2*	-0.112	0.162	0.490	0.126	0.137	0.359	-0.287	0.155	0.065
33	rs516246	*FUT2*	-0.011	0.074	0.879	-0.061	0.062	0.324	0.017	0.069	0.809
34	rs11203202	*UBASH3A*	-0.113	0.074	0.127	-0.076	0.062	0.221	-0.097	0.071	0.172

†Models were adjusted for covariates as described in Section 7.3.2 by growth parameter.

Table 7-3 | Summary statistics of T1D SNP-age interactions for longitudinal measures of SDS for weight, height, HC, BMI and skinfold thickness (SFT) between birth and 24 months of age†.

	SNP	Gene	Weight SDS			Height SDS			HC SDS			BMI SDS			SFT SDS		
			Estimate	SE	*P* value	Estimate	SE	*P* value	Estimate	SE	*P* value	Estimate	SE	*P* value	Estimate	SE	*P* value
1	rs2476601	*PTPN22*	0.046	0.028	0.096	0.027	0.027	0.322	0.032	0.036	0.369	0.035	0.048	0.466	0.046	0.026	0.071
2	rs3024505	*IL10*	0.027	0.024	0.267	0.043	0.024	0.076	-0.008	0.032	0.798	0.004	0.042	0.918	0.055	0.022	0.013
3	rs2111485	*IFIH1*	0.042	0.016	0.008	0.040	0.015	0.009	0.051	0.021	0.014	0.026	0.027	0.350	0.008	0.015	0.598
4	rs3087243	*CTLA4*	-0.019	0.016	0.234	0.005	0.016	0.732	-0.009	0.021	0.682	-0.043	0.028	0.121	-0.001	0.015	0.928
5	rs4588	*GC*	0.007	0.017	0.666	0.015	0.016	0.363	0.034	0.022	0.124	-0.010	0.029	0.736	0.009	0.016	0.556
6	rs75793288	*IL2*	0.002	0.016	0.898	-0.010	0.016	0.525	0.013	0.021	0.524	0.002	0.028	0.949	0.005	0.015	0.722
7	rs17426593	*HLA-DR4*	-0.011	0.020	0.571	-0.028	0.020	0.158	-0.057	0.027	0.033	0.009	0.035	0.800	0.025	0.019	0.191
8	rs2187668	*HLA-DR3*	0.039	0.022	0.078	0.058	0.022	0.007	0.074	0.029	0.011	-0.010	0.038	0.787	-0.044	0.021	0.034
9	rs7454108	*HLA-DQ8*	-0.021	0.028	0.456	-0.068	0.027	0.012	-0.060	0.037	0.103	0.072	0.048	0.137	0.025	0.026	0.329
10	rs3135388	*HLA-DQB1*06:02*	0.022	0.022	0.331	0.026	0.022	0.229	-0.010	0.030	0.725	0.028	0.039	0.476	-0.039	0.021	0.062
11	rs72928038	*BACH2*	0.009	0.021	0.678	0.008	0.021	0.712	-0.039	0.028	0.169	0.016	0.038	0.672	-0.017	0.020	0.398
12	rs6920220	*TNFAIP3*	-0.009	0.019	0.653	0.009	0.019	0.655	-0.016	0.025	0.540	0.022	0.034	0.528	-0.036	0.018	0.044
13	rs6476839	*GLIS3*	-0.006	0.016	0.723	0.019	0.016	0.214	-0.046	0.021	0.024	-0.036	0.028	0.196	-0.027	0.015	0.075
14	rs61839660	*IL2RA*	0.021	0.025	0.395	0.017	0.025	0.488	0.058	0.032	0.071	0.045	0.044	0.304	0.034	0.024	0.152
15	rs11258747	*PRKCQ*	0.025	0.018	0.158	-0.002	0.018	0.900	0.027	0.023	0.242	0.063	0.031	0.044	-0.018	0.017	0.298
16	rs12416116	*RNLS*	-0.027	0.017	0.104	-0.016	0.017	0.342	-0.038	0.022	0.087	-0.035	0.029	0.230	0.013	0.016	0.402
17	rs10741657	*CYP2R1*	0.027	0.016	0.086	0.014	0.016	0.356	0.002	0.021	0.912	0.063	0.028	0.023	0.014	0.015	0.348
18	rs12794714	*CYP2R1*	0.016	0.015	0.298	-0.021	0.015	0.162	0.002	0.020	0.906	0.069	0.027	0.010	0.006	0.015	0.667
19	rs689	*INS*	0.016	0.017	0.336	0.014	0.017	0.422	-0.018	0.023	0.435	-0.027	0.030	0.365	0.035	0.016	0.033
20	rs694739	*BAD*	-0.031	0.016	0.044	-0.013	0.015	0.400	0.002	0.020	0.920	-0.009	0.027	0.746	-0.053	0.014	<0.0001
21	rs12785878	*DHCR7*	-0.040	0.018	0.026	-0.019	0.018	0.280	-0.045	0.023	0.052	-0.030	0.031	0.341	-0.013	0.017	0.447
22	rs11170466	*ITGB7*	0.019	0.036	0.596	-0.004	0.036	0.906	0.014	0.047	0.762	0.034	0.064	0.589	-0.004	0.033	0.899
23	rs705704	*ERBB3*	0.014	0.016	0.392	-0.005	0.016	0.741	-0.007	0.021	0.745	0.054	0.028	0.056	0.016	0.016	0.318
24	rs10877012	*CYP27B1*	-0.028	0.016	0.087	-0.015	0.016	0.366	-0.051	0.022	0.018	-0.030	0.028	0.292	-0.014	0.016	0.386
25	rs653178	*SH2B3*	0.037	0.015	0.012	0.036	0.015	0.012	0.006	0.019	0.752	0.000	0.026	0.993	0.065	0.014	<0.0001
26	rs56994090	*DLK1*	-0.025	0.016	0.111	-0.034	0.016	0.031	0.029	0.021	0.156	-0.023	0.027	0.398	0.000	0.015	0.979
27	rs34593439	*CTSH*	-0.014	0.024	0.544	-0.014	0.024	0.552	0.022	0.031	0.477	-0.024	0.042	0.574	0.014	0.022	0.538
28	rs12927355	*CLEC16A*	-0.045	0.016	0.006	-0.055	0.016	0.001	-0.027	0.022	0.220	-0.059	0.028	0.039	0.001	0.016	0.945
29	rs151234	*IL27*	-0.032	0.023	0.173	-0.051	0.023	0.029	0.006	0.030	0.849	0.004	0.041	0.917	-0.008	0.021	0.725
30	rs12453507	*ORMDL3*	0.031	0.016	0.054	0.024	0.016	0.125	0.035	0.021	0.088	0.042	0.028	0.137	0.011	0.015	0.456
31	rs1893217	*PTPN2*	-0.017	0.020	0.404	-0.019	0.020	0.336	0.000	0.027	0.992	0.025	0.035	0.479	-0.005	0.020	0.805
32	rs34536443	*TYK2*	-0.004	0.034	0.898	-0.025	0.033	0.438	-0.002	0.043	0.957	-0.012	0.058	0.834	-0.016	0.032	0.617
33	rs516246	*FUT2*	0.005	0.016	0.729	-0.022	0.015	0.150	0.008	0.021	0.711	0.032	0.027	0.244	0.037	0.015	0.012
34	rs11203202	*UBASH3A*	-0.006	0.016	0.687	-0.012	0.015	0.457	-0.037	0.020	0.067	-0.014	0.027	0.603	-0.017	0.015	0.272

†Models were adjusted for covariates as described in Section 7.3.2 by growth parameter.

Tables 7-4 and 7-5 display the summary statistics — estimate of the interaction between the SNP and age of infant, standard error (SE) and *P* value — of the T1D SNP-age interaction from multivariate linear regressions of longitudinal measures of growth parameters, taken from birth to 3 months and from birth to 24 months respectively, on genetic risk scores after adjusting for covariates as described for individual SNPs. There was evidence that the development of skinfold thickness in infancy is significantly influenced by the non-HLA genetic risk score constructed from 29 T1D susceptibility SNPs.

Table 7-4 | Summary statistics of T1D genetic risk score-age interactions for longitudinal measures of SDS for weight, height and skinfold thickness (SFT) between birth and 3 months of age†.

Genetic Risk Score	Weight SDS			Height SDS			SFT SDS		
	Estimate	SE	*P* value	Estimate	SE	*P* value	Estimate	SE	*P* value
HLA									
HLA Risk_2level	0.111	0.215	0.604	-0.107	0.179	0.550	0.048	0.197	0.807
HLA Risk_3level	0.015	0.085	0.855	-0.058	0.071	0.416	-0.066	0.080	0.405
Unweighted									
nonHLA RiskSum	1.178	0.911	0.197	-0.521	0.785	0.508	-0.050	0.901	0.956
VitD RiskSum	0.320	0.310	0.303	-0.328	0.261	0.210	-0.115	0.295	0.696
Weighted									
nonHLA w-RiskScore	0.949	3.360	0.778	-0.950	2.896	0.743	2.629	3.271	0.422
VitD w-RiskScore	-1.150	2.856	0.688	-1.440	2.412	0.551	-1.910	2.712	0.482
Combined RiskScore	-0.003	0.025	0.897	-0.014	0.022	0.511	-0.026	0.024	0.278

†Models were adjusted for covariates as described in Section 7.3.2 by growth parameter.

Table 7-5 | Summary statistics of T1D genetic risk score-age interactions for longitudinal measures of SDS for weight, height, HC, BMI and skinfold thickness (SFT) between birth and 24 months of age†.

Genetic Risk Score	Weight SDS			Height SDS			HC SDS			BMI SDS			SFT SDS		
	Estimate	SE	*P* value	Estimate	SE	*P* value	Estimate	SE	*P* value	Estimate	SE	*P* value	Estimate	SE	*P* value
HLA															
HLA Risk_2level	0.065	0.045	0.145	0.080	0.044	0.065	0.089	0.060	0.138	0.032	0.077	0.677	0.041	0.041	0.317
HLA Risk_3level	0.026	0.018	0.146	0.015	0.017	0.391	0.028	0.024	0.234	0.033	0.031	0.279	-0.013	0.017	0.422
Unweighted															
nonHLA RiskSum	-0.100	0.197	0.611	-0.216	0.195	0.268	-0.138	0.264	0.601	0.042	0.347	0.905	0.167	0.193	0.388
VitD RiskSum	-0.089	0.067	0.183	-0.077	0.065	0.240	-0.125	0.087	0.152	-0.009	0.115	0.939	-0.033	0.063	0.601
Weighted															
nonHLA w-RiskScore	0.985	0.721	0.173	0.283	0.714	0.692	-0.100	0.970	0.918	0.282	1.270	0.824	2.067	0.694	0.003
VitD w-RiskScore	-1.322	0.603	0.029	-0.726	0.591	0.220	-1.988	0.794	0.012	-1.210	1.043	0.246	-0.590	0.576	0.306
Combined RiskScore	0.009	0.005	0.085	0.006	0.005	0.246	0.001	0.007	0.918	0.010	0.009	0.264	-0.003	0.005	0.588

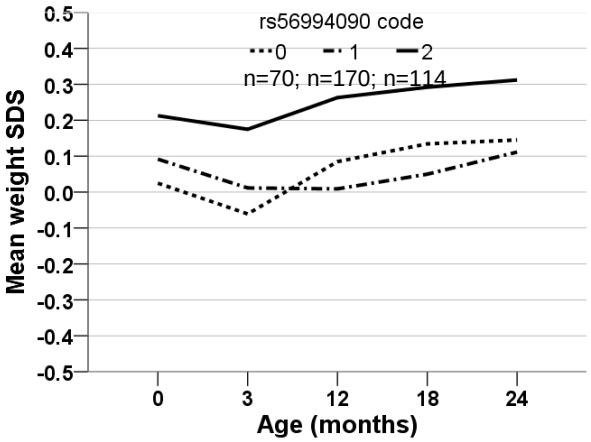
†Models were adjusted for covariates as described in Section 7.3.2 by growth parameter.

7.3.3 Plots of growth by genotypes

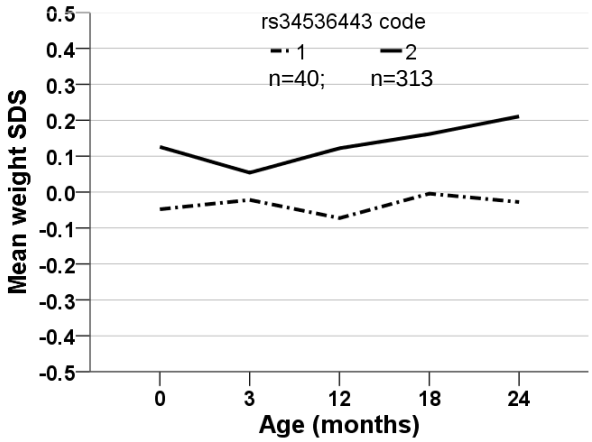
Trajectories of mean SD scores of anthropometric traits in infancy were plotted by genotypic group for the very significant associations by excluding cases with missing values in order to follow the same subjects (Figures 7-2 to 7-4).

Figure 7-2 | Weight SDS trajectories by T1D-risk genotypic group in a. *DLK1*, b. *TYK2*, c. *GC*, d. *IFIH1*, *e. CLEC16A.* Genotype frequencies are listed in legends.

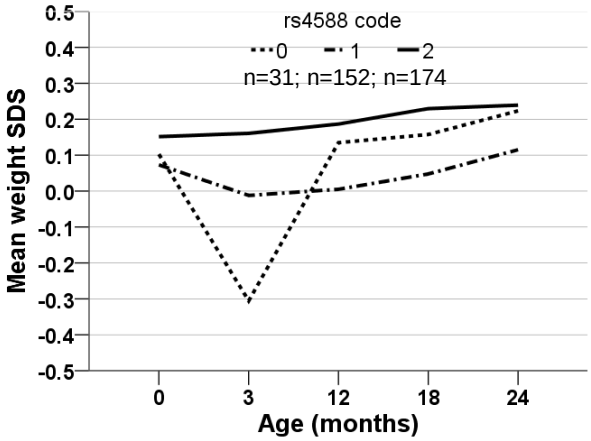
a. *DLK1*



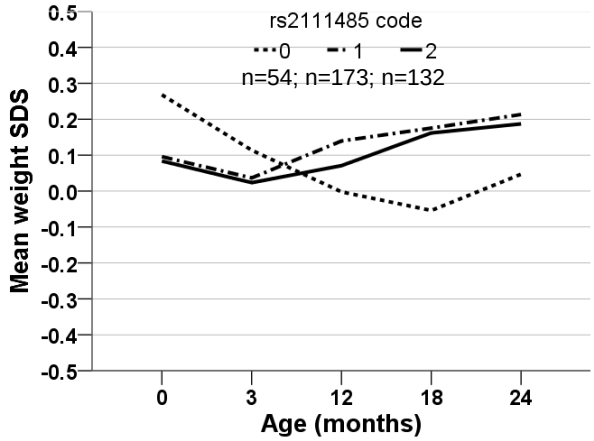
b. *TYK2*



c. *GC*



d. *IFIH1*



e. *CLEC16A*

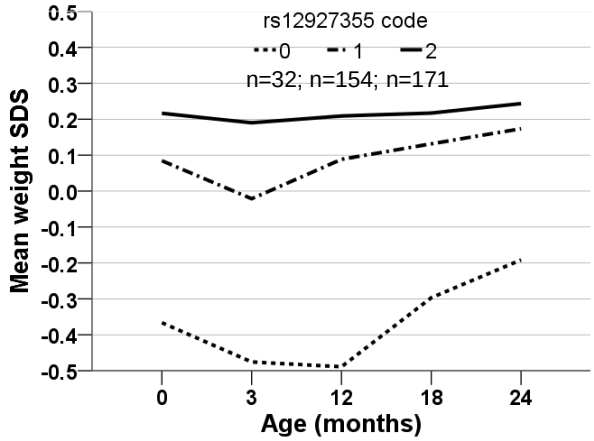
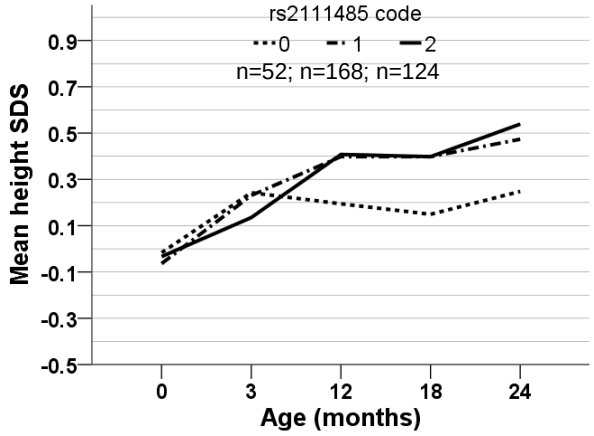
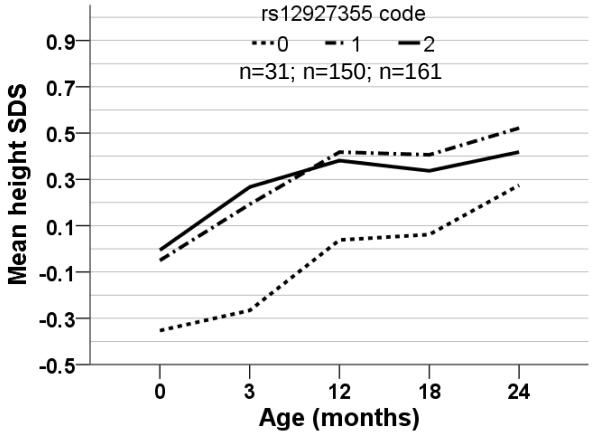


Figure 7-3 | Height SDS trajectories by T1D-risk genotypic group in a. *IFIH1*, b. *CLEC16A*, c. *BACH2*, d. *HLA-DR3*. Genotype frequencies are listed in legends.

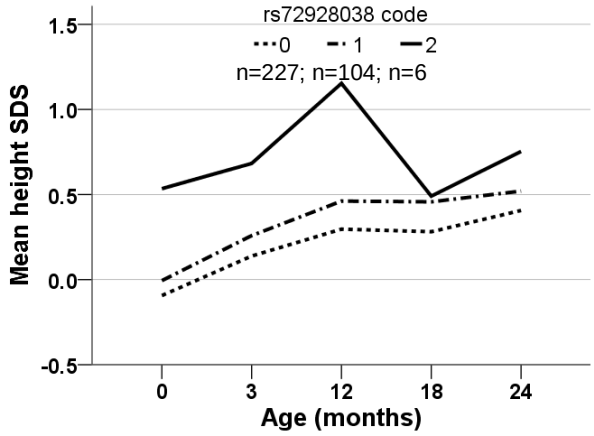
a. *IFIH1*



b. *CLEC16A*



c. *BACH2*



d. *HLA-DR3*

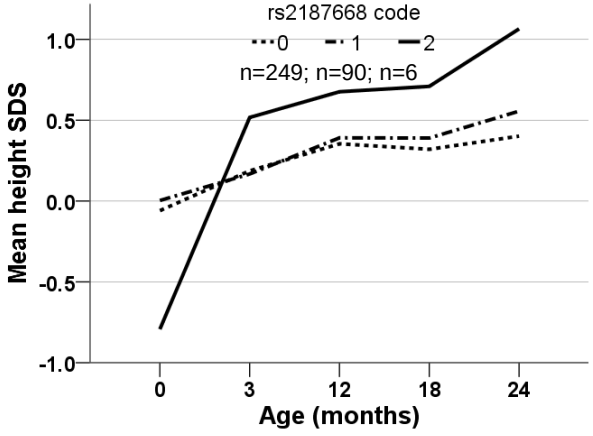
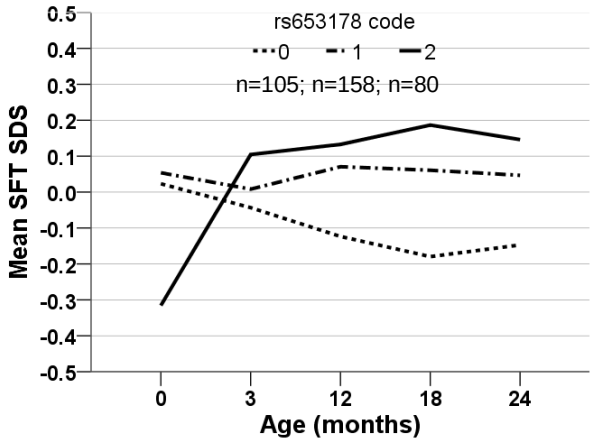
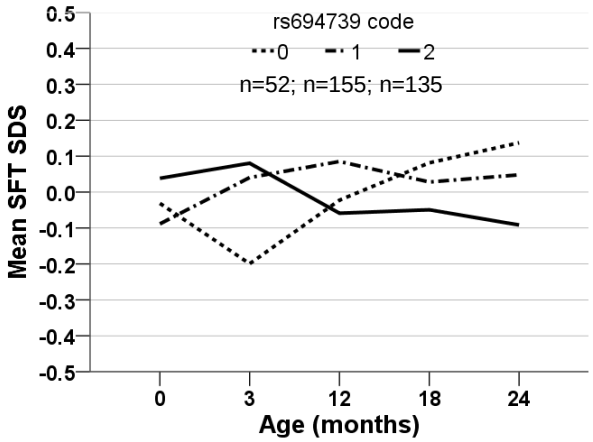


Figure 7-4 | Skinfold thickness (SFT) SDS trajectories by T1D-risk genotypic group in a. *SH2B3*, b. *BAD*. Genotype frequencies are listed in legends.

a. *SH2B3*



b. *BAD*



7.4 DISCUSSION

7.4.1 The advantage of longitudinal analysis

This chapter has elevated the search for an association between each selected T1D susceptibility SNP (or genetic risk score) and somatic growth in infancy by taking advantage of the longitudinal anthropometric phenotypes in the CBGS cohort. Application of a linear MEM to repeated measures has illustrated, as claimed by a prior study in Alzheimer's disease, the power gains from longitudinal analysis of measures at various time points, by virtue of using more data, vs. cross-sectional analysis at a single time point[547]. Specifically, the longitudinal analysis in the CBGS strengthened the genetic influences that were detected in the cross-sectional regressions, exemplified by associations between skinfold thickness in infancy and SNPs in *BAD* (P=2.4x10^{-4} vs. P=0.009) and *SH2B3* (P=7.0x10^{-6} vs. P=0.004), thus reducing doubts about the robustness of these relationships. It also surfaced SNPs in association with growth phenotypes that were not identified by the cross-sectional analysis, e.g. *IFIH1* and longitudinal weight (P=0.008) or height (P=0.009). In contrast, longitudinal analysis did not corroborate the cross-sectional association between *TYK2* and weight or *BACH2* and height, possibly suggestive of a time-variable effect of these genetic variants.

Comparison of longitudinal analyses across different time spans of repeated measures (first 3 months vs. first 24 months) possibly conveys the dynamic nature of genetic effects; the significant associations between *GC* and repeated measures of skinfold thickness (P=0.008) or weight (P=0.016) during the first 3 months of life disappeared when the analysis was conducted across the first 24 months of life. This implicit time-dependent associations between SNPs and phenotypes invokes the hypothesis that T1D susceptibility SNPs influence gene expression, which might be switched on/off during a critical time of development. It is currently accepted that the location of most T1D susceptibility loci in non-coding regions of the human genome provides clear evidence that alteration of gene expression, rather than protein sequence, constitutes the primary basis of genetic predisposition to T1D[508]. Moulder and Lahesmaa[552] investigated the influence of SNPs on differential gene expression using the Immunochip and reported that where *cis* eQTL (*cis* acting expressed quantitative trait loci) effect was detected included several genes of the type I IFN activated network that had been previously found in children with T1D-associated autoimmunity. The authors concluded that the innate immune response to type I IFN might be present throughout the progression of the disease and even before the appearance of autoantibodies[552], underscoring the long presymptomatic stage of the disease.

7.4.2 Interpretation of SNP-growth associations

7.4.2.1 Innate immunity and early growth

The rs2111485 in *IFIH1* (G>A) was found to positively associate with longitudinal measures of weight and height during infancy in the CBGS, i.e. accelerated growth per risk allele, which reconciles with the findings of the systematic literature review. The *IFIH1* encodes interferon-induced with helicase C domain protein 1 (IFIH1), a cytoplasmic receptor of viral nucleic acids (double-stranded RNA) which plays a key role in sensing viral infection and activating a cascade of antiviral responses, such as the induction of type I IFN and proinflammatory cytokines. The mRNA expression of *IFIH1* in the human islets is upregulated by enterovirus infection[88]. Risk alleles associate with the development of autoantibodies targeting β-cells, whereas decreased expression of the gene protects against T1D[553]. The finding here puts forward the possibility that pleiotropy explains the link between early growth and risk of infections, complementing the discussion on *TYK2* in Chapter 6. The absence of a longitudinal association between growth and *TYK2*, albeit a significant association with the baseline value (at birth), may be interpreted as evidence that its action is restricted to foetal growth. Adjacencies are drawn between *TYK2* and *IFIH1*: they are both implicated in type I IFN and cytokine pathways that modulate antiviral responses in the β-cell, they are expressed in pancreatic islets, and their protective alleles were found to associate with slower foetal and postnatal growth respectively in the CBGS. The fact that the minor alleles of both *IFIH1* and *TYK2* are protective against T1D via a reduction of their immunological activity underscores the notion that an excessive inflammatory response to viral infections contributes to autoimmunity and ensuing T1D; in contrast, if β-cells were directly destroyed by the virus, reduced activity of these responses would be deleterious[88]. It follows, and along the lines of the discussion in Chapter 6, that both genes mediate independent effects on growth via molecular interactions with growth factors. These lines of evidence converge on the possibility that the phenotypes of rapid growth and risk of T1D are the end results of biological events partially mediated by virus infection-regulating genes, exemplified by *TYK2* and *IFIH1* which act *in utero* and postnatally respectively, and whose expression in the islets is upregulated by viruses as a mechanism to protect the slow-replicating vulnerable β-cell. My findings in the CBGS have the following implications: i) the association between growth and susceptibility to T1D is not causal, ii) the previously reported associations between T1D and enterovirus infections *in utero*[56,554] or in early life[55,57-59] are biologically credible but genetically distinct, iii) interindividual differences in the innate immune system in early life, in conjunction with immunomodulatory exogenous factors (e.g. virus), are critical in the aetiology of T1D, thus giving justice to the vaccine approach as a preventative strategy for the disease.

7.4.2.2 *CLEC16A* and gains in infancy

The common polymorphism rs12927355 (C>T) pinpointed to *CLEC16A* on Chr16p13, and known to confer protection against T1D, correlated with longitudinal measures of SD scores of weight (*P*=0.006) and height (*P*=0.001) in infancy and showed borderline association with weight SDS at 12 months (Chapter 6). Graphical representation shows distinct weight trajectories from birth by genotypic group of this SNP (Figure 7-2) characterised by i) apparent differential weight, i.e. homozygous for the protective allele had less weight and homozygous for the risk allele weighed more than the other groups, attaining statistical significance at 12 months (overall per-risk allele increase was 0.1 SDS, *P*=0.047); and ii) different rates of weight gain, i.e. homozygous for the protective allele showed rapid weight gain from 12 months of age onwards vs. carriers of two risk alleles who gained weight in line with the reference.

CLEC16A encodes a membrane-associated endosomal protein and is expressed in various immune cell types. Islet transcriptomics revealed its expression in pancreatic islets[101]. Functional studies have shown that pancreas-specific deletion of *CLEC16A* in islets from mice resulted in abnormal mitochondria, characterised by reduced oxygen consumption and ATP concentration, both of which are required for normal β-cell function; these experimental findings suggested that *CLEC16A* controls β-cell function and prevents diabetes by regulating mitophagy[555]. Patients harbouring a diabetogenic polymorphism in this gene have reduced islet *CLEC16A* expression and reduced insulin secretion[555]. This proposed mechanism appears to reconcile with the finding in the CBGS: carriers of the T1D susceptibility allele had significantly slower rate of growth albeit weighing significantly more at 12 months. A possible scenario for this observation lies in the T1D-protective polymorphism (minor allele) upregulating expression of the gene in mid-infancy, reflected in the abrupt change in the weight of minor homozygotes at 12 months. This debate is perplexed by doubts cast about the correct identification of the candidate gene at this locus; *CLEC16A* is proximal to *DEXI*, which is the gene that has been pinpointed by the fine mapping study of T1D loci[102]. Of relevance to this thesis, the finding here surfaces that abrupt changes in insulin secretion have pronounced effects on rates of growth which are not immediately reflected in somatic size at a single point in time. This lack of synchronisation possibly accounts for the discrepant findings in the literature on T1D and cross-sectional measures of size.

7.4.2.3 Chr6 and height gains

The longitudinal analysis failed to reinforce the cross-sectional association found between the predisposing T1D SNP rs72928038 (G>A) in *BACH2* and longitudinal measures of height SDS at 12 or 18 months (Chapter 6). The discordance could suggest that change in height (i.e. slope) does not differ by number of copies of risk allele carried by the infant at this locus. This

explanation is plausible judging by the near-parallel trajectories of height between genotyping groups (Figure 7-3). The small count (n=6) of homozygous for the minor allele might account for the exaggerated height observed amongst this genotypic group at 12 months of age; yet it appears that height SDS gets progressively higher for each additional T1D risk allele carried. *BACH2* (BTB Domain and CNC Homolog 2) encodes a transcription factor of the Basic Region-Leucine Zipper family that is involved in both innate and adaptive immunity[556]. It is one of strongly suspected causal T1D genes initially thought to be selectively expressed in B cells[23]. More recent evidence points to its expression and function in β-cells, and upregulation by proinflammatory cytokines, whilst itself regulates expression of *PTPN2*[557]. Experimental studies in human islets and primary rat β-cells ascribed an anti-apoptotic function to *BACH2* in β-cells[557]. Another SNP harboured on *BACH2* (rs11755527, r^2=0.194), which was also associated with T1D, was previously found to strongly associate with Coeliac disease[558] and antibodies to Grave's disease[559]. *BACH2* is one of five genes identified as a potential biomarker for risk stratification of antibody positive subjects[552,560]. Collectively, these findings provide strong ground for a role of *BACH2* in the development of autoantibodies and invoke the hypothesis that higher-than-average height during infancy might be a biomarker of autoimmunity. In comparison, the genetic variant in *PTPN2* on Chr18p11, which encodes a phosphatase that modulates β-cell apoptosis by modulating local IFN production *after* a viral infection[561], accelerates progression to T1D *after* the appearance of islet autoantibodies[562]. Taken together, the T1D susceptibility variant in *BACH2* influences development of autoimmunity and height in infancy, doing justice to the prediction made in 1992 by Blom et al.[409] that rapid linear gain 'is a marker of a physiological mechanism that affects both growth and the pathogenesis of T1D'.

The other T1D risk locus that is a suspect for the accelerated linear growth in infancy is also harboured on Chr6. The rs2187668 tagging the widely incriminated HLA-DR3 was associated with longitudinal measures of height SDS in the CBGS (*P*=0.007), reflected in both higher measures of height at single ages and faster changes over time (Figure 7-3). The literature on the association between risk of T1D and height gains is scarce. One study found that length SDS at birth correlated with high-risk HLA[244] but the finding was not replicated by a prospective Italian study[405]. A combined registry-based study from Norway and Denmark found no significant association between height gains from birth to 12 months and T1D development. It would be interesting if future studies consider islet autoimmunity as the outcome, which is more proximal to the timing of events explored in this thesis. Notably, longitudinal analysis by sex in the CBGS for this SNP unveiled a much stronger association between linear growth in infancy and the rs2187668 in boys (Interaction Estimate=0.142, $P=3.1\text{x}10^{-7}$) vs. girls (Interaction Estimate=-0.062, *P*=0.066). This finding concurs with observational studies which reported that height gains between probands and controls were

more marked for boys than girls[393,409]. Considering the modest male to female excess in the incidence of T1D was found to be restricted to the HLA-DR3 type[520], the results here ignite the hypothesis that the HLA-DR3 exerts a sex-specific effect on the development of autoantibodies which is phenotypically manifested by rapid linear growth. The mechanisms behind the influence of the HLA-DR3 on linear growth are unknown but associations with IGF-1 levels merit investigation (Chapter 8).

7.4.2.4 Immune genes and skinfold thickness

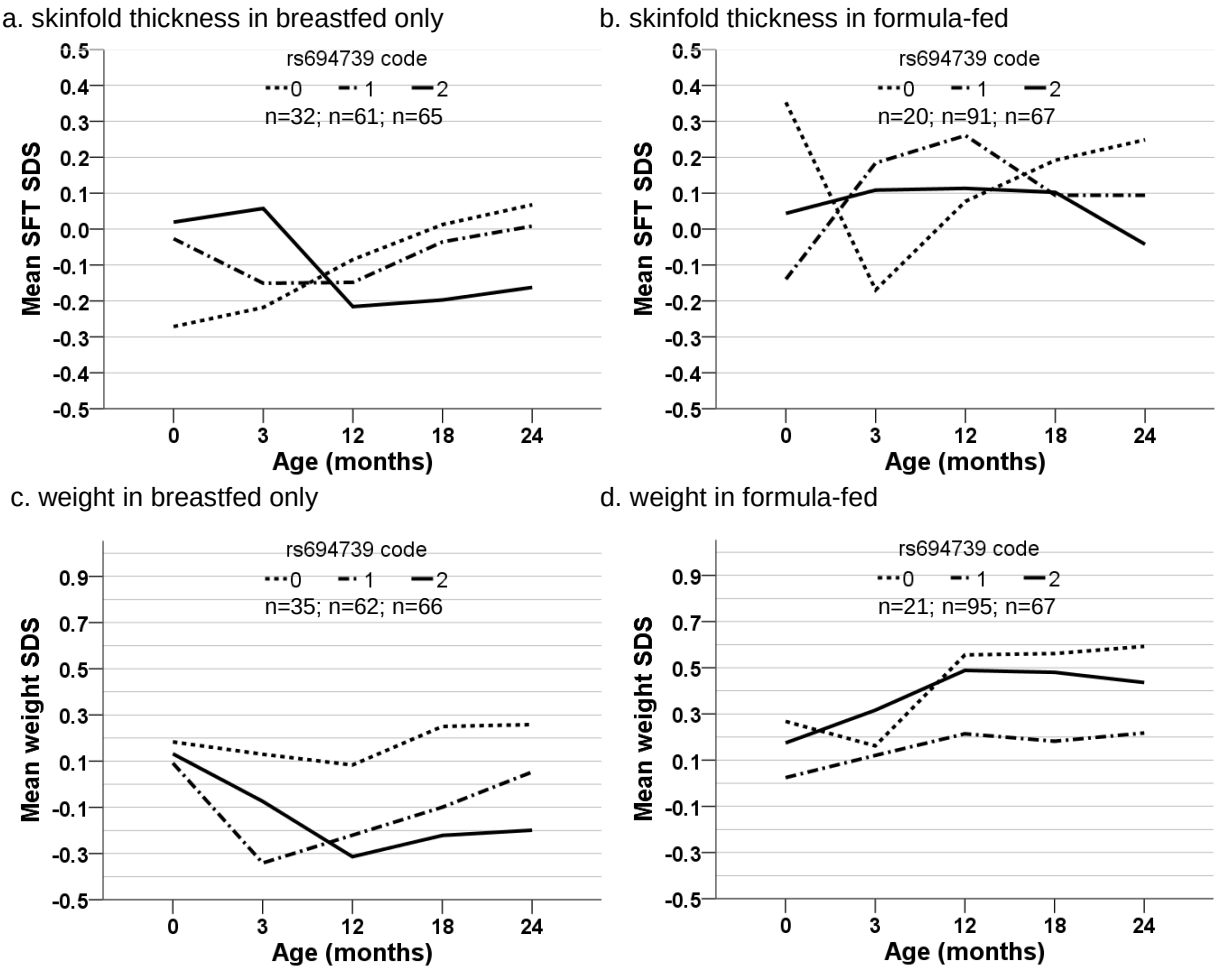
This is the first study to establish associations between T1D susceptibility variants on islet-expressed genes and adiposity proxied by the average value of the SD scores of skinfold thicknesses measured at four anatomical sites (triceps, subscapular, flank, and quadriceps).

In the CBGS, the rs653178 in *SH2B3* on Chr12q24 showed one of the strongest associations with skinfold thickness SDS (Interaction Estimate=0.065, P=7.0x10^{-6}) amongst all pairwise associations conducted in the cohort to-date. Children homozygous for the rs653178 T1D risk allele consistently had higher adiposity levels from 3 months onwards, whereas homozygous for the major allele consistently showed lower adiposity (Figure 7-4). The *SH2B3* is an immune gene with pleiotropic effects. The most compelling evidence of its involvement in growth comes from a clinical case reporting that a germline mutation in *SH2B3* resulting in complete loss of its protein led to the development of growth retardation-mild developmental delay, in addition to increased risk of autoimmune hepatitis and B-cell acute lymphoblastic leukaemia[563]. *SH2B3* encodes an SH2-containing protein with adaptor functions on a variety of signalling pathways. It specifically encodes LNK, an important negative regulator of cell-signalling events relevant to a number of receptors, including the TCR and the MPL (myeloproliferative leukaemia virus), which is a receptor for thrombopoietin on platelets[23]. These pleiotropic effects possibly account for the GWAS-level associations between the rs3184504 in *SH2B3* (a T1D-associated SNP in linkage with the index SNP rs653178, r^2=0.99[102]) and platelet count[564], cardiovascular disease[564], hypertension[565], coronary artery disease and ischaemic stroke[566]. In this context, my finding reinforces the message that multiple associations observed with a variant are explained by pleiotropy, with the implication that T1D complications might have a genetic background. Importantly, the result here sheds light on the 'early origins' of adult disease; consistently increased subcutaneous fat in infancy may qualify as a marker of genetic factors with critical roles in cardiovascular-related pathways.

The T1D predisposing SNP rs694739 (A>G) in *BAD* on Chr11q13 also demonstrated an association with skinfold thickness SDS (Interaction Estimate=-0.053, P=2.4x10^{-4}), yet there was no clear differentiation of trajectories by genotypic group (Figure 7-4). *BAD* encodes a pro-apoptotic BcL-2 member with bifunctional activities in glucose-stimulated insulin secretion

and physiologic adaptation of β-cell mass during high-fat feeding[567]. Amongst exclusively breastfed children, the trajectories of skinfold thickness SDS by the rs694739 genotypic groups acquired an inverse SNP dose-dependent effect from 12 months of age, also reflected in weight, which were not evident in genotypically similar infants on formula milk (Figure 7-5).

Figure 7-5 | Trajectories by T1D-risk genotypic group in *BAD* of the SDS for a. skinfold thickness (SFT) in breastfed only, b. skinfold thickness in formula-fed, c. weight in breastfed only, d. weight in formula-fed infants. Genotype frequencies are listed in legends.



This inverse association between skinfold thickness and T1D risk rs694739 alleles reached statistical significance in the cross-sectional multivariate analysis across all genotyped children at 24 months of age (Beta=-0.132, *P*=0.009, Chapter 6). What could account for this phenomenon which appears to be restricted to breastfed infants? My suggested explanation is drawn from the discussion on nutrients in Chapter 1; triglycerides do not cross the placenta during gestation[269] and post-delivery capacity for fat oxidation falls behind fat intake[267], pointing to poor programming of fat metabolism in early life. Is it possible that the diabetogenic rs694739 polymorphism downregulates expression of *BAD*, resulting in reduced capacity for β-cell expansion in response to the high-fat feeding from breast milk, which leads to lower

insulin secretion and thus reduced capacity to store fat in adipose tissue, in parallel with an increase in the vulnerability of the stressed β-cell that increases risk of T1D? This study did not measure gene expression. Nevertheless, despite SNPs representing 'fixed' traits of the genome, we may infer that the SNP is the effector of gene expression (the missing variable), which brings about the differential outcomes of skinfold thickness measured over time. This line of thinking leads me to further hypothesise that breast-milk, or one of its constituents, might regulate expression of *BAD* or other genes that program β-cells early in life with compensatory mechanisms for high-fat feeding, resulting in protective effects against β-cell destruction.

