Imperial College London

Department of Epidemiology and Biostatistics

Genetic and environmental correlates of growth patterns leading to obesity

Shikta Das

A thesis submitted for the degree of Doctor of Philosophy

at Imperial College London

Acknowledgements

There are a few individuals without whom this thesis would have been impossible. I would like to take this opportunity to thank them from my heart for being so supportive of me at all times.

I would like to thank my supervisor Professor Marjo-Riitta Jarvelin for her constant encouragement and patience during my PhD. Her enthusiasm for my work gave me the moral support to carry on, even when things looked too difficult. I am proud to be her 23rd PhD student. She has taught me research skills as well as very important career and life-skills. I also would like to thank for her husband Mr. Lauri Lakio for his encouragement and support. I would like to thank Dr Paul O'Reilly for his patience in helping me to grasp key epidemiological concepts and the finer nuances of thesis writing. He was always available when I needed guidance. His feedback on the drafts and revisions has helped me immensely in completing this thesis. He has made an invaluable contribution to my thesis effort by teaching me to be independent in my thinking, research and writing. I would also like to thank Dr Lachlan Coin for his encouragement throughout my PhD. He has been extremely supportive of my work and a great inspiration.

I am grateful to my MSc supervisor from Birkbeck College, Dr Adrian Shepherd, for planting the idea of a PhD in my head and helping me turn it into a tangible ambition. He provided honest advice when I needed it most and helped me chose the right PhD rather than any PhD.

I am indebted to Dr Ulla Sovio for coaching me so effectively about key epidemiological skills during my initial days of PhD research. She has been a role model for me. I cannot thank her enough for paving the way for me, as well as other researchers, with her novel methodology. Dr Penny has been very helpful with my queries – and I have had lots of them, over the years. Adai; I thank you for being my mentor and for taking the time to explain to me the basics of genetics. I am grateful for your constant drive to keep me focused on finishing; and for motivating me when I needed it most.

I cannot thank my gang of girls in Room 171 enough - Anokhi, Rachel, Daniela and Mai. Your unlimited supply of gossip, hugs, coffee and chocolate kept me going through the literature reviews on obesity, and much else. You gave me the emotional support I needed during my ups and downs. I am thankful to James McGovern and the others in the administration and support teams for accommodating my requests for meetings and other support.

I had the privilege to work with NFBC teams in Finland and the UK and get engaged with various teams working on early growth and genetics (Early Growth Genetics Consortium, EGG). I would like to extend my special thanks to Dr. Rob Taal for his great and extremely professional companionship while working on two EGG projects. He was a very supportive co-partner for both my papers and a very kind soul. Rob has been a fantastic colleague and I look forward to collaborating with him again in the future.

Data collection, management and analyses has been supported by the grants from Academy of Finland (Grant no. 1114194, Grant no. 12926901), University Hospital Oulu and Unit of Primary Care, Biocenter, University of Oulu, and the European Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643 and EU FP7- EurHEALTHAgeing - European Research on Developmental, Birth and Genetic Determinants of Ageing. Grant no. 24000522).

Of course, my PhD would not have been possible but for the funding from the Medical Research Council Medical Research Council, UK (G0500539, G0600705, PrevMetSyn/SALVE, PS0476). The MRC studentship and the Centenary Early Career Award 2012/2013; afforded me the opportunity to expand my research interests further and into areas that are not necessarily traditional but hold potentially significant and tangible benefits to the society.

I am grateful to my friends, who have become my extended family in this country. Thank you to Sheela for always being there for everything, for stepping in to be with Raaghav when I could not. For the countless cups of strong tea on Fridays while I moaned and for telling off my husband when he demanded more time than a PhD candidate wife could possibly give him. Thank you Pam and Vinita for affectionately looking after my son when I could not be around. Thank you to Sakshi for being a loving older sister to Raaghav, while I was busy writing. To my gang of mums from Dulwich Prep London - Monica, Sairah and Harleen – thank you for always being helpful and supportive in difficult times.

I would also like to thank my parents (Dr. Alok & Rekha) for raising me a happy child and encouraging me to work towards a PhD since my early years. I would also like to thank them for always giving me positive thoughts and for inculcating in me a sense of inquisitiveness, an essential skill for a scientist.