Of note, both the cross-sectional and longitudinal associations with skinfold thickness SDS were in opposite direction between the variants in *BAD* (lower levels and slower change of adiposity by T1D risk allele) vs. *SH2B3* (higher levels and higher rate of change by T1D risk allele), which underscore the different mechanisms operating in T1D. Despite the discordance, the stronger associations for both SNPs detected with skinfold thickness vs. weight invokes the possibility that the widely-researched measure of weight has masked the effect of adiposity. Weight SDS in the entire CBGS cohort of 1,660 children correlated strongly with skinfold thickness SDS (*rho*=0.462, *P*<0.0001), as did their changes between birth and 12 months (*rho*=0.507, *P*<0.0001). Chapter 8 probes for possible mechanisms behind these associations.

7.4.3 Genetic risk scores and early growth

The combined genetic risk score I constructed was comparable with the one independently generated and applied by another research group for prediction of T1D risk. Here, I explored its effect on growth in infancy. This triangulation would answer the question if growth is on the causal pathway of the disease. However, there was no evidence of an association in the CBGS, indicating that the putative relationship between rapid growth in early life and risk of developing T1D is not aetiological. Instead, the non-HLA genetic risk score associated with skinfold thickness, which is the recurring phenotype in the association analyses of this study.

7.4.4 Limitations

In first three months of life, 'growth rates change so rapidly that size measurements at a single time point are not as reliable as measurements at later time points'[568], necessitating frequent measurements during this period in a longitudinal study. As this was an exploratory exercise, I did not correct for multiple testing, but qualitatively considered the requirement for a stringent level of significance in the Discussion. In contrast to the widely-used metric of weight, the use of skinfold thickness is limited by the absence of reference charts; publication of children's skinfold reference data remains a research priority for assessment of relative fatness[264].

Chapter 8

T1D LOCI AND ENDOCRINE FACTORS

8.1 CONTEXT

Susceptibility to T1D, a disease characterised by islet inflammation and immune-mediated destruction of pancreatic β-cells, has a strong genetic element ascribed to >50 loci of varying effects. The evolving consensus is that most genes that are associated with these risk-increasing SNPs are expressed in β-cells, raising the exciting possibility that a large fraction of T1D risk is determined by the islet transcriptome, which regulates the development, function, apoptosis, antiviral responses and innate immunity of β-cells[508,569]. Genetic variants affecting target organs, such as pancreatic islets, may shape disease-specific pathology[349]. Conceivably, these variants and their candidate genes could shed light on mechanisms behind the archetypal phenotype of accelerated growth in infancy observed in prediabetic children. The effect of these variants on growth could be mediated directly via hormone(s) acting postnatally. Additionally, emerging evidence casts lights on the involvement of microbiota in growth in infancy: a causal relationship was established between the immature gut microbiota of undernourished infants and impaired growth phenotypes[570]. An interesting series of novel correlations identified in the CBGS (Chapters 6 and 7) between early growth and T1D susceptibility SNPs — including variants mapped to genes expressed in the islets or affecting the microbiome — have been woven with existing biological facts about the candidate genes to shape hypotheses about the molecular mediators behind these associations (Table 8-1). Here, I took advantage of the CBGS cohort and applied the statistical methods I developed previously to evaluate associations between the biology-screened T1D susceptibility SNPs and cross-sectional or longitudinal whole blood levels of endocrine regulators of somatic growth in infancy, namely C-peptide (the fragment that results from the proteolytic cleavage of proinsulin to insulin and has a longer half-life than insulin), IGF-1, and leptin.

Table 8-1 | Hypothesised mediators of early growth and risk of T1D.

Locus	Type	Hypothesised mediator of growth	Rationale
GLIS3	Non-immune	Insulin / β-cell	It encodes a transcription factor with involvement in pancreatic **β-cell development**.
PTPN2	Immune	Insulin / β-cell	It encodes a phosphatase that modulates **β-cell apoptosis**.
CLEC16A	Immune	Insulin / β-cell	It controls **β-cell function** by regulating mitophagy.
BAD	Immune	Insulin / β-cell	Its product regulates the physiologic adaptation of **β-cell mass** during high-fat feeding and plays a physiological role in glucose-stimulated insulin secretion.
		Leptin	The proapoptotic activity of its product is induced by affecting the level of heterodimerisation of Bcl-x proteins, also implicated in the mechanism behind the association between leptin and autoimmunity.
SH2B3	Immune	IGF-1	It has an established phenotype of **growth in humans**.
FUT2	Immune	Microbiome	Associates with gut microbiota in healthy individuals.
PTPN22	Immune	Microbiome	A study found that *PTPN22* did not significantly associate with T1D among those born with caesarean section, which in tandem with the finding that children born by caesarean section have compositionally different colonisation vs. children born vaginally[571,572], could suggest that *PTPN22*, involved in the activation of T cells, operates via interaction between intestinal flora and host T cells[573].

8.2 METHODS

8.2.1 Study design and measures

The details of the CBGS study and subjects, SNP selection and genotyping, genetic scores and quality controls were described in previous chapters. The outcomes of this study were whole blood levels of IGF-1, C-peptide and leptin, which were measured from DBS samples routinely taken at the study visits and assayed as described in Appendix II.

8.2.2 Statistical analysis

Cross-sectional and longitudinal association analyses were performed between whole blood concentrations of IGF-1 measured at ages 3, 12, 18 and 24 months, C-peptide measured at ages 3 and 12 months or leptin measured at ages 3 and 12 months (explained variable), and individual T1D susceptibility SNPs or genetic risk scores (explanatory variable), using multivariate linear regression as described in Chapters 6 and 7, with adjustment for preselected covariates (Table 3-21) by co-entering those variables in the model. Analyses of endocrine factors were restricted to 586 children after excluding extremes of growth trajectories (maternal T1D, <36 weeks of gestation, twins). Table 8-2 lists the counts and percentages of non-missing observations for hormones by age. For IGF-1 levels, 25 babies had complete data for four visits, 65 for three visits, 91 for two visits, 125 for one visit and 280 for none. For C-peptide levels, 19 babies had data for two visits, 141 for one visit, and 426 for none. For leptin levels, 48 babies had data for two visits, 121 for one visit, and 417 for none.

Table 8-2 | Counts (%) of genotyped infants included in the analysis who had measured levels of IGF 1, C-peptide and leptin by age.

	IGF-1	C-peptide	Leptin
Birth			
3 months	212 (36.2)	123 (21.0)	121 (20.6)
12 months	166 (28.3)	56 (9.6)	96 (16.4)
18 months	135 (23.0)		
24 months	89 (15.2)		

8.3 RESULTS

8.3.1 Descriptive statistics

Summary statistics of blood hormone levels by age and sex of children included in the analyses are presented in Table 8-3.

Table 8-3 | Count (N), Mean ± SD and range for levels of IGF-1, C-peptide and leptin by age and sex of infants included in the analysis.

	Boys			Girls		
	N	**Mean ± SD**	**Range**	**N**	**Mean ± SD**	**Range**
IGF-1 (ng/ml)						
3-month	114	49.4 ± 20.2	8.0, 107.0	98	48.2 ± 18.7	19.0, 119.0
12-month	95	42.7 ± 17.8	10.0, 87.0	71	55.8 ± 26.3	14.0, 153.0
18-month	70	51.2 ± 22.5	20.0, 120.0	65	63.5 ± 22.8	23.0, 133.0
24-month	49	56.4 ± 28.7	21.0, 145.0	40	65.3 ± 26.3	24.0, 120.0
C-peptide (pmol/L)						
3-month	63	701.6 ± 386.2	124.9, 2,061.0	60	665.5 ± 426.1	177.9, 2,685.0
12-month	36	602.3 ± 343.8	164.5, 1,500.8	20	655.6 ± 363.9	143.3, 1,588.0
Leptin (ng/ml)						
3-month	60	2.0 ± 1.1	0.6, 5.5	61	3.4 ± 2.0	0.6, 9.6
12-month	56	1.1 ± 0.6	0.6, 3.2	40	1.6 ± 1.1	0.6, 5.6

8.3.2 T1D SNPs and hormone levels: cross-sectional associations

Tables 8-4 to 8-6 display the summary statistics (standardised Beta, *P* value) of multivariate linear regressions of the blood concentration of each hormone by age on individual SNPs. As leptin levels show a distinct sex-dimorphic effect, the analysis for this hormone was repeated by sex (Table 8-7). There was strong evidence of an association between IGF-1 levels at 12 months and T1D susceptibility SNPs tagging the *HLA-DQ8* (*P*=0.003) and in *CTSH* (*P*=0.005), and between IGF-1 levels at 24 months and the SNP tagging the *HLA-DR3* (*P*=0.009). Significant associations were detected between leptin at 3 months and the SNP in *DHCR7* in girls only (*P*=0.004).

Table 8-4 | Associations between T1D SNPs and IGF-1 levels (ng/ml) at ages 3, 12, 18, 24 months†.

	SNP	Gene	Birth		3 months		12 months		18 months		24 months	
			Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
1	rs2476601	*PTPN22*			0.013	0.852	0.061	0.432	0.076	0.377	-0.188	0.102
2	rs3024505	*IL10*			0.042	0.550	-0.042	0.598	0.053	0.546	-0.114	0.344
3	rs2111485	*IFIH1*			-0.066	0.350	-0.118	0.139	0.038	0.668	0.026	0.826
4	rs3087243	*CTLA4*			0.100	0.158	0.004	0.959	0.125	0.147	0.024	0.835
5	rs4588	*GC*			0.166	0.017	0.108	0.171	0.074	0.395	0.028	0.814
6	rs75793288	*IL2*			-0.008	0.906	0.001	0.987	0.057	0.506	0.088	0.461
7	rs17426593	*HLA-DR4*			0.012	0.864	-0.166	0.035	-0.064	0.465	-0.043	0.719
8	rs2187668	*HLA-DR3*			-0.042	0.551	0.034	0.663	0.183	0.030	0.296	0.009
9	rs7454108	*HLA-DQ8*			-0.022	0.759	-0.227	0.003	-0.025	0.772	-0.161	0.163
10	rs3135388	*HLA-DQB1*06:02*			0.009	0.893	-0.022	0.780	0.091	0.287	0.083	0.483
11	rs72928038	*BACH2*			0.101	0.155	-0.022	0.780	0.077	0.381	0.013	0.910
12	rs6920220	*TNFAIP3*			-0.016	0.824	-0.001	0.992	-0.040	0.644	0.125	0.310
13	rs6476839	*GLIS3*			-0.060	0.395	-0.092	0.238	-0.053	0.542	0.004	0.976
14	rs61839660	*IL2RA*			-0.085	0.225	-0.145	0.064	-0.149	0.084	-0.194	0.095
15	rs11258747	*PRKCQ*			0.096	0.171	-0.054	0.492	-0.037	0.671	0.072	0.535
16	rs12416116	*RNLS*			0.022	0.762	0.032	0.688	0.030	0.730	0.062	0.594
17	rs10741657	*CYP2R1*			-0.004	0.955	0.085	0.283	0.139	0.109	0.164	0.171
18	rs12794714	*CYP2R1*			0.073	0.304	-0.037	0.635	-0.034	0.696	0.296	0.011
19	rs689	*INS*			-0.140	0.048	-0.001	0.985	-0.006	0.947	-0.269	0.022
20	rs694739	*BAD*			0.021	0.764	-0.023	0.771	-0.059	0.504	0.092	0.435
21	rs12785878	*DHCR7*			0.056	0.426	0.030	0.705	0.059	0.488	0.063	0.587
22	rs11170466	*ITGB7*			0.081	0.255	0.018	0.822	0.039	0.652	-0.187	0.101
23	rs705704	*ERBB3*			0.054	0.455	0.007	0.931	-0.010	0.911	0.019	0.874
24	rs10877012	*CYP27B1*			0.065	0.355	0.030	0.705	-0.030	0.730	0.153	0.188
25	rs653178	*SH2B3*			-0.032	0.654	0.197	0.012	-0.028	0.749	0.215	0.065
26	rs56994090	*DLK1*			0.096	0.167	-0.010	0.898	-0.085	0.326	0.058	0.620
27	rs34593439	*CTSH*			0.088	0.208	-0.217	0.005	0.093	0.277	-0.242	0.038
28	rs12927355	*CLEC16A*			-0.142	0.042	-0.176	0.024	-0.053	0.537	0.139	0.232
29	rs151234	*IL27*			-0.052	0.457	-0.059	0.447	-0.012	0.892	0.039	0.737
30	rs12453507	*ORMDL3*			0.003	0.965	0.063	0.421	-0.029	0.741	0.185	0.114
31	rs1893217	*PTPN2*			-0.079	0.259	-0.001	0.994	0.130	0.128	0.001	0.994
32	rs34536443	*TYK2*			0.123	0.082	-0.013	0.871	0.116	0.183	0.087	0.465
33	rs516246	*FUT2*			-0.100	0.161	-0.156	0.045	0.031	0.723	-0.066	0.571
34	rs11203202	*UBASH3A*			0.052	0.463	-0.144	0.069	-0.123	0.158	-0.263	0.026

†Models were adjusted for sex and type of milk feeding at age 3 months.

Table 8-5 | Associations between T1D SNPs and C-peptide levels (pmol/L) at ages 3 and 12 months†.

	SNP	Gene	Birth		3 months		12 months		18 months		24 months	
			Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
1	rs2476601	*PTPN22*			0.157	0.089	0.268	0.064				
2	rs3024505	*IL10*			-0.029	0.762	-0.164	0.246				
3	rs2111485	*IFIH1*			-0.144	0.128	-0.012	0.933				
4	rs3087243	*CTLA4*			0.030	0.749	-0.042	0.771				
5	rs4588	*GC*			-0.014	0.879	0.027	0.854				
6	rs75793288	*IL2*			0.011	0.909	0.006	0.969				
7	rs17426593	*HLA-DR4*			0.044	0.641	0.186	0.195				
8	rs2187668	*HLA-DR3*			0.099	0.284	-0.016	0.910				
9	rs7454108	*HLA-DQ8*			0.134	0.152	0.192	0.192				
10	rs3135388	*HLA-DQB1*06:02*			0.040	0.664	0.146	0.302				
11	rs72928038	*BACH2*			-0.073	0.438	0.138	0.329				
12	rs6920220	*TNFAIP3*			0.043	0.645	0.035	0.805				
13	rs6476839	*GLIS3*			0.002	0.985	0.058	0.682				
14	rs61839660	*IL2RA*			0.004	0.963	-0.123	0.387				
15	rs11258747	*PRKCQ*			-0.035	0.707	-0.034	0.819				
16	rs12416116	*RNLS*			-0.071	0.450	-0.046	0.751				
17	rs10741657	*CYP2R1*			0.036	0.697	-0.111	0.434				
18	rs12794714	*CYP2R1*			0.060	0.518	-0.175	0.215				
19	rs689	*INS*			-0.041	0.664	0.115	0.423				
20	rs694739	*BAD*			-0.170	0.067	-0.045	0.753				
21	rs12785878	*DHCR7*			-0.009	0.926	0.105	0.457				
22	rs11170466	*ITGB7*			0.033	0.728	-0.098	0.500				
23	rs705704	*ERBB3*			0.004	0.968	0.183	0.198				
24	rs10877012	*CYP27B1*			-0.021	0.821	-0.161	0.250				
25	rs653178	*SH2B3*			0.064	0.499	0.099	0.488				
26	rs56994090	*DLK1*			0.053	0.574	-0.155	0.277				
27	rs34593439	*CTSH*			0.090	0.332	0.088	0.535				
28	rs12927355	*CLEC16A*			0.063	0.498	0.102	0.471				
29	rs151234	*IL27*			0.062	0.502	-0.162	0.252				
30	rs12453507	*ORMDL3*			-0.073	0.430	0.026	0.856				
31	rs1893217	*PTPN2*			0.077	0.409	0.086	0.545				
32	rs34536443	*TYK2*			0.020	0.827	0.082	0.574				
33	rs516246	*FUT2*			0.129	0.168	0.024	0.869				
34	rs11203202	*UBASH3A*			-0.054	0.564	-0.008	0.953				

†Models were adjusted for type of milk feeding at age 3 months.

Table 8-6 | Associations between T1D SNPs and leptin levels (ng/ml) at ages 3 and 12 months†.

	SNP	Gene	Birth		3 months		12 months		18 months		24 months	
			Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
1	rs2476601	*PTPN22*			0.002	0.980	0.045	0.654				
2	rs3024505	*IL10*			0.022	0.805	-0.095	0.357				
3	rs2111485	*IFIH1*			0.018	0.843	-0.116	0.248				
4	rs3087243	*CTLA4*			0.038	0.673	0.121	0.231				
5	rs4588	*GC*			0.088	0.325	-0.068	0.508				
6	rs75793288	*IL2*			-0.083	0.349	0.071	0.485				
7	rs17426593	*HLA-DR4*			-0.054	0.551	-0.132	0.199				
8	rs2187668	*HLA-DR3*			-0.162	0.067	0.039	0.703				
9	rs7454108	*HLA-DQ8*			-0.029	0.743	-0.036	0.727				
10	rs3135388	*HLA-DQB1*06:02*			-0.010	0.910	0.115	0.253				
11	rs72928038	*BACH2*			-0.001	0.988	0.062	0.544				
12	rs6920220	*TNFAIP3*			-0.093	0.295	0.025	0.808				
13	rs6476839	*GLIS3*			0.017	0.849	-0.054	0.596				
14	rs61839660	*IL2RA*			0.084	0.350	0.007	0.946				
15	rs11258747	*PRKCQ*			-0.019	0.835	-0.157	0.128				
16	rs12416116	*RNLS*			0.004	0.965	0.204	0.051				
17	rs10741657	*CYP2R1*			0.115	0.205	0.141	0.164				
18	rs12794714	*CYP2R1*			0.160	0.073	0.211	0.036				
19	rs689	*INS*			0.007	0.937	0.136	0.182				
20	rs694739	*BAD*			0.129	0.148	-0.003	0.975				
21	rs12785878	*DHCR7*			0.230	0.009	0.021	0.835				
22	rs11170466	*ITGB7*			0.168	0.064	-0.073	0.495				
23	rs705704	*ERBB3*			-0.011	0.904	0.228	0.023				
24	rs10877012	*CYP27B1*			0.108	0.224	-0.039	0.701				
25	rs653178	*SH2B3*			-0.093	0.298	-0.031	0.759				
26	rs56994090	*DLK1*			-0.015	0.869	-0.059	0.564				
27	rs34593439	*CTSH*			0.086	0.336	-0.094	0.358				
28	rs12927355	*CLEC16A*			-0.097	0.273	0.166	0.101				
29	rs151234	*IL27*			-0.152	0.085	0.045	0.655				
30	rs12453507	*ORMDL3*			-0.013	0.884	0.117	0.250				
31	rs1893217	*PTPN2*			0.135	0.129	0.051	0.616				
32	rs34536443	*TYK2*			-0.010	0.916	0.062	0.542				
33	rs516246	*FUT2*			-0.183	0.040	-0.110	0.287				
34	rs11203202	*UBASH3A*			-0.103	0.255	0.083	0.414				

†Models were adjusted for sex and parity.

Table 8-7 | Associations between T1D SNPs and leptin levels (ng/ml) by sex at ages 3 and 12 months†.

			Male				Female			
	SNP	Gene	3 months		12 months		3 months		12 months	
			Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
1	rs2476601	*PTPN22*	-0.078	0.576	-0.037	0.792	0.059	0.675	0.072	0.644
2	rs3024505	*IL10*	-0.209	0.125	-0.171	0.228	0.164	0.251	0.037	0.814
3	rs2111485	*IFIH1*	-0.013	0.923	0.013	0.928	0.039	0.783	-0.231	0.129
4	rs3087243	*CTLA4*	-0.126	0.371	-0.051	0.720	0.135	0.330	0.254	0.094
5	rs4588	*GC*	0.032	0.816	0.133	0.345	0.143	0.302	-0.172	0.265
6	rs75793288	*IL2*	-0.120	0.376	0.071	0.614	-0.079	0.575	0.068	0.662
7	rs17426593	*HLA-DR4*	0.033	0.812	-0.055	0.704	-0.127	0.368	-0.165	0.285
8	rs2187668	*HLA-DR3*	-0.037	0.783	0.143	0.309	-0.298	0.031	-0.026	0.867
9	rs7454108	*HLA-DQ8*	0.054	0.692	-0.109	0.441	-0.094	0.502	0.009	0.954
10	rs3135388	*HLA-DQB1*06:02*	0.021	0.876	0.175	0.215	-0.030	0.831	0.051	0.740
11	rs72928038	*BACH2*	0.062	0.658	0.004	0.975	-0.039	0.791	0.156	0.320
12	rs6920220	*TNFAIP3*	-0.121	0.383	-0.077	0.591	-0.099	0.490	0.097	0.533
13	rs6476839	*GLIS3*	-0.216	0.109	0.026	0.851	0.143	0.307	-0.026	0.879
14	rs61839660	*IL2RA*	0.148	0.278	0.033	0.814	0.064	0.650	-0.058	0.703
15	rs11258747	*PRKCQ*	-0.096	0.487	-0.195	0.189	0.017	0.904	-0.063	0.686
16	rs12416116	*RNLS*	0.153	0.268	0.175	0.224	-0.064	0.646	0.211	0.209
17	rs10741657	*CYP2R1*	0.279	0.042	-0.012	0.933	0.050	0.723	0.240	0.114
18	rs12794714	*CYP2R1*	0.293	0.032	0.032	0.821	0.120	0.387	0.306	0.045
19	rs689	*INS*	-0.036	0.796	0.151	0.293	0.029	0.836	0.184	0.231
20	rs694739	*BAD*	0.057	0.681	-0.165	0.249	0.203	0.144	0.128	0.413
21	rs12785878	*DHCR7*	0.042	0.765	0.196	0.164	0.384	0.004	-0.058	0.708
22	rs11170466	*ITGB7*	0.061	0.653	0.085	0.563	0.240	0.089	-0.163	0.320
23	rs705704	*ERBB3*	-0.050	0.716	0.333	0.015	0.014	0.923	0.058	0.732
24	rs10877012	*CYP27B1*	0.011	0.933	-0.219	0.124	0.186	0.186	0.037	0.809
25	rs653178	*SH2B3*	-0.094	0.496	-0.066	0.646	-0.114	0.412	-0.134	0.394
26	rs56994090	*DLK1*	0.130	0.336	-0.220	0.114	-0.106	0.452	0.067	0.670
27	rs34593439	*CTSH*	0.080	0.557	0.025	0.860	0.105	0.447	-0.189	0.221
28	rs12927355	*CLEC16A*	-0.059	0.667	-0.108	0.446	-0.138	0.320	0.303	0.048
29	rs151234	*IL27*	0.031	0.823	0.093	0.509	-0.278	0.040	0.030	0.848
30	rs12453507	*ORMDL3*	-0.041	0.762	0.244	0.078	0.002	0.989	-0.089	0.593
31	rs1893217	*PTPN2*	0.071	0.605	0.138	0.338	0.195	0.163	-0.083	0.593
32	rs34536443	*TYK2*	0.074	0.590	-0.090	0.527	-0.048	0.738	0.126	0.431
33	rs516246	*FUT2*	-0.061	0.662	-0.135	0.340	-0.288	0.038	-0.085	0.592
34	rs11203202	*UBASH3A*	-0.003	0.985	0.085	0.549	-0.191	0.174	0.054	0.727

†Models were adjusted for parity.

Tables 8-8 to 8-10 display summary statistics (standardised Beta, *P* value) of regressions of hormone levels at each age on genetic risk scores after adjusting for covariates as described for individual SNPs. The borderline associations detected with genetic risk scores appear to be driven by the respective associations with individual SNPs.

Table 8-8 | Associations between T1D genetic risk scores and IGF-1 levels (ng/ml) at ages 3, 12, 18, 24 months†.

Genetic Risk Score	Birth		3 months		12 months		18 months		24 months	
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
HLA										
HLA Risk_2level			-0.014	0.838	-0.165	0.034	0.171	0.044	-0.114	0.329
HLA Risk_3level			-0.052	0.463	-0.159	0.042	0.128	0.144	0.087	0.470
Unweighted										
nonHLA RiskSum			0.062	0.383	-0.121	0.126	0.017	0.846	0.114	0.341
VitD RiskSum			0.157	0.025	0.062	0.432	0.035	0.693	0.253	0.028
Weighted										
nonHLA w-RiskScore			-0.050	0.478	-0.084	0.290	0.018	0.839	-0.145	0.225
VitD w-RiskScore			0.110	0.119	0.059	0.456	-0.005	0.955	0.207	0.074
Combined RiskScore			-0.033	0.647	-0.124	0.116	0.118	0.172	0.024	0.842

†Models were adjusted for sex and type of milk feeding at age 3 months.

Table 8-9 | Associations between T1D genetic risk scores and C-peptide levels (pmol/L) at ages 3 and 12 months†.

Genetic Risk Score	Birth	3 months		12 months		18 months		24 months	
		Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
HLA									
HLA Risk_2level		-0.018	0.844	-0.075	0.605				
HLA Risk_3level		0.173	0.062	0.143	0.311				
Unweighted									
nonHLA RiskSum		0.020	0.833	0.042	0.768				
VitD RiskSum		0.016	0.861	-0.100	0.491				
Weighted									
nonHLA w-RiskScore		0.041	0.659	0.160	0.259				
VitD w-RiskScore		-0.027	0.773	-0.134	0.351				
Combined RiskScore		0.137	0.143	0.210	0.135				

†Models were adjusted for type of milk feeding at age 3 months.

Table 8-10 | Associations between T1D genetic risk scores and leptin levels (ng/ml) by sex at ages 3 and 12 months†.

Genetic Risk Score	Male				Female			
	3 months		12 months		3 months		12 months	
	Beta	*P* value	Beta	*P* value	Beta	*P* value	Beta	*P* value
HLA								
HLA Risk_2level	0.094	0.489	0.197	0.158	-0.184	0.182	-0.166	0.283
HLA Risk_3level	-0.006	0.967	0.042	0.770	-0.282	0.044	-0.042	0.786
Unweighted								
nonHLA RiskSum	-0.052	0.711	0.059	0.677	0.167	0.247	0.176	0.263
VitD RiskSum	0.229	0.101	0.061	0.667	0.405	0.003	0.042	0.784
Weighted								
nonHLA w-RiskScore	-0.081	0.561	0.091	0.514	0.122	0.392	0.185	0.241
VitD w-RiskScore	0.086	0.533	-0.106	0.456	0.334	0.015	0.020	0.899
Combined RiskScore	-0.002	0.987	0.175	0.211	-0.119	0.399	0.105	0.496

†Models were adjusted for parity.

8.3.3 T1D SNPs and hormone levels: longitudinal associations

Table 8-11 displays the summary statistics — estimate of the interaction between the SNP and age of infant, standard error (SE) and *P* value — of longitudinal association analyses between repeated measures of the blood concentration of each hormone and individual SNPs. Longitudinal analysis strengthened the previously detected association between IGF-1 levels and the SNP tagging the *HLA-DR3* (*P*=0.002) and identified a new association between IGF-1 and *UBASH3A* (*P*=0.005). Except for a borderline association with the vitamin D genetic score, possibly driven by the association with the individual SNP in *DHCR7*, there was no evidence of a relationship between genetic risk scores and changes in levels of hormones (Table 8-12).

Table 8-11 | Summary statistics of T1D SNP-age interactions for longitudinal measures of levels of IGF-1 between 3 and 24 months of age, and C-peptide and leptin between 3 and 12 months of age†.

	SNP	Gene	IGF-1 (ng/ml)			C-peptide (pmol/L)			Leptin (ng/ml)		
			Estimate	SE	*P* value	Estimate	SE	*P* value	Estimate	SE	*P* value
1	rs2476601	*PTPN22*	-2.930	1.918	0.127	121.721	125.143	0.334	0.127	0.405	0.754
2	rs3024505	*IL10*	-0.572	1.644	0.728	-99.980	135.710	0.463	-0.296	0.389	0.449
3	rs2111485	*IFIH1*	1.316	1.197	0.272	68.756	82.298	0.407	-0.106	0.250	0.673
4	rs3087243	*CTLA4*	-0.154	1.165	0.895	-50.496	87.046	0.564	-0.112	0.246	0.650
5	rs4588	*GC*	-1.324	1.373	0.335	1.563	98.455	0.987	-0.146	0.290	0.615
6	rs75793288	*IL2*	1.822	1.170	0.120	-1.835	99.528	0.985	0.422	0.271	0.123
7	rs17426593	*HLA-DR4*	-1.573	1.635	0.336	114.474	104.406	0.279	0.135	0.337	0.691
8	rs2187668	*HLA-DR3*	5.618	1.827	0.002	-107.713	113.309	0.345	0.687	0.352	0.054
9	rs7454108	*HLA-DQ8*	-2.911	2.197	0.186	32.644	166.280	0.845	0.056	0.468	0.905
10	rs3135388	*HLA-DQB1*06:02*	0.928	1.654	0.575	56.149	108.971	0.610	0.235	0.321	0.468
11	rs72928038	*BACH2*	-0.710	1.709	0.678	102.278	103.031	0.326	-0.069	0.329	0.834
12	rs6920220	*TNFAIP3*	1.945	1.429	0.174	28.645	96.193	0.767	0.374	0.302	0.218
13	rs6476839	*GLIS3*	1.109	1.205	0.358	50.699	91.503	0.581	-0.137	0.273	0.617
14	rs61839660	*IL2RA*	-1.631	1.744	0.350	-72.227	117.778	0.544	-0.318	0.368	0.392
15	rs11258747	*PRKCQ*	0.307	1.281	0.811	81.250	103.529	0.434	-0.076	0.329	0.819
16	rs12416116	*RNLS*	-0.452	1.312	0.730	42.806	105.085	0.685	0.285	0.300	0.344
17	rs10741657	*CYP2R1*	2.473	1.312	0.060	-116.795	83.445	0.168	-0.092	0.271	0.736
18	rs12794714	*CYP2R1*	1.121	1.192	0.347	-139.745	88.906	0.122	-0.072	0.271	0.792
19	rs689	*INS*	-1.111	1.378	0.420	68.533	92.432	0.461	0.205	0.280	0.467
20	rs694739	*BAD*	0.525	1.178	0.656	69.405	79.731	0.388	-0.409	0.253	0.111
21	rs12785878	*DHCR7*	0.848	1.359	0.533	47.447	100.614	0.639	-0.540	0.288	0.064
22	rs11170466	*ITGB7*	-4.130	2.650	0.120	-54.813	164.038	0.739	-1.257	0.490	0.012
23	rs705704	*ERBB3*	-0.597	1.203	0.620	79.999	83.428	0.343	0.386	0.234	0.104
24	rs10877012	*CYP27B1*	0.015	1.289	0.991	-78.957	84.910	0.358	-0.252	0.257	0.333
25	rs653178	*SH2B3*	2.665	1.198	0.027	68.968	82.223	0.405	0.157	0.259	0.544
26	rs56994090	*DLK1*	-1.060	1.184	0.371	-85.143	84.311	0.317	-0.077	0.245	0.755
27	rs34593439	*CTSH*	-3.149	1.866	0.092	-24.722	114.277	0.830	-0.401	0.369	0.283
28	rs12927355	*CLEC16A*	2.110	1.221	0.085	-12.147	81.739	0.882	0.449	0.263	0.091
29	rs151234	*IL27*	0.542	1.606	0.736	-111.642	115.214	0.338	0.432	0.353	0.226
30	rs12453507	*ORMDL3*	1.040	1.242	0.403	50.525	80.863	0.536	0.145	0.245	0.557
31	rs1893217	*PTPN2*	1.050	1.466	0.474	18.765	116.067	0.872	-0.460	0.319	0.154
32	rs34536443	*TYK2*	-0.122	2.721	0.964	-5.339	156.967	0.973	0.255	0.539	0.638
33	rs516246	*FUT2*	0.807	1.102	0.464	-41.290	76.994	0.594	0.475	0.231	0.043
34	rs11203202	*UBASH3A*	-3.366	1.184	0.005	-33.313	89.785	0.711	0.534	0.250	0.036

†IGF-1 models were adjusted for sex and type of milk feeding at age 3 months. C-peptide models were adjusted for type of milk feeding at age 3 months. Leptin models were adjusted for sex and parity.

Table 8-12 | Summary statistics of T1D genetic risk score-age interactions for longitudinal measures of levels of IGF-1 between 3 and 24 months of age, and C-peptide and leptin between 3 and 12 months of age†.

Genetic Risk Score	IGF-1 (ng/ml)			C-peptide (pmol/L)			Leptin (ng/ml)		
	Estimate	SE	*P* value	Estimate	SE	*P* value	Estimate	SE	*P* value
HLA									
HLA Risk_2level	3.264	4.345	0.453	-180.5	331.6	0.587	0.502	0.791	0.527
HLA Risk_3level	2.097	1.491	0.160	-59.5	103.4	0.567	0.474	0.296	0.113
Unweighted									
nonHLA RiskSum	9.706	14.828	0.513	429.0	998.3	0.670	0.971	3.283	0.769
VitD RiskSum	1.031	4.854	0.832	-395.5	373.3	0.293	-2.393	1.116	0.036
Weighted									
nonHLA w-RiskScore	-21.978	54.420	0.687	3257.4	3534.5	0.363	8.784	11.669	0.455
VitD w-RiskScore	3.431	45.056	0.939	-2998.5	3297.7	0.368	-16.826	9.813	0.092
Combined RiskScore	0.319	0.412	0.439	5.6	26.7	0.836	0.134	0.083	0.114

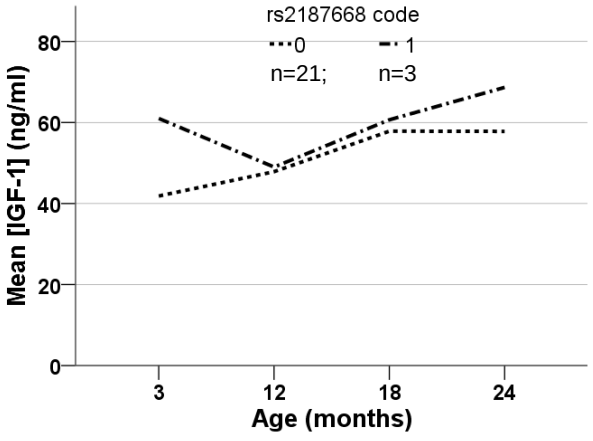
†IGF-1 models were adjusted for sex and type of milk feeding at age 3 months. C-peptide models were adjusted for type of milk feeding at age 3 months. Leptin models were adjusted for sex and parity.

8.3.4 Plots of IGF-1 by genotypes

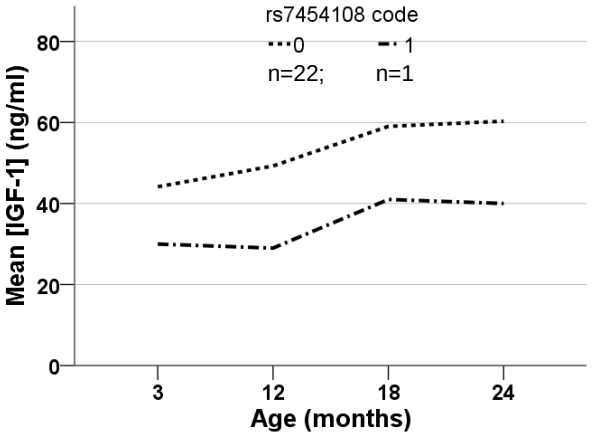
Dynamic or static plots of the mean IGF-1 concentration by T1D-risk genotypic group were generated for significantly-associated SNPs (Figure 8-1). Trajectories were plotted only for infants who completed follow-up, albeit the minimal counts of subjects by genotypic groups for some SNPs owing to the rarity of alleles and small sample who had measured IGF-1 levels.

Figure 8-1 | IGF-1 levels by T1D-risk genotypic group in a. *HLA-DR3*, b. *HLA-DQ8*, c. *CYP2R1*, d. GC, e. *SH2B3*, f. *UBASH3A,* g. *FUT2*, h. *CTSH*. Genotype frequencies are listed in legends.

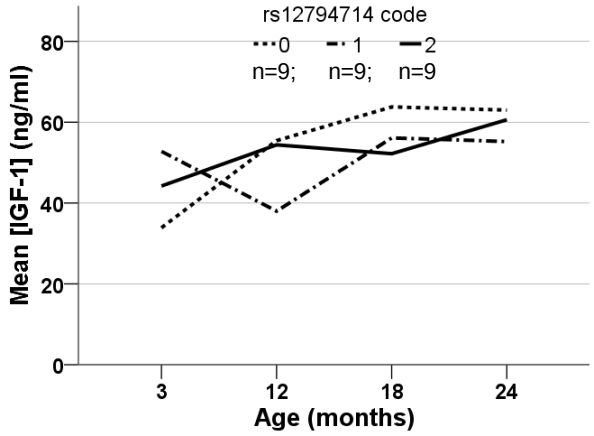
a. *HLA-DR3*



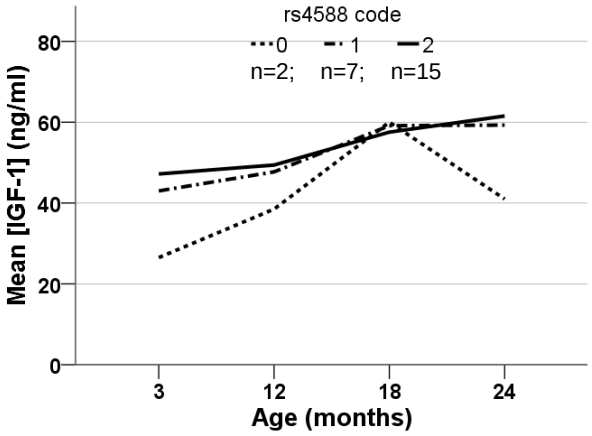
b. *HLA-DQ8*



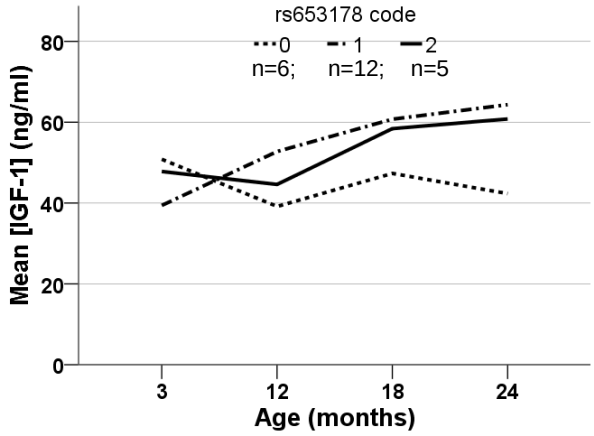
c. *CYP2R1*



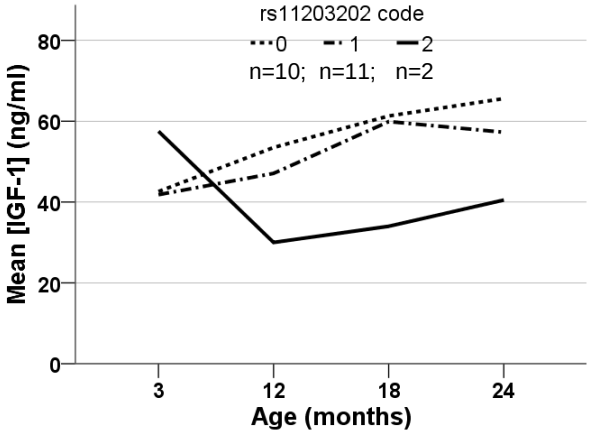
d. *GC*



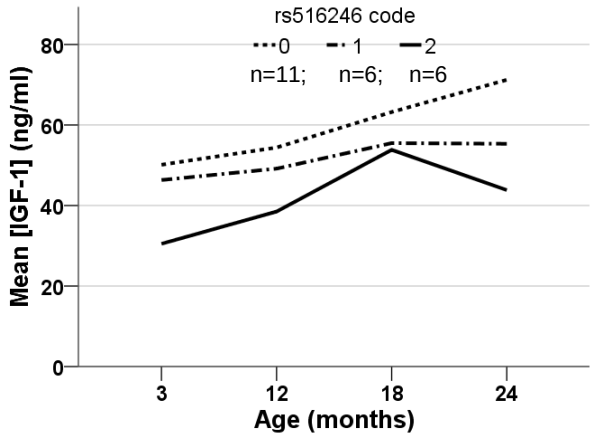
e. *SH2B3*



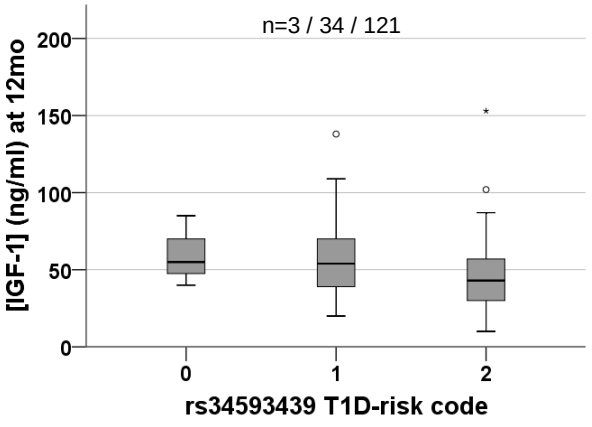
e. *UBASH3A*



g. *FUT2*



h. *CTSH*



8.4 DISCUSSION

8.4.1 IGF-1 in association with T1D SNPs

8.4.1.1 HLA

Homozygosity for the T1D high-risk alleles of the rs7454108 tagging the *HLA-DQ8* associated with lower IGF-1 levels at 12 months (Beta=-0.227, *P*=0.003). The finding here echoes the published association between decreased IGF-1 levels and T1D high-risk HLA genotypes in the DIABIMMUNE cohort[421]. In contrast, my study also showed that the high-risk *HLA-DR3* allele positively associated with IGF-1 levels at ages 18 (Beta=0.183, *P*=0.030) and 24 months (Beta=0.296, *P*=0.009), in reconciliation with the association I found in the CBGS between the *HLA-DR3* minor allele and increased height at 24 months (*P*=0.021, Chapter 6) or linear gains in infancy (*P*=0.007, Chapter 7). This analysis underscores the heterogeneity between the aetiology of 'DR4-associated' and 'DR3-associated' disease[520]. Despite the small sample size, which accounts for the absence of homozygous for high-risk HLA alleles and not allowing for sex-specific analysis, the study provides emerging evidence that the *HLA-DR3* associates with rapid height gains in infancy, more strongly in boys than girls, possibly mediated by IGF-1.

IGF-1 is believed to promote bone formation by inducing osteoblastogenesis and reducing apoptosis of osteoblasts, which are responsible for synthesising the bone matrix (in contrast, osteocytes influence bone structure and osteoclasts promote bone resorption)[574]. It is currently accepted that systemic IGF-1 maintains cortical bone integrity (vs. locally produced IGF-1 which contributes to trabecular bone integrity)[574]. It has been known that IGF-1 and GH stimulate production of proinflammatory cytokines (TNF-α, IL-1β and IL-6) in human osteoblasts[575] as well as immune cells[576,577], and children with chronic inflammatory conditions show elevated antiinflammatory and proinflammatory cytokines in parallel with abnormalities in the GH/IGF-1 axis[578]. In the CBGS, the timing of the HLA-DR3 effects on height gains and IGF-1 levels in mid-infancy might be explained by the timing of GHR expression, which occurs after 6 months of age[299]. Taken together, it is plausible that the HLA-DR3 influences production/bioactivity of IGF-1, resulting in production of proinflammatory cytokines in bone cells and immune cells, which independently affect the outcomes of linear growth and autoimmunity in mid to late infancy. Replication in well-powered samples that allow for sex-specific analysis, and possibly measurement of growth hormone, would be of the essence.

8.4.1.2 *SH2B3*

A significant association was found between the SNP in *SH2B3* and IGF-1 levels at 12 months (Beta=0.197, *P*=0.012). However, it is dubious whether this result is sufficient evidence that IGF-1 mediates the strong significant associations between the SNP in *SH2B3* and cross-sectional measures of skinfold thickness at 18 and 24 months (Chapter 6) or its longitudinal trajectory in infancy (Chapter 7). Nevertheless, considering that the adipose tissue harbours immune cells that regulate inflammation and insulin resistance through cytokine and chemokine secretion[579], I advocate that there is biological plausibility to replicate the association between IGF-1 and the rs653178 in larger population samples and identify the effect of *SH2B3* on IGF-1 by functional studies. My advocacy is reinforced by the strong association of the rs653178 with a range of autoimmune conditions of the joints[102]. In triangulation, systemic and local interactions between IGF-1 and inflammatory cytokines are involved in the pathophysiology of arthritis, which justify the therapeutic effect observed after exogenous administration of GH and IGF-1 to patients with a juvenile form of degenerative arthritis[305,580]. An association between *SH2B3* and IGF-1, if present, could aid our understanding of the involvement of IGF-1 in the development of autoimmunity. GWAS on the development of islet autoimmunity as the outcome measure are lacking, hence cohort studies are instrumental for identifying predisposing genetic variants. The TEDDY study has announced that three regions were associated with the risk of developing any persistent confirmed islet autoantibody, one of which was near *SH2B3*[581].

8.4.1.3 *UBASH3A*

UBASH3A on Chr21q22 encodes Ubiquitin-associated and SH3 domain-containing protein A, which negatively regulates T-cell signalling[582]. There is emerging evidence to suggest that *UBASH3A* is a causal gene for T1D by reducing negative regulation of production of the T1D-protective IL-2[582,583]. Todd[583] contends that this newfound mechanism of action reinforces the disease model according to which the major aetiological pathways of T1D are deficiencies in IL-2 signalling and functions of Treg, which are wholly dependent on IL-2. Non-coding polymorphisms in *UBASH3A* that predispose to T1D have been associated with its increased transcription, which leads to reduced expression of *IL2* by effector T cells[582]. In the CBGS the T1D-predisposing allele (G) on the rs11203202 (C>G) was found to negatively associate with longitudinal measures of IGF-1 levels during infancy (*P*=0.005). In addition to the slower *rate* of change of IGF-1 levels for each T1D-predisposing allele, Figure 8-1 depicts that IGF-1 *levels* were lower amongst homozygous for the T1D-predisposing allele vs. other genotypic groups, but the difference reached statistical significance only at 24 months (*P*=0.026). Taken together, these lines of evidence suggest that the T1D-predisposing polymorphism rs11203202 in *UBASH3A* associates with *slower* change and *lower* levels in IGF-1 during

infancy, and molecularly associates with *increased UBASH3A* transcription which leads to *reduced* transcription of *IL2*. Even though the latter finding was made for the rs112030203 predisposing allele as the variant that upregulates *UBASH3A* transcription, it could also apply to the designated rs112030202 (r^2=0.45) which was reported as the index SNP. The question that ensues is whether IGF-1 is on the causal IL-2-mediated pathway for the development of T1D or merely an association with genetic risk of T1D. Almost 20 years ago, it was shown that IGF-1 increased transcription and protein synthesis of IL-2, with proliferative effects on T lymphocytes[584]. One possible scenario would be that *UBASH3A* suppresses IGF-1 production in infancy, which in turn decreases transcription and protein synthesis of IL-2 (aetiological). Alternatively, *UBASH3A* might exert independent effects on IGF-1 and IL-2 (association). The DAISY study previously found that *UBASH3A* is a robust predictor of progression to islet autoimmunity and diabetes[585]. In this context, my finding in the CBGS could unveil a previously unrecognised causal role of IGF-1 in the development of islet autoimmunity, or suggest that the slow change in IGF-1 levels postnatally serves as a biomarker of disease risk.

8.4.1.4 *CTSH*

The T1D-associated variant rs34593439 on Chr15q25 is mapped to *CTSH*, whose product is Cathepsin H, a lysosomal protein involved in protein degradation in most cell types. In β-cells, Cathepsin H has the dual role in i) cytokine-induced apoptosis and ii) insulin synthesis by acting as a positive regulator of insulin transcription[541]. *CTSH* expression is suppressed by cytokines in islets and β-cells, and carriers of variants causing reduced *CTSH* expression have poorer β-cell function and a faster disease progression than carriers of the variant linked to high *CTSH* expression[541]. In the CBGS, the T1D susceptibility variant within this gene negatively associated with IGF-1 levels at 12 months (*P*=0.005). The findings on both *CTSH* and *UBASH3A* cast light on a negative association between IGF-1 levels in infancy and risk of T1D, which, judging by the functions of the implicated genes, is probably mediated by cytokines.

8.4.1.5 *FUT2*

The association detected in the CBGS between the rs516246 in *FUT2* and IGF-1 levels at 12 months merits attention albeit its nominal significance (*P*=0.045). *FUT2* on Chr19q13 encodes fucosyltransferase 2, also known as secretor transferase, which is an enzyme expressed in epithelial tissue and determines the capacity to secret blood type antigens in body fluids ('secretor'). Homozygous for the non-functional alleles ('non-secretors') do not secrete ABO antigens in body fluids, which has implications for health, such as resistance or susceptibility to specific infections and increased risk for a range of autoimmune diseases, including T1D[586]. Reportedly, 20% of individuals with European ancestry are non-secretors[587]; in the CBGS, 25% of children were inferred to have the 'non-secretor' type. The identification of *FUT2* as a T1D

candidate gene is interesting considering that its expression is modulated by the intestinal microbiota, casting light on the gut as a critical component in the development of autoimmunity. The intestinal bacterial community in non-secretors is characterised by a reduction in the diversity and abundance of Bifidobacteria in faecal samples vs. secretors[588]. In mice, the expression of *FUT2* in the small intestine is absent before weaning and in germ-free animals[589].

In the CBGS, a borderline negative association was detected between the T1D risk-increasing allele on *FUT2* and IGF-1 levels at 12 months, but there was no evidence of an association with anthropometric measures. Yet, triangulation of prior findings indirectly suggests a role of FUT2 in growth. Studies in mice independently demonstrated that microbiota promoted somatic growth by facilitating production and bioactivity of IGF-1[590], and vitamin B12 was found to play a role in regulating postweaning growth and bone formation[591]. A recent GWAS established that inactivation of FUT2 was associated with higher levels of vitamin B12 levels when bound to the carrier protein haptocorrin which is only taken by the liver[592]. Could the putative association between growth and B12 be mediated by FUT2? This is not unlikely; the putative role of vitamin B12 in obesity was found to be an observational association mediated by FUT2 and not causal[593].

8.4.1.6 Is IGF-1 on the causal pathway of T1D?

The involvement of IGF-1 in the pathogenesis of T1D remains elusive to-date. Notwithstanding the lack of a genetic association between IGF-1 and T1D in large-scale studies, it is believed that a decrease in IGF-1 concentrations after birth may relate to the heightened risk of islet autoantibody seroconversion between 6 months and 24 months of age[112,113,115], which could be mediated by IGF-1 replacement therapies that delay onset of autoimmune diabetes[355,594]. However, a study in German prospective cohorts of at-risk children found that concentrations of IGF-1 and IGFBP-3 did not differ between children with and without later islet autoimmunity[355]. My analyses in the CBGS provide credence to a possible association between slower rate of IGF-1 changes, which add to cumulative lower levels of IGF-1, and risk of T1D, exemplified by the association with the putative causal *UBASH3A*. However, the associations with IGF-1 fail to explain the entrenched links between risk of T1D and either breastfeeding or accelerated infancy weight gains. This argument is reinforced by i) the observation that IGF-1 levels were lower in exclusively breastfed vs. formula-fed infants in the CBGS (Chapter 3), which does not reconcile with the finding that exclusive breastfeeding has protective effects, albeit weak, against T1D[595] (unless any possible protective effects conferred by breast-feeding are not mediated by IGF-1); and ii) the inconsistency between T1D susceptibility SNPs that associate with weight and T1D susceptibility SNPs that associate with IGF-1. The novel findings in this study support that the role of IGF-1 in T1D might be searched in the production of IL-2. Alternatively, it has long been known that IGF-1 is endogenously

produced in gastrointestinal secretions, which along with the presence of IGF1R in enterocytes and the expression of IGF-1 mRNA in all regions of the gastrointestinal tract, support its role in increasing intestinal mucosal growth[596], and in effect, mucosal immunity.

8.4.2 C-peptide

There was no evidence of a genetic association with C-peptide in the CBGS. This could be attributed to the sensitivity of insulin secretion to feeding, which was not considered in the study design, i.e. there was wide variability in the timing of blood sampling with respect to the timing of the child's feeding.

8.4.3 Vitamin D metabolism milieu

8.4.3.1 Context

Vitamin D is implicated in the development of T1D, providing justification for the seasonality of diagnosis[78,79] and the 'north-south geographic gradient'[80,81]. Genetic studies identified five SNPs mapped to genes in the vitamin D metabolism milieu (*CYP27B1, DHCR7*, *CYP2R1*, *GC*) that associate with risk of T1D[75-77]. Two of these, the rs12785878 in *DHCR7* and the rs10741657 in *CYP2R1*, showed GWAS-level associations with serum levels of 25(OH)D[529].

8.4.3.2 *GC*

The rs4588 (C>A), a common non-synonymous variant (Thr436Lys) within the *GC* gene that was previously found to confer borderline protection against T1D (OR=0.95, *P*=0.050)[77], significantly correlated with IGF-1 levels at 3 months (Beta=0.166, *P*=0.017) in the CBGS. *GC* encodes DBP, also known as gc-globulin, the binding protein and main carrier of 25(OH)D, which is the main storage form and circulating metabolite of vitamin D synthesised by P450-catalysed hydroxylation in the liver. The levels of 25(OH)D in the serum serve as the primary indicator of vitamin D status because of its long plasma half-life which spans 2 to 3 weeks[597]. Standard analytical methods measure total 25(OH)D comprising free and DBP-bound forms. An inverse relationship exists between DBP levels and bioavailability of vitamin D: higher levels of DBP-bound 25(OH)D is an indicator of decreased vitamin D function[598], possibly owing to the reduced availability of DBP-bound vitamin D to target tissues[599].

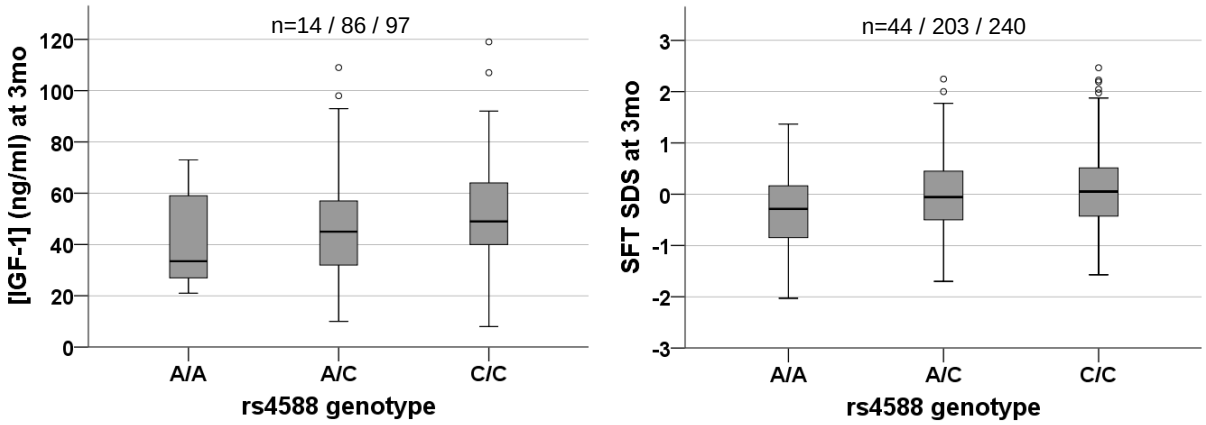
The common rs4588 polymorphism has been established as a strong genetic determinant of circulating DBP and 25(OH)D (independent of its direct effects on DBP) in healthy children aged 6 to 36 months old[600]. At a molecular level, the substitution of a lysine residue for the threonine residue at position 436 eliminates the O-glycosylation site from the

DBP molecule; this structural change affects the half-life of the protein[600]. DBP is produced in the liver and cleared by the kidneys. It serves as the ligand for the megalin/cubulin complex that facilitates entry of bound 25(OH)D into renal tubular cells[601], implying that the *GC* polymorphism may affect levels of 25(OH)D beyond its effect on DBP levels[600].

Considering all lines of evidence, the finding here translates as IGF-1 levels in infancy decrease with progressive substitution of threonine for lysine in the rs4588 (Figure 8-2a), which is known to correlate with suppressed levels of circulating 25(OH)D independent of the direct effect on DBP. In the CBGS, the rs4588 also showed significant associations in the same direction, with weight SDS (*P*=0.026) and skinfold thickness SDS (*P*=0.002, Figure 8-2b), specifically in the triceps (data not shown).

Figure 8-2 | Measures of a. IGF-1 levels and b. skinfold thickness (SFT) SDS by the rs4588 genotypic group at age 3 months. Genotype frequencies are listed in legends (AA / AC / CC).

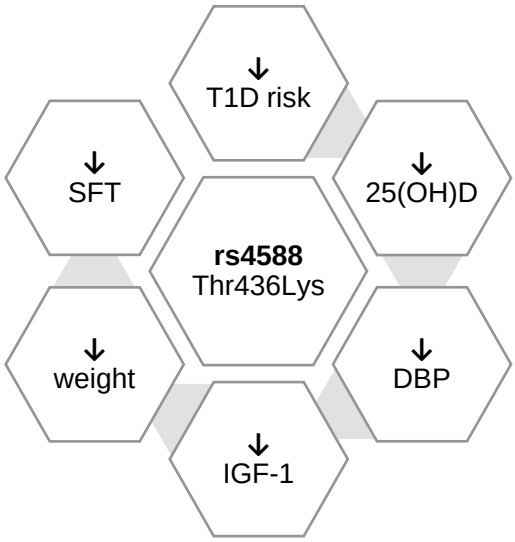
a. IGF-1 levels b. skinfold thickness SDS



Poor vitamin D status has been implicated in the pathogenesis of T1D[602] but it remains unclear whether it is a confounder or mediator in the causal pathway of T1D development. Studies reported an association between lower 25(OH)D levels and risk of T1D in children and young adults[603,604] which was not replicated by the DAISY study[605]. Lower concentrations of DBP were found in patients with T1D compared with their first-degree relatives[606]. A data-driven candidate gene study (eGWAS) discovered that DBP, which is expressed in pancreatic α-cells, is a potential islet autoantigen based on evidence that patients with T1D showed higher levels and frequency of serum autoantibodies against DBP (DBP-Ab) which inversely correlated with 25(OH)D levels vs. controls[607]. A recent study found that lower DBP levels particularly in the third trimester of pregnancy were associated with development of T1D in the offspring[608]. Another study during pregnancy detected no relationship between maternal and foetal 25(OH)D or birth weight but reported that women with lower 25(OH)D in early pregnancy had higher insulin resistance[609]. Figure 8-3 summarises established associations, including the ones found here, with the rs4588. The recently-reported associations of serum levels of

25(OH)D and polymorphisms in the *GC* and *VDR* in the TEDDY study of at-risk children[610] point to the immunolomodulatory effect of vitamin D. Proposed mechanisms for the protective effects of vitamin D include actions on immune cells (downregulates expression of molecules impacting T-cell activation, induces expression of IL-10 secreting Treg cells, imprints T cells with a Treg profile[611]) and direct protection of β-cells from proinflammatory cytokines[612].

Figure 8-3 | Novel and reported associations with the rs4588 polymorphism in infancy or adulthood.



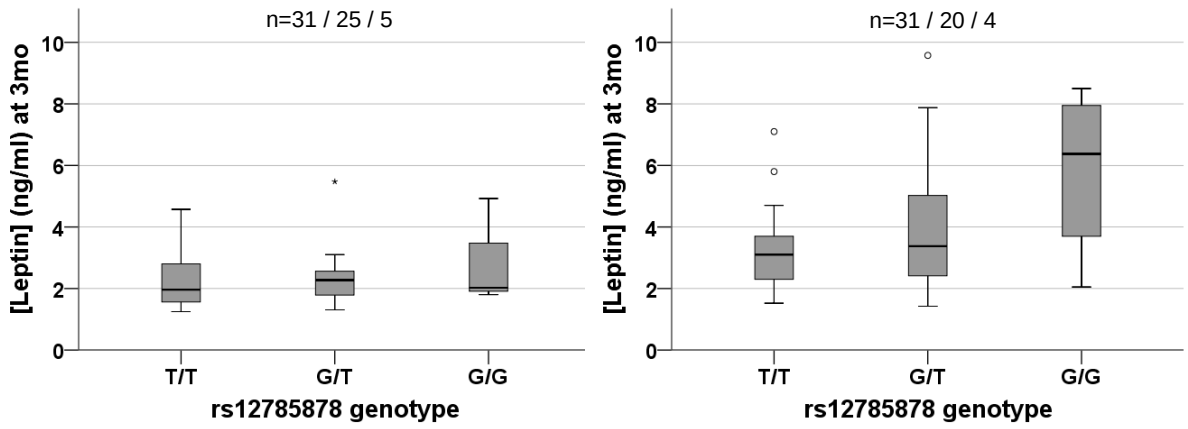
The independent associations found here between the rs4588 and both IGF-1 and skinfold thickness SDS at 3 months may provide insights into mechanisms of growth in early life. Cross-sectional analysis by sex unveiled that the relationship between the rs4588 and skinfold thickness was evident in boys (P=0.005, n=281) but not in girls (P=0.124, n=236), which is unlikely explained by differences in statistical power. However, this SNP did not show a sex-based association with IGF-1 levels. In the CBGS, measures of skinfold thickness at 3 months significantly correlated with IGF-1 and leptin at the same age (Chapter 3). The molecular interplay between IGF-1 and leptin has been demonstrated experimentally in animal tissues: a bidirectional crosstalk exists where IGF-1 induces phosphorylation (activation) of leptin receptor and leptin phosphorylates IGF1R[613]. Leptin was found to regulate hepatic IGF-1 production independent of GH[614]. In the clinical setting, acquired and congenital hypoleptinaemia is followed by a decrease in IGF-1 levels which are restored by leptin treatment[615]. Thus, there is ground to hypothesise that leptin mediates the sex-specific association with adiposity found in the CBGS. However, the rs4588 did not show evidence of a relationship with leptin levels at 3 months. Unless sample size limits associations with leptin, these lines of thinking open up the possibility that the rs4588 influences levels of IGF-1 and adiposity independently and mediates its effect on the latter via a male-sex hormone. Conclusively, the findings suggest that phenotypes which associate with this genetic variant feature lower levels of vitamin D and higher adiposity in boys.

8.4.3.3 *DHCR7*

Inverse associations between adiposity measures (e.g. BMI and body fat content) and serum 25(OH)D were identified in cross-sectional studies in children, adolescents[616], as well as obese and healthy adults[617-620]. The underlying mechanistic pathway is unclear but has been attributed to i) decreased exposure to sunlight due to restricted mobility, ii) sequestration in white adipose tissue, the major storage site for vitamin D, or iii) negative feedback from increased levels of 25(OH)D on hepatic synthesis of 25(OH)D[617,621]. A prospective longitudinal study later elucidated that the associations between adiposity measures and 25(OH)D levels were mediated by leptin[622]. The finding resonated with a prior *in vitro* tissue culture model that demonstrated that leptin secretion by human adipose tissue was negatively and powerfully controlled by 25(OH)D[623]. The negative regulatory effect of 25(OH)D on leptin concentrations was reinforced by a recent study which reported that women with levels of 25(OH)D >100nM showed relatively lower leptin levels across a broad range of BMI[624]. Leptin is a hormone produced by adipose tissue and secreted into the bloodstream, whose circulating levels are determined by body fat mass, thus qualifying as a candidate mediator of the associations between adiposity and 25(OH)D. My study lends itself to this investigation as it provides access to a number of vitamin D metabolism-related SNPs.

The rs12785878 (T>G) within *DHCR7* on Chr11q12 associated with islet autoimmunity[625], T1D[77] and multiple sclerosis[626]. In the CBGS, leptin levels at 3 months were found to be significantly higher with progressive addition of the autoimmunity risk-increasing G allele (vitamin D-decreasing allele) of the rs12785878 (*P*=0.009) in a SNP dose-dependent manner (Figure 8-4). This finding has not been reported previously. The association was evident in girls but absent in boys despite their equal proportions in the sample analysed.

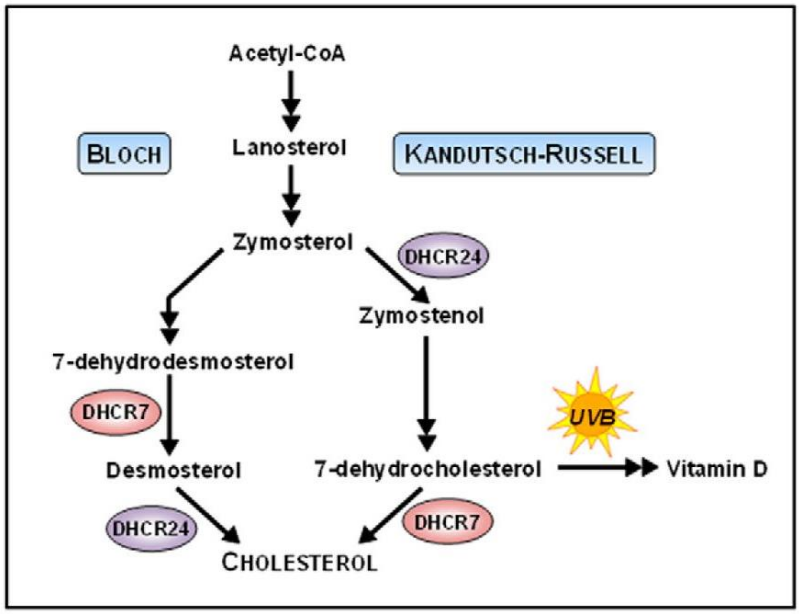
Figure 8-4 | Leptin levels at age 3 months by the rs12785878 genotypic group in a. males, b. females.



The *DHRC7* encodes dehydrocholesterol reductase, which acts as a 'switch' in the cross-talk of cholesterol synthesis and vitamin D production; vitamin D impedes cholesterol

production by reducing DHCR7 protein expression in a dose-dependent manner, and conversely, the highly labile nature of DHCR7 is further destabilised by cholesterol driving the flux towards vitamin D synthesis (Figure 8.5)[627]. It is proposed that polymorphisms in *DHCR7* linked to higher vitamin D status are a recent evolutionary adaptation that exerted selective pressure during early human migration to Northern latitudes of low sunlight as a protective mechanism against vitamin D insufficiency[628]. My finding in the CBGS points to the genetic constitution of the vitamin D metabolism milieu playing an important role in the 'north-south geographic gradient' of the incidence of autoimmune diabetes.

Figure 8-5 | 7-DHCR7 is a substrate to cholesterol or vitamin D. From Prahbu et al.[627].



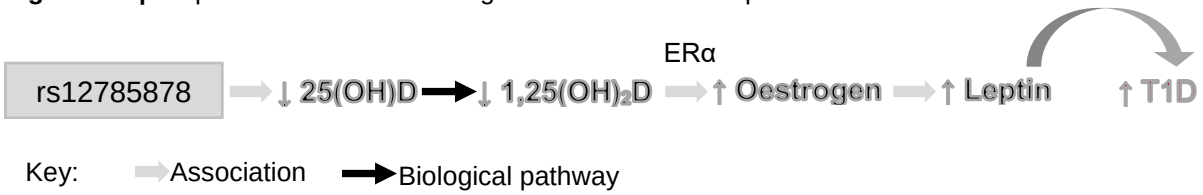
8.4.3.4 Proposed mechanism

The novel finding of the sex-specific association between leptin levels and the rs1278747 in *DHCR7* has the potential to unveil a non-previously reported mechanism behind the triangle of vitamin D-leptin-T1D. A prior systematic literature review of observational studies concurred on an inverse relationship between leptin and 25(OH)D, albeit not reflected in intervention studies of high heterogeneity[629]. Association studies fall short of providing a definitive answer to the question 'which is the cart, which is the horse'. An *in vitro* study demonstrated that leptin secretion by human adipose tissue was negatively and powerfully controlled by 25(OH)D[623]. Interestingly, a molecular study in ovarian cancer cells discovered that the anti-leptin activity of vitamin D is mediated by downregulating the human telomerase reverse transcriptase mRNA expression and cell growth, which occurs through oestrogen receptor-alpha activation (ERα)[630]. Of note, women with T1D have a significantly higher risk of ovarian cancer[631], which could point to ERα as a candidate underlying factor shared by these comorbidities. The ERα is a 'classical' oestrogen receptor discovered in 1958, which structurally belongs to the nuclear

receptor superfamily and functionally acts as a ligand-activated transcription factor that mediates the known effects of oestrogens[632]. It is encoded by a gene on Chr6q25. ERα is expressed in most cells of the immune system[633] as well as in human β-cells[634]. Tiano and Mauvais-Jarvis[634] reviewed the effect of oestrogen receptors on the β-cell, and deciphered the presence of sexual dimorphism in diabetes with β-cell failure in rodent models, that is female animals were protected from the development of the disease. The authors argued that activation of ERα promotes β-cell survival and insulin biosynthesis[634], with the female steroid 17β-oestradiol affecting pancreatic β-cell function in mammals by acting through all ER receptors[635]. Vitamin D plays a role in breast cancer and has been found to suppress the production of leptin, which is involved in downstream cell signalling pathways in oestrogen-dependent cell types[636]. The involvement of oestrogen in the leptin-vitamin D pathway could explain the sex-dimorphic nature of my finding, but the mechanism remains elusive. The abovementioned molecular analysis showed that vitamin D suppresses leptin activity whereas the results here allude to vitamin D suppressing leptin production.

Recently, an interventional study in women reported that high blood levels of vitamin D associated with reduced oestrogen independent of weight loss[637]. I identified a molecular study that might provide context and direction to this reported association. In breast adipose tissue, calcitriol downregulated the expression of ERα[638]. Oestrogen was previously shown to increase the production of leptin in women[639]. These lines of evidence lead me to suggest that the rs12785878 minor allele associates with reduced levels of 25(OH)D, the precursor of calcitriol in the liver, with ensuing lower levels of calcitriol resulting in reduced downregulation of ERα that leads to increased oestrogen activity and leptin production (Figure 8-6).

Figure 8-6 | Proposed mechanism linking the rs12785878 to leptin.



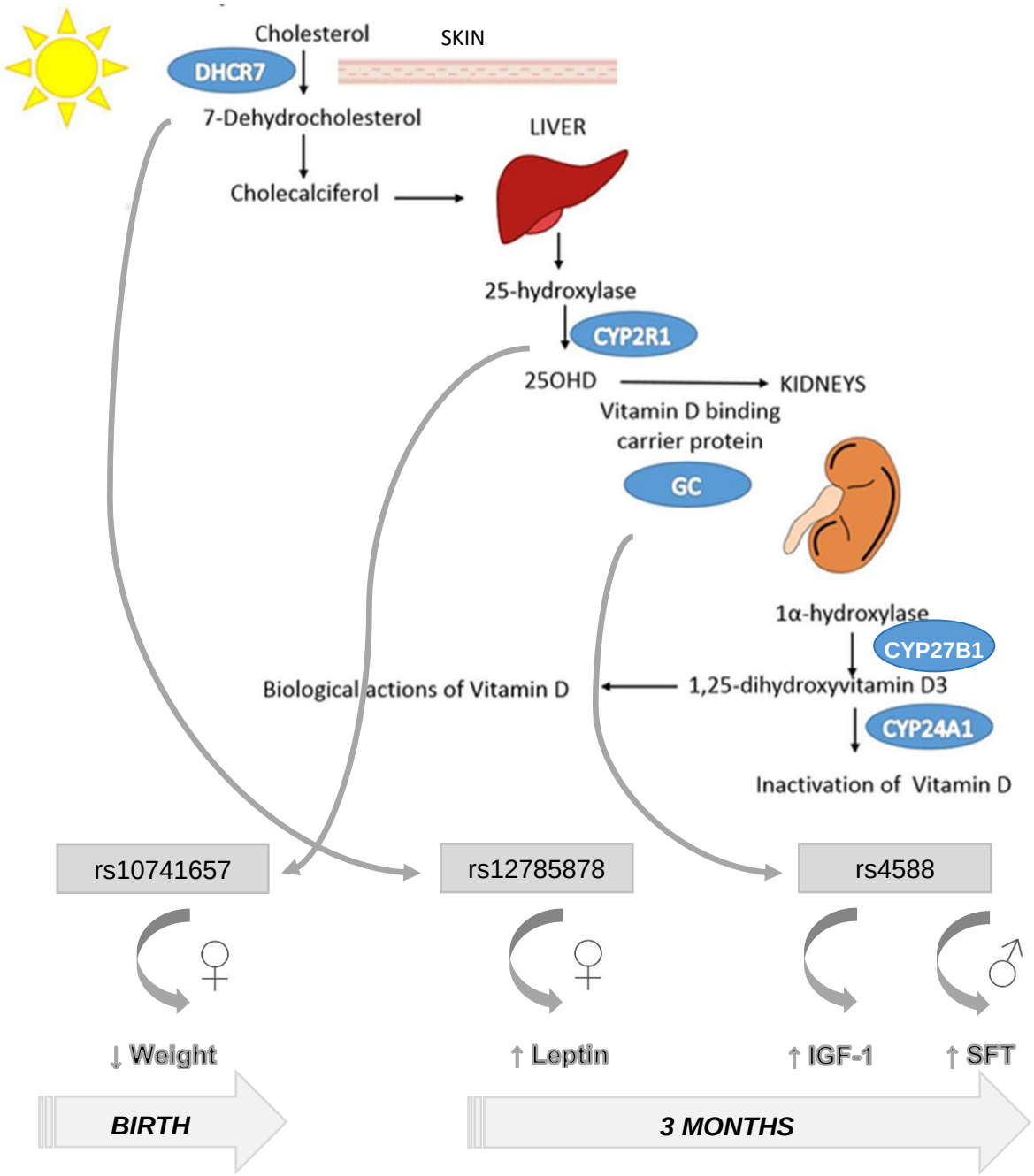
Interestingly, the T allele (vitamin D-increasing allele) of the rs12785878 in *DHCR7* was previously associated with greater decreases in insulin and HOMA-IR (P<0.002) in response to high-protein diets in adults[640]. Considering that leptin decreases insulin synthesis and secretion by pancreatic β-cells[641], I posit that the effect on insulin secretion and IR observed in that study might be mediated by leptin. However, could these lines of evidence drawn from studies in adults explain the existence of sex differences before puberty? In fact, a surge of sex hormones is observed within the first few months of life in both male and female infants, a process known as 'mini-puberty' and marked by an increase in testosterone in boys and an increase in testosterone and oestradiol in girls[642], reinforcing the possibility that oestrogen is implicated in the association between leptin and the rs12785878 in the early life of females.

8.4.3.5 The emerging role of vitamin D in early growth

Leptin is a cytokine-like hormone produced in the fat tissue that modulates the adaptive immune system. Elevated leptin levels increase susceptibility to autoimmunity and chronic inflammation by an increase and expansion of $CD4^+$ T-cells, decrease in the proportion of Treg cells, and increased secretion of proinflammatory cytokines[643]. A study in adults found that levels of 25(OH)D negatively correlated with leptin in women and positively correlated with IGF-1 in men[644]. It is thus reasonable to question if other genetic variants in the vitamin D milieu show a male-biased association with IGF-1 levels. This study has not identified such an association but detected a male-specific effect of the rs4588 in the *GC* gene on adiposity, which also associated with IGF-1 across boys and girls. On closer inspection, the additional association found here between the rs10741657 (known to correlate with 25(OH)D at GWAS-level) and birth weight (Chapter 6) was evident in girls (P=0.018) but not in boys (P=0.266). As said, previous observational studies reported an inverse U-shaped association between birth weight and vitamin D levels in the neonate, and the risk of SGA was reduced at moderate concentrations[525]. An inverse U-shaped relationship is characteristic of the association between IGF-1 and BMI, in adults at least, meaning IGF-1 positively associates with BMI in normal weight subjects, but is low in both thin and obese individuals[528,645]. Conceivably, tightly regulated biological processes across the board might favour a narrow window of physiological levels of endocrine regulators. Alternatively, it is possible that the U-shaped relationship observed for vitamin D and birth weight masks the involvement of IGF-1 early in life. Ong et al.[646] previously identified strong positive correlations between cord blood IGF-1 and size at birth in non-primiparous pregnancies. That opens up the possibility that the association between IGF-1 and size at birth is not linear across all types of pregnancies.

Collectively, the novel associations found here between polymorphisms in the vitamin D metabolism milieu and infancy adiposity or growth regulators (Figure 8-7) convey a possible role of this prohormone in growth pathways that involve IGF-1 and leptin in a sex-specific manner. Sex hormones are believed to mediate many of the sex-based differences in the immune response of females, who are known to show more resilient immune reactions than males in whom inflammation is more severe[647]. Sex-dimorphic differences of immunity also exist in infancy, with scarce evidence suggesting that boys develop more robust innate immunity than girls[392]. The results of my study imply that an altered immune-endocrine axis plays a critical role in the genetic susceptibility to T1D and cast light on leptin, whose production might be influenced by vitamin D in a sex-specific manner during a critical window of endocrine and immune development in humans.

Figure 8-7 | Associations found for SNPs in the vitamin D metabolism pathway (direction of change is with respect to the T1D-risk increasing allele). Illustration adapted from Mokry et al[626].



8.4.4 Limitations

Hormones were measured from DBS by adaptation of standard serum assays with slightly different detection limits. The samples used were stored at -20 °C and subjected to some freeze-thaw cycles. Linear longitudinal analyses on C-peptide and leptin were conducted on two timepoints, questioning the validity of their model fit. Blood levels of hormones, specifically IGF-1, were not available at birth to test for associations with SNPs in *TYK2* and *CYP2R1.*

Chapter 9

FINAL DISCUSSION

9.1 CONTRIBUTION TO RESEARCH

This study was inspired by the association between T1D development in childhood and presymptomatic growth alterations in early life, with the dual objective of identifying novel genetic contributors to physiological early growth as well as phenotypic markers of genetic risk of T1D. The study was strengthened by access to a range of longitudinal anthropometric measures and growth regulators, along with a panel of 34 biology-screened T1D susceptibility loci, which allowed for insightful biological relationships to emerge (Figures 9-1 and 9-2). Overall, the study concept, design and outcomes have underscored the pivotal role of genetic studies — especially when the rationale is underpinned by biological knowledge — in illuminating molecular pathways in health and disease and reducing commonly observed associations to pleiotropy. The concept of pleiotropy is one we still grapple with, possibly predicated on the conventional belief which isolates biological processes by anatomical systems and chronological age; genetics unveil that molecular pathways have no anatomic barriers and biological processes commence as early as *in utero* and evolve on the continuum of life at a slow rate.

With respect to growth, the study confirmed the intrauterine growth-limiting effects of twinship, short gestational length, smoking during pregnancy, primiparity, as well as the persistent impact of breastfeeding — most frequently practised by mothers with higher academic achievements — on containing excesses of gains throughout infancy. It corroborated the sex-dimorphic nature of growth trajectory, particularly adiposity, in correlation with known endocrine regulators. Arguably, the main contribution made is casting light on the possibility of a cross-talk between vitamin D and leptin, and possibly other growth regulators, which is speculatively mediated by sex hormones. The novel strong significant association between adiposity and *SH2B3*, a locus known to be implicated in cardiovascular-related pathways, informs of possible biomarkers of at-risk individuals as early as infancy and does justice to the concept of the early origins of adult disease.

With respect to the genetic risk of T1D, this study conveyed the message that immune candidate genes, known to predispose to islet autoimmunity, associate with anthropometric measures in infancy, such as height gains and adiposity. The investigation here makes the strong case that the positive association between risk of T1D and birth weight is probably the most consistent amongst all anthropometric measures. Beyond birth, significant associations between any index of growth and T1D loci have discordant directions of effect. Even though the study was not sufficiently powered to detect minuscule genetic effects, it illuminated *TYK2*, a strongly suspected T1D causal gene, as a potential contributor to alterations in birth weight. Birth weight by viable genotypes at this locus follows a bimodal distribution, with children at genetically lower risk of T1D weighing on average 0.22 SDS less than their at-risk counterparts. Such magnitude of effect is in line with the contention that differences in birth weight as a risk factor for T1D are within physiological range. Beyond infancy, adiposity takes the preponderance over weight in associations with genetic susceptibility to T1D, hinting at the common denominator of cytokine signalling between adipose tissue (obesity) and development of autoimmunity. Umbilical cord blood leptin is currently accepted as a biomarker of neonatal fat mass[323], addressing the limitation of birth weight as an anthropometric index that does not differentiate between fat and non-fat mass. Conceivably, leptin in infancy could be investigated as a risk factor for autoimmune diabetes. Contrary to expectations, significant associations were detected with polymorphisms shared by other autoimmune diseases, implying that rapid early growth is a phenotype not unique to T1D, or environmental factors (including epigenetic effects and viruses) account for the accelerated growth in prediabetic children. The results here highlight the critical importance of the first 3 months of life, in growth and risk of disease, when T1D susceptibility polymorphisms in non-coding regions are possibly differentially expressed via interaction with the environment.

Figure 9-1 | Negative logarithm of *P* values of longitudinal associations (birth to 24 mo of age) between SNP and skinfold thickness SDS (y-axis) vs. SNP and weight SDS (x-axis) for each of 34 SNPs in the study; size of bubble represents the OR for T1D risk alleles at each of the indicated loci.

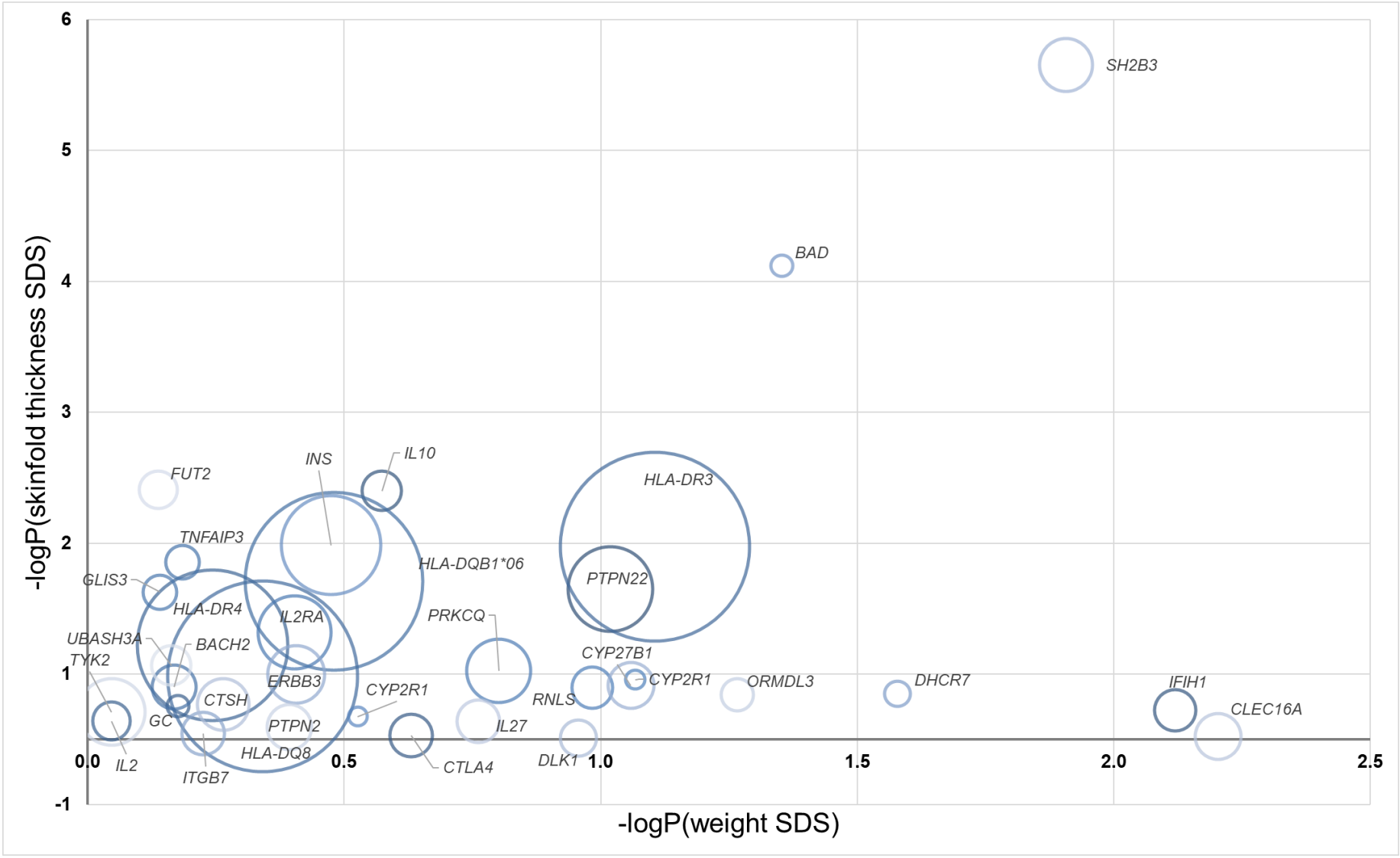
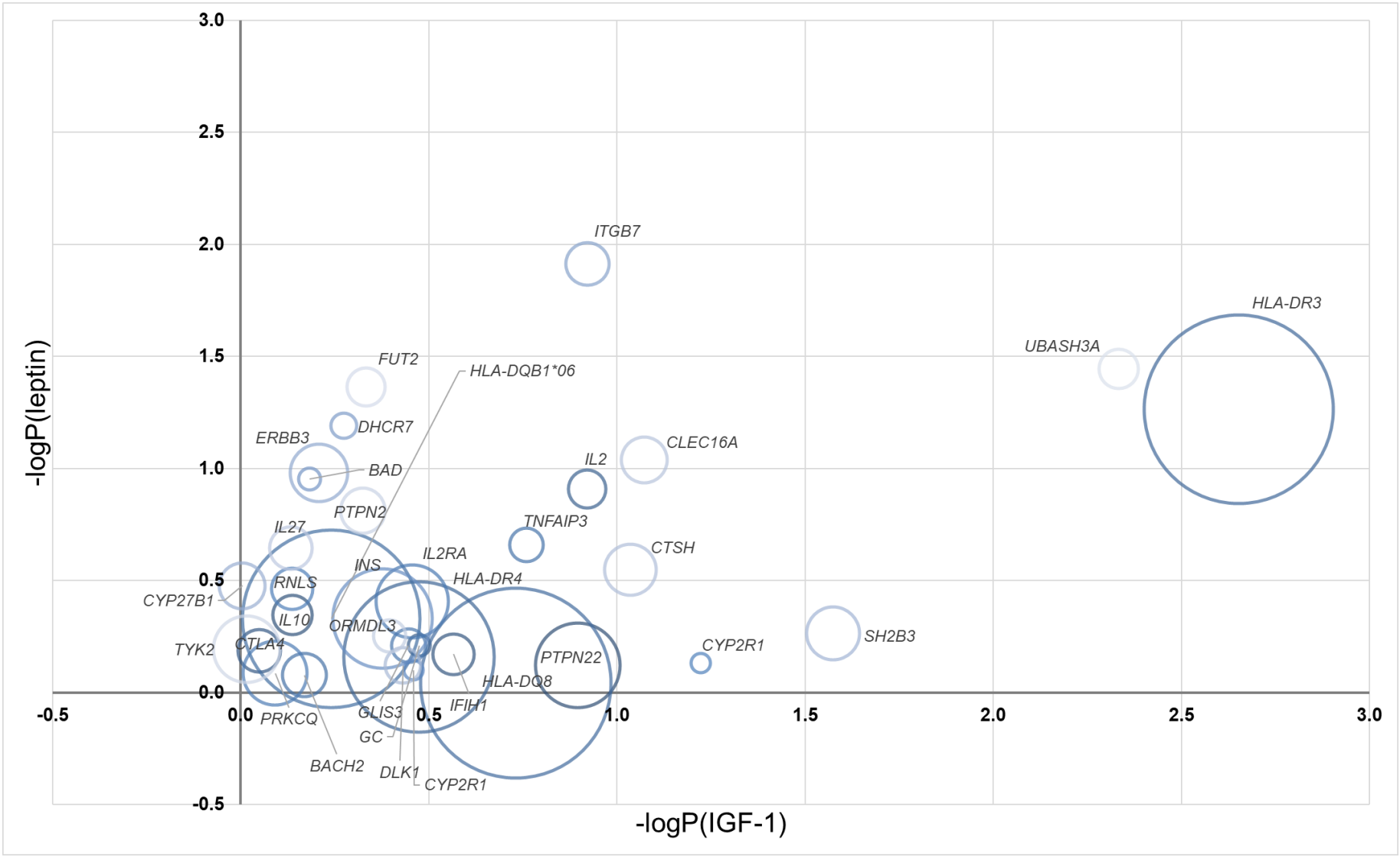


Figure 9-2 | Negative logarithm of *P* values of longitudinal associations between SNP and leptin levels from 3 to 12 mo of age (y-axis) vs. SNP and IGF-1 levels from 3 to 24 mo (x-axis) for each of 34 SNPs in the study; size of bubble represents the OR for T1D risk alleles at each of the indicated loci.



9.2 IMPLICATIONS

The wealth of data in this thesis has served to corroborate the idea that T1D is a heterogeneous disease. At the first level, pathogenic mechanisms might operate differentially between males and females, evidenced by the sex-dimorphic associations identified between growth and polymorphisms in the *HLA-DR3*, *DHCR7*, *CYP2R1* and *GC*. At the next level, different susceptibility loci map out to different pathways, largely exemplified by the differences between the DR4- and DR3-associated phenotypes. Phenotypes clustered by gender and SNPs identified in this thesis allow for a preliminary arrangement of genetic elements that mediate the disease with the potential of taking investigations to the next level of precision medicine.

The findings of this study have implications beyond T1D. It provides ground for adding to the biological plausibility of the recommendation for vitamin D supplementation in individuals at risk of autoimmune diseases, particularly multiple sclerosis, which also associates with the rs12785878 in *DHCR7*. Based on the novel dose-dependent association between this SNP (known to also associate with vitamin D levels) and leptin levels, it is reasonable to hypothesise that vitamin D supplementation might exert dual effects by increasing serum vitamin D levels (protection enhancement) and possibly suppressing leptin (risk aversion). Considering the sex-dimorphic nature of the association found here, the question that merits investigation is whether men at risk of multiple sclerosis would benefit from vitamin D supplementation to the same extent as women at risk, which could be addressed by interventional studies assessing outcomes by sex.

9.3 FUTURE DIRECTION

It is becoming clear that the field of immune-mediated diseases transitions from GWAS to the post-GWAS era where the goal is to define causal variants and understand their effect on transcriptional regulation and function by cell types[435]. Promising possibilities arising from this study include i) epigenetics to decipher when and how gene expression is regulated in infancy, ii) functional studies to probe for mechanisms behind the effect of T1D candidate genes on adiposity and leptin, and the mediating role of sex hormones, and iii) population studies to assess the effect of adiposity during infancy and track its development alongside the development of islet autoimmunity by sex. The progress made by this thesis in fine-tuning polymorphisms of importance and elucidating the genetic elements of the sex-dimorphic nature of early life phenotypes of T1D genetic risk could hopefully contribute to deciphering the incipient foetal origins of the disease.

APPENDIX I

A-I-1a | Summary of single-centred prospective cohort studies of islet autoimmunity (IA) and/or T1D.

Cohort	Centre	Years	Risk*	Description
BABYDIAB	Germany	1989-2000	F	Examined natural history of IA in 1,650 children born to a parent with T1D.
BABYDIET	Germany	1989-2006	F, G	Randomised 150 newborns with a first-degree relative with T1D and T1D risk HLA genotypes to gluten exposure at 6 or 12 mo of age.
DIPP	Finland	1994-	G	The Type 1 Diabetes Prediction and Prevention recruits newborns based on genetic screening (moderate or high risk) and follows them up until 15 yr of age.
DiPiS	Sweden	2000-2004	G	The Diabetes Prediction in Skane is a population-based study in Skane (most southern province of Sweden) that recruited 35,658 children born during the designated period for HLA typing and identified > 2,500 at risk children that were followed annually.
DIABFIN	Italy	1999-	G	Population-based cohort study that recruits neonates in three cities in continental Italy (Rome, Milan, Genoa) to identify those with high and moderate HLA risk genotypes for T1D and assess environmental factors contributing to T1D.
DAISY	USA (Colorado)	1993-2009	F, G	The Diabetes Autoimmunity Study in the Young recruited newborns carrying T1D high-risk HLA genotypes, and siblings or offspring of people with T1D after 9 mo of age.
AUSTRALIAN BABY DIAB	Australia (South)	1990s-	F	Birth cohort recruiting infants who have a first-degree relative with T1D and follows them from birth with 6-month reviews.
ENDIA	Australia	2013-	F	Pregnancy/birth cohort that investigates the determinants of T1D in at-risk children by recruiting 1,400 unborn infants less than 6 mo of age with a first-degree relative with T1D. It investigates relationships between genotype, the development of IA, and prenatal and postnatal environmental factors.

* F: familial, G: genetic

Table A-I-1b | Summary of multi-centred prospective cohort studies of IA and/or T1D.

Cohort	Centre	Years	Risk*	Description
DIABIMMUNE	Finland, Estonia, Karelia	2008-	G	Birth cohort of infants with increased genetic risk of autoimmune disease followed up from birth up to the age of 3 yr and tested for organ-specific autoantibodies, allergies, infections, gut microbiota and nutritional factors.
TEDDY	Finland, Germany, Sweden, USA	2004-2010	F, G	The Environmental Determinants of Diabetes in the Young was a multicentre cohort across six clinical centres with the purpose of identifying environmental factors that trigger or protect against islet autoimmunity. It recruited 7,751 children from the general population and 917 with a first-degree relative with T1D.
TRIGR	US, Canada, Australia, Europe (12)	2002-2007	G	The Trial to Reduce IDDM in the Genetically at Risk was an international double-blinded trial to test whether hydrolysed infant formula compared to cow's milk-based formula decreases risk of T1D in children with increased genetic susceptibility.

* F: familial, G: genetic

Table A-I-2 | Association studies of T1D susceptibility HLA and size at birth (1/2). BW, birth weight; BL, birth length, HrBW, high relative birth weight.

First author Year	Country Study	N	Birth years	Design	No of HLA groups	Results
Stene 2001	Norway	969	1982-1998	Retrospective (mouth swab)	5	– Homozygous DQB1*06:02 had highest BW vs. DQ2/8 who had lowest BW. Mean ΔBW=354 g (95% CI 105 to 604).
Aroviita 2004	Finland	1,263	1999-2001	Retrospective (cord blood)	1	– DRB1*13 was more frequent among highest BW vs. lowest BW. Δ(median BW)=40 g, *P*=0.044.
Larsson 2005	Sweden DiPiS	16,709	2000-2003	Prospective (cord blood)	7	– DQ2/8, DQ8/604, DQ8/X had higher frequency of HrBW vs. other HLA types (OR=1.20 [95% CI 1.09 to 1.60], *P*=0.0006). – DQB1*06:03 had increased frequency of HrBW (OR=1.13 [95% CI 1.02 to 1.28], *P*=0.004).
Stene 2006	Norway	471 cases; 1,369 controls	1985-1999	Case-control (cord blood)	4	– Hint of DQB1*06:02 carriers having higher BW. – Trend of low BW and T1D risk, adjusted for *HLA*, *INS*. – No interaction between BW and *HLA* (*P*=0.62) or *INS* (*P*=0.83).
Larsson 2007	Sweden DiPiS	2,848	2000-2004	Prospective (cord blood)	7	– DQ8/DQ2 carriers had higher frequency of HrBW. If infections in 2nd or 3rd trimesters, the OR was aggravated (OR=5.24 [95% CI 1.75 to 15.7], *P*=0.003).

First author Year	Country Study	N	Birth years	Design	No of HLA groups	Results
Locatelli 2007	Italy DIABIN	4,349	1999-	Prospective (cord blood)	3	– No independent association between HLA and BW or BL. – Length of gestation inversely correlated with HLA risk.
Larsson 2008	Sweden DiPiS	58 cases; 155 controls	2000-2004	Prospective (cord blood)	2	– High-risk HLA increased BL SDS (*P*<0.010) but no significant association with BW SDS.
Jarvinen 2008	Finland	342	1987-1999	Prospective	684 haplotypes	– Carriers of Finnish HLA haplotype (n=20) had higher BW (3,925 g, SD=446) vs. non-carriers (3,676 g, SD=439). – DR13 carriers tended to be heavier vs. non-carriers.
Stene 2011	Norway	456-463 cases; 1,377-1,401 controls		Case-Control (mouth swab)	5	– Relative risks conferred by *HLA*, *INS*, *PTPN22* were independent of perinatal factors except for *PTPN22* and mode of delivery.
Sterner 2011	TEDDY	5,461	2004-2010	Prospective	4	– Swedish infants were longer at birth if carrying DQ2/8 (*P*=0.023) or DQ8/8 (*P*=0.046) vs. DQ4/8.
Peet 2012	DIABIMMUNE	7,931	2008-2010	Prospective (cord)	4	– No association between T1D HLA risk and BW.
Peet 2014	DIABIMMUNE	235 (Est) + 261 (Fin)	2008-2010	Prospective	4	– High-risk HLA carriers had lower BW SDS vs. controls at 12 mo (0.37 SDS [95% CI 0.09 to 0.66] vs. 0.92 SDS [95% CI 0.78 to 1.1], *P*=0.001). – Length SDS: lower in high-risk HLA at 12 mo and 24 mo.

Table A-I-3 | Association studies of postnatal growth and IA and/or T1D (1/5).

First author Year	Country Study	N	Design	BW	ΔW	ΔH	Δ BMI	Results
Baum 1975	UK Children's Diabetic Clinic in Oxford	25 cases; 80 controls	Retrospective case-control		↑			– Prediabetic males (8.71 kg, SD=1.20, n=21) were heavier at 6 mo vs. controls (8.10 kg, SD=0.99, n=40), *P*<0.05; females (10.65 kg, SD=0.88, n=9) were heavier at 1 yr of age vs. controls (9.83 kg, SD=0.99, n=40), *P*<0.05.
Blom 1992	Sweden Swedish Childhood Diabetes Register	337 cases, 517 controls	Retrospective		↔	↑		– Mean height SDS after birth was consistently higher in prediabetic vs. referent boys (albeit overlap of 95% CI), but no consistent differences in girls. – Weight SDS showed similar patterns for diabetic and referent children of both sexes.
Johansson 1994	Sweden 5 paediatric departments (1974-1988)	297 cases, 192 controls	Retrospective		↑	↔		– Weight gain from birth to 30 mo was greater in probands vs. referents (*P*<0.02) with the difference up to 18 mo and 30 mo was more marked in girls (0.54 vs. 0.21 SD and 0.56 vs. 0.14 SD, *P*<0.05) than in boys (0.60 vs. 0.44 SD and 0.60 SD vs. 0.44 SD, non-significant).
Hypponen 1999	Finland Childhood Diabetes in Finland (1986-89)	435 cases 386 controls	Retrospective nationwide	↔	↑			– T1D girls grew faster vs. controls in infancy; ΔW increased from 111 g (95% CI 0 to 218), *P*=0.04 at 1 mo to 286 g (95% CI 123 to 450), *P*=0.0006 at 7 mo after birth. For boys, the ΔW was stable in infancy (95 g [95% CI -2 to 205], *P*=0.09).

First author Year	Country Study	N	Design	BW	ΔW	ΔH	Δ BMI	Results
Hypponen 2000	Finland Childhood Diabetes in Finland (1986-89)	586 cases 571 controls	Retrospective nationwide	↔	↑	↑		– Prediabetic boys and girls were heavier than controls: <2 yr of age, ΔrW in boys was 1.0% (-0.2 to 2.1, *P*=0.09), and in girls 1.5% (0.3 to 2.6, *P*=0.01); from 2 to 9.9 yr of age, ΔrW in boys was 2.7% (1.4 to 4.1, *P*<0.001), in girls was at max 5.6% (2.9 to 8.2, *P*<0.001) at 8 yr of age. – ΔrH between diabetic and control boys was 0.26 SDS (0.12 to 0.41, *P*<0.001) for < 2 yr of age; and 0.23 SDS (0.08 to 0.38, *P*=0.003) from 2 to 9.9 yr of age. In girls, ΔrH was 0.19 SDS (0.01 to 0.37, *P*=0.03) from 6 mo to 1.9 yr of age; and 0.28 SDS (0.12 to 0.44, *P*<0.001) from 2 to 9.9 yr of age].
Bruining 2000	Netherlands	91 cases 125 healthy siblings 2,151 controls	Retrospective family & population-based	↔		↑	↑	– BMI from birth to 1 yr of age increased in probands (+0.31 SDS, *P*=0.002) vs. siblings (0.10 SDS, *P*=0.17). – Length from 1 to 3 yr of age increased in probands (0.49 SDS, *P*=0.001) vs. siblings (0.47 SDS, *P*=0.001).
EURODIAB 2002	Europe (5 centres)	616 cases 1,616 controls	Retrospective		↑	↑	↑	– Height SDS and Weight SDS increased in probands from 1 mo after birth to a max of 0.32 (*P*<0.001) and 0.41 (*P*<0.001) respectively between 1 to 2 yr of age. – BMI: largest Δ of 0.27 SDS from 1 to 2 yr of age (*P*=0.002).

First author Year	Country Study	N	Design	BW	ΔW	ΔH	Δ BMI	Results
Svensson 2007	Denmark	490 cases 696 controls	Prospective population-based	↔	↔	↔	↔	– No association between T1D and linear increase in any growth parameter.
Ljungkrantz 2008	Sweden Diabetes and environment around Baltic Sea	316 cases 1:2 controls	Case-control	↔	↑	↑	↔	– Prediabetics had more pronounced weight and height gains vs. controls with the differences being more pronounced for girls. Children who developed diabetes <5 yr of age, gained more weight than controls between 3mo to 3 yr after birth (0.76 vs. 0.33 SDS, *P*<0.01). – Prediabetics had similar BMI from birth to 3 yr of age vs. controls.
Lammi 2009	Finland	218 case-control pairs	Retrospective record-linkage				↑	– Max BMI <3 years of age per 1 kg/m^2 (OR=1.19 [95% CI 1.03 to 1.36], *P*=0.02).
Couper 2009	Australia Australian Baby Diab Study	548 subjects (46 T1D)	Prospective		↑	↔	↑	– Weight z-score was a predictor of IA, even after controlling for HLA (HR=1.43 [95% CI 1.10 to 1.84] per unit increase in z, *P*=0.007). – BMI z-score over time was a predictor of IA, even after controlling for HLA (HR=1.29 [95% CI 1.01 to 1.67] per unit increase in z, *P*=0.04).

First author Year	Country Study	N	Design	BW	ΔW	ΔH	Δ BMI	Results
Kharagjitsingh 2010	Netherlands 5 childhood diabetes clinics in Rotterdam	213 cases 255 healthy siblings	Retrospective family-controlled	↔	↑	↑	↔	– ΔW in cases vs. siblings in1st yr after birth: 0.27 vs. 0.02, *P*=0.04. – ΔH in cases vs. siblings in 1st yr after birth: 0.15 vs. -0.09, *P*=0.02.
Carlsson 2012	Sweden Better Diabetes Diagnosis (2005-)	2,294 Ab-positive, T1D cases	Prospective population-based				HLA	– The proportion of highest risk HLA-DQ genotype decreased with increasing BMI (*P*<0.0004). – Lower risk DQ genotypes were associated with increased % of patients who were overweight (*P*<0.0001).
Peet 2014	Finland + Estonia DIABIMMUNE	235 (Est) + 261 (Fin)	Prospective	↔	HLA	HLA		– Weight SDS in high-risk HLA vs. controls at 12 mo (0.37 [95% CI 0.09 to 0.66] vs. 0.92 [95% CI 0.78 to 1.1], *P*=0.001). – Length SDS was lower in high-risk HLA at 12 mo (0.36 [95% CI 0.03 to 0.70] vs. 0.83 [95% CI 0.67 to 0.96], *P*=0.01), and 24 mo (-0.11 [95% CI 0.45 to 0.20] vs. 0.47 [95% CI 0.28 to 0.65], *P*=0.035).
Beyerlein 2014	Germany BABYDIAB + BABYDIET	1,011	Prospective				↑	– Age at infant BMI peak was associated with IA (HR=0.60 [95% CI 0.41 to 0.87] per 2SD increase in age). – ΔBMI peak and rebound was associated with IA, unadjusted (HR=1.52 [95% CI 1.04 to 2.22] per 2SD increase in age).

First author Year	Country Study	N	Design	BW	ΔW	ΔH	Δ BMI	Results
Magnus 2015	Denmark (DNBC birth cohort) Norway (MoBa birth cohort)	99,832	Prospective population-based	↔	↑	↔		– Weight gain in the 1st yr of life positively associated with T1D ($HR_{pooled\ adj}$=1.24 per 1 SD [95% CI 1.09 to 1.41]). – Length gain in the 1st yr after birth did not associate ($HR_{pooled\ adj}$=1.06 per 1 SD [95% CI 0.86 to 1.32]).
Yassouridis 2016	Germany BABYDIAB + BABYDIET	2,236 (191 with IA)	Prospective			↑	↑	– IA was associated with rapidly increasing BMI SDS until the age of 3 yr (adjusted OR 2.02 [95% CI 1.03 to 3.73]). – High height SDS at birth followed by a decrease to average values after 3 yr of age was protective against IA (OR 0.16 [95% CI 0.01 to 0.62]).
Larsson 2016	Finland, Germany, Sweden, US TEDDY	575	Prospective		↑			– Development of IA (n=575) was related to weight z-scores at 12 mo, (HR 1.16 per 1.14 kg in males or per 1.02 kg in females [95% CI 1.06 to 1.27], *P*<0.001), but not at 24 or 36 mo. – No association found with either weight or height and T1D (n=169).

APPENDIX II

Details of DNA extraction

DNA was extracted from cord blood, capillary blood or buccal samples, using a chloroform-based method, quantified using Picogreen, and stored at a normalised concentration of 50 ng/µl at 4 °C in the MRC Compound Unit at the Institute of Metabolic Sciences, Addenbrooke's Hospital, Cambridge, UK.

Details of hormone assays (adapted from Prentice[424])

IGF-1 and IGFBP-3 concentrations were measured from DBS in Tüebingen, Germany. IGF-1 was extracted from two 3.2-mm blood-spot disks using 400 µL acidifying buffer and quantified with a radioimmunoassay according to the manufacturer's instructions (Mediagnost, Tüebingen, Germany). IGFBP-3 was extracted from a single blood-spot disk in 1 ml of a buffer at neutral pH and measured with a specific radioimmunoassay as described elsewhere[648].

Laboratory analyses for C-peptide and leptin were done by Peter Barker and Keith Burling at the Core Biochemical Assay Laboratory, Cambridge Biomedical Research Centre. C-peptide levels were measured using the Mercodia Ultrasensitive C-peptide enzyme linked immunosorbent assay (ELISA) for plasma adapted for DBS based on the adaptation of the Mercodia insulin assay for DBS described previously[649]. A blood-spot disk was transferred into the assay plate coated with mouse monoclonal anti-C-peptide, followed by addition of 25 µL zero calibrator and 50 µL assay, and incubation with mixing at 4 °C for 16 hours. Once the plate reached room temperature, the blood spot was removed. After washing, 100 µL of diluted enzyme conjugate, prepared according to manufacturer's instructions, were added to the plate and incubated for 1 h. Following washing with a buffer, 200 µL substrate TMB was added and the plate incubated in the dark for half an hour before stopping the reaction with 50 µL stop solution and reading the absorbances at 450 nm with the Victor plate reader (Perkin Elmer).

Leptin levels were determined from a single blood-spot disk eluted in 200 µL DELFIA Multibuffer (Perkin Elmer) and measured using DELFIA® (dissociation-enhanced lanthanide fluorescence immunoassay) time-resolved fluorescence technology according to the manufacturer's instructions. As a result of the lower detection limit (1.25 ng/ml) set higher than the serum detection limit, in order to counteract the reduced blood volume provided by the DBS samples (1/25th of the equivalent serum volume), 43 samples and 115 samples were below the detection limit at 3 and 12 months respectively (P. Barker, personal communication, 2018).

Table A-II-1 | Kolmogorov-Smirnov tests for anthropometric measures by age.

	Statistic	df	*P* value
Gestation (wk)	0.1	1,660	<0.0001
Birth			
Weight (kg)	0.034	1,657	<0.0001
Weight SDS	0.019	1,657	0.142
Height (cm)	0.065	1,599	<0.0001
Height SDS	0.023	1,599	0.040
HC (cm)	0.038	1,602	<0.0001
HC SDS	0.025	1,602	0.025
BMI (kg/m^2)	0.027	1,596	0.090
BMI SDS	0.030	1,586	0.002
SFT SDS	0.069	1,592	<0.0001
3 months			
Weight (kg)	0.020	1,340	0.200
Weight SDS	0.021	1,340	0.149
Height (cm)	0.036	1,335	<0.0001
Height SDS	0.023	1,335	0.091
SFT SDS	0.024	1,342	0.069
12 months			
Weight (kg)	0.035	1,182	0.001
Weight SDS	0.022	1,181	0.200
Height (cm)	0.022	1,175	0.200
Height SDS	0.016	1,174	0.200
HC (cm)	0.058	1,179	<0.0001
HC SDS	0.030	1,178	0.017
BMI (kg/m^2)	0.035	1,175	0.002
BMI SDS	0.016	1,174	0.200
SFT SDS	0.034	1,177	0.003
18 months			
Weight (kg)	0.038	1,094	0.001
Weight SDS	0.023	1,093	0.200
Height (cm)	0.031	1,096	0.015
Height SDS	0.014	1,095	0.200
HC (cm)	0.053	1,088	<0.0001
HC SDS	0.025	1,087	0.118
BMI (kg/m^2)	0.034	1,092	0.004
BMI SDS	0.019	1,091	0.200
SFT SDS	0.052	1,095	<0.0001
24 months			
Weight (kg)	0.046	1,051	<0.0001
Weight SDS	0.018	1,050	0.200
Height (cm)	0.019	1,043	0.200
Height SDS	0.019	1,042	0.200
HC (cm)	0.044	1,051	<0.0001
HC SDS	0.029	1,050	0.041
BMI (kg/m^2)	0.037	1,037	0.002
BMI SDS	0.017	1,036	0.200
SFT SDS	0.059	1,057	<0.0001

Table A-II-2 | Kolmogorov-Smirnov tests for blood levels of endocrine factors by age.

	Statistic	**df**	***P* value**
IGF-1 (ng/ml)	0.064	569	<0.0001
3-month	0.084	387	<0.0001
12-month	0.070	300	0.001
18-month	0.105	164	<0.0001
24-month			
IGFBP-3 (ng/ml)			
3-month	0.045	566	0.007
12-month	0.058	377	0.004
18-month	0.047	289	0.200
24-month	0.057	155	0.200
C-peptide (pmol/L)			
3-month	0.074	317	<0.0001
12-month	0.108	144	<0.0001
Leptin (ng/ml)			
3-month	0.145	311	<0.0001
12-month	0.268	238	<0.0001

Table A-II-3 | Pairwise associations with Y=Weight (*rho* ≡ Spearman's coefficient; Δ ≡ mean difference for t-test; F ≡ F-statistic for ANOVA).

Variable		Birth			12 months			24 months		
Name	**Type**	***rho***	**Δ / F**	***P* value**	***rho***	**Δ / F**	***P* value**	***Rho***	**Δ / F**	***P* value**
Newborn index										
Gestational Age	Continuous	0.541		<0.0001	0.144		<0.0001	0.122		<0.0001
Infant Sex (baseline: males)	Binary		0.142	<0.0001		0.739	<0.0001		0.641	<0.0001
Pregnancy influences										
Maternal T1D (baseline: non-T1D)	Binary		-0.703	0.004		-1.108	0.038		-1.779	0.036
Parity (baseline: primiparous)	Binary		-0.163	<0.0001		-0.035	0.619		0.075	0.414
Smoking (baseline: non-smoking)	Binary		0.128	0.033		-0.465	0.017		-0.687	0.008
Maternal height	Continuous	0.216		<0.0001	0.244		<0.0001	0.239		<0.0001
Maternal pre-pregnancy weight	Continuous	0.216		<0.0001	0.230		<0.0001	0.266		<0.0001
Maternal pre-pregnancy BMI	Continuous	0.143		<0.0001	0.138		<0.0001	0.175		<0.0001
Maternal pregnancy weight gain	Continuous	0.143		<0.0001	0.035		0.354	0.031		0.429
Maternal age	Continuous	0.035		0.202	-0.086		0.003	-0.041		0.184
Paternal age	Continuous	0.042		0.134	-0.068		0.026	-0.047		0.143
Post-pregnancy influences										
Birth delivery	Categorical		2.894	0.021		5.539	<0.0001		2.286	0.058
3-month feeding (baseline: breast milk only)	Binary					-0.519	<0.0001		-0.331	<0.0001
Birth weight SDS	Continuous				0.452		<0.0001	0.437		<0.0001
Birth length SDS	Continuous				0.351		<0.0001	0.343		
Sociodemographic influences										
Ethnicity (baseline: White Caucasian)	Binary		0.009	0.907		-0.076	0.687		-0.072	0.775
Index of multiple deprivation	Continuous	-0.037		0.258	0.074		0.035	0.076		0.034
Marital status	Categorical		0.960	0.383		0.509	0.601		0.466	0.627
Maternal education	Categorical		1.079	0.341		5.332	0.005		0.790	0.455

Table A-II-4 | Pairwise associations with Y=Height (*rho* ≡ Spearman's coefficient; Δ ≡ mean difference for t-test; F ≡ F-statistic for ANOVA).

Variable		Birth			12 months			24 months		
Name	**Type**	***rho***	**Δ / F**	***P* value**	***rho***	**Δ / F**	***P* value**	***Rho***	**Δ / F**	***P* value**
Newborn index										
Gestational Age	Continuous	0.419		<0.0001	0.211		<0.0001	0.171		<0.0001
Infant Sex (baseline: males)	Binary		0.929	<0.0001		1.781	<0.0001		1.438	<0.0001
Pregnancy influences										
Maternal T1D (baseline: non-T1D)	Binary		-0.363	0.780		-2.450	0.057		-1.950	0.422
Parity (baseline: primiparous)	Binary		-0.433	0.001		0.082	0.634		0.423	0.051
Smoking (baseline: non-smoking)	Binary		0.161	0.585		-0.871	0.063		-1.594	0.009
Maternal height	Continuous	0.203		<0.0001	0.309		<0.0001	0.313		<0.0001
Maternal pre-pregnancy weight	Continuous	0.149		<0.0001	0.210		<0.0001	0.217		<0.0001
Maternal pre-pregnancy BMI	Continuous	0.074		0.012	0.092		0.004	0.088		0.009
Maternal pregnancy weight gain	Continuous	0.082		0.017	0.036		0.333	0.052		0.191
Maternal age	Continuous	0.043		0.122	-0.075		0.011	-0.060		0.053
Paternal age	Continuous	0.002		0.942	-0.063		0.039	-0.050		0.123
Post-pregnancy influences										
Birth delivery	Categorical		8.082	<0.0001		1.382	0.238		1.994	0.093
3-month feeding (baseline: breast milk only)	Binary					-1.093	<0.0001		-0.962	<0.0001
Birth weight SDS	Continuous				0.430		<0.0001	0.392		<0.0001
Birth length SDS	Continuous				0.351		<0.0001	0.343		<0.0001
Sociodemographic influences										
Ethnicity (baseline: White Caucasian)	Binary		-0.117	0.747		0.207	0.651		-0.300	0.611
Index of multiple deprivation	Continuous	-0.050		0.133	0.030		0.394	0.090		0.012
Marital status	Categorical		0.265	0.768		0.245	0.783		0.521	0.594
Maternal education	Categorical		1.644	0.194		1.954	0.143		0.003	0.997

Table A-II-5 | Pairwise associations with Y=Δweight SDS (*rho* ≡ Spearman's coefficient; Δ ≡ mean difference for t-test; F ≡ F-statistic for ANOVA).

Variable		Birth-12 months			Birth-24 months		
Name	**Type**	***rho***	**Δ / F**	***P* value**	***rho***	**Δ / F**	***P* value**
Newborn index							
Gestation	Continuous						
Infant Sex (baseline: males)	Binary						
Pregnancy influences							
Maternal T1D (baseline: non-T1D)	Binary		1.835	<0.0001		1.603	0.388
Parity (baseline: primiparous)	Binary		0.363	<0.0001		0.400	<0.0001
Smoking (baseline: non-smoking)	Binary		-0.650	<0.0001		-0.590	0.002
Maternal height	Continuous	0.069		0.027	0.037		0.268
Maternal pre-pregnancy weight	Continuous	0.040		0.212	0.037		0.264
Maternal pre-pregnancy BMI	Continuous	0.003		0.920	0.009		0.796
Maternal pregnancy weight gain	Continuous	-0.030		0.431	-0.037		0.353
Maternal age	Continuous	-0.144		<0.0001	-0.094		0.003
Paternal age	Continuous	-0.118		<0.0001	-0.086		0.007
Post-pregnancy influences							
Birth delivery	Categorical		5.281	<0.0001		5.861	<0.0001
3-month feeding (baseline: breast milk only)	Binary		-0.496	<0.0001		-0.263	<0.0001
Birth weight SDS	Continuous	-0.487		<0.0001	-0.527		<0.0001
Birth length SDS	Continuous	-0.204		<0.0001	-0.211		<0.0001
Sociodemographic influences							
Ethnicity (baseline: White Caucasian)	Binary		-0.156	0.390		-0.085	0.655
Index of multiple deprivation	Continuous	0.129		<0.0001	0.125		<0.0001
Marital status	Categorical		1.141	0.320		1.576	0.207
Maternal education	Categorical		7.750	<0.0001		1.828	0.162

Table A-II-6 | Pairwise associations with Y=Δheight SDS (*rho* ≡ Spearman's coefficient; Δ ≡ mean difference for t-test; F ≡ F-statistic for ANOVA).

Variable		Birth-12 months			Birth-24 months		
Name	**Type**	***rho***	**Δ / F**	***P* value**	***rho***	**Δ / F**	***P* value**
Newborn index							
Gestation	Continuous						
Infant Sex (baseline: males)	Binary						
Pregnancy influences							
Maternal T1D (baseline: non-T1D)	Binary		0.927	0.077		0.9995	0.192
Parity (baseline: primiparous)	Binary		0.274	<0.0001		0.340	<0.0001
Smoking (baseline: non-smoking)	Binary		-0.466	0.007		-0.555	0.005
Maternal height	Continuous	0.128		<0.0001	0.092		0.006
Maternal pre-pregnancy weight	Continuous	0.054		0.097	0.005		0.872
Maternal pre-pregnancy BMI	Continuous	-0.002		0.960	-0.050		0.152
Maternal pregnancy weight gain	Continuous	-0.049		0.203	-0.043		0.282
Maternal age	Continuous	-0.128		<0.0001	-0.106		0.001
Paternal age	Continuous	-0.087		0.005	-0.060		0.069
Post-pregnancy influences							
Birth delivery	Categorical		1.191	0.313		3.255	0.012
3-month feeding (baseline: breast milk only)	Binary		-0.466	<0.0001		-0.378	<0.0001
Birth weight SDS	Continuous	-0.273		<0.0001	-0.305		<0.0001
Birth length SDS	Continuous	-0.436		<0.0001	-0.482		<0.0001
Sociodemographic influences							
Ethnicity (baseline: White Caucasian)	Binary		-0.070	0.682		-0.173	0.368
Index of multiple deprivation	Continuous	0.087		0.016	0.133		<0.0001
Marital status	Categorical		1.155	0.315		0.899	0.407
Maternal education	Categorical		6.106	0.002		0.912	0.402

Table A-II-7 | Pairwise associations with Y=HC SDS (*rho* ≡ Spearman's coefficient; Δ ≡ mean difference for t-test; F ≡ F-statistic for ANOVA).

Variable		Birth			12 months			24 months		
Name	**Type**	***rho***	**Δ / F**	***P* value**	***rho***	**Δ / F**	***P* value**	***Rho***	**Δ / F**	***P* value**
Newborn index										
Gestation	Continuous									
Infant Sex (baseline: males)	Binary									
Pregnancy influences										
Maternal T1D (baseline: non-T1D)	Binary		-1.113	0.023		0.002	0.998		-0.466	0.536
Parity (baseline: primiparous)	Binary		-0.230	<0.0001		-0.687	0.283		-0.003	0.964
Smoking (baseline: non-smoking)	Binary		0.159	0.156		-0.024	0.891		-0.150	0.432
Maternal height	Continuous	0.188		<0.0001	0.196		<0.0001	0.191		<0.0001
Maternal pre-pregnancy weight	Continuous	0.208		<0.0001	0.188		<0.0001	0.166		<0.0001
Maternal pre-pregnancy BMI	Continuous	0.133		<0.0001	0.106		0.001	0.091		0.007
Maternal pregnancy weight gain	Continuous	0.077		0.024	0.048		0.203	0.019		0.627
Maternal age	Continuous	0.046		0.100	0.007		0.824	0.033		0.285
Paternal age	Continuous	0.052		0.065	0.023		0.443	0.029		0.371
Post-pregnancy influences										
Birth delivery	Categorical		12.805	<0.0001		6.261	<0.0001		2.431	0.046
3-month feeding (baseline: breast milk only)	Binary					-0.014	0.831		0.151	0.026
Weight SDS at respective age	Continuous	0.624		<0.0001	0.447		<0.0001	0.466		<0.0001
Head Circumference SDS at birth	Continuous				0.516		<0.0001	0.512		<0.0001
Sociodemographic influences										
Ethnicity (baseline: White Caucasian)	Binary		0.083	0.547		0.001	0.997		0.102	0.570
Index of multiple deprivation	Continuous	-0.099		0.003	-0.013		0.704	-0.008		0.828
Marital status	Categorical		3.565	0.029		0.039	0.961		0.168	0.846
Maternal education	Categorical		0.736	0.479		1.440	0.238		4.380	0.013

Table A-II-8 | Pairwise associations with Y=BMI SDS (*rho* ≡ Spearman's coefficient; Δ ≡ mean difference for t-test; F ≡ F-statistic for ANOVA).

Variable		Birth			12 months			24 months		
Name	**Type**	***rho***	**Δ / F**	***P* value**	***rho***	**Δ / F**	***P* value**	***Rho***	**Δ / F**	***P* value**
Newborn index										
Gestation	Continuous									
Infant Sex (baseline: males)	Binary									
Pregnancy influences										
Maternal T1D (baseline: non-T1D)	Binary		-3.047	<0.0001		-0.483	0.289		-0.992	0.128
Parity (baseline: primiparous)	Binary		-0.336	<0.0001		-0.071	0.237		-0.047	0.414
Smoking (baseline: non-smoking)	Binary		0.404	0.004		-0.290	0.080		0.350	0.036
Maternal height	Continuous	0.093		0.001	0.088		0.005	0.051		0.125
Maternal pre-pregnancy weight	Continuous	0.171		<0.0001	0.170		<0.0001	0.192		<0.0001
Maternal pre-pregnancy BMI	Continuous	0.138		<0.0001	0.151		<0.0001	0.189		<0.0001
Maternal pregnancy weight gain	Continuous	0.059		0.087	0.016		0.677	-0.033		0.412
Maternal age	Continuous	0.039		0.158	-0.063		0.033	-0.004		0.897
Paternal age	Continuous	0.062		0.028	-0.031		0.315	-0.002		0.951
Post-pregnancy influences										
Birth delivery	Categorical		20.840	<0.0001		6.676	<0.0001		3.370	0.009
3-month feeding (baseline: breast milk only)	Binary					-0.269	<0.0001		-0.017	0.768
Weight SDS at respective age	Continuous	0.737		<0.0001	0.770		<0.0001	0.721		<0.0001
BMI SDS at birth	Continuous				0.268		<0.0001	0.275		<0.0001
Sociodemographic influences										
Ethnicity (baseline: White Caucasian)	Binary		0.101	0.561		-0.168	0.296		-0.063	0.696
Index of multiple deprivation	Continuous	-0.049		0.140	0.046		0.199	0.006		0.871
Marital status	Categorical		0.638	0.528		0.407	0.665		0.847	0.429
Maternal education	Categorical		0.329	0.720		3.423	0.033		1.647	0.193

Table A-II-9 | Pairwise associations with Y=Skinfold thickness (SFT) SDS (*rho* ≡ Spearman's coefficient; Δ ≡ mean difference for t-test; F ≡ F-statistic for ANOVA).

Variable		Birth			12 months			24 months		
Name	**Type**	***rho***	**Δ / F**	***P* value**	***rho***	**Δ / F**	***P* value**	***rho***	**Δ / F**	***P* value**
Newborn index										
Gestation	Continuous									
Infant Sex (baseline: males)	Binary									
Pregnancy influences										
Maternal T1D (baseline: non-T1D)	Binary		-1.291	<0.0001		-0.039	0.911		0.521	0.358
Parity (baseline: primiparous)	Binary		-0.290	<0.0001		0.076	0.094		0.080	0.108
Smoking (baseline: non-smoking)	Binary		-0.054	0.589		-0.048	0.703		-0.131	0.348
Maternal height	Continuous	0.053		0.064	-0.007		0.817	0.013		0.688
Maternal pre-pregnancy weight	Continuous	0.096		0.001	0.060		0.058	0.092		0.006
Maternal pre-pregnancy BMI	Continuous	0.089		0.003	0.083		0.010	0.098		0.004
Maternal pregnancy weight gain	Continuous	0.010		0.780	-0.039		0.295	0.002		0.965
Maternal age	Continuous	0.033		0.235	-0.046		0.120	-0.024		0.433
Paternal age	Continuous	0.005		0.860	-0.005		0.872	-0.026		0.413
Post-pregnancy influences										
Birth delivery	Categorical		1.596	0.173		5.275	<0.0001		2.065	0.083
3-month feeding (baseline: breast milk only)	Binary		0.093	0.061		-0.260	<0.0001		-0.065	0.198
Weight SDS at respective age	Continuous	0.462		<0.0001	0.488		<0.0001	0.455		<0.0001
STF SDS at birth	Continuous				0.094		0.002	0.104		0.001
Sociodemographic influences										
Ethnicity (baseline: White Caucasian)	Binary		0.136	0.263		0.020	0.875		0.183	0.182
Index of multiple deprivation	Continuous	-0.047		0.159	0.058		0.104	0.033		0.359
Marital status	Categorical		2.562	0.078		0.869	0.420		0.388	0.679
Maternal education	Categorical		0.841	0.432		2.386	0.093		0.645	0.525

Table A-II-10 | Pairwise associations with Y=[IGF-1] (*rho* ≡ Spearman's coefficient; Δ ≡ mean difference for t-test; F ≡ F-statistic for ANOVA). *1 case

Variable		$[IGF\text{-}1]_{3months}$			$[IGF\text{-}1]_{12months}$			$[IGF\text{-}1]_{24months}$		
Name	**Type**	***rho***	**Δ / F**	***P* value**	***rho***	**Δ / F**	***P* value**	***Rho***	**Δ / F**	***P* value**
Newborn index										
Gestation	Continuous	-0.057		0.175	-0.092		0.070	-0.119		0.130
Infant Sex (baseline: males)	Binary		3.892	0.021		-11.057	<0.0001		-8.725	0.039
Pregnancy influences										
Maternal T1D (baseline: non-T1D)	Binary		-13.25	0.351		-13.53	0.248		-37.05	0.174*
Parity (baseline: primiparous)	Binary		2.848	0.093		7.382	0.002		3.838	0.374
Smoking (baseline: non-smoking)	Binary		-1.024	0.831		-15.68	0.034		1.736	0.889
Maternal height	Continuous	-0.034		0.451	-0.057		0.304	-0.035		0.671
Maternal pre-pregnancy weight	Continuous	0.063		0.166	-0.013		0.820	-0.035		0.693
Maternal pre-pregnancy BMI	Continuous	0.092		0.046	0.016		0.772	-0.014		0.871
Maternal pregnancy weight gain	Continuous	-0.090		0.096	-0.013		0.845	0.112		0.261
Maternal age	Continuous	-0.001		0.988	-0.026		0.612	-0.003		0.967
Paternal age	Continuous	0.007		0.875	-0.060		0.264	-0.146		0.070
Post-pregnancy influences										
Birth delivery	Categorical		2.538	0.039		1.192	0.314		0.041	0.997
3-month feeding (baseline: breast milk only)	Binary		-9.949	<0.0001		-6.202	0.011		-3.750	0.396
Weight SDS at respective age	Continuous	-0.312		<0.0001	0.224		<0.0001	0.282		<0.0001
$[IGF\text{-}1]_{3months}$	Continuous				-0.246		<0.0001	0.325		0.003
Sociodemographic influences										
Ethnicity (baseline: White Caucasian)	Binary		-5.769	0.251		2.743	0.733		-4.301	0.778
Index of multiple deprivation	Continuous	0.127		0.011	0.030		0.605	-0.051		0.551
Marital status	Categorical		0.177	0.837		0.579	0.561		1.295	0.277
Maternal education	Categorical		4.619	0.011		6.009	0.003		2.087	0.129

Table A-II-11 | Pairwise associations with Y=[C-peptide] (*rho* ≡ Spearman's coefficient; Δ ≡ mean difference for t-test; F ≡ F-statistic for ANOVA). *1 case

Variable		$[C\text{-peptide}]_{3months}$			$[C\text{-peptide}]_{12months}$		
Name	**Type**	***rho***	**Δ / F**	***P* value**	***Rho***	**Δ / F**	***P* value**
Newborn index							
Gestation	Continuous	-0.148		0.008	-0.090		0.281
Infant Sex (baseline: males)	Binary		25.983	0.547		8.708	0.871
Pregnancy influences							
Maternal T1D (baseline: non-T1D)	Binary		-408.8	0.286*		N/A	N/A
Parity (baseline: primiparous)	Binary		48.319	0.274		108.10	0.042
Smoking (baseline: non-smoking)	Binary		101.70	0.460		-77.26	0.638
Maternal height	Continuous	0.071		0.240	0.168		0.061
Maternal pre-pregnancy weight	Continuous	0.032		0.607	0.099		0.276
Maternal pre-pregnancy BMI	Continuous	0.008		0.899	0.029		0.755
Maternal pregnancy weight gain	Continuous	0.056		0.443	0.075		0.481
Maternal age	Continuous	-0.063		0.269	-0.071		0.403
Paternal age	Continuous	-0.006		0.917	-0.173		0.044
Post-pregnancy influences							
Birth delivery	Categorical		0.194	0.942		1.280	0.281
3-month feeding (baseline: breast milk only)	Binary		-112.60	0.012		1.294	0.981
Weight SDS at respective age	Continuous	0.239		<0.0001	-0.185		0.027
$[C\text{-peptide}]_{3months}$	Continuous				-0.127		0.399
Sociodemographic influences							
Ethnicity (baseline: White Caucasian)	Binary		-65.497	0.514		18.090	0.881
Index of multiple deprivation	Continuous	0.097		0.198	0.046		0.653
Marital status	Categorical		0.094	0.911		0.313	0.732
Maternal education	Categorical		0.948	0.390		2.726	0.074

Table A-II-12 | Pairwise associations with Y=[Leptin] (*rho* ≡ Spearman's coefficient; Δ ≡ mean difference for t-test; F ≡ F-statistic for ANOVA). *1 case

Variable		$[Leptin]_{3months}$			$[Leptin]_{12months}$		
Name	**Type**	***rho***	**Δ / F**	***P* value**	***rho***	**Δ / F**	***P* value**
Newborn index							
Gestation	Continuous	-0.022		0.702	-0.230		<0.0001
Infant Sex (baseline: males)	Binary		-0.760	0.002		-0.338	0.003
Pregnancy influences							
Maternal T1D (baseline: non-T1D)	Binary		0.196	0.929*		-0.084	0.924*
Parity (baseline: primiparous)	Binary		-0.414	0.105		0.207	0.068
Smoking (baseline: non-smoking)	Binary		-1.278	0.649		-0.851	0.029
Maternal height	Continuous	-0.025		0.688	-0.038		0.593
Maternal pre-pregnancy weight	Continuous	-0.019		0.755	0.015		0.833
Maternal pre-pregnancy BMI	Continuous	0.006		0.927	0.046		0.524
Maternal pregnancy weight gain	Continuous	-0.024		0.742	-0.040		0.633
Maternal age	Continuous	0.022		0.697	-0.153		0.019
Paternal age	Continuous	-0.108		0.069	-0.012		0.865
Post-pregnancy influences							
Birth delivery	Categorical		1.404	0.233		0.786	0.535
3-month feeding (baseline: breast milk only)	Binary		0.140	0.582		-0.115	0.326
Weight SDS at respective age	Continuous	0.349		<0.0001	0.198		0.002
$[Leptin]_{3months}$	Continuous				-0.410		<0.0001
Sociodemographic influences							
Ethnicity (baseline: White Caucasian)	Binary		-0.480	0.395		-0.346	0.138
Index of multiple deprivation	Continuous	0.062		0.422	0.101		0.194
Marital status	Categorical		1.593	0.205		4.336	0.014
Maternal education	Categorical		0.397	0.673		0.922	0.401

Table A-II-13 | Endocrine blood levels by a. twinship, b. gestational length, c. maternal T1D.

	Singleton			Gestation			Maternal T1D		
	yes	no		≥ 36 wk	< 36 wk		no	yes	
	Mean ± SD (N)	Mean ± SD (N)	*P* value	Mean ± SD (N)	Mean ± SD (N)	*P* value	Mean ± SD (N)	Mean ± SD (N)	*P* value
IGF-1 (ng/ml)									
3-month	50.3 ± 20.1 (559)	48.8 ± 14.7 (10)	0.934	50.1 ± 19.9 (561)	61.0 ± 27.0 (8)	0.282	50.3 ± 20.0 (567)	63.5 ± 21.9 (2)	0.301
12-month	51.3 ± 23.3 (383)	54.0 ± 14.3 (4)	0.587	50.8 ± 23.1 (377)	71.7 ± 21.0 (10)	0.004	51.2 ± 23.1 (383)	64.8 ± 35.8 (4)	0.358
18-month	58.7 ± 24.3 (296)	58.5 ± 18.6 (4)	0.942	58.3 ± 24.2 (292)	70.1 ± 23.3 (8)	0.163	58.6 ± 24.2 (299)	89.0 (1)	0.200
24-month	N/A	N/A	N/A	62.0 ± 27.1 (161)	74.3 ± 29.3 (3)	0.432	62.0 ± 27.0 (163)	99.0 (1)	0.220
IGFBP-3 (ng/ml)									
3-month	1,704.4 ± 385.0 (557)	1,602.4 ± 210.7 (9)	0.435	1,702.6 ± 379.3 (559)	1,711.7 ± 654.4 (7)	0.759	1,701.8 ± 383.2 (564)	1,974.3 ± 135.3 (2)	0.189
12-month	1,965.0 ± 500.2 (374)	1,649.7 ± 235.8 (3)	0.196	1,958.8 ± 500.0 (367)	2,098.8 ± 498.7 (10)	0.421	1,958.8 ± 500.0 (373)	2,310.8 ± 307.2 (4)	0.082
18-month	2,003.0 ± 473.9 (286)	1,899.3 ± 380.8 (3)	0.726	1,997.9 ± 473.9 (218)	2,140.4 ± 431.5 (8)	0.308	2,001.2 ± 473.3 (288)	2,208.6 (1)	0.588
24-month	N/A	N/A	N/A	2,054.1 ± 473.9 (152)	2,242.2 ± 666.5 (3)	0.603	2,051.8 ± 473.3 (154)	2,968.7 (1)	0.103
C-peptide (pmol/L)									
3-month	805.6 ± 437.5 (309)	733.8 ± 246.6 (8)	0.841	804.5 ± 434.3 (312)	759.0 ± 437.0 (5)	0.834	802.3 ± 433.5 (316)	1,266.0 (1)	0.259
12-month	584.5 ± 320.4 (141)	734.7 ± 252.8 (3)	0.272	N/A	N/A	N/A	N/A	N/A	N/A
Leptin (ng/ml)									
3-month	2.9 ± 2.2 (304)	3.0 ± 2.2 (7)	0.998	2.9 ± 2.2 (307)	2.9 ± 2.2 (4)	0.760	2.9 ± 2.2 (310)	2.8 (1)	0.862
12-month	1.3 ± 0.9 (236)	1.7 ± 0.0 (2)	0.210	N/A	N/A	N/A	1.3 ± 0.9 (237)	1.4 (1)	0.782

Table A-II-14 | Rationale for gene selection based on biological knowledge.

Gene	Molecular function	Phenotype: human	Phenotype: mouse	Pathways & Interactions
PTPN22			growth/size/body	
IL10			growth/size/body	
IFIH1	growth/inhibition of tumour cell lines			
CTLA4				*PTPN22* signalling
IL2			growth/size/body	
TNFAIP3			growth/size/body	
GLIS3		Hypothyroidism	growth/size/body	
IL2RA			growth/size/body	*IL2* signalling
PRKCQ				
INS	growth factor			
BAD	link between growth factor receptor signalling and apoptotic pathways		growth/size/body	
ITGB7	gut-associated			
CYP27B1			growth/size/body	
ERBB3	epidermal-growth factor receptor		growth/size/body	
SH2B3	signalling activities by growth factor and cytokine receptors	growth retardation-mild developmental delay		
DLK1	epidermal growth factor		growth/size/body	
CTSH			growth/size/body	
CLEC16A			growth/size/body	
PTPN2	epidermal growth factor receptor with roles in growth factor mediated cell signalling		growth/size/body	
TYK2	gene ontology: GHR binding		growth/size/body	
FUT2	affects intestinal microbiota composition			
UBASH3A				interacts with c-CBL but binds to ubiquitin via its UBA domain

ACKNOWLEDGEMENTS

I am grateful to the MRC, the Raymond and Beverley Sackler fund and the School of Clinical Medicine for funding my PhD studies and project. I particularly thank the Director, Professor John Sinclair, and Committee of the MRC/Sackler Prize PhD Programme for granting me the award, the gratitude for which has been an impetus to bring this mission to completion. I am sincerely thankful to the University of Cambridge for offering an array of training courses for researchers, which honed my skills and allowed me to independently conduct all the statistical and genetic workings required for this project. I also thank David Dunger for giving me the opportunity to pursue this project and providing access to resources and expertise, Ken Ong for readily providing the DNA samples and offering reliable expert guidance on statistical genetics, Clive Petry for providing clear answers to my questions on human genetics and thoughtful advice, Philippa Prentice for guiding me through the CBGS and paving the way with her work, Tengyao Wang for sharing his impressive statistical knowledge and augmenting my understanding of the subject, Neil Walker and Oliver Burren for generously providing instrumental insights into the genetics of the disease, Dr Nick Holmes for graciously giving me permission to attend some excellent lectures in Immunology at the University of Cambridge, Andy Dunham at LGC for his efficient work and duly answering my queries, the CBGS research nurses — Suzanne, Amie and Karen — for conducting the study visits, Diane Wingate and Karen Whitehead for processing the biological specimen, Peter Barker for assaying the hormones and patiently answering my questions, Jane Horsford for her valued assistance, Burak Salgin for instilling his enthusiasm in clinical research which was a driving force for pursuing the PhD, Professor Mike Lovett whose timely inspiring teaching of Medical Genomics provided solid knowledge and was motivational for embarking on the final challenge of writing this thesis, and all the participating children and parents of the CBGS for their time and contribution to this research.

Finally, my heartfelt thanks go to my family for their love and support, my parents, Andreas and Smirni, my siblings, Agathi and Antonis, and importantly, my beloved little nephews — Andreas, Giorgos and Stephanos — for being an inspiration for this journey of discovery and growth.

REFERENCES

1. Drayer N. Height of diabetic children at onset of symptoms. *Arch Dis Child*. 1974;49(8):616-20.

2. Tattersall R, Pyke D. Growth in Diabetic Children. *Lancet*. 1973:1105-9.

3. Karam JH, Grodsky GM, Forsham PH. Letter: Weight-gain in infancy and development of juvenile diabetes mellitus. *Lancet*. 1976;1(7949):45.

4. Baum JD, Ounsted M, Smith MA. Weight gain in infancy and subsequent development of diabetes mellitus in childhood. *Lancet*. 1975:866.

5. Gorsuch AN, Lister J, Dean BM, Spencer KM, McNally JM, Bottazzo GF, et al. Evidence for a long prediabetic period in type I (insulin-dependent) diabetes mellitus. *Lancet*. 1981;318(8260):1363-5.

6. Leslie RD, Elliott RB. Early environmental events as a cause of IDDM. Evidence and implications. *Diabetes*. 1994;43(7):843-50.

7. Patterson CC, Dahlquist GG, Gyürüs E, Green A, Soltész G. Incidence trends for childhood type 1 diabetes in Europe during 1989-2003 and predicted new cases 2005-20: a multicentre prospective registration study. *Lancet*. 2009;373(9680):2027-33.

8. Dahlquist G. The aetiology of type 1 diabetes: an epidemiological perspective. *Acta Paediatr Suppl*. 1998;425:5-10.

9. Soltész G. Diabetes in the young: a paediatric and epidemiological perspective. *Diabetologia*. 2003;46(4):447-54.

10. Haynes A, Bower C, Bulsara M, Finn J, Jones T, Davis E. Perinatal risk factors for childhood Type 1 diabetes in Western Australia--a population-based study (1980-2002). *Diabet Med*. 2007;24(5):564-70.

11. Ziegler AG, Hummel M, Schenker M, Bonifacio E. Autoantibody appearance and risk for development of childhood diabetes in offspring of parents with type 1 diabetes: the 2-year analysis of the German BABYDIAB Study. *Diabetes*. 1999;48(3):460-8.

12. Naserke HE, Bonifacio E, Ziegler AG. Prevalence, characteristics and diabetes risk associated with transient maternally acquired islet antibodies and persistent islet antibodies in offspring of parents with type 1 diabetes. *J Clin Endocrinol Metab*. 2001;86(10):4826-33.

13. Lindberg B, Ivarsson S-A, Landin-Olsson M, Sundkvist G, Svanberg L, Lernmark Å. Islet autoantibodies in cord blood from children who developed Type I (insulin-dependent) diabetes mellitus before 15 years of age. *Diabetologia*. 1999;42(2):181-7.

14. Larsson HE, Lynch K, Lernmark B, Nilsson A, Hansson G, Almgren P, et al. Diabetes-associated HLA genotypes affect birthweight in the general population. *Diabetologia*. 2005;48(8):1484-91.

15. Stene LC, Magnus P, Rønningen KS, Joner G. Diabetes-associated HLA-DQ genes and birth weight. *Diabetes*. 2001;50(12):2879-82.

16. Dunger DB, Ong KKL, Huxtable SJ, Sherriff A, Woods KA, Ahmed ML, et al. Association of the INS VNTR with size at birth. *Nat Genet*. 1998;19:98.

17. Ong KK, Petry CJ, Barratt BJ, Ring S, Cordell HJ, Wingate DL, et al. Maternal-fetal interactions and birth order influence insulin variable number of tandem repeats allele class associations with head size at birth and childhood weight gain. *Diabetes*. 2004;53(4):1128-33.

18. Moore GE, Ishida M, Demetriou C, Al-Olabi L, Leon LJ, Thomas AC, et al. The role and interaction of imprinted genes in human fetal growth. *Philos Trans R Soc Lond B Biol Sci*. 2015;370(1663):20140074.

19. Atkinson MA. The pathogenesis and natural history of type 1 diabetes. *Cold Spring Harb Perspect Med*. 2012;2:a007641.

20. Noble JA, Erlich HA. Genetics of type 1 diabetes. *Cold Spring Harb Perspect Med*. 2012;2(1):a007732.

21. Steck AK, Rewers MJ. Genetics of type 1 diabetes. *Clin Chem*. 2011;57:176-85.

22. Rotter J, Landaw E. Measuring the genetic contribution of a single locus to a multilocus disease. *Clin Genet*. 1984;26(6):529-42.

23. Todd JA. Etiology of type 1 diabetes. *Immunity*. 2010;32(4):457-67.

24. Lambert AP, Gillespie KM, Thomson G, Cordell HJ, Todd JA, Gale EA, et al. Absolute risk of childhood-onset type 1 diabetes defined by human leukocyte antigen class II genotype: a population-based study in the United Kingdom. *J Clin Endocrinol Metab*. 2004;89(8):4037-43.

25. Roep BO, Atkinson M, von Herrath M. Satisfaction (not) guaranteed: re-evaluating the use of animal models of type 1 diabetes. *Nat Rev Immunol*. 2004;4(12):989-97.

26. Storling J, Brorsson CA. Candidate genes expressed in human islets and their role in the pathogenesis of type 1 diabetes. *Curr Diab Rep*. 2013;13(5):633-41.

27. Bennett ST, Lucassen AM, Gough SC, Powell EE, Undlien DE, Pritchard LE, et al. Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nat Genet*. 1995;9(3):284-92.

28. Bennett ST, Wilson AJ, Cucca F, Nerup J, Pociot F, McKinney PA, et al. IDDM2-VNTR-encoded susceptibility to type 1 diabetes: dominant protection and parental transmission of alleles of the insulin gene-linked minisatellite locus. *J Autoimmun*. 1996;9(3):415-21.

29. Pugliese A, Miceli D. The insulin gene in diabetes. *Diabetes Metab Res Rev*. 2002;18(1):13-25.

30. She JX, Bui MM, Tian XH, Muir A, Wakeland EK, Zorovich B, et al. Additive susceptibility to insulin-dependent diabetes conferred by HLA-DQB1 and insulin genes (Abstract). *Autoimmunity*. 1994;18(3):195-203.

31. Pociot F, Lernmark Å. Genetic risk factors for type 1 diabetes. *Lancet*. 2016;387(10035):2331-9.

32. Redondo MJ, Yu L, Hawa M, Mackenzie T, Pyke DA, Eisenbarth GS, et al. Heterogeneity of type I diabetes: analysis of monozygotic twins in Great Britain and the United States. *Diabetologia*. 2001;44(3):354-62.

33. Dib S, Gomes M. Etiopathogenesis of type 1 diabetes mellitus: prognostic factors for the evolution of residual beta cell function. *Diabetol Metab Syndr*. 2009;1:25.

34. Mehdi AM, Hamilton-Williams EE, Cristino A, Ziegler A, Bonifacio E, Le Cao KA, et al. A peripheral blood transcriptomic signature predicts autoantibody development in infants at risk of type 1 diabetes. *JCI Insight*. 2018;3(5).

35. Alper CA, Awdeh Z. Incomplete penetrance of MHC susceptibility genes: prospective analysis of polygenic MHC-determined traits. *Tissue Antigens*. 2000;56(3):199-206.

36. Larsen CE, Alper CA. The genetics of HLA-associated disease. *Curr Opin Immunol*. 2004;16(5):660-7.

37. Knip M, Simell O. Environmental triggers of type 1 diabetes. *Cold Spring Harb Perspect Med*. 2011;3:1-15.

38. Green A, Patterson CC. Trends in the incidence of childhood-onset diabetes in Europe 1989–1998. *Diabetologia*. 2001;44(3):B3-B8.

39. DIAMOND Project Group. Incidence and trends of childhood Type 1 diabetes worldwide 1990-1999. *Diabet Med*. 2006;23(8):857-66.

40. Hermann R, Knip M, Veijola R, Simell O, Laine AP, Akerblom HK, et al. Temporal changes in the frequencies of HLA genotypes in patients with Type 1 diabetes--indication of an increased environmental pressure? *Diabetologia*. 2003;46(3):420-5.

41. Vehik K, Hamman RF, Lezotte D, Norris JM, Klingensmith GJ, Rewers M, et al. Trends in high-risk HLA susceptibility genes among Colorado youth with type 1 diabetes. *Diabetes Care*. 2008;31(7):1392-6.

42. Fourlanos S, Varney MD, Tait BD, Morahan G, Honeyman MC, Colman PG, et al. The rising incidence of type 1 diabetes is accounted for by cases with lower-risk human leukocyte antigen genotypes. *Diabetes Care*. 2008;31(8):1546-9.

43. Resic-Lindehammer S, Larsson K, Ortqvist E, Carlsson A, Cederwall E, Cilio CM, et al. Temporal trends of HLA genotype frequencies of type 1 diabetes patients in Sweden from 1986 to 2005 suggest altered risk. *Acta Diabetol*. 2008;45(4):231-5.

44. Bach J-F. The effect of infections on susceptibility to autoimmune and allergic diseases. *N Engl J Med*. 2002;347(12):911-20.

45. Bonifacio E, Pflüger M, Marienfeld S, Winkler C, Hummel M, Ziegler AG. Maternal type 1 diabetes reduces the risk of islet autoantibodies: Relationships with birthweight and maternal HbA1c. *Diabetologia*. 2008;51(7):1245-52.

46. TEDDY Study Group. The Environmental Determinants of Diabetes in the Young (TEDDY) study: study design. *Pediatr Diabetes*. 2007;8(5):286-98.

47. TRIGR Study Group. Study design of the Trial to Reduce IDDM in the Genetically at Risk (TRIGR). *Pediatr Diabetes*. 2007;8(3):117-37.

48. Penno MA, Couper JJ, Craig ME, Colman PG, Rawlinson WD, Cotterill AM, et al. Environmental determinants of islet autoimmunity (ENDIA): a pregnancy to early life cohort study in children at-risk of type 1 diabetes. *BMC Pediatr*. 2013;13:124.

49. Rewers M, Bugawan TL, Norris JM, Blair A, Beaty B, Hoffman M, et al. Newborn screening for HLA markers associated with IDDM: Diabetes Autoimmunity Study in the Young (DAISY). *Diabetologia*. 1996;39(7):807.

50. Hahl J, Simell T, Ilonen J, Knip M, Simell O. Costs of predicting IDDM. *Diabetologia*. 1998;41(1):79-85.

51. Kupila A, Muona P, Simell T, Arvilommi P, Savolainen H, Hamalainen AM, et al. Feasibility of genetic and immunological prediction of type I diabetes in a population-based birth cohort. *Diabetologia*. 2001;44(3):290-7.

52. Fohlman J, Friman G. Is juvenile diabetes a viral disease? (Abstract). *Ann Med*. 1993;25(6):569-74.

53. Trucco M, LaPorte R. Exposure to superantigens as an immunogenetic explanation of type I diabetes mini-epidemics (Abstract). *J Pediatr Endocrinol Metab*. 1995;8(1):3-10.

54. Yoon JW, Jun HS. Viruses cause type 1 diabetes in animals. *Ann N Y Acad Sci*. 2006;1079:138-46.

55. Banatvala JE, Bryant J, Schernthaner G, Borkenstein M, Schober E, Brown D, et al. Coxsackie B, mumps, rubella, and cytomegalovirus specific IgM responses in patients with juvenile-onset insulin-dependent diabetes mellitus in Britain, Austria, and Australia. *Lancet*. 1985;1(8443):1409-12.

56. Dahlquist G, Frisk G, Ivarsson SA, Svanberg L, Forsgren M, Diderholm H. Indications that maternal coxsackie B virus infection during pregnancy is a risk factor for childhood-onset IDDM. *Diabetologia*. 1995;38(11):1371-3.

57. Ray CG, Palmer JP, Crossley JR, Williams RH. Coxsackie B virus antibody responses in juvenile-onset diabetes mellitus. *Clin Endocrinol*. 1980;12(4):375-8.

58. Krogvold L, Edwin B, Buanes T, Frisk G, Skog O, Anagandula M, et al. Detection of a low-grade enteroviral infection in the islets of langerhans of living patients newly diagnosed with type 1 diabetes. *Diabetes*. 2015;64(5):1682-7.

59. Dotta F, Censini S, van Halteren AG, Marselli L, Masini M, Dionisi S, et al. Coxsackie B4 virus infection of beta cells and natural killer cell insulitis in recent-onset type 1 diabetic patients. *Proc Natl Acad Sci USA*. 2007;104(12):5115-20.

60. Ylipaasto P, Klingel K, Lindberg AM, Otonkoski T, Kandolf R, Hovi T, et al. Enterovirus infection in human pancreatic islet cells, islet tropism in vivo and receptor involvement in cultured islet beta cells. *Diabetologia*. 2004;47(2):225-39.

61. Rosu V, Ahmed N, Paccagnini D, Gerlach G, Fadda G, Hasnain SE, et al. Specific immunoassays confirm association of Mycobacterium avium subsp. paratuberculosis with type-1 but not type-2 diabetes mellitus. *PLoS One*. 2009;4(2):e4386.

62. Vaarala O, Klemetti P, Juhela S, Simell O, Hyoty H, Ilonen J. Effect of coincident enterovirus infection and cows' milk exposure on immunisation to insulin in early infancy. *Diabetologia*. 2002;45(4):531-4.

63. Vaarala O, Knip M, Paronen J, Hamalainen AM, Muona P, Vaatainen M, et al. Cow's milk formula feeding induces primary immunization to insulin in infants at genetic risk for type 1 diabetes. *Diabetes*. 1999;48(7):1389-94.

64. Knip M, Åkerblom HK, Becker D, Dosch H-M, Dupre J, Fraser W, et al. Hydrolyzed Infant Formula and Early β-Cell Autoimmunity: A Randomized Clinical Trial. *JAMA*. 2014;311(22):2279-87.

65. Choi YJ, Kim MK, Jeong SJ. Vitamin D deficiency in infants aged 1 to 6 months. *Korean J Pediatr*. 2013;56(5):205-10.

66. Wall CR, Grant CC, Jones I. Vitamin D status of exclusively breastfed infants aged 2–3 months. *Arch Dis Child*. 2013;98(3):176.

67. Darmawikarta D, Chen Y, Lebovic G, Birken CS, Parkin PC, Maguire JL. Total duration of breastfeeding, vitamin D supplementation, and serum levels of 25-Hydroxyvitamin D. *Am J Public Health*. 2016;106(4):714-9.

68. Ward LM, Gaboury I, Ladhani M, Zlotkin S. Vitamin D-deficiency rickets among children in Canada. *CMAJ*. 2007;177(2):161-6.

69. Wagner CL, Greer FR. Prevention of rickets and vitamin D deficiency in infants, children, and adolescents. *Pediatrics*. 2008;122(5):1142-52.

70. Rigby WF. The immunobiology of vitamin D. *Immunol Today*. 1988;9(2):54-8.

71. Mathieu C, Laureys J, Sobis H, Vandeputte M, Waer M, Bouillon R. 1,25-Dihydroxyvitamin D3 prevents insulitis in NOD mice. *Diabetes*. 1992;41(11):1491-5.

72. Aranow C. Vitamin D and the Immune System. *J Investig Med*. 2011;59(6):881-6.

73. IOM (Institute of Medicine). *Dietary Reference Intakes for Calcium and Vitamin D*. Washington, DC: National Academies Press (US); 2011.

74. Jones G, Strugnell SA, DeLuca HF. Current understanding of the molecular actions of vitamin D. *Physiol Rev*. 1998;78(4):1193-231.

75. Bailey R, Cooper J, Zeitels L, Smyth D. Association of the vitamin D metabolism gene CYP27B1 with type 1 diabetes. *Diabetes*. 2007;56(10):2616-21.

76. Ramos-Lopez E, Bruck P, Jansen T, Herwig J, Badenhoop K. CYP2R1 (vitamin D 25-hydroxylase) gene is associated with susceptibility to type 1 diabetes and vitamin D levels in Germans. *Diabetes Metab Res Rev*. 2007;23(8):631-6.

77. Cooper JD, Smyth DJ, Walker NM, Stevens H, Burren OS, Wallace C, et al. Inherited variation in vitamin D genes is associated with predisposition to autoimmune disease type 1 diabetes. *Diabetes*. 2011;60(5):1624-31.

78. Levy-Marchal C, Patterson C, Green A. Variation by age group and seasonality at diagnosis of childhood IDDM in Europe. The EURODIAB ACE Study Group (Abstract). *Diabetologia*. 1995;38(7):823-30.

79. Moltchanova EV, Schreier N, Lammi N, Karvonen M. Seasonal variation of diagnosis of Type 1 diabetes mellitus in children worldwide. *Diabet Med*. 2009;26(7):673-8.

80. Dahlquist G, Mustonen L. Childhood onset diabetes--time trends and climatological factors. *Int J Epidemiol*. 1994;23(6):1234-41.

81. Knip M, Veijola R, Virtanen SM, Hyöty H, Vaarala O, Åkerblom HK. Environmental Triggers and Determinants of Type 1 Diabetes. *Diabetes*. 2005;54(suppl 2):S125.

82. Holick MF. Diabetes and the vitamin D connection. *Curr Diab Rep*. 2008;8(5):393-8.

83. Suarez L, Barrett-Connor E. Seasonal variation in fasting plasma glucose levels in man. *Diabetologia*. 1982;22(4):250-3.

84. Richer MJ, Horwitz MS. Viral infections in the pathogenesis of autoimmune diseases: focus on type 1 diabetes. *Front Biosci*. 2008;13:4241-57.

85. Gamble DR, Taylor KW. Seasonal incidence of diabetes mellitus. *Br Med J*. 1969;3(5671):631-3.

86. Kramna L, Kolarova K, Oikarinen S, Pursiheimo JP, Ilonen J, Simell O, et al. Gut virome sequencing in children with early islet autoimmunity. *Diabetes Care*. 2015;38(5):930-3.

87. Lee HS, Briese T, Winkler C, Rewers M, Bonifacio E, Hyoty H, et al. Next-generation sequencing for viruses in children with rapid-onset type 1 diabetes. *Diabetologia*. 2013;56(8):1705-11.

88. Op de Beeck A, Eizirik DL. Viral infections in type 1 diabetes mellitus — why the β cells? *Nat Rev Endocrinol*. 2016;12(5):263-73.

89. Larsson HE, Lynch K, Lernmark B, Hansson G, Lernmark Å, Ivarsson SA, et al. Relationship between increased relative birthweight and infections during pregnancy in children with a high-risk diabetes HLA genotype. *Diabetologia*. 2007;50(6):1161-9.

90. Fasano A. Intestinal permeability and its regulation by zonulin: diagnostic and therapeutic implications. *Clin Gastroenterol Hepatol*. 2012;10(10):1096-100.

91. Fasano A, Not T, Wang W, Uzzau S, Berti I, Tommasini A, et al. Zonulin, a newly discovered modulator of intestinal permeability, and its expression in coeliac disease. *Lancet*. 2000;355(9214):1518-9.

92. Sapone A, De Magistris L, Pietzak M, Clemente MG, Tripathi A, Cucca F, et al. Zonulin upregulation is associated with increased gut permeability in subjects with type 1 diabetes and their relatives. *Diabetes*. 2006;55(5):1443-9.

93. Vaarala O, Atkinson Ma, Neu J. The "perfect storm" for type 1 diabetes: The complex interplay between intestinal microbiota, gut permeability, and mucosal immunity. *Diabetes*. 2008;57(10):2555-62.

94. Savilahti E, Ormala T, Saukkonen T, Sandini-Pohjavuori U, Kantele JM, Arato A, et al. Jejuna of patients with insulin-dependent diabetes mellitus (IDDM) show signs of immune activation. *Clin Exp Immunol*. 1999;116(1):70-7.

95. Westerholm-Ormio M, Vaarala O, Pihkala P, Ilonen J, Savilahti E. Immunologic activity in the small intestinal mucosa of pediatric patients with type 1 diabetes. *Diabetes*. 2003;52(9):2287-95.

96. Picker LJ, Martin RJ, Trumble A, Newman LS, Collins PA, Bergstresser PR, et al. Differential expression of lymphocyte homing receptors by human memory/effector T cells in pulmonary versus cutaneous immune effector sites. *Eur J Immunol*. 1994;24(6):1269-77.

97. Evangelou M, Smyth DJ, Fortune MD, Burren OS, Walker NM, Guo H, et al. A method for gene-based pathway analysis using genomewide association study summary statistics reveals nine new type 1 diabetes associations. *Genet Epidemiol*. 2014;38(8):661-70.

98. Paronen J, Klemetti P, Kantele JM, Savilahti E, Perheentupa J, Akerblom HK, et al. Glutamate decarboxylase-reactive peripheral blood lymphocytes from patients with IDDM express gut-specific homing receptor alpha4beta7-integrin. *Diabetes*. 1997;46(4):583-8.

99. Atkinson MA, Eisenbarth GS. Type 1 diabetes: new perspectives on disease pathogenesis and treatment. *Lancet*. 2001;358(9277):221-9.

100. Eizirik DL, Colli ML, Ortis F. The role of inflammation in insulitis and β-cell loss in type 1 diabetes. *Nat Rev Endocrinol*. 2009;5:219.

101. Eizirik DL, Sammeth M, Bouckenooghe T, Bottu G, Sisino G, Igoillo-Esteve M, et al. The human pancreatic islet transcriptome: Expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. *PLoS Genet*. 2012;8(3).

102. Onengut-Gumuscu S, Chen W-M, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet*. 2015;47:381–6.

103. Foulis AK. Pancreatic pathology in type 1 diabetes in human. *Novartis Found symposium*. 2008;292:2-13.

104. Pukel C, Baquerizo H, Rabinovitch A. Destruction of rat islet cell monolayers by cytokines. *Diabetes*. 1988;37(January):133-6.

105. Ziegler AG, Nepom GT. Prediction and pathogenesis in type 1 diabetes. *Immunity*. 2010;32(4):468-78.

106. Chase HP, Cooper S, Osberg I, Stene LC, Barriga K, Norris J, et al. Elevated C-reactive protein levels in the development of type 1 diabetes. *Diabetes*. 2004;53(10):2569-73.

107. Verge CF, Gianani R, Kawasaki E, Yu L, Pietropaolo M, Jackson RA, et al. Prediction of type I diabetes in first-degree relatives using a combination of insulin, GAD, and ICA512bdc/IA-2 autoantibodies. *Diabetes*. 1996;45(7):926-33.

108. Insel RA, Dunne JL, Atkinson MA, Chiang JL, Dabelea D, Gottlieb PA, et al. Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association. *Diabetes Care*. 2015;38(10):1964-74.

109. LaGasse JM, Brantley MS, Leech NJ, Rowe RE, Monks S, Palmer JP, et al. Successful prospective prediction of type 1 diabetes in schoolchildren through multiple defined autoantibodies: an 8-year follow-up of the Washington State Diabetes Prediction Study. *Diabetes Care*. 2002;25(3):505-11.

110. Bingley PJ. Clinical applications of diabetes antibody testing. *J Clin Endocrinol Metab*. 2010;95(1):25-33.

111. Wenzlau JM, Juhl K, Yu L, Moua O, Sarkar SA, Gottlieb P, et al. The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. *Proc Natl Acad Sci USA*. 2007;104(43):17040-5.

112. Parikka V, Nanto-Salonen K, Saarinen M, Simell T, Ilonen J, Hyoty H, et al. Early seroconversion and rapidly increasing autoantibody concentrations predict prepubertal manifestation of type 1 diabetes in children at genetic risk. *Diabetologia*. 2012;55(7):1926-36.

113. Ziegler AG, Bonifacio E. Age-related islet autoantibody incidence in offspring of patients with type 1 diabetes. *Diabetologia*. 2012;55(7):1937-43.

114. Ziegler AG, Rewers M, Simell O, et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *JAMA*. 2013;309(23):2473-9.

115. Krischer JP, Lynch KF, Schatz DA, Ilonen J, Lernmark A, Hagopian WA, et al. The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. *Diabetologia*. 2015;58(5):980-7.

116. Vehik K, Dabelea D. The changing epidemiology of type 1 diabetes: why is it going through the roof? *Diabetes Metab Res Rev*. 2011;27(1):3-13.

117. Concannon P, Rich SS, Nepom GT. Genetics of type 1A diabetes. *N Engl J Med*. 2009;360(16):1646-54.

118. Hummel M, Bonifacio E, Schmid S, Walter M, Knopff A, Ziegler AG. Brief communication: early appearance of islet autoantibodies predicts childhood type 1 diabetes in offspring of diabetic parents. *Ann Intern Med*. 2004;140(11):882-6.

119. Pollanen PM, Lempainen J, Laine AP, Toppari J, Veijola R, Vahasalo P, et al. Characterisation of rapid progressors to type 1 diabetes among children with HLA-conferred disease susceptibility. *Diabetologia*. 2017;60(7):1284-93.

120. Lynch KF, Lee HS, Torn C, Vehik K, Krischer JP, Larsson HE, et al. Gestational respiratory infections interacting with offspring HLA and CTLA-4 modifies incident beta-cell autoantibodies. *J Autoimmun*. 2018;86:93-103.

121. Ilonen J, Lempainen J, Hammais A, Laine AP, Harkonen T, Toppari J, et al. Primary islet autoantibody at initial seroconversion and autoantibodies at diagnosis of type 1 diabetes as markers of disease heterogeneity. *Pediatr Diabetes*. 2018;19(2):284-92.

122. Ilonen J, Kiviniemi M, Lempainen J, Simell O, Toppari J, Veijola R, et al. Genetic susceptibility to type 1 diabetes in childhood - estimation of HLA class II associated disease risk and class II effect in various phases of islet autoimmunity. *Pediatr Diabetes*. 2016;17 Suppl 22:8-16.

123. Mbunwe E, Van der Auwera BJ, Weets I, Van Crombrugge P, Crenier L, Coeckelberghs M, et al. In antibody-positive first-degree relatives of patients with type 1 diabetes, HLA-A*24 and HLA-B*18, but not HLA-B*39, are predictors of impending diabetes with distinct HLA-DQ interactions. *Diabetologia*. 2013;56(9):1964-70.

124. Lipponen K, Gombos Z, Kiviniemi M, Siljander H, Lempainen J, Hermann R, et al. Effect of HLA class I and class II alleles on progression from autoantibody positivity to overt type 1 diabetes in children with risk-associated class II genotypes. *Diabetes*. 2010;59(12):3253-6.

125. Chen YG, Cabrera SM, Jia S, Kaldunski ML, Kramer J, Cheong S, et al. Molecular signatures differentiate immune states in type 1 diabetic families. *Diabetes*. 2014;63(11):3960-73.

126. Tai N, Wong FS, Wen L. The role of the innate immune system in destruction of pancreatic beta cells in NOD mice and humans with type I diabetes. *J Autoimmun*. 2016;71:26-34.

127. Skog O, Korsgren O. Aetiology of type 1 diabetes: physiological growth in children affects disease progression. *Diabetes Obes Metab*. 2018;20(4):775-85.

128. Korsgren S, Molin Y, Salmela K, Lundgren T, Melhus Å, Korsgren O. On the Etiology of Type 1 Diabetes: A New Animal Model Signifying a Decisive Role for Bacteria Eliciting an Adverse Innate Immunity Response. *Am J Pathol*. 2012;181(5):1735-48.

129. Skowera A, Ellis RJ, Varela-Calvino R, Arif S, Huang GC, Van-Krinks C, et al. CTLs are targeted to kill beta cells in patients with type 1 diabetes through recognition of a glucose-regulated preproinsulin epitope. *J Clin Invest*. 2008;118(10):3390-402.

130. Fattovich G, Giustina G, Favarato S, Ruol A. A survey of adverse events in 11,241 patients with chronic viral hepatitis treated with alfa interferon. *J Hepatol*. 1996;24(1):38-47.

131. Strachan DP. Hay fever, hygiene, and household size. *Br Med J*. 1989;299(6710):1259.

132. Viskari H, Koskela P, Lonnrot M, Luonuansuu S, Reunanen A, Baer M. Can changing epidemiology of enterovirus infections explain the increasing incidence of type 1 diabetes. *Diabetes Care*. 2000;23:414–6.

133. Nathanson N, Kew OM. From emergence to eradication: the epidemiology of poliomyelitis deconstructed. *Am J Epidemiol*. 2010;172(11):1213-29.

134. Hyoty H. Viruses in type 1 diabetes. *Pediatr Diabetes*. 2016;17 Suppl 22:56-64.

135. Sioofy-Khojine AB, Lehtonen J, Nurminen N, Laitinen OH, Oikarinen S, Huhtala H, et al. Coxsackievirus B1 infections are associated with the initiation of insulin-driven autoimmunity that progresses to type 1 diabetes. *Diabetologia*. 2018;61(5):1193-202.

136. Uusitalo U, Lee HS, Andren Aronsson C, Vehik K, Yang J, Hummel S, et al. Early Infant Diet and Islet Autoimmunity in the TEDDY Study. *Diabetes Care*. 2018;41(3):522-30.

137. Torres-Castro I, Arroyo-Camarena UD, Martinez-Reyes CP, Gomez-Arauz AY, Duenas-Andrade Y, Hernandez-Ruiz J, et al. Human monocytes and macrophages undergo M1-type inflammatory polarization in response to high levels of glucose. *Immunol Lett*. 2016;176:81-9.

138. Stentz FB, Umpierrez GE, Cuervo R, Kitabchi AE. Proinflammatory cytokines, markers of cardiovascular risks, oxidative stress, and lipid peroxidation in patients with hyperglycemic crises. *Diabetes*. 2004;53(8):2079-86.

139. Stentz FB, Kitabchi AE. Hyperglycemia-induced activation of human T-lymphocytes with de novo emergence of insulin receptors and generation of reactive oxygen species. *Biochem Biophys Res Commun*. 2005;335(2):491-5.

140. Tripathy D, Mohanty P, Dhindsa S, Syed T, Ghanim H, Aljada A, et al. Elevation of free fatty acids induces inflammation and impairs vascular reactivity in healthy subjects. *Diabetes*. 2003;52(12):2882-7.

141. Maciolek JA, Pasternak JA, Wilson HL. Metabolism of activated T lymphocytes. *Curr Opin Immunol*. 2014;27:60-74.

142. Matveyenko AV, Butler PC. Relationship between beta-cell mass and diabetes onset. *Diabetes Obes Metab*. 2008;10 Suppl 4(0 4):23-31.

143. Leighton E, Sainsbury CAR, Jones GC. A Practical Review of C-Peptide Testing in Diabetes. *Diabetes Ther*. 2017;8(3):475-87.

144. Hellerström C, Swenne I, Andersson A. Islet Cell Replication and Diabetes. In: Lefèbvre P, Pipeleers D, editors. *The Pathology of the Endocrine Pancreas in Diabetes*. Heidelberg, Germany: Springer Verlag; 1988. p. 141-70.

145. Williams AJ, Thrower SL, Sequeiros IM, Ward A, Bickerton AS, Triay JM, et al. Pancreatic volume is reduced in adult patients with recently diagnosed type 1 diabetes. *J Clin Endocrinol Metab*. 2012;97(11):E2109-13.

146. Henquin JC, Ravier MA, Nenquin M, Jonas JC, Gilon P. Hierarchy of the beta-cell signals controlling insulin secretion. *Eur J Clin Invest*. 2003;33(9):742-50.

147. Olofsson CS, Gopel SO, Barg S, Galvanovskis J, Ma X, Salehi A, et al. Fast insulin secretion reflects exocytosis of docked granules in mouse pancreatic B-cells. *Pflugers Arch*. 2002;444(1-2):43-51.

148. Koskinen MK, Helminen O, Matomaki J, Aspholm S, Mykkanen J, Makinen M, et al. Reduced beta-cell function in early preclinical type 1 diabetes. *Eur J Endocrinol*. 2016;174(3):251-9.

149. Koskinen MK, Lempainen J, Loyttyniemi E, Helminen O, Hekkala A, Harkonen T, et al. Class II HLA genotype association with first phase insulin response is explained by islet autoantibodies. *J Clin Endocrinol Metab*. 2018;103(8):2870-2878.

150. Meier JJ, Butler AE, Saisho Y, Monchamp T, Galasso R, Bhushan A, et al. Beta-cell replication is the primary mechanism subserving the postnatal expansion of beta-cell mass in humans. *Diabetes*. 2008;57(6):1584-94.

151. Bjork E, Kampe O, Karlsson FA, Pipeleers DG, Andersson A, Hellerstrom C, et al. Glucose regulation of the autoantigen GAD65 in human pancreatic islets. *J Clin Endocrinol Metab*. 1992;75(6):1574-6.

152. Bakay M, Pandey R, Hakonarson H. Genes involved in type 1 diabetes: An update. *Genes*. 2013;4:499-521.

153. Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. *PLoS Comput Biol*. 2012;8(12).

154. Visser M, Kayser M, Palstra RJ. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome Res*. 2012;22(3):446-55.

155. Beck S, Trowsdale J. The human major histocompatability complex: lessons from the DNA sequence. *Annu Rev Genomics Hum Genet*. 2000;1:117-37.

156. de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet*. 2006;38(10):1166-72.

157. Traherne JA, Horton R, Roberts AN, Miretti MM, Hurles ME, Stewart CA, et al. Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history. *PLoS Genet*. 2006;2(1):e9.

158. Warner CM, Tyas DA, Goldstein C, Comiskey M, Cohen J, Brenner CA. Genotyping: the HLA system and embryo development. *Reprod Biomed Online*. 2002;4(2):133-9.

159. Zheng ML, Niu LY. [Association of soluble human leukocyte antigen DQB1 with recurrent spontaneous abortion] (Abstract). *Zhonghua Fu Chan Ke Za Zhi*. 2006;41(3):152-4.

160. Honda K, Takakuwa K, Hataya I, Yasuda M, Kurabayashi T, Tanaka K. HLA-DQB1 and HLA-DPB1 genotypes in severe preeclampsia (Abstract). *Obstet Gynecol*. 2000;96(3):385-9.

161. Järvinen TM, Harjutsalo V, Kinnunen L, Miettinen ME, Tuomilehto-Wolf E, Tuomilehto J. A population-specific diabetogenic haplotype HLA-A2,Cw1,B56,DR4,DQ8 is associated with high birthweight in Finnish diabetic families. *Genes and Immunity*. 2008;9(3):207-13.

162. Penn DJ, Damjanovich K, Potts WK. MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proceedings of the National Academy of Sciences*. 2002;99(17):11260-64.

163. European Group for Blood and Marrow Transplantation (EBMT). HLA MANUAL: A Guide to the completion of the EBMT Form. Available from: www.ebmt.org.

164. Sette A, Buus S, Colon S, Smith JA, Miles C, Grey HM. Structural characteristics of an antigen required for its interaction with Ia and recognition by T cells. *Nature*. 1987;328:395.

165. Todd JA, Bell JI, McDevitt HO. HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature*. 1987;329(6140):599-604.

166. Rotter JI, Rimoin DL. Diabetes mellitus: the search for genetic markers (Abstract). *Diabetes Care*. 1979;2(2):215-26.

167. Hodge SE, Rotter JI, Lange KL. A three-allele model for heterogeneity of juvenile onset insulin-dependent diabetes. *Ann Hum Genet*. 1980;43(4):399-409.

168. Noble JA, Valdes AM, Cook M, Klitz W, Thomson G, Erlich HA. The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. *Am J Hum Genet*. 1996;59(5):1134-48.

169. Horn GT, Bugawan TL, Long CM, Manos MM, Erlich HA. Sequence analysis of HLA class II genes from insulin-dependent diabetic individuals. *Hum Immunol*. 1988;21(4):249-63.

170. Shankarkumar U. The Human Leukocyte Antigen (HLA) system. *Int J Hum Genet*. 2004;4(2):91-103.

171. Noble JA, Valdes AM, Varney MD, Carlson JA, Moonsamy P, Fear AL, et al. HLA class I and genetic susceptibility to type 1 diabetes: results from the Type 1 Diabetes Genetics Consortium. *Diabetes*. 2010;59(11):2972-9.

172. Pugliese A, Gianani R, Moromisato R, Awdeh ZL, Alper CA, Erlich HA, et al. HLA-DQB1*0602 Is Associated With Dominant Protection From Diabetes Even Among Islet Cell Antibody–Positive First-Degree Relatives of Patients with IDDM. *Diabetes*. 1995;44(6):608.

173. Redondo MJ, Grant SF, Davis A, Greenbaum C. Dissecting heterogeneity in paediatric Type 1 diabetes: association of TCF7L2 rs7903146 TT and low-risk human leukocyte antigen (HLA) genotypes. *Diabet Med*. 2017;34(2):286-90.

174. Risch N. Assessing the role of HLA-linked and unlinked determinants of disease. *Am J Hum Genet*. 1987;40(1):1-14.

175. Kim MS, Polychronakos C. Immunogenetics of type 1 diabetes. *Horm Res*. 2005;64(4):180-8.

176. Pociot F, Akolkar B, Concannon P, Erlich HA, Julier C, Morahan G, et al. Genetics of type 1 diabetes: What's next? *Diabetes*. 2010;59(July):1561-71.

177. Mangalam AK, Taneja V, David CS. HLA class II molecules influence susceptibility versus protection in inflammatory diseases by determining the cytokine profile. *J Immunol*. 2013;190(2):513-8.

178. Thude H, Rosenhahn S, Hunger-Dathe W, Muller UA, Barz D. A transmembrane protein-tyrosine phosphatase receptor type C (CD45) exon A point mutation (77 C to G) is not associated with the development of type 1 diabetes mellitus in a German population. *Eur J Immunogenet*. 2004;31(6):245-7.

179. Thude H, Weissenborn S, Vilser C, Muller UA, Kloos C, Wolf G, et al. No association between transmembrane protein-tyrosine-phosphatase receptor type C (CD45) exon A 77C>G transversion and Hashimoto's thyroiditis in a German population. *Hum Immunol*. 2010;71(2):220-3.

180. Ramanujam R, Pirskanen R, Hammarstrom L. The CD45 77C/G allele is not associated with myasthenia gravis - a reassessment of the potential role of CD45 in autoimmunity. *BMC Res Notes*. 2010;3:292.

181. Chistiakov DA, Voronova NV, Chistiakov PA. The crucial role of IL-2/IL-2RA-mediated immune regulation in the pathogenesis of type 1 diabetes, an evidence coming from genetic and animal model studies. *Immunol Lett*. 2008;118(1):1-5.

182. Torn C, Hadley D, Lee HS, Hagopian W, Lernmark A, Simell O, et al. Role of type 1 diabetes-associated SNPs on risk of autoantibody positivity in the TEDDY study. *Diabetes*. 2015;64(5):1818-29.

183. Pietrobelli A, Agosti M. Nutrition in the first 1000 days: ten practices to minimize obesity emerging from published science. *Int J Environ Res Public Health*. 2017;14(12).

184. Comerford KB, Ayoob KT, Murray RD, Atkinson SA. The role of avocados in maternal diets during the periconceptional period, pregnancy, and lactation. *Nutrients*. 2016;8(5).

185. Cole TJ. Secular trends in growth. *Proc Nutr Soc*. 2000;59(2):317-24.

186. Wilcox AJ. On the importance--and the unimportance--of birthweight. *Int J Epidemiol*. 2001;30(6):1233-41.

187. Read JS, Clemens JD, Klebanoff MA. Moderate low birth weight and infectious disease mortality during infancy and childhood. *Am J Epidemiol*. 1994;140:721-33.

188. Samuelsen SO, Magnus P, Bakketeig LS. Birth weight and mortality in childhood in Norway. *Am J Epidemiol*. 1998;148(10):983-91.

189. Hviid A, Melbye M. The impact of birth weight on infectious disease hospitalization in childhood. *Am J Epidemiol*. 2007;165(7):756-61.

190. Hales CN, Barker DJ. Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Diabetologia*. 1992;35(7):595-601.

191. Hales CN, Barker DJ. The thrifty phenotype hypothesis. *Br Med Bull*. 2001;60:5-20.

192. Kermack WO, McKendrick AG, McKinlay PL. Death-rates in Great Britain and Sweden. Some general regularities and their significance. *Int J Epidemiol*. 2001;30(4):678-83.

193. Almond D, Currie J. Killing Me Softly: The Fetal Origins Hypothesis. *J Econ Perspect*. 2011;25(3):153-72.

194. Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature*. 2010;465(7299):721-7.

195. McKinney PA, Parslow R, Gurney K, Law G, Bodansky HJ, Williams DRR. Antenatal risk factors for childhood diabetes mellitus, a case-control study of medical record data in Yorkshire, UK. *Diabetologia*. 1997;40(8):933-9.

196. Ievins R, Roberts SE, Goldacre MJ. Perinatal factors associated with subsequent diabetes mellitus in the child: Record linkage study. *Diabet Med*. 2007;24(6):664-70.

197. Royal College of Obstetricians and Gynaecologists. *The investigation and management of the small-for-gestational-age fetus (Green-top Guideline No.31)*. 2013. Available from: https://www.rcog.org.uk.

198. Clayton PE, Cianfarani S, Czernichow P, Johannsson G, Rapaport R, Rogol A. Management of the child born small for gestational age through to adulthood: a consensus statement of the International Societies of Pediatric Endocrinology and the Growth Hormone Research Society. *J Clin Endocrinol Metab*. 2007;92(3):804-10.

199. Cameron N, Hawley NL. Should the UK use WHO growth charts? *Paediatr Child Health*. 2010;20(4):151-6.

200. Wright C, Booth I, Buckler J, Cameron N, Cole T, Healy M, et al. Growth reference charts for use in the United Kingdom. *Arch Dis Child*. 2002;86(1):11-4.

201. Cole TJ, Freeman JV, Preece MA. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Stat Med*. 1998;17(4):407-29.

202. WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards: Growth velcity based on weight, length and head circumference. *World Health Organization*. 2009.

203. Aroviita P, Partanen J, Sistonen P, Teramo K, Kekomäki R. High birth weight is associated with human leukocyte antigen (HLA) DRB1*13 in full-term infants. *Eur J Immunogenet*. 2004;31(1):21-6.

204. Algert CS, McElduff a, Morris JM, Roberts CL. Perinatal risk factors for early onset of Type 1 diabetes in a 2000-2005 birth cohort. *Diabet Med*. 2009;26(12):1193-7.

205. Khashan AS, Kenny LC, Lundholm C, Kearney PM, Gong T, McNamee R, et al. Gestational Age and Birth Weight and the Risk of Childhood Type 1 Diabetes: A Population-Based Cohort and Sibling Design Study. *Diabetes Care*. 2015;38(12):2308-15.

206. Weber DR, Leonard MB, Zemel BS. Body composition analysis in the pediatric population. *Pediatr Endocrinol Rev*. 2012;10(1):130-9.

207. Lamb MM, Yin X, Zerbe GO, Klingensmith GJ, Dabelea D, Fingerlin TE, et al. Height growth velocity, islet autoimmunity and type 1 diabetes development: The Diabetes Autoimmunity Study in the Young. *Diabetologia*. 2009;52(10):2064-71.

208. Keys A, Fidanza F, Karvonen MJ, Kimura N, Taylor HL. Indices of relative weight and obesity. *Int J Epidemiol*. 2014;43(3):655-65.

209. Lawler-Heavner J, Cruickshanks KJ, Hay WW, Gay EC, Hamman RF. Birth size and risk of insulin-dependent diabetes mellitus (IDDM). *Diabetes Res Clin Pract*. 1994;24(3):153-9.

210. Beyerlein A, Fahrmeir L, Mansmann U, Toschke AM. Alternative regression models to assess increase in childhood BMI. *BMC Med Res Methodol*. 2008;8:59.

211. Mayr A, Hothorn T, Fenske N. Prediction intervals for future BMI values of individual children - a non-parametric approach by quantile boosting. *BMC Med Res Methodol*. 2012;12:6.

212. Hypponen E, Kenward MG, Virtanen SM, Piitulainen A, Virta-Autio P, Tuomilehto J, et al. Infant feeding, early weight gain, and risk of type 1 diabetes. Childhood Diabetes in Finland (DiMe) Study Group. *Diabetes Care*. 1999;22(12):1961-5.

213. Amorim M, Melo AN. Revisiting head circumference of Brazilian newborns in public and private maternity hospitals. *Arq Neuropsiquiatr*. 2017;75(6):372-80.

214. Davies DP. Size at birth and growth in the first year of life of babies who are overweight and underweight at birth. *Proc Nutr Soc*. 1980;39(1):25-33.

215. Ong KK, Petry CJ, Emmett PM, Sandhu MS, Kiess W, Hales CN, et al. Insulin sensitivity and secretion in normal children related to size at birth, postnatal growth, and plasma insulin-like growth factor-I levels. *Diabetologia*. 2004;47:1064-70.

216. Smith GD. Assessing intrauterine influences on offspring health outcomes: can epidemiological studies yield robust findings? *Basic Clin Pharmacol Toxicol*. 2008;102(2):245-56.

217. Ong KKL, Preece MA, Emmett PM, Ahmed ML, Dunger DB. Size at birth and early childhood growth in relation to maternal smoking, parity and infant breast-feeding: longitudinal birth cohort study and analysis. *Pediatr Res*. 2002;52(6):863-7.

218. Beaty TH, Skjaerven R, Breazeale DR, Liang KY. Analyzing sibship correlations in birth weight using large sibships from Norway. *Genet Epidemiol*. 1997;14(4):423-33.

219. Hutcheon JA, Platt RW, Meltzer SJ, Egeland GM. Is birth weight modified during pregnancy? Using sibling differences to understand the impact of blood glucose, obesity, and maternal weight gain in gestational diabetes. *Am J Obstet Gynecol*. 2006;195(2):488-94.

220. Carr-Hill R, Campbell DM, Hall MH, Meredith A. Is birth weight determined genetically? *Br Med J*. 1987;295(6600):687-9.

221. Brooke OG, Anderson HR, Bland JM, Peacock JL, Stewart CM. Effects on birth weight of smoking, alcohol, caffeine, socioeconomic factors, and psychosocial stress. *Br Med J*. 1989;298(6676):795-801.

222. Wilcox MA, Chang AM, Johnson IR. The effects of parity on birthweight using successive pregnancies. *Acta Obstet Gynecol Scand*. 1996;75(5):459-3.

223. Yu Z, Han S, Zhu J, Sun X, Ji C, Guo X. Pre-pregnancy body mass index in relation to infant birth weight and offspring overweight/obesity: a systematic review and meta-analysis. *PLoS One*. 2013;8(4).

224. Ong KK, Diderholm B, Salzano G, Wingate D, Hughes IA, MacDougall J, et al. Pregnancy insulin, glucose, and BMI contribute to birth outcomes in nondiabetic mothers. *Diabetes Care*. 2008;31(11):2193-7.

225. Dougherty CR, Jones AD. The determinants of birth weight. *Am J Obstet Gynecol*. 1982;144(2):190-200.

226. Catalano PM, Kirwan JP. Maternal factors that determine neonatal size and body fat. *Curr Diab Rep*. 2001;1(1):71-7.

227. Addo OY, Stein AD, Fall CH, Gigante DP, Guntupalli AM, Horta BL, et al. Maternal height and child growth patterns. *J Pediatr*. 2013;163(2):549-54.e1.

228. Zhang X, Cnattingius S, Platt RW, Joseph KS, Kramer MS. Are babies born to short, primiparous, or thin mothers “normally” or “abnormally” small? *J Pediatr*. 2007;150(6):603-7.e3.

229. Zhang G, Bacelis J, Lengyel C, Teramo K, Hallman M, Helgeland Ø, et al. Assessing the causal relationship of maternal height on birth size and gestational age at birth: a Mendelian randomization analysis. *PLoS Med*. 2015;12(8):1-23.

230. Sterner Y, Törn C, Lee HS, Larsson H, Winkler C, McLeod W, et al. Country-specific birth weight and length in type 1 diabetes high-risk HLA genotypes in combination with prenatal characteristics. *J Perinatol*. 2011;31(12):764-9.

231. Hummel M, Marienfeld S, Huppmann M, Knopff A, Voigt M, Bonifacio E, et al. Fetal growth is increased by maternal type 1 diabetes and HLA DR4-related gene interactions. *Diabetologia*. 2007;50(4):850-8.

232. Hill DJ, Milner RD. Insulin as a growth factor. *Pediatr Res*. 1985;19(9):879-86.

233. Sermer M, Naylor CD, Gare DJ, Kenshole AB, Ritchie JW, Farine D, et al. Impact of increasing carbohydrate intolerance on maternal-fetal outcomes in 3637 women without gestational diabetes. The Toronto Tri-Hospital Gestational Diabetes Project. *Am J Obstet Gynecol*. 1995;173(1):146-56.

234. Parretti E, Mecacci F, Papini M, Cioni R, Carignani L, Mignosa M, et al. Third-trimester maternal glucose levels from diurnal profiles in nondiabetic pregnancies: correlation with sonographic parameters of fetal growth. *Diabetes Care*. 2001;24(8):1319-23.

235. Scholl TO, Chen X, Gaughan C, Smith WK. Influence of maternal glucose level on ethnic differences in birth weight and pregnancy outcome. *Am J Epidemiol*. 2002;156(6):498-506.

236. Crozier SR, Inskip HM, Godfrey KM, Cooper C, Harvey NC, Cole Z, et al. Weight gain in pregnancy and childhood body composition: Findings from the Southampton Women's Survey. *American Journal of Clinical Nutrition*. 2010;91(6):1745-51.

237. Thilaganathan B. Association of higher maternal blood pressure with lower infant birthweight. *Hypertension*. 2016;67(3):499.

238. Tyrrell J, Richmond RC, Palmer TM, Feenstra B, Rangarajan J, Metrustry S, et al. Genetic evidence for causal relationships between maternal obesity-related traits and birth weight. *JAMA*. 2016;315(11):1129-40.

239. Hamulka J, Zielinska MA, Chadzynska K. The combined effects of alcohol and tobacco use during pregnancy on birth outcomes. *Rocz Panstw Zakl Hig*. 2018;69(1):45-54.

240. Knudsen VK, Orozova-Bekkevold IM, Mikkelsen TB, Wolff S, Olsen SF. Major dietary patterns in pregnancy and fetal growth. *Eur J Clin Nutr*. 2008;62(4):463-70.

241. Nahum GG, Stanislaw H. Relationship of paternal factors to birth weight. *J Reprod Med*. 2003;48(12):963-8.

242. Ramakrishnan U, Martorell R, Schroeder DG, Flores R. Role of intergenerational effects on linear growth. *J Nutr*. 1999;129(2):544S-9S.

243. Pietiläinen KH, Kaprio J, Räsänen M, Winter T, Rissanen A, Rose RJ. Tracking of body size from birth to late adolescence: contributions of birth length, birth weight, duration of gestation, parents' body size, and twinship. *Am J Epidemiol*. 2001;154(1):21-9.

244. Larsson HE, Hansson G, Carlsson A, Cederwall E, Jonsson B, Jönsson B, et al. Children developing type 1 diabetes before 6 years of age have increased linear growth independent of HLA genotypes. *Diabetologia*. 2008;51(9):1623-30.

245. Peet A, M HA, Kool P, Ilonen J, Knip M, Tillmann V. Early postnatal growth in children with HLA-conferred susceptibility to type 1 diabetes. *Diabetes Metab Res Rev*. 2014;30:60-8.

246. Botton J, Heude B, Maccario J, Ducimetière P, Charles M-A, FLVS Study Group. Postnatal weight and height growth velocities at different ages between birth and 5 y and body composition in adolescent boys and girls. *Am J Clin Nutr*. 2008;87(6):1760-8.

247. U.S National Library of Medicine. MEdical Subject Headings (MESH) Database. Available from: www.nlm.nih.gov.

248. Widdowson EM. Harmony of Growth. *Lancet*. 1972:i:901-5.

249. Karlberg J. A biologically-oriented mathematical model (ICP) for human growth. *Acta Paediatr*. 1989;78(s350):70-94.

250. Tse WY, Hindmarsh PC, Brook CGD. The Infancy-Childhood-Puberty model of growth: clinical aspects. *Acta Paediatr*. 1989;78(s356):38-43.

251. Bailey R. Physical development and growth. In: Salkind N, editor. *Encyclopedia of human development*. Thousand Oaks: SAGE Publications, Inc; 2005.

252. Haymond M, Kappelgaard A-M, Czernichow P, Biller BMK, Takano K, Kiess W, et al. Early recognition of growth abnormalities permitting early intervention. *Acta Paediatr*. 2013;102(8):787-96.

253. Tanner JM. Issues and advances in adolescent growth and development. *J Adolesc Health Care*. 1987;8(6):470-8.

254. Hauspie RC, Bergman P, Bielicki T, Susanne C. Genetic variance in the pattern of the growth curve for height: A longitudinal analysis of male twins. *Ann Hum Biol*. 1994;21(4):347-62.

255. Johnson L, Llewellyn CH, van Jaarsveld CHM, Cole TJ, Wardle J. Genetic and environmental influences on infant growth: prospective analysis of the gemini twin birth cohort. *PLoS One*. 2011;6(5):e19918.

256. Butte NF. Energy requirements of infants. *Public Health Nutr*. 2005;8(7a):953-67.

257. Kuzawa CW. Adipose tissue in human infancy and childhood: an evolutionary perspective. *Am J Phys Anthropol*. 1998;Suppl 27:177-209.

258. Smith ER, Hurt L, Chowdhury R, Sinha B, Fawzi W, Edmond KM. Delayed breastfeeding initiation and infant survival: A systematic review and meta-analysis. *PLoS One*. 2017;12(7):e0180722.

259. Hauspie RC. Growth And Development. In: Rudan P, editor. *Physical (Biological) Anthropology*. Oxford, United Kingdom: Encyclopedia of Life Support Systems; 2009.

260. Prader A, Tanner JM, von Harnack GA. Catch-up growth following illness or starvation: An example of developmental canalization in man. *J Pediatr*. 1963;62(5):646-59.

261. Prader A. Catch-up growth. *Postgrad Med J*. 1978;54 Suppl 1:133-46.

262. Fields DA, Demerath EW, Pietrobelli A, Chandler-Laney PC. Body composition at 6 months of life: comparison of air displacement plethysmography and dual-energy X-ray absorptiometry. *Obesity*. 2012;20(11):2302-6.

263. Chomtho S, Wells JC, Williams JE, Davies PS, Lucas A, Fewtrell MS. Infant growth and later body composition: evidence from the 4-component model. *Am J Clin Nutr*. 2008;87(6):1776-84.

264. Wells JC, Fewtrell MS. Measuring body composition. *Arch Dis Child*. 2006;91(7):612-7.

265. Butte NF, Hopkinson JM, Wong WW, Smith EO, Ellis KJ. Body composition during the first 2 years of life: an updated reference. *Pediatr Res*. 2000;47(5):578-85.

266. Fomon SJ, Haschke F, Ziegler EE, Nelson SE. Body composition of reference children from birth to age 10 years. *Am J Clin Nutr*. 1982;35(5 Suppl):1169-75.

267. Jordan PN, Hall KD. Dynamic coordination of macronutrient balance during infant growth: insights from a mathematical model. *Am J Clin Nutr*. 2008;87(3):692-703.

268. Gale C, Thomas EL, Jeffries S, Durighel G, Logan KM, Parkinson JR, et al. Adiposity and hepatic lipid in healthy full-term, breastfed, and formula-fed human infants: a prospective short-term longitudinal cohort study. *Am J Clin Nutr*. 2014;99(5):1034-40.

269. Herrera E, Amusquivar E. Lipid metabolism in the fetus and the newborn. *Diabetes Metab Res Rev*. 2000;16(3):202-10.

270. EFSA Panel on Dietetic Products Nutrition and Allergies (NDA). Scientific Opinion on the appropriate age for introduction of complementary feeding of infants. *EFSA Journal*. 2009;7(12):1423.

271. Stordal K, Lundeby KM, Brantsaeter AL, Haugen M, Nakstad B, Lund-Blix NA, et al. Breast-feeding and infant hospitalization for infections: large cohort and sibling analysis. *J Pediatr Gastroenterol Nutr*. 2017;65(2):225-31.

272. Nicklaus S, Boggio V, Chabanet C, Issanchou S. A prospective study of food variety seeking in childhood, adolescence and early adult life. *Appetite*. 2005;44(3):289-97.

273. Cooke LJ, Wardle J, Gibson EL, Sapochnik M, Sheiham A, Lawson M. Demographic, familial and trait predictors of fruit and vegetable consumption by pre-school children. *Public Health Nutr*. 2004;7(2):295-302.

274. Bell KA, Wagner CL, Feldman H, Shypailo R, Belfort MB. Associations of infant feeding with trajectories of body composition and growth. *Am J Clin Nutr*. 2017;106(2):491-8.

275. Rzehak P, Oddy WH, Mearin ML, Grote V, Mori TA, Szajewska H, et al. Infant feeding and growth trajectory patterns in childhood and body composition in young adulthood. *Am J Clin Nutr*. 2017;106(2):568-80.

276. Ong KL, Ahmed ML, Emmett PM, Preece MA, Dunger DB. Association between postnatal catch-up growth and obesity in childhood: prospective cohort study. *BMJ*. 2000;320(7240):967.

277. Gale C, Logan KM, Santhakumaran S, Parkinson JR, Hyde MJ, Modi N. Effect of breastfeeding compared with formula feeding on infant body composition: a systematic review and meta-analysis. *Am J Clin Nutr*. 2012;95(3):656-69.

278. Giannì Maria L, Roggero P, Morlacchi L, Garavaglia E, Piemontese P, Mosca F. Formula-fed infants have significantly higher fat-free mass content in their bodies than breastfed babies. *Acta Paediatr*. 2014;103(7):e277-e81.

279. Li R, Magadia J, Fein SB, Grummer-Strawn LM. Risk of bottle-feeding for rapid weight gain during the first year of life. *Arch Pediatr Adolesc Med*. 2012;166(5):431-6.

280. Wood CT, Skinner AC, Yin HS, Rothman RL, Sanders LM, Delamater AM, et al. Bottle Size and Weight Gain in Formula-Fed Infants. *Pediatrics*. 2016;138(1).

281. Prentice P, Ong KK, Schoemaker MH, van Tol EA, Vervoort J, Hughes IA, et al. Breast milk nutrient content and infancy growth. *Acta Paediatr*. 2016;105(6):641-7.

282. Hornell A, Lagstrom H, Lande B, Thorsdottir I. Protein intake from 0 to 18 years of age and its relation to health: a systematic literature review for the 5th Nordic Nutrition Recommendations. *Food Nutr Res*. 2013;57.

283. Koletzko B, von Kries R, Closa R, Escribano J, Scaglioni S, Giovannini M, et al. Can infant feeding choices modulate later obesity risk? *Am J Clin Nutr*. 2009;89(5):1502s-8s.

284. Jensen RG. Lipids in human milk. *Lipids*. 1999;34(12):1243-71.

285. Martysiak-Zurowska D, Zoralska K, Zagierski M, Szlagtys-Sidorkiewicz A. Fatty acid composition in breast milk of women from Gdansk and the surrounding district in the course of lactation (Abstract). *Med Wieku Rozwoj*. 2011;15(2):167-77.

286. Goran MI, Martin AA, Alderete TL, Fujiwara H, Fields DA. Fructose in Breast Milk Is Positively Associated with Infant Body Composition at 6 Months of Age. *Nutrients*. 2017;9(2).

287. Huang LL, Yang F, Xiong F. [Association of leptin, adiponectin, and ghrelin in breast milk with the growth of infants with exclusive breastfeeding] (Abstract). *Zhongguo Dang Dai Er Ke Za Zhi*. 2018;20(2):91-6.

288. Rifas-Shiman SL, Fleisch A, Hivert MF, Mantzoros C, Gillman MW, Oken E. First and second trimester gestational weight gains are most strongly associated with cord blood levels of hormones at delivery important for glycemic control and somatic growth. *Metabolism*. 2017;69:112-9.

289. Kalhan SC, Schwartz R, Adam PAJ. Placental Barrier to Human Insulin-I125 in Insulin-Dependent Diabetic Mothers. *J Clin Endocrinol Metab*. 1975;40(1):139-42.

290. Rastogi GK, Letarte J, Fraser TR. Immunoreactive insulin content of 203 pancreases from foetuses of healthy mothers. *Diabetologia*. 1970;6:445-6.

291. Saltiel AR, Kahn CR. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature*. 2001;414(December):799-806.

292. Grasso S, Palumbo G, Rugolo S, Cianci A, Tumino G, Reitano G. Human fetal insulin secretion in response to maternal glucose and leucine administration. *Pediatr Res*. 1980;14:782-3.

293. Shima K, Price S, Foà PP. Serum insulin concentration and birth weight in human infants. *Proc Soc Exp Biol Med*. 1966;121(1):55-9.

294. Pegorier JP, Prip-Buus C, Duee PH, Girard J. Hormonal control of fatty acid oxidation during the neonatal period. *Diabetes Metab*. 1992;18:156–60.

295. Henquin J-C, Nenquin M. Dynamics and regulation of insulin secretion in pancreatic islets from normal young children. *PLoS One*. 2016;11(11):e0165961.

296. Farquhar JW. The child of the diabetic woman. *Arch Dis Child*. 1959;34:76-96.

297. Heitz PU, Klöppel G, Häcki WH, Polak JM, Pearse AGE. Nesidioblastosis: the pathologic basis of persistent hyperinsulinemic hypoglycemia in infants: morphologic and quantitative analysis of

seven cases based on specific immunostaining and electron microscopy. *Diabetes*. 1977;26(7):632-42.

298. Roe F. Beckwith's syndrome with hyperplasia extreme. *Pediatrics*. 1973;52(3).

299. Karlberg J, Jalil F, Lam B, Low L, Yeung CY. Linear growth retardation in relation to the three phases of growth. *Eur J Clin Nutr*. 1994;48 Suppl 1:S25-44.

300. Wit JM, Camacho-Hubner C. Endocrine regulation of longitudinal bone growth. *Endocr Dev*. 2011;21:30-41.

301. Guevara-Aguirre J. Insulin-like growth factor I--an important intrauterine growth factor. *N Engl J Med*. 1996;335(18):1389-91.

302. Stewart CE, Rotwein P. Growth, differentiation, and survival: multiple physiological functions for insulin-like growth factors. *Physiol Rev*. 1996;76(4):1005-26.

303. Daughaday WH, Rotwein P. Insulin-like growth factors I and II. Peptide, messenger ribonucleic acid and gene structures, serum, and tissue concentrations. *Endocr Rev*. 1989;10(1):68-91.

304. Fowlkes JL. Insulinlike growth factor-binding protein proteolysis an emerging paradigm in insulinlike growth factor physiology. *Trends Endocrinol Metab*. 1997;8(8):299-306.

305. Heemskerk VH, Daemen MA, Buurman WA. Insulin-like growth factor-1 (IGF-1) and growth hormone (GH) in immunity and inflammation. *Cytokine Growth Factor Rev*. 1999;10(1):5-14.

306. Hellstrom A, Ley D, Hansen-Pupp I, Hallberg B, Lofqvist C, van Marter L, et al. Insulin-like growth factor 1 has multisystem effects on foetal and preterm infant development. *Acta Paediatr*. 2016;105(6):576-86.

307. Wit JM, Walenkamp MJ. Role of insulin-like growth factors in growth, development and feeding. *World Rev Nutr Diet*. 2013;106:60-5.

308. Pfaffle R, Kies W, Klammt J. From GHRH to IGF-1 and downstream: clinical phenotypes and biological mechanisms. *Pediatr Endocrinol Rev*. 2011;9 Suppl 1:529-34.

309. Hwa V, Fang P, Derr MA, Fiegerlova E, Rosenfeld RG. IGF-I in human growth: lessons from defects in the GH-IGF-I axis. *Nestle Nutr Inst Workshop Ser*. 2013;71:43-55.

310. Laron Z. Insulin-like growth factor 1 (IGF-1): a growth hormone. *Mol Pathol*. 2001;54(5):311-6.

311. Ong KK, Langkamp M, Ranke MB, Whitehead K, Hughes IA, Acerini CL, et al. Insulin-like growth factor I concentrations in infancy predict differential gains in body length and adiposity: the Cambridge Baby Growth Study. *Am J Clin Nutr*. 2009;90(1):156-61.

312. Zhang Y, Proenca R, Maffei M, Barone M, Leopold L, Friedman JM. Positional cloning of the mouse obese gene and its human homologue. *Nature*. 1994;372(6505):425-32.

313. Frederich RC, Hamann A, Anderson S, Löllmann B, Lowell BB, Flier JS. Leptin levels reflect body lipid content in mice: Evidence for diet-induced resistance to leptin action. *Nat Med*. 1995;1:1311.

314. Considine RV, Sinha MK, Heiman ML, Kriauciunas A, Stephens TW, Nyce MR, et al. Serum immunoreactive-leptin concentrations in normal-weight and obese humans. *N Engl J Med*. 1996;334(5):292-5.

315. Bouret SG, Draper SJ, Simerly RB. Trophic action of leptin on hypothalamic neurons that regulate feeding. *Science*. 2004;304(5667):108.

316. Friedman JM, Halaas JL. Leptin and the regulation of body weight in mammals. *Nature*. 1998;395:763.

317. Kelesidis T, Kelesidis I, Chou S, Mantzoros CS. Narrative review: the role of leptin in human physiology: emerging clinical applications. *Ann Intern Med*. 2010;152(2):93-100.

318. Licinio J, Mantzoros C, Negrao AB, Cizza G, Wong ML, Bongiorno PB, et al. Human leptin levels are pulsatile and inversely related to pituitary-adrenal function. *Nat Med*. 1997;3(5):575-9.

319. Sinha MK, Ohannesian JP, Heiman ML, Kriauciunas A, Stephens TW, Magosin S, et al. Nocturnal rise of leptin in lean, obese, and non-insulin-dependent diabetes mellitus subjects. *J Clin Invest*. 1996;97(5):1344-7.

320. Farooqi IS, Jebb SA, Langmack G, Lawrence E, Cheetham CH, Prentice AM, et al. Effects of recombinant leptin therapy in a child with congenital leptin deficiency. *N Engl J Med*. 1999;341(12):879-84.

321. Farooqi IS, Matarese G, Lord GM, Keogh JM, Lawrence E, Agwu C, et al. Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency. *J Clin Invest*. 2002;110(8):1093-103.

322. Myers MG, Cowley MA, Munzberg H. Mechanisms of leptin action and leptin resistance. *Annu Rev Physiol*. 2008;70:537-56.

323. Hauguel-de Mouzon S, Lepercq J, Catalano P. The known and unknown of leptin in pregnancy. *Am J Obstet Gynecol*. 2006;194(6):1537-45.

324. Christou H, Serdy S, Mantzoros CS. Leptin in Relation to Growth and Developmental Processes in the Fetus. *Semin Reprod Med*. 2002;20(02):123-30.

325. Sivan E, Lin WM, Homko CJ, Reece EA, Boden G. Leptin is present in human cord blood. *Diabetes*. 1997;46(5):917.

326. Pardo IM, Geloneze B, Tambascia MA, Pereira JL, Barros Filho AA. [Leptin as a marker of sexual dimorphism in newborn infants] (Abstract). *J Pediatr (Rio J)*. 2004;80(4):305-8.

327. Karakosta P, Chatzi L, Plana E, Margioris A, Castanas E, Kogevinas M. Leptin levels in cord blood and anthropometric measures at birth: a systematic review and meta-analysis. *Paediatr Perinat Epidemiol*. 2011;25(2):150-63.

328. Ong KK, Ahmed ML, Sherriff A, Woods KA, Watts A, Golding J, et al. Cord blood leptin is associated with size at birth and predicts infancy weight gain in humans. ALSPAC Study Team. Avon Longitudinal Study of Pregnancy and Childhood. *J Clin Endocrinol Metab*. 1999;84(3):1145-8.

329. Parker M, Rifas-Shiman SL, Belfort MB, Taveras EM, Oken E, Mantzoros C, et al. Gestational glucose tolerance and cord blood leptin levels predict slower weight gain in early infancy. *J Pediatr*. 2011;158(2):227-33.

330. Chaoimh CN, Murray DM, Kenny LC, Irvine AD, Hourihane JO, Kiely M. Cord blood leptin and gains in body weight and fat mass during infancy. *Eur J Endocrinol*. 2016;175(5):403-10.

331. Javaid MK, Godfrey KM, Taylor P, Robinson SM, Crozier SR, Dennison EM, et al. Umbilical cord leptin predicts neonatal bone mass. *Calcif Tissue Int*. 2005;76(5):341-7.

332. Shen W, Velasquez G, Chen J, Jin Y, Heymsfield SB, Gallagher D, et al. Comparison of the relationship between bone marrow adipose tissue and volumetric bone mineral density in children and adults. *J Clin Densitom*. 2014;17(1):163-9.

333. Cioffi JA, Shafer AW, Zupancic TJ, Smith-Gbur J, Mikhail A, Platika D, et al. Novel B219/OB receptor isoforms: possible role of leptin in hematopoiesis and reproduction. *Nat Med*. 1996;2(5):585-9.

334. Bennett BD, Solar GP, Yuan JQ, Mathias J, Thomas GR, Matthews W. A role for leptin and its cognate receptor in hematopoiesis. *Curr Biol*. 1996;6(9):1170-80.

335. Hennessy E, Alberman E. Intergenerational influences affecting birth outcome. I. Birthweight for gestational age in the children of the 1958 British birth cohort. *Paediatr Perinat Epidemiol*. 1998;12 Suppl 1:45-60.

336. Horikoshi M, Beaumont RN, Day FR, Warrington NM, Kooijman MN, Fernandez-Tajes J, et al. Genome-wide associations for birth weight and correlations with adult disease. *Nature*. 2016;538(7624):248-52.

337. Peters J. The role of genomic imprinting in biology and disease: an expanding view. *Nature Reviews Genetics*. 2014;15:517.

338. McGrath J, Solter D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell*. 1984;37(1):179-83.

339. Surani MA, Barton SC, Norris ML. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature*. 1984;308(5959):548-50.

340. Baker J, Liu JP, Robertson EJ, Efstratiadis A. Role of insulin-like growth factors in embryonic and postnatal growth. *Cell*. 1993;75(1):73-82.

341. Smith AC, Choufani S, Ferreira JC, Weksberg R. Growth regulation, imprinted genes, and chromosome 11p15.5. *Pediatr Res*. 2007;61:43R.

342. Petry CJ, Seear RV, Wingate DL, Acerini CL, Ong KK, Hughes IA, et al. Maternally transmitted foetal H19 variants and associations with birth weight. *Hum Genet*. 2011;130(5):663-70.

343. St-Pierre J, Hivert M-F, Perron P, Poirier P, Guay S-P, Brisson D, et al. IGF2 DNA methylation is a modulator of newborn's fetal growth and development. *Epigenetics*. 2012;7(10):1125-32.

344. Mansell T, Novakovic B, Meyer B, Rzehak P, Vuillermin P, Ponsonby AL, et al. The effects of maternal anxiety during pregnancy on IGF2/H19 methylation in cord blood. *Transl Psychiatry*. 2016;6:e765.

345. Dunger D, Ahmed L, Ong K. Growth and body composition in type 1 diabetes mellitus. *Horm Res*. 2002;58 Suppl 1:66-71.

346. Kharagjitsingh AV, de Ridder MAJ, Alizadeh BZ, Veeze HJ, Bruining GJ, Roep BO, et al. Genetic correlates of early accelerated infant growth associated with juvenile-onset type 1 diabetes. *Pediatr Diabetes*. 2012;13(3):266-71.

347. Goris A, Liston A. The immunogenetic architecture of autoimmune disease. *Cold Spring Harbor Perspectives in Biology*. 2012;4(3).

348. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. *Int J Epidemiol*. 2013;42(4):1134-44.

349. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015;518(7539):337-43.

350. Hu JX, Thomas CE, Brunak S. Network biology concepts in complex disease comorbidities. *Nat Rev Genet*. 2016;17(10):615-29.

351. Patterson C, Guariguata L, Dahlquist G, Soltész G, Ogle G, Silink M. Diabetes in the young – a global view and worldwide estimates of numbers of children with type 1 diabetes. *Diabetes Res Clin Pract*. 2014;103(2):161-75.

352. Gale EA. The rise of childhood type 1 diabetes in the 20th century. *Diabetes*. 2002;51(12):3353-61.

353. Bonifacio E, Beyerlein A, Hippich M, Winkler C, Vehik K, Weedon MN, et al. Genetic scores to stratify risk of developing multiple islet autoantibodies and type 1 diabetes: A prospective study in children. *PLoS Med*. 2018;15(4):e1002548.

354. Hermann R, Turpeinen H, Laine A, Veijola R, Knip M, Simell O, et al. HLA DR-DQ-encoded genetic determinants of childhood-onset type 1 diabetes in Finland: an analysis of 622 nuclear families. *Tissue Antigens*. 2003;62(2):162-9.

355. Beyerlein A, Thiering E, Pflueger M, Bidlingmaier M, Stock J, Knopff A, et al. Early infant growth is associated with the risk of islet autoimmunity in genetically susceptible children. *Pediatr Diabetes*. 2014(1):534-42.

356. Larsson HE, Vehik K, Haller MJ, Liu X, Akolkar B, Hagopian W, et al. Growth and Risk for Islet Autoimmunity and Progression to Type 1 Diabetes in Early Childhood: The Environmental Determinants of Diabetes in the Young Study. *Diabetes*. 2016;65(7):1988-95.

357. Ponsonby A-L, Pezic A, Cochrane J, Cameron FJ, Pascoe M, Kemp A, et al. Infant anthropometry, early life infection, and subsequent risk of type 1 diabetes mellitus: a prospective birth cohort study. *Pediatr Diabetes*. 2011;12(4 Pt 1):313-21.

358. Frayling TM, Hattersley AT, Smith GD, Ben-Shlomo Y. Conflicting results on variation in the IGFI gene highlight methodological considerations in the design of genetic association studies. *Diabetologia*. 2002;45(11):1605-6.

359. Ioannidis JPA, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet*. 2001;29:306.

360. Laakso M, Pyorala K. Age of onset and type of diabetes. *Diabetes Care*. 1985;8(2):114-7.

361. Dahlquist GG, Pundziūte-Lyckå A, Nyström L. Birthweight and risk of type 1 diabetes in children and young adults: a population-based register study. *Diabetologia*. 2005;48(6):1114-7.

362. Robertson L, Harrild K. Maternal and neonatal risk factors for childhood type 1 diabetes: a matched case-control study. *BMC public health*. 2010;10:281.

363. Kyvik KO, Bache I, Green A, Beck-Nielsen H, Buschard K. No association between birth weight and Type 1 diabetes mellitus--a twin-control study. *Diabet Med*. 2000;17(2):158-62.

364. Kharagjitsingh AV, De Ridder MAJ, Roep BO, Koeleman BPC, Bruining GJ, Veeze HJ. Revisiting infant growth prior to childhood onset type 1 diabetes. *Clin Endocrinol*. 2010;72(5):620-4.

365. Stene LC, Magnus P, Lie RT, Søvik O, Joner G. Birth weight and childhood onset type 1 diabetes: population based cohort study. *BMJ*. 2001;322(April):889-92.

366. Cardwell CR, Stene LC, Joner G, Davis EA, Cinek O, Rosenbauer J, et al. Birthweight and the risk of childhood-onset type 1 diabetes: A meta-analysis of observational studies using individual patient data. *Diabetologia*. 2010;53(4):641-51.

367. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet*. 2003;33(2):177-82.

368. Harder T, Roepke K, Diller N, Stechling Y, Dudenhausen JW, Plagemann A. Birth weight, early weight gain, and subsequent risk of type 1 diabetes: systematic review and meta-analysis. *Am J Epidemiol*. 2009;169(12):1428-36.

369. Dahlquist G, Bennich SS, Källén B. Intrauterine growth pattern and risk of childhood onset insulin dependent (type I) diabetes: population based case-control study. *Br Med J*. 1996;313:1174-7.

370. Goldacre RR. Associations between birthweight, gestational age at birth and subsequent type 1 diabetes in children under 12: a retrospective cohort study in England, 1998-2012. *Diabetologia*. 2018;61(3):616-25.

371. Yassouridis C, Leisch F, Winkler C, Ziegler AG, Beyerlein A. Associations of growth patterns and islet autoimmunity in children with increased risk for type 1 diabetes: a functional analysis approach. *Pediatr Diabetes*. 2017;18(2):103-10.

372. Dahlquist GG, Patterson C, Soltesz G. Perinatal risk factors for childhood type 1 diabetes in Europe. The EURODIAB Substudy 2 Study Group. *Diabetes Care*. 1999;22(10):1698-702.

373. Tai T, Wang C, Lin L, Lee L, Tsai S, Chen C. A case-control study on risk factors for type 1 diabetes in Taipei city. *Diabetes Res Clin Pract*. 1998;42:197-203.

374. Wei JN, Li HY, Chang CH, Sung FC, Li CY, Lin CC, et al. Birth weight and type 1 diabetes among schoolchildren in Taiwan-A population-based case-controlled study. *Diabetes Res Clin Pract*. 2006;74(3):309-15.

375. Garcia Cuartero B, Gonzalez Vergaz A, Herranz S, Vazquez MC, Carvajal O, Carpintero P, et al. Low birth weight is a risk factor for type 1 diabetes (Abstract). *An Pediatr (Barc)*. 2009;70(6):542-6.

376. Borras V, Freitas A, Castell C, Gispert R, Jané M. Type 1 diabetes and perinatal factors in Catalonia (Spain). *Pediatr Diabetes*. 2011;12(4 PART 2):419-23.

377. Malcova H, Sumnik Z, Drevinek P, Venhacova J, Lebl J, Cinek O. Absence of breast-feeding is associated with the risk of type 1 diabetes: A case-control study in a population with rapidly increasing incidence. *Eur J Pediatr*. 2006;165(2):114-9.

378. Waldhoer T, Rami B, Schober E. Perinatal risk factors for early childhood onset type 1 diabetes in Austria - a population-based study (1989-2005). *Pediatr Diabetes*. 2008;9(3 Pt 1):178-81.

379. Persson E, Waernbaum I. Estimating a marginal causal odds ratio in a case-control design: Analyzing the effect of low birth weight on the risk of type 1 diabetes mellitus. *Stat Med*. 2013;32(14):2500-12.

380. Dahlquist G, Kallen B. Maternal-child blood group incompatibility and other perinatal events increase the risk for early-onset Type 1 (insulin-dependent) diabetes mellitus. *Diabetologia*. 1992;35:671-5.

381. Hathout EH, Hartwick N, Fagoaga OR, Colacino AR, Sharkey J, Racine M, et al. Clinical, autoimmune, and HLA characteristics of children diagnosed with type 1 diabetes before 5 years of age. *Pediatrics*. 2003;111(4):860-3.

382. Leete P, Willcox A, Krogvold L, Dahl-Jorgensen K, Foulis AK, Richardson SJ, et al. Differential insulitic profiles determine the extent of beta-cell destruction and the age at onset of type 1 diabetes. *Diabetes*. 2016;65(5):1362-9.

383. Wadsworth EJK, Shield JPH, Hunt LP, Baum JD. A case-control study of environmental factors associated with diabetes in the under 5s. *Diabet Med*. 1997;14(5):390-6.

384. Johansson C, Samuelsson U, Ludvigsson J. A high weight gain early in life is associated with an increased risk of Type 1 (insulin-dependent)diabetes mellitus. *Diabetologia*. 1994;37(1):91-4.

385. Khan N, Couper JJ. Low-birth-weight infants show earlier onset of IDDM. *Diabetes Care*. 1994;17(7):653-6.

386. Kotani Y, Yokota I, Kagami S, Amemiya S, Matsuura N, Sasaki N, et al. Relatively small birth size and accelerated early growth of Japanese type 1 diabetic children with younger onset. *Clin Pediatr Endocrinol*. 2006;15(2):73-8.

387. Dahlquist G, Blom L, Lonnberg G. The Swedish Childhood Diabetes Study - a multivariate analysis of risk determinants for diabetes in different age groups. *Diabetologia*. 1991;34:757-62.

388. Palmer JP, Helgvist S, Spinas GA, Molvig J, Mandrup-Poulsen T, Andersen HU, et al. Interaction fo beta-cell activity and IL-1 concentration and exposure time in isolated rat islets of Langherans. *Diabetes*. 1989;38(10):1211-6.

389. Hill DE. Effect of insulin on fetal growth. *Semin Perinatol*. 1978;2(4):319-28.

390. Podar T, Onkamo P, Forsen T, Karvonen M, Tuomilehto-Wolf E, Tuomilehto J. Neonatal anthropometric measurements and risk of childhood-onset type 1 diabetes. DiMe Study Group. *Diabetes Care*. 1999;22(12):2092-4.

391. Chandra RK, Ali SK, Kutty KM, Chandra S. Thymus-dependent lymphocytes and delayed hypersensitivity in low birth weight infants. *Biol Neonate*. 1977;31(1-2):15-8.

392. Klein SL, Flanagan KL. Sex differences in immune responses. *Nature Reviews Immunology*. 2016;16:626.

393. Hypponen E, Virtanen SM, Kenward MG, Knip M, Akerblom HK, The Childhood Diabetes in Finland Study Group. Obesity, increased linear growth, and risk of type 1 diabetes in children. *Diabetes Care*. 2000;23(12):1755-60.

394. Blom L, Dahlquist G, Nystrom L, Sandstrom A, Wall S. The Swedish childhood diabetes study - social and perinatal determinants for diabetes in childhood. *Diabetologia*. 1989;32:7-13.

395. Kyvik KO, Green A, Svendsen A, Mortensen K. Breast feeding and the development of type 1 diabetes mellitus. *Diabet Med*. 1992;9(3):233-5.

396. Ljungkrantz M, Ludvigsson J, Samuelsson U. Type 1 diabetes: Increased height and weight gains in early childhood. *Pediatr Diabetes*. 2008;9(Part II):50-6.

397. Bache I, Bock T, Volund A, Buschard K. Previous maternal abortion, longer gestation, and younger maternal age decrease the risk of type 1 diabetes among male offspring. *Diabetes Care*. 1999;22(7):1063-5.

398. Svensson J, Carstensen B, Mortensen HB, Borch-Johnsen K. Growth in the first year of life and the risk of type 1 diabetes in a Danish population. *Paediatr Perinat Epidemiol*. 2007;21(1):44-8.

399. Eurodiab Substudy 2 Study Group. Rapid early growth is associated with increased risk of childhood type 1. *Diabetes Care*. 2002;25(10):1755-60.

400. Rolland-Cachera MF, Brambilla P, Manzoni P, Akrout M, Sironi S, Del Maschio A, et al. Body composition assessed on the basis of arm circumference and triceps skinfold thickness: a new index validated in children by magnetic resonance imaging. *Am J Clin Nutr*. 1997;65(6):1709-13.

401. Court S, Parkin M, Roberts DF, Wentzel J. HLA antigens and growth in diabetic children (Abstract). *Ann Hum Biol*. 1982;9(4):329-36.

402. Ober C. The maternal-fetal relationship in human pregnancy: an immunogenetic perspective. *Exp Clin Immunogenet*. 1992;9(1):1-14.

403. Stene LC, Thorsby PM, Berg JP, Ronningen KS, Undlien DE, Joner G. The relation between size at birth and risk of type 1 diabetes is not influenced by adjustment for the insulin gene (-23HphI) polymorphism or HLA-DQ genotype. *Diabetologia*. 2006;49(9):2068-73.

404. Peet A, Koo P, Ilonen J, Knip M, Tillmann V. Birth weight in newborn infants with different diabetes-associated HLA genotypes in three neighbouring countries: Finland, Estonia and Russian Karelia. *Diabetes Metab Res Rev*. 2012;28(6):455-61.

405. Locatelli M, Buzzetti R, Galgani A, Montemari AL, Khazrai M, Petrone A, et al. Length of gestation and gender are associated with HLA genotypes at risk for Type 1 diabetes (Italian DIABFIN 3). *Diabet Med*. 2007;24(8):916-9.

406. Bruining GJ. Association between infant growth before onset of juvenile type-1 diabetes and autoantibodies to IA-2 Netherlands Kolibrie study group of childhood diabetes. *Lancet*. 2000;356(9230):655-6.

407. Magnus MC, Olsen SF, Granström C, Joner G, Skrivarhaug T, Svensson J, et al. Infant growth and risk of childhood-onset type 1 diabetes in children from 2 Scandinavian birth cohorts. *JAMA Pediatrics*. 2015;169(12):e153759-e.

408. Couper JJ, Beresford S, Hirte C, Baghurst PA, Pollard A, Tait BD, et al. Weight gain in early life predicts risk of islet autoimmunity in children with a first-degree relative with type 1 diabetes. *Diabetes Care*. 2009;32(1):94-9.

409. Blom L, Persson LA, Dahlquist G. A high linear growth is associated with an increased risk of childhood diabetes mellitus. *Diabetogia*. 1992;35:528-33.

410. Maas JA, Mook-Kanamori DO, Ay L, Steegers EA, van Duijn CM, Hofman A, et al. Insulin VNTR and IGF-1 promoter region polymorphisms are not associated with body composition in early childhood: the generation R study. *Horm Res Paediatr*. 2010;73(2):120-7.

411. Nucci AM, Becker DJ, Virtanen SM, Cuthberston D, Softness B, Huot C, et al. Growth differences between North American and European children at risk for type 1 diabetes. *Pediatr Diabetes*. 2012;13(5):419-25.

412. Wormser D. *Introduction and scope of epidemiology*. [Lecture]. Translational Medicine and Therapeutics. University of Cambridge. 29th October 2012.

413. Freedman DS, Wang J, Maynard LM, Thornton JC, Mei Z, Pierson Jr RN, et al. Relation of BMI to fat and fat-free mass among children and adolescents. *Int J Obes (Lond)*. 2004;29:1.

414. Lammi N, Moltchanova E, Blomstedt PA, Tuomilehto J, Eriksson JG, Karvonen M. Childhood BMI trajectories and the risk of developing young adult-onset diabetes. *Diabetologia*. 2009;52(3):408-14.

415. Sovio U, Mook-Kanamori DO, Warrington NM, Lawrence R, Briollais L, Palmer CN, et al. Association between common variation at the FTO locus and changes in body mass index from infancy to late childhood: the complex nature of genetic association through growth and development. *PLoS Genet*. 2011;7(2):e1001307.

416. Rolland-Cachera MF, Deheeger M, Bellisle F, Sempe M, Guilloud-Bataille M, Patois E. Adiposity rebound in children: a simple indicator for predicting obesity. *Am J Clin Nutr*. 1984;39(1):129-35.

417. Viner RM, Hindmarsh PC, Taylor B, Cole TJ. Childhood body mass index (BMI), breastfeeding and risk of Type 1 diabetes: findings from a longitudinal national birth cohort. *Diabet Med*. 2008;25(9):1056-61.

418. Mericq V, Ong KK, Bazaes R, Pena V, Avila A, Salazar T, et al. Longitudinal changes in insulin sensitivity and secretion from birth to age three years in small- and appropriate-for-gestational-age children. *Diabetologia*. 2005;48(12):2609-14.

419. Benevento D, Bizzarri C, Patera IP, Rava L, Schiaffini R, Ciampalini P, et al. Birth weight influences the clinical phenotype and the metabolic control of patients with type 1 diabetes (T1D). *Diabetes Metab Res Rev*. 2013;29:60-5.

420. Winkler C, Marienfeld S, Zwilling M, Bonifacio E, Ziegler AG. Is islet autoimmunity related to insulin sensitivity or body weight in children of parents with type 1 diabetes? *Diabetologia*. 2009;52(10):2072-8.

421. Peet A, Hamalainen AM, Kool P, Ilonen J, Knip M, Tillmann V. Circulating IGF1 and IGFBP3 in relation to the development of β-cell autoimmunity in young children. *European Journal of Endocrinology*. 2015;173(2):129-37.

422. Frystyk J. Free insulin-like growth factors - measurements and relationships to growth hormone secretion and glucose homeostasis. *Growth Horm IGF Res*. 2004;14(5):337-75.

423. Prentice P, Acerini CL, Eleftheriou A, Hughes IA, Ong KK, Dunger DB. Cohort Profile: the Cambridge Baby Growth Study (CBGS). *Int J Epidemiol*. 2016;45(1):35-35g.

424. Prentice P. Development and evaluation of biomarkers of infant nutrition, growth and body composition [PhD dissertation]: University of Cambridge; 2015.

425. Barton B. *Medical statistics : a guide to SPSS, data analysis and critical appraisal.* Peat JK, editor. Chichester, West Sussex; Hoboken, NJ: Wiley Blackwell/BMJ Books; 2014.

426. van Dommelen P, de Gunst M, van der Vaart A, van Buuren S, Boomsma D. Growth references for height, weight and body mass index of twins aged 0-2.5 years. *Acta Paediatr*. 2008;97(8):1099-104.

427. Mook-Kanamori DO, van Beijsterveldt CE, Steegers EA, Aulchenko YS, Raat H, Hofman A, et al. Heritability estimates of body size in fetal life and early childhood. *PLoS One*. 2012;7(7):e39901.

428. Neu J, Rushing J. Cesarean versus vaginal delivery: long-term infant outcomes and the hygiene hypothesis. *Clin Perinatol*. 2011;38(2):321-31.

429. Hansen-Pupp I, Löfqvist C, Polberger S, Niklasson A, Fellman V, Hellström A, et al. Influence of Insulin-Like Growth Factor I and Nutrition During Phases of Postnatal Growth in Very Preterm Infants. *Pediatr Res*. 2011;69:448.

430. Serrao F, Papacci P, Costa S, Giannantonio C, Cota F, Vento G, et al. Effect of early expressed human milk on insulin-like growth factor 1 and short-term outcomes in preterm infants. *PLoS One*. 2016;11(12):e0168139.

431. Engstrom E, Niklasson A, Wikland KA, Ewald U, Hellstrom A. The role of maternal factors, postnatal nutrition, weight gain, and gender in regulation of serum IGF-I among preterm infants. *Pediatr Res*. 2005;57(4):605-10.

432. Prevost T. *Getting started with SPSS.* [Course Material]. Translational Medicine and Therapeutics. University of Cambridge. November 2012.

433. Cole TJ. Conditional reference charts to assess weight gain in British infants. *Arch Dis Child*. 1995;73(1):8.

434. Burren O, Adlem E, Achuthan P, Christensen M, Coulson R, Todd J. T1DBase: Update 2011, organization and presentation of large-scale data sets for type 1 diabetes research. *Nucleic Acids Res*. 2011;39:997-1001.

435. Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat Rev Genet*. 2013;14:661.

436. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet*. 2011;43(12):1193-201.

437. Winkler C, Krumsiek J, Buettner F, Angermüller C, Giannopoulou EZ, Theis FJ, et al. Feature ranking of type 1 diabetes susceptibility genes improves prediction of type 1 diabetes. *Diabetologia*. 2014;57(12):2521-9.

438. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet*. 2009;41(6):703-7.

439. Bradfield JP, Qu HQ, Wang K, Zhang H, Sleiman PM, Kim CE, et al. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet*. 2011;7(9).

440. Barker JM, Triolo TM, Aly TA, Baschal EE, Babu SR, Kretowski A, et al. Two single nucleotide polymorphisms identify the highest-risk diabetes HLA genotype: potential for rapid screening. *Diabetes*. 2008;57(11):3152-5.

441. Nguyen C, Varney MD, Harrison LC, Morahan G. Definition of high-risk type 1 diabetes HLA-DR and HLA-DQ types using only three single nucleotide polymorphisms. *Diabetes*. 2013;62(6):2135-40.

442. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*. 2017;45(Database issue: D896-D901).

443. Kelley J, de Bono B, Trowsdale J. IRIS: a database surveying known human immune system genes. *Genomics*. 2005;85(4):503-11.

444. LGC Group. *SNPviewer User guide*. 2014.

445. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661-78.

446. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Mägi R, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc*. 2014;9(5):1192-212.

447. Tabangin ME, Woo JG, Martin LJ. The effect of minor allele frequency on the likelihood of obtaining false positives. *BMC Proc*. 2009;3(Suppl 7):S41-S.

448. Lunetta KL. Genetic association studies. *Circulation*. 2008;118(1):96-101.

449. Howson JM, Walker NM, Clayton D, Todd JA, Type 1 Diabetes Genetics C. Confirmation of HLA class II independent type 1 diabetes associations in the major histocompatibility complex including HLA-B and HLA-A. *Diabetes Obes Metab*. 2009;11 Suppl 1:31-45.

450. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg Equilibrium. *Am J Hum Genet*. 2005;76:887-3.

451. Bout J. Stepwise regression: when to use it? Available from: www.goodsciencebadscience.nl.

452. Scott R. *Mendelian randomisation*. [Lecture]. University of Cambridge, Department of Public Health and Primary Care. 28th November 2016.

453. Hill WG, Goddard ME, Visscher PM. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet*. 2008;4(2):e1000008.

454. Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M, Tuomilehto J. Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. *Diabetes*. 2003;52(4):1052-5.

455. Burgess S, Thompson SG. *Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation*. USA: Taylor & Francis Group; 2015.

456. Oram RA, Patel K, Hill A, Shields B, McDonald TJ, Jones A, et al. A type 1 diabetes genetic risk score can aid discrimination between type 1 and type 2 diabetes in young adults. *Diabetes Care*. 2016;39(3):337-44.

457. Patel K, Oram R, Flanagan S, De Franco E, Colclough K, shepherd M, et al. Type 1 diabetes genetic risk score: a novel tool to discriminate monogenic and type 1 diabetes. *Diabetes*. 2016;65:2094-99.

458. Thomas NJ, Jones SE, Weedon MN, Shields BM, Oram RA, Hattersley AT. Frequency and phenotype of type 1 diabetes in the first six decades of life: a cross-sectional, genetically stratified survival analysis from UK Biobank. *Lancet Diabetes Endocrinol*. 2018;6(2):122-9.

459. International Multiple Sclerosis Genetics C, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, et al. Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med*. 2007;357(9):851-62.

460. Redondo MJ, Kawasaki E, Mulgrew CL, Noble JA, Erlich HA, Freed BM, et al. DR- and DQ-associated protection from type 1A diabetes: comparison of DRB1*1401 and DQA1*0102-DQB1*0602*. *The Journal of Clinical Endocrinology & Metabolism*. 2000;85(10):3793-7.

461. Bell GI, Selby MJ, Rutter WJ. The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *Nature*. 1982;295(5844):31-5.

462. Bell GI, Horita S, Karam JH. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes*. 1984;33(2):176-83.

463. Kennedy GC, German MS, Rutter WJ. The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nat Genet*. 1995;9(3):293-8.

464. Bennett ST, Todd JA. Human type 1 diabetes and the insulin gene: principles of mapping polygenes. *Annu Rev Genet*. 1996;30:343-70.

465. Vafiadis P, Bennett ST, Colle E, Grabs R, Goodyer CG, Polychronakos C. Imprinted and genotype-specific expression of genes at the IDDM2 locus in pancreas and leucocytes. *Journal of Autoimmunity*. 1996;9(3):397-403.

466. Cocozza S, Riccardi G, Monticelli A, Capaldo B, Genovese S, Krogh V, et al. Polymorphism at the 5‘ end flanking region of the insulin gene is associated with reduced insulin secretion in healthy individuals. *Eur J Clin Invest*. 1988;18(6):582-6.

467. Ibanez L, Ong K, Potau N, Marcos MV, de Zegher F, Dunger D. Insulin gene variable number of tandem repeat genotype and the low birth weight, precocious pubarche, and hyperinsulinism sequence. *J Clin Endocrinol Metab*. 2001;86(12):5788-93.

468. Mitchell SM, Hattersley AT, Knight B, Turner T, Metcalf BS, Voss LD, et al. Lack of support for a role of the insulin gene variable number of tandem repeats minisatellite (INS-VNTR) locus in fetal growth or type 2 diabetes-related intermediate traits in United Kingdom populations. *J Clin Endocrinol Metab*. 2004;89(1):310-7.

469. Ong KKL, Phillips DI, Fall C, Poulton J, Bennett ST, Golding J, et al. The insulin gene VNTR, type 2 diabetes and birth weight. *Nat Genet*. 1999;21:262.

470. Landmann E, Geller F, Schilling J, Rudloff S, Foeller-Gaudier E, Gortner L. Absence of the wild-type allele (192 base pairs) of a polymorphism in the promoter region of the IGF-I gene but not a polymorphism in the insulin gene variable number of tandem repeat locus is associated with accelerated weight gain in infancy. *Pediatrics*. 2006;118(6):2374-9.

471. Vu-Hong TA, Durand E, Deghmoun S, Boutin P, Meyre D, Chevenne D, et al. The INS VNTR locus does not associate with smallness for gestational age (SGA) but interacts with SGA to increase insulin resistance in young adults. *J Clin Endocrinol Metab*. 2006;91(6):2437-40.

472. Bennett AJ, Sovio U, Ruokonen A, Martikainen H, Pouta A, Taponen S, et al. Variation at the Insulin Gene VNTR (Variable Number Tandem Repeat) Polymorphism and Early Growth. *Diabetes*. 2004;53(8):2126.

473. Mook-Kanamori DO, Miranda Geelhoed JJ, Steegers EA, Witteman JC, Hofman A, Moll HA, et al. Insulin gene variable number of tandem repeats is not associated with weight from fetal life until infancy: the Generation R Study. *Eur J Endocrinol*. 2007;157(6):741-8.

474. Stawerska R, Szałapska M, Borowiec M, Antosik K, Młynarski W, Lewiński A. The influence of INS VNTR class III allele on auxological parameters, glucose, insulin, lipids, and adipocytokines secretion in prepubertal children born small for gestational age. *Endokrynol Pol*. 2016;67(6):585-91.

475. Lindsay RS, Hanson RL, Wiedrich C, Knowler WC, Bennett PH, Baier LJ. The insulin gene variable number tandem repeat class I/III polymorphism is in linkage disequilibrium with birth weight but not type 2 diabetes in the Pima population. *Diabetes*. 2003;52(1):187-93.

476. Petry CJ, Rayco-Solon P, Fulford AJC, Stead JDH, Wingate DL, Ong KK, et al. Common polymorphic variation in the genetically diverse African insulin gene and its association with size at birth. *Hum Genet*. 2009;126(3):375-84.

477. Osada H, Seki K, Sekiya S. Genetic variations within the insulin gene region are associated with accelerated fetal growth. *Tohoku J Exp Med*. 2007;212(1):27-34.

478. Moore GE, Abu-Amero SN, Bell G, Wakeling EL, Kingsnorth A, Stanier P, et al. Evidence that insulin is imprinted in the human yolk sac. *Diabetes*. 2001;50(1):199-203.

479. Giannoukakis N, Deal C, Paquette J, Goodyer CG, Polychronakos C. Parental genomic imprinting of the human IGF2 gene. *Nat Genet*. 1993;4(1):98-101.

480. Vafiadis P, Bennett ST, Todd JA, Grabs R, Polychronakos C. Divergence between genetic determinants of IGF2 transcription levels in leukocytes and of IDDM2-encoded susceptibility to type 1 diabetes. *J Clin Endocrinol Metab*. 1998;83(8):2933-9.

481. Vafiadis P, Grabs R, Goodyer CG, Colle E, Polychronakos C. A functional analysis of the role of IGF2 in IDDM2-encoded susceptibility to type 1 diabetes. *Diabetes*. 1998;47(5):831-6.

482. Barratt BJ, Payne F, Lowe CE, Hermann R, Healy BC, Harold D, et al. Remapping the insulin gene/IDDM2 locus in type 1 diabetes. *Diabetes*. 2004;53(7):1884-9.

483. Stead JDH, Jeffreys AJ. Structural analysis of insulin minisatellite alleles reveals unusually large differences in diversity between Africans and non-Africans. *Am J Hum Genet*. 2002;71(6):1273-84.

484. La Batide-Alanore A, Tregouet DA, Jaquet D, Bouyer J, Tiret L. Familial aggregation of fetal growth restriction in a French cohort of 7,822 term births between 1971 and 1985. *Am J Epidemiol*. 2002;156(2):180-7.

485. Jaquet D, Swaminathan S, Alexander Greg R, Czernichow P, Collin D, Salihu Hamisu M, et al. Significant paternal contribution to the risk of small for gestational age. *BJOG*. 2005;112(2):153-9.

486. Heude B, Petry CJ, Pembrey M, Dunger DB, Ong KK. The insulin gene variable number of tandem repeat: associations and interactions with childhood body fat mass and insulin secretion in normal children. *J Clin Endocrinol Metab*. 2006;91(7):2770-5.

487. Sandhu MS, Heude B, Young EH, Luben R, Luan Ja, Khaw K-T, et al. INS VNTR class genotype and indexes of body size and obesity. *Diabetes*. 2005;54(9):2812.

488. Le Stunff C, Fallin D, Bougneres P. Paternal transmission of the very common class I INS VNTR alleles predisposes to childhood obesity. *Nat Genet*. 2001;29(1):96-9.

489. Weaver JU, Kopelman PG, Hitman GA. Central obesity and hyperinsulinaemia in women are associated with polymorphism in the 5' flanking region of the human insulin gene. *Eur J Clin Invest*. 1992;22(4):265-70.

490. Waterworth DM, Bennett ST, Gharani N, McCarthy MI, Hague S, Batty S, et al. Linkage and association of insulin gene VNTR regulatory polymorphism with polycystic ovary syndrome. *Lancet*. 1997;349(9057):986-90.

491. Bazaes RA, Petry CJ, Ong KK, Ávila A, Dunger DB, Mericq MV. Insulin gene VNTR genotype is associated with insulin sensitivity and secretion in infancy. *Clin Endocrinol*. 2003;59(5):599-603.

492. Bouatia-Naji N, De Graeve F, Bronner G, Lecoeur C, Vatin V, Durand E, et al. INS VNTR is not associated with childhood obesity in 1,023 families: a family-based study. *Obesity*. 2008;16(6):1471-5.

493. Hansen SK, Gjesing AP, Rasmussen SK, Glumer C, Urhammer SA, Andersen G, et al. Large-scale studies of the HphI insulin gene variable-number-of-tandem-repeats polymorphism in relation to Type 2 diabetes mellitus and insulin release. *Diabetologia*. 2004;47(6):1079-87.

494. Ahmed S, Bennett ST, Huxtable SJ, Todd JA, Matthews DR, Gough SC. INS VNTR allelic variation and dynamic insulin secretion in healthy adult non-diabetic Caucasian subjects. *Diabet Med*. 1999;16(11):910-7.

495. Scott RA, Scott LJ, Magi R, Marullo L, Gaulton KJ, Kaakinen M, et al. An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes*. 2017;66(11):2888-902.

496. Thorsby PM, Berg JP, Birkeland KI. Insulin gene variable number of tandem repeats is associated with increased fat mass during adolescence in non-obese girls. *Scand J Clin Lab Invest*. 2005;65(2):163-8.

497. Vafiadis P, Ounissi-Benkalha H, Palumbo M, Grabs R, Rousseau M, Goodyer CG, et al. Class III alleles of the variable number of tandem repeat insulin polymorphism associated with silencing of thymic insulin predispose to type 1 diabetes. *J Clin Endocrinol Metab*. 2001;86(8):3705-10.

498. Paquette J, Giannoukakis N, Polychronakos C, Vafiadis P, Deal C. The INS 5' variable number of tandem repeats is associated with IGF2 expression in humans. *J Biol Chem*. 1998;273(23):14158-64.

499. Pugliese A, Zeller M, Fernandez A, Jr., Zalcberg LJ, Bartlett RJ, Ricordi C, et al. The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. *Nat Genet*. 1997;15(3):293-7.

500. Fan Y, Rudert William A, Grupillo M, He J, Sisino G, Trucco M. Thymus-specific deletion of insulin induces autoimmune diabetes. *The EMBO Journal*. 2009;28(18):2812-24.

501. Kyewski B, Derbinski J. Self-representation in the thymus: an extended view. *Nature Reviews Immunology*. 2004;4:688.

502. Song L, Luo J, Peng Q, Wang J, Xie L, He Y, et al. Lack of association of INS VNTR polymorphism with polycystic ovary syndrome: a meta-analysis. *J Assist Reprod Genet*. 2014;31(6):675-81.

503. Cai CQ, Zhang T, Breslin MB, Giraud M, Lan MS. Both polymorphic variable number of tandem repeats and autoimmune regulator modulate differential expression of insulin in human thymic epithelial cells. *Diabetes*. 2011;60(1):336-44.

504. Sparks AE, Chen C, Breslin MB, Lan MS. Functional domains of Autoimmune Regulator (AIRE) modulate INS-VNTR transcription in human thymic epithelial cells. *J Biol Chem*. 2016;291(21):11313-22.

505. Le Fur S, Auffray C, Letourneur F, Cruaud C, Le Stunff C, Bougnères P. Heterogeneity of class I INS VNTR allele association with insulin secretion in obese children. *Physiol Genomics*. 2006;25(3):480-4.

506. Královičová J, Gaunt TR, Rodriguez S, Wood PJ, Day INM, Vořechovský I. Variants in the human insulin gene that affect pre-mRNA splicing. *Diabetes*. 2006;55(1):260.

507. Groop L, Pociot F. Genetics of diabetes – Are we missing the genes or the disease? *Mol Cell Endocrinol*. 2014;382(1):726-39.

508. Storling J, Pociot F. Type 1 diabetes candidate genes linked to pancreatic islet cell inflammation and beta-cell apoptosis. *Genes*. 2017;8(2):72.

509. Bergholdt R, Brorsson C, Palleja A, Berchtold LA, Fløyel T, Bang-Berthelsen CH, et al. Identification of novel type 1 diabetes candidate genes by integrating genome-wide association data, protein-protein interactions, and human pancreatic islet gene expression. *Diabetes*. 2012;61(4):954.

510. Brorsson CA, Nielsen LB, Andersen ML, Kaur S, Bergholdt R, Hansen L, et al. Genetic risk score modelling for disease progression in new-onset type 1 diabetes patients: increased genetic load of islet-expressed and cytokine-regulated candidate genes predicts poorer glycemic control. *J Diabetes Res*. 2016;2016:9570424.

511. Ahuja R, Pinyol R, Reichenbach N, Custer L, Klingensmith J, Kessels MM, et al. Cordon-bleu is an actin nucleation factor and controls neuronal morphology. *Cell*. 2007;131(2):337-50.

512. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790-7.

513. Pugliese A, Boulware D, Yu L, Babu S, Steck AK, Becker D, et al. The HLA-DRB1*15:01-DQA1*01:02-DQB1*06:02 haplotype protects autoantibody-positive relatives from type 1 diabetes throughout the stages of disease progression. *Diabetes*. 2016; 65(4):1109-19.

514. Thomson G, Robinson WP, Kuhner MK, Joe S, MacDonald MJ, Gottschall JL, et al. Genetic heterogeneity, modes of inheritance, and risk estimates for a joint study of Caucasians with insulin-dependent diabetes mellitus. *Am J Hum Genet*. 1988;43(6):799-816.

515. Karnes JH, Bastarache L, Shaffer CM, Gaudieri S, Xu Y, Glazer AM, et al. Phenome-wide scanning identifies multiple diseases and disease severity phenotypes associated with HLA variants. *Sci Transl Med*. 2017;9(389).

516. Ngo ST, Steyn FJ, McCombe PA. Gender differences in autoimmune disease. *Front Neuroendocrinol*. 2014;35(3):347-69.

517. Weets I, Van Autreve J, Van der Auwera BJ, Schuit FC, Du Caju MV, Decochez K, et al. Male-to-female excess in diabetes diagnosed in early adulthood is not specific for the immune-mediated form nor is it HLA-DQ restricted: possible relation to increased body mass index. *Diabetologia*. 2001;44(1):40-7.

518. Kagnoff MF. Celiac disease: pathogenesis of a model immunogenetic disease. *J Clin Invest*. 2007;117(1):41-9.

519. Megiorni F, Pizzuti A. HLA-DQA1 and HLA-DQB1 in Celiac disease predisposition: practical implications of the HLA molecular typing. *J Biomed Sci*. 2012;19(1):88.

520. Cucca F, Goy JV, Kawaguchi Y, Esposito L, Merriman ME, Wilson AJ, et al. A male-female bias in type 1 diabetes and linkage to chromosome Xp in MHC HLA-DR3-positive patients. *Nat Genet*. 1998;19(3):301-2.

521. Krischer JP, Cuthbertson DD, Greenbaum C. Male sex increases the risk of autoimmunity but not type 1 diabetes. *Diabetes Care*. 2004;27(8):1985-90.

522. Faas S, Trucco M. The genes influencing the susceptibility to IDDM in humans. *J Endocrinol Invest*. 1994;17(7):477-95.

523. Wilkin TJ, Murphy MJ. The gender insulin hypothesis: why girls are born lighter than boys, and the implications for insulin resistance. *Int J Obes (Lond)*. 2006;30:1056.

524. Graham J, Hagopian WA, Kockum I, Li LS, Sanjeevi CB, Lowe RM, et al. Genetic effects on age-dependent onset and islet cell autoantibody markers in type 1 diabetes. *Diabetes*. 2002;51(5):1346-55.

525. Sanjeevi CB, Lybrand TP, DeWeese C, Landin-Olsson M, Kockum I, Dahlquist G, et al. Polymorphic amino acid variations in HLA-DQ are associated with systematic physical property changes and occurrence of IDDM. Members of the Swedish Childhood Diabetes Study. *Diabetes*. 1995;44(1):125-31.

526. Berends AL, Steegers EA, Isaacs A, Aulchenko YS, Liu F, de Groot CJ, et al. Familial aggregation of preeclampsia and intrauterine growth restriction in a genetically isolated population in The Netherlands. *Eur J Hum Genet*. 2008;16:1437.

527. Keller A, Händel MN, Frederiksen P, Jacobsen R, Cohen AS, McGrath JJ, et al. Concentration of 25-hydroxyvitamin D from neonatal dried blood spots and the relation to gestational age, birth weight and Ponderal Index: the D-tect study. *Br J Nutr*. 2018;119(12):1416-23.

528. Zhu P, Tong S-L, Hu W-B, Hao J-H, Tao R-X, Huang K, et al. Cord blood 25-hydroxyvitamin D and fetal growth in the China-Anhui Birth Cohort Study. *Sci Rep*. 2015;5:14930.

529. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet*. 2010;376(9736):180-8.

530. Laborda J. The role of the epidermal growth factor-like protein dlk in cell differentiation. *Histol Histopathol*. 2000;15(1):119-29.

531. Cleaton MA, Dent CL, Howard M, Corish JA, Gutteridge I, Sovio U, et al. Fetus-derived DLK1 is required for maternal metabolic adaptations to pregnancy and is associated with fetal growth restriction. *Nat Genet*. 2016;48(12):1473-80.

532. Babon Jeffrey J, Lucet Isabelle S, Murphy James M, Nicola Nicos A, Varghese Leila N. The molecular regulation of Janus kinase (JAK) activation. *Biochem J*. 2014;462(1):1.

533. Richardson SJ, Morgan NG, Foulis AK. Pancreatic pathology in type 1 diabetes mellitus. *Endocr Pathol*. 2014;25(1):80-92.

534. Marroqui L, Dos Santos RS, Floyel T, Grieco FA, Santin I, Op de Beeck A, et al. TYK2, a candidate gene for type 1 diabetes, modulates apoptosis and the innate immune response in human pancreatic beta-cells. *Diabetes*. 2015;64(11):3808-17.

535. Dendrou CA, Cortes A, Shipman L, Evans HG, Attfield KE, Jostins L, et al. Resolving TYK2 locus genotype-to-phenotype differences in autoimmunity. *Sci Transl Med*. 2016;8(363):363ra149-363ra149.

536. Hellgren G, Jansson JO, Carlsson LM, Carlsson B. The growth hormone receptor associates with Jak1, Jak2 and Tyk2 in human liver. *Growth Horm IGF Res*. 1999;9(3):212-8.

537. Chia DJ. Minireview: mechanisms of growth hormone-mediated gene regulation. *Mol Endocrinol*. 2014;28(7):1012-25.

538. von Horn H, Ekstrom C, Ellis E, Olivecrona H, Einarsson C, Tally M, et al. GH is a regulator of IGF2 promoter-specific transcription in human liver. *J Endocrinol*. 2002;172(3):457-65.

539. Kreins AY, Ciancanelli MJ, Okada S, Kong XF, Ramirez-Alejo N, Kilic SS, et al. Human TYK2 deficiency: Mycobacterial and viral infections without hyper-IgE syndrome. *J Exp Med*. 2015;212(10):1641-62.

540. Carlsson A, Kockum I, Lindblad B, Engleson L, Nilsson A, Forsander G, et al. Low risk HLA-DQ and increased body mass index in newly diagnosed type 1 diabetes children in the Better Diabetes Diagnosis study in Sweden. *Int J Obes (Lond)*. 2012;36(5):718-24.

541. Floyel T, Brorsson C, Nielsen LB, Miani M, Bang-Berthelsen CH, Friedrichsen M, et al. CTSH regulates beta-cell function and disease progression in newly diagnosed type 1 diabetes patients. *Proc Natl Acad Sci U S A*. 2014;111(28):10305-10.

542. Hagopian W, Lee HS, Liu E, Rewers M, She JX, Ziegler AG, et al. Co-occurrence of Type 1 Diabetes and Celiac Disease Autoimmunity. *Pediatrics*. 2017;140(5).

543. Ma Y, Mazumdar M, Memtsoudis SG. Beyond repeated-measures analysis of variance: advanced statistical methods for the analysis of longitudinal data in anesthesia research. *Reg Anesth Pain Med*. 2012;37(1):99-105.

544. Fitzmaurice GM, Ravichandran C. A primer in longitudinal data analysis. *Circulation*. 2008;118(19):2005-10.

545. Twisk J, Smidt N, de Vente W. Applied analysis of recurrent events: a practical overview. *J Epidemiol Community Health*. 2005;59(8):706-10.

546. Wang S, Gao W, Ngwa J, Allard C, Liu CT, Cupples LA. Comparing baseline and longitudinal measures in association studies. *BMC Proc*. 2014;8(Suppl 1):S84.

547. Xu Z, Shen X, Pan W, Alzheimer's Disease Neuroimaging Initiative. Longitudinal analysis is more powerful than cross-sectional analysis in detecting genetic association with neuroimaging phenotypes. *PLoS One*. 2014;9(8):e102312.

548. Kaslow RA, Ostrow DG, Detels R, Phair JP, Polk BF, Rinaldo CRJ. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *Am J Epidemiol*. 1987;126(2):310-8.

549. Luan J, Kerner B, Zhao JH, Loos RJ, Sharp SJ, Muthen BO, et al. A multilevel linear mixed model of the association between candidate genes and weight and body mass index using the Framingham longitudinal family data. *BMC Proc*. 2009;3 Suppl 7:S115.

550. Kincaid C. Guidelines for Selecting the Covariance Structure in Mixed Model Analysis. *SAS Technical Papers* [Internet]. 2005; Paper 198-30. Available from: https://support.sas.com/resources/papers.

551. Gauderman WJ, Macgregor S, Briollais L, Scurrah K, Tobin M, Park T, et al. Longitudinal data analysis in pedigree studies. *Genet Epidemiol*. 2003;25 Suppl 1:S18-28.

552. Moulder R, Lahesmaa R. Early signs of disease in type 1 diabetes. *Pediatr Diabetes*. 2016;17 Suppl 22:43-8.

553. Colli ML, Moore F, Gurzov EN, Ortis F, Eizirik DL. MDA5 and PTPN2, two candidate genes for type 1 diabetes, modify pancreatic beta-cell responses to the viral by-product double-stranded RNA. *Hum Mol Genet*. 2010;19(1):135-46.

554. Hyoty H, Hiltunen M, Knip M, Laakkonen M, Vahasalo P, Karjalainen J, et al. A prospective study of the role of coxsackie B and other enterovirus infections in the pathogenesis of IDDM. Childhood Diabetes in Finland (DiMe) Study Group. *Diabetes*. 1995;44(6):652-7.

555. Soleimanpour SA, Gupta A, Bakay M, Ferrari AM, Groff DN, Fadista J, et al. The diabetes susceptibility gene Clec16a regulates mitophagy. *Cell*. 2014;157(7):1577-90.

556. Zhou Y, Wu H, Zhao M, Chang C, Lu Q. The Bach family of transcription factors: a comprehensive review. *Clin Rev Allergy Immunol*. 2016;50(3):345-56.

557. Marroqui L, Santin I, Dos Santos RS, Marselli L, Marchetti P, Eizirik DL. BACH2, a candidate risk gene for type 1 diabetes, regulates apoptosis in pancreatic beta-cells via JNK1 modulation and crosstalk with the candidate gene PTPN2. *Diabetes*. 2014;63(7):2516-27.

558. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet*. 2010;42(4):295-302.

559. Plagnol V, Howson JMM, Smyth DJ, Walker N, Hafler JP, Wallace C, et al. Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. *PLoS Genet*. 2011;7(8).

560. Jin Y, Sharma A, Bai S, Davis C, Liu H, Hopkins D, et al. Risk of type 1 diabetes progression in islet autoantibody-positive children can be further stratified using expression patterns of multiple genes implicated in peripheral blood lymphocyte activation and function. *Diabetes*. 2014;63(7):2506-15.

561. Long SA, Cerosaletti K, Wan JY, Ho JC, Tatum M, Wei S, et al. An autoimmune-associated variant in PTPN2 reveals an impairment of IL-2R signaling in CD4(+) T cells. *Genes Immun*. 2011;12(2):116-25.

562. Lempainen J, Laine AP, Hammais A, Toppari J, Simell O, Veijola R, et al. Non-HLA gene effects on the disease process of type 1 diabetes: From HLA susceptibility to overt disease. *J Autoimmun*. 2015;61:45-53.

563. Perez-Garcia A, Ambesi-Impiombato A, Hadler M, Rigo I, LeDuc CA, Kelly K, et al. Genetic loss of SH2B3 in acute lymphoblastic leukemia. *Blood*. 2013;122(14):2425-32.

564. Soranzo N, Spector TD, Mangino M, Kuhnel B, Rendon A, Teumer A, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet*. 2009;41(11):1182-90.

565. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, et al. Genome-wide association study of blood pressure and hypertension. *Nat Genet*. 2009;41:677.

566. Dichgans M, Malik R, Konig IR, Rosand J, Clarke R, Gretarsdottir S, et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. *Stroke*. 2014;45(1):24-36.

567. Danial NN, Walensky LD, Zhang CY, Choi CS, Fisher JK, Molina AJ, et al. Dual role of proapoptotic BAD in insulin secretion and beta cell survival. *Nat Med*. 2008;14(2):144-53.

568. Howe LD, Tilling K, Lawlor DA. Accuracy of height and weight data from child health records. *Arch Dis Child*. 2009;94(12):950.

569. Santin I, Eizirik DL. Candidate genes for type 1 diabetes modulate pancreatic islet inflammation and beta-cell apoptosis. *Diabetes Obes Metab*. 2013;15 Suppl 3:71-81.

570. Blanton LV, Charbonneau MR, Salih T, Barratt MJ, Venkatesh S, Ilkaveya O, et al. Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. *Science*. 2016;351(6275).

571. Gronlund MM, Lehtonen OP, Eerola E, Kero P. Fecal microflora in healthy infants born by different methods of delivery: permanent changes in intestinal flora after cesarean delivery. *J Pediatr Gastroenterol Nutr*. 1999;28(1):19-25.

572. Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, et al. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics*. 2006;118(2):511-21.

573. Stene LC, Rønningen KS, Undlien DE, Joner G. Does the relative risk for type 1 diabetes conferred by HLA-DQ, INS, and PTPN22 polymorphisms vary with maternal age, birth weight, or cesarean section? *Pediatr Diabetes*. 2011;12(2):91-4.

574. Ahmed SF, Farquharson C. The effect of GH and IGF1 on linear growth and skeletal development and their modulation by SOCS proteins. *J Endocrinol*. 2010;206(3):249-59.

575. Swolin D, Ohlsson C. Growth hormone increases interleukin-6 produced by human osteoblast-like cells. *J Clin Endocrinol Metab*. 1996;81(12):4329-33.

576. Renier G, Clement I, Desfaits AC, Lambert A. Direct stimulatory effect of insulin-like growth factor-I on monocyte and macrophage tumor necrosis factor-alpha production. *Endocrinology*. 1996;137(11):4611-8.

577. Uronen-Hansson H, Allen ML, Lichtarowicz-Krynska E, Aynsley-Green A, Cole TJ, Höidén-Guthenberg I, et al. Growth hormone enhances proinflammatory cytokine production by monocytes in whole blood. *Growth Horm IGF Res*. 2003;13(5):282-6.

578. Wong SC, Dobie R, Altowati MA, Werther GA, Farquharson C, Ahmed SF. Growth and the Growth Hormone-Insulin Like Growth Factor 1 Axis in Children With Chronic Inflammation: Current Evidence, Gaps in Knowledge, and Future Directions. *Endocr Rev*. 2016;37(1):62-110.

579. Makki K, Froguel P, Wolowczuk I. Adipose tissue in obesity-related inflammation and insulin resistance: cells, cytokines, and chemokines. *ISRN Inflamm*. 2013;2013:139239.

580. Davies UM, Jones J, Reeve J, Camacho-Hubner C, Charlett A, Ansell BM, et al. Juvenile rheumatoid arthritis. Effects of disease activity and recombinant human growth hormone on insulin-like growth factor 1, insulin-like growth factor binding proteins 1 and 3, and osteocalcin. *Arthritis Rheum*. 1997;40(2):332-40.

581. Sharma A, Liu X, Hadley D, Hagopian W, Chen WM, Onengut-Gumuscu S, et al. Identification of non-HLA genes associated with development of islet autoimmunity and type 1 diabetes in the prospective TEDDY cohort. *J Autoimmun*. 2018;89:90-100.

582. Ge Y, Paisie TK, Newman JRB, McIntyre LM, Concannon P. UBASH3A mediates risk for type 1 diabetes through inhibition of T-cell receptor–induced NF-κB signaling. *Diabetes*. 2017;66(7):2033-43.

583. Todd JA. Evidence that UBASH3 is a causal gene for type 1 diabetes. *Eur J Hum Genet*. 2018;26(7):925-7.

584. Brocardo MG, Schillaci R, Galeano A, Radrizzani M, White V, Guerrico AG, et al. Early effects of insulin-like growth factor-1 in activated human T lymphocytes. *J Leukoc Biol*. 2001;70(2):297-305.

585. Steck AK, Wong R, Wagner B, Johnson K, Liu E, Romanos J, et al. Effects of non-HLA gene polymorphisms on development of islet autoimmunity and type 1 diabetes in a population with high-risk HLA-DR,DQ genotypes. *Diabetes*. 2012;61:753-8.

586. Smyth DJ, Cooper JD, Howson JMM, Clarke P, Downes K, Mistry T, et al. FUT2 nonsecretor status links type 1 diabetes susceptibility and resistance to infection. *Diabetes*. 2011;60(11):3081-4.

587. Kelly RJ, Rouquier S, Giorgi D, Lennon GG, Lowe JB. Sequence and expression of a candidate for the human Secretor blood group alpha(1,2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. *J Biol Chem*. 1995;270(9):4640-9.

588. Wacklin P, Makivuokko H, Alakulppi N, Nikkila J, Tenkanen H, Rabina J, et al. Secretor genotype (FUT2 gene) is strongly associated with the composition of Bifidobacteria in the human intestine. *PLoS One*. 2011;6(5):e20113.

589. Maroni L, van de Graaf SF, Hohenester SD, Oude Elferink RP, Beuers U. Fucosyltransferase 2: a genetic risk factor for primary sclerosing cholangitis and Crohn's disease--a comprehensive review. *Clin Rev Allergy Immunol*. 2015;48(2-3):182-91.

590. Schwarzer M, Makki K, Storelli G, Machuca-Gayet I, Srutkova D, Hermanova P, et al. Lactobacillus plantarum strain maintains growth of infant mice during chronic undernutrition. *Science*. 2016;351(6275):854-7.

591. Roman-Garcia P, Quiros-Gonzalez I, Mottram L, Lieben L, Sharan K, Wangwiwatsin A, et al. Vitamin B(12)–dependent taurine synthesis regulates growth and bone mass. *J Clin Invest*. 2014;124(7):2988-3002.

592. Velkova A, Diaz JEL, Pangilinan F, Molloy AM, Mills JL, Shane B, et al. The FUT2 secretor variant p.Trp154Ter influences serum vitamin B12 concentration via holo-haptocorrin, but not holo-transcobalamin, and is associated with haptocorrin glycosylation. *Hum Mol Genet*. 2017;26(24):4975-88.

593. Allin KH, Friedrich N, Pietzner M, Grarup N, Thuesen BH, Linneberg A, et al. Genetic determinants of serum vitamin B12 and their relation to body mass index. *European Journal of Epidemiology*. 2017;32(2):125-34.

594. Chen W, Salojin KV, Mi QS, Grattan M, Meagher TC, Zucker P, et al. Insulin-like growth factor (IGF)-I/IGF-binding protein-3 complex: therapeutic efficacy and mechanism of protection against type 1 diabetes. *Endocrinology*. 2004;145(2):627-38.

595. Cardwell CR, Stene LC, Ludvigsson J, Rosenbauer J, Cinek O, Svensson J, et al. Breast-feeding and childhood-onset type 1 diabetes: a pooled analysis of individual participant data from 43 observational studies. *Diabetes Care*. 2012;35(11):2215-25.

596. Burrin DG. Is milk-borne insulin-like growth factor-I essential for neonatal development? *J Nutr*. 1997;127(5 Suppl):975s-9s.

597. Holick MF. Resurrection of vitamin D deficiency and rickets. *J Clin Invest*. 2006;116(8):2062-72.

598. Bhan I. Vitamin D Binding Protein and Bone Health. *Int J Endocrinol*. 2014(3):561214.

599. Powe CE, Evans MK, Wenger J, Zonderman AB, Berg AH, Nalls M, et al. Vitamin D–binding protein and vitamin D status of Black Americans and White Americans. *N Engl J Med*. 2013;369(21):1991-2000.

600. Carpenter TO, Zhang JH, Parra E, Ellis BK, Simpson C, Lee WM, et al. Vitamin D binding protein is a key determinant of 25-hydroxyvitamin D levels in infants and toddlers. *J Bone Miner Res*. 2013;28(1):213-21.

601. Nykjaer A, Dragun D, Walther D, Vorum H, Jacobsen C, Herz J, et al. An endocytic pathway essential for renal uptake and activation of the steroid 25-(OH) vitamin D3. *Cell*. 1999;96(4):507-15.

602. Danescu LG, Levy S, Levy J. Vitamin D and diabetes mellitus. *Endocrine*. 2009;35(1):11-7.

603. Munger KL, Levin LI, Massa J, Horst R, Orban T, Ascherio A. Preclinical serum 25-hydroxyvitamin D levels and risk of type 1 diabetes in a cohort of US military personnel. *Am J Epidemiol*. 2013;177(5):411-9.

604. Eurodiab Substudy 2 Study Group. Vitamin D supplement in early childhood and risk for Type I (insulin-dependent) diabetes mellitus. The EURODIAB Substudy 2 Study Group. *Diabetologia*. 1999;42(1):51-4.

605. Simpson M, Brady H, Yin X, Seifert J, Barriga K, Hoffman M, et al. No association of vitamin D intake or 25-hydroxyvitamin D levels in childhood with risk of islet autoimmunity and type 1 diabetes: the Diabetes Autoimmunity Study in the Young (DAISY). *Diabetologia*. 2011;54(11):2779-88.

606. Blanton D, Han Z, Bierschenk L, Linga-Reddy MV, Wang H, Clare-Salzler M, et al. Reduced serum vitamin D-binding protein levels are associated with type 1 diabetes. *Diabetes*. 2011;60(10):2566-70.

607. Kodama K, Zhao Z, Toda K, Yip L, Fuhlbrigge R, Miao D, et al. Expression-based genome-wide association study links vitamin D–binding protein with autoantigenicity in type 1 diabetes. *Diabetes*. 2016;65(5):1341-9.

608. Sorensen IM, Joner G, Jenum PA, Eskild A, Brunborg C, Torjesen PA, et al. Vitamin D-binding protein and 25-hydroxyvitamin D during pregnancy in mothers whose children later developed type 1 diabetes. *Diabetes Metab Res Rev*. 2016;32(8):883-90.

609. Walsh JM, McGowan CA, Kilbane M, McKenna MJ, McAuliffe FM. The relationship between maternal and fetal vitamin D, insulin resistance, and fetal growth. *Reprod Sci*. 2013;20(5):536-41.

610. Norris JM, Lee HS, Frederiksen B, Erlund I, Uusitalo U, Yang J, et al. Plasma 25-hydroxyvitamin D concentration and risk of islet autoimmunity. *Diabetes*. 2018;67(1):146-54.

611. Takiishi T, Van Belle T, Gysemans C, Mathieu C. Effects of vitamin D on antigen-specific and non-antigen-specific immune modulation: relevance for type 1 diabetes. *Pediatr Diabetes*. 2013;14(2):81-9.

612. Wolden-Kirk H, Overbergh L, Christesen HT, Brusgaard K, Mathieu C. Vitamin D and diabetes: its importance for beta cell and immune function. *Mol Cell Endocrinol*. 2011;347(1-2):106-20.

613. Saxena NK, Taliaferro-Smith L, Knight BB, Merlin D, Anania FA, O'Regan RM, et al. Bidirectional crosstalk between leptin and insulin-like growth factor-I signaling promotes invasion and migration of breast cancer cells via transactivation of epidermal growth factor receptor. *Cancer Res*. 2008;68(23):9712-22.

614. Ajuwon KM, Kuske JL, Ragland D, Adeola O, Hancock DL, Anderson DB, et al. The regulation of IGF-1 by leptin in the pig is tissue specific and independent of changes in growth hormone. *J Nutr Biochem*. 2003;14(9):522-30.

615. Erlandsson MC, Doria Medina R, Töyrä Silfverswärd S, Bokarewa MI. Smoking functions as a negative regulator of IGF1 and impairs adipokine network in patients with rheumatoid arthritis. *Mediators Inflamm*. 2016;2016:3082820.

616. Alemzadeh R, Kichler J, Babar G, Calhoun M. Hypovitaminosis D in obese children and adolescents: relationship with adiposity, insulin sensitivity, ethnicity, and season. *Metabolism*. 2008;57(2):183-91.

617. Bell NH, Epstein S, Greene A, Shary J, Oexmann MJ, Shaw S. Evidence for alteration of the vitamin D-endocrine system in obese subjects. *J Clin Invest*. 1985;76(1):370-3.

618. Parikh SJ, Edelman M, Uwaifo GI, Freedman RJ, Semega-Janneh M, Reynolds J, et al. The relationship between obesity and serum 1,25-dihydroxy vitamin D concentrations in healthy adults. *J Clin Endocrinol Metab*. 2004;89(3):1196-9.

619. Rodriguez-Rodriguez E, Navia B, Lopez-Sobaler AM, Ortega RM. Vitamin D in overweight/obese women and its relationship with dietetic and anthropometric variables. *Obesity*. 2009;17(4):778-82.

620. Konradsen S, Ag H, Lindberg F, Hexeberg S, Jorde R. Serum 1,25-dihydroxy vitamin D is inversely associated with body mass index. *Eur J Nutr*. 2008;47(2):87-91.

621. Wortsman J, Matsuoka LY, Chen TC, Lu Z, Holick MF. Decreased bioavailability of vitamin D in obesity. *Am J Clin Nutr*. 2000;72(3):690-3.

622. Ding C, Parameswaran V, Blizzard L, Burgess J, Jones G. Not a simple fat-soluble vitamin: changes in serum 25-(OH)D levels are predicted by adiposity and adipocytokines in older adults. *J Intern Med*. 2010;268(5):501-10.

623. Menendez C, Lage M, Peino R, Baldelli R, Concheiro P, Dieguez C, et al. Retinoic acid and vitamin D(3) powerfully inhibit in vitro leptin secretion by human adipose tissue. *J Endocrinol*. 2001;170(2):425-31.

624. Maetani M, Maskarinec G, Franke AA, Cooney RV. Association of leptin, 25-hydroxyvitamin D, and parathyroid hormone in women. *Nutr Cancer*. 2009;61(2):225-31.

625. Frederiksen BN, Kroehl M, Fingerlin TE, Wong R, Steck AK, Rewers M, et al. Association between vitamin D metabolism gene polymorphisms and risk of islet autoimmunity and progression to type 1 diabetes: the diabetes autoimmunity study in the young (DAISY). *J Clin Endocrinol Metab*. 2013;98(11):E1845-51.

626. Mokry LE, Ross S, Ahmad OS, Forgetta V, Smith GD, Leong A, et al. Vitamin D and risk of multiple sclerosis: A Mendelian randomization study. *PLoS Med*. 2015;12(8):e1001866.

627. Prabhu AV, Luu W, Sharpe LJ, Brown AJ. Cholesterol-mediated degradation of 7-dehydrocholesterol reductase switches the balance from cholesterol to vitamin D synthesis. *J Biol Chem*. 2016;291(16):8363-73.

628. Kuan V, Martineau AR, Griffiths CJ, Hyppönen E, Walton R. DHCR7 mutations linked to higher vitamin D status allowed early human migration to Northern latitudes. *BMC Evol Biol*. 2013;13(1):144.

629. Hajimohammadi M, Shab-Bidar S, Neyestani TR. Vitamin D and serum leptin: a systematic review and meta-analysis of observational studies and randomized controlled trials. *Eur J Clin Nutr*. 2017;71(10):1144-53.

630. Kasiappan R, Sun Y, Lungchukiet P, Quarni W, Zhang X, Bai W. Vitamin D suppresses leptin stimulation of cancer growth through microRNA. *Cancer Res*. 2014;74(21):6194-204.

631. Carstensen B, Read SH, Friis S, Sund R, Keskimaki I, Svensson AM, et al. Cancer incidence in persons with type 1 diabetes: a five-country study of 9,000 cancers in type 1 diabetic individuals. *Diabetologia*. 2016;59(5):980-8.

632. Enmark E, Gustafsson JA. Oestrogen receptors - an overview. *J Intern Med*. 1999;246(2):133-8.

633. Kovats S. Estrogen receptors regulate innate immune cells and signaling pathways. *Cell Immunol*. 2015;294(2):63-9.

634. Tiano JP, Mauvais-Jarvis F. Importance of oestrogen receptors to preserve functional beta-cell mass in diabetes. *Nat Rev Endocrinol*. 2012;8(6):342-51.

635. Deroo BJ, Korach KS. Estrogen receptors and human disease. *The Journal of Clinical Investigation*. 2006;116(3):561-70.

636. Vona-Davis L, Rose DP. Adipokines as endocrine, paracrine, and autocrine factors in breast cancer risk and progression. *Endocr Relat Cancer*. 2007;14(2):189-206.

637. Mason C, De Dieu Tapsoba J, Duggan C, Imayama I, Wang C-Y, Korde LA, et al. Effects of vitamin D supplementation during weight loss on sex hormones in postmenopausal women. *Menopause*. 2016;23(6):645-52.

638. Krishnan AV, Swami S, Feldman D. Vitamin D and breast cancer: inhibition of estrogen synthesis and signaling. *J Steroid Biochem Mol Biol*. 2010;121(1):343-8.

639. Shimizu H, Shimomura Y, Nakanishi Y, Futawatari T, Ohtani K, Sato N, et al. Estrogen increases in vivo leptin production in rats and human subjects. *J Endocrinol*. 1997;154(2):285-92.

640. Qi Q, Zheng Y, Huang T, Rood J, Bray GA, Sacks FM, et al. Vitamin D metabolism-related genetic variants, dietary protein intake and improvement of insulin resistance in a 2 year weight-loss trial: POUNDS Lost. *Diabetologia*. 2015;58(12):2791-9.

641. Paz-Filho G, Mastronardi C, Wong M-L, Licinio J. Leptin therapy, insulin sensitivity, and glucose homeostasis. *Indian J Endocrinol Metab*. 2012;16(Suppl 3):S549-S55.

642. Copeland KC, Chernausek S. Mini-puberty and growth. *Pediatrics*. 2016;138(1):e20161301.

643. Procaccini C, Pucino V, Mantzoros CS, Matarese G. Leptin in autoimmune diseases. *Metabolism*. 2015;64(1):92-104.

644. Gomez JM, Maravall FJ, Gomez N, Navarro MA, Casamitjana R, Soler J. Relationship between 25-(OH) D3, the IGF-I system, leptin, anthropometric and body composition variables in a healthy, randomly selected population. *Horm Metab Res*. 2004;36(1):48-53.

645. Schneider HJ, Saller B, Klotsche J, Marz W, Erwa W, Wittchen HU, et al. Opposite associations of age-dependent insulin-like growth factor-I standard deviation scores with nutritional state in normal weight and obese subjects. *Eur J Endocrinol*. 2006;154(5):699-706.

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

647. Fairweather D, Frisancho-Kiss S, Rose NR. Sex differences in autoimmune disease from a pathological perspective. *Am J Pathol*. 2008;173(3):600-9.

648. Blum WF, Ranke MB, Kietzmann K, Gauggel E, Zeisel HJ, Bierich JR. A specific radioimmunoassay for the growth hormone (GH)-dependent somatomedin-binding protein: its use for diagnosis of GH deficiency. *J Clin Endocrinol Metab*. 1990;70(5):1292-8.

649. Martin RM, Patel R, Zinovik A, Kramer MS, Oken E, Vilchuck K, et al. Filter paper blood spot enzyme linked immunoassay for insulin and application in the evaluation of determinants of child insulin resistance. *PLoS One*. 2012;7(10):e46752.