I would like to send a big thank you to my in-laws (Rajesh and Mithilesh) for visiting us every summer through my PhD and giving me beautiful memories for life. Thank you for keeping me well fed and for looking after our son so lovingly when I was chained to my desk at St Mary's.

I have no words to thank my husband (Saurabh) who has been my mentor, confidante and friend, in both good and hard times through this PhD and my little six year old hero (Raaghav). Raaghav has been wonderful and almost never complained; even when he could not see his mummy for several days and nights while she was working. Saurabh – thank you for being a great father and even a mother to our son by stepping into my shoes when I could not be there for him. I cannot thank enough my close friends who have been my mental support all through my PhD.

I would also like to extend my thanks to a lot of people who are not around with me but are always in my thoughts.

Declaration of Originality

I declare the work presented in this thesis to be my own except where collaboration is specifically indicated in the text.



Shikta Das

October 2013

We confirm that this is an accurate presentation of this work.

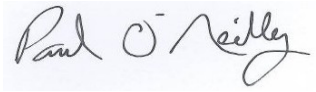
Name of Supervisor: Prof Marjo-Riitta Jarvelin

Signature …  **…**

Name of Supervisor: Dr Lachlan Coin

Signature……..  **…….**

Name of Supervisor: Dr Paul O'Reilly

Signature …  **…**

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

υγιή μητέρα, υγιές μωρό

Healthy mother, healthy baby

Abstract

The intrauterine period is a vulnerable period of development. Any adverse environment can permanently change the body's organ structure and function, expressed as an increased disease risk later in life. Studies show that variability in growth patterns in early life is associated with obesity and other cardiovascular diseases in adulthood, but the genetic and environmental determinants of these processes are largely unknown.

The main objectives of this study were to identify genetic and environmental pre- and postnatal factors associated with early growth in infancy and childhood and later metabolic outcomes in adulthood from the Northern Finland Birth Cohorts (NFBCs).

Several maternal and paternal factors, such as height, smoking, parity and pre-eclampsia, had direct association with faster postnatal height growth, some of which had their association mediated by size-at-birth variables. It was observed that an obesogenic environment *in utero* and during a child's growth exerts a 'programming' effect on the glucose-insulin axis as well as other cardio-vascular risk factors in adolescence. Moreover, the study shows that Leukocyte Telomere Length (LTL) at 31 years, a marker for aging, is inversely associated with multiple measures of adiposity in both men and women, and that a BMI increase in women from childhood to adulthood is associated with shorter telomeres at age 31. Two new genetic variants in/near *SBNO1* and *HMGA2* genes are associated with infant head circumference, which may indicate influence of brain growth and neurodevelopment via early life. Variants in/near *LEPR-LEPROT*, *FTO*, *TFAP2B* and *GNPDA2* showed an age-dependent association with adiposity in early childhood, while three loci (*FTO*, *TFAP2B* and *GNPDA2*) had their effect on adult adiposity mediated by early growth phenotypes.

This study emphasises the clinical importance of early growth markers as they may inform public health policy aimed at improving the pre-pregnancy environment and to monitor childhood growth during the first few years of development.

Abbreviations

AIC	Akaike information criterion
ALSPAC	Avon Longitudinal Study of Parents and Children
AP	Adiposity peak
AR	Adiposity rebound
BAI	Body adiposity index (BAI)
BIC	Bayesian information criterion
BMI	Body mass index
CBGS	Cambridge Baby Growth Study
COPSAC	Copenhagen Study on Asthma in Children
CVD	Cardio Vascular Disease
DBP	Diastolic blood pressure
DNA	Deoxyribonucleic acid
DNBC	Danish National Birth Cohort
DOHaD	Developmental origin of health and disease
EFSOCH	The Exeter Family Study of Childhood Health
EGG	Early growth and genetics consortium
G x E	Gene-environment
GA	Gestational age
GenR	The Generation R Study
GWAS	Genome-wide association study
GWG	Gestational weight gain
GWLS	Genome-wide linkage studies
HBCS	Helsinki Birth Cohort Study
HC	Head circumference
HDL	High density lipoprotein
HMA	Homeostasis model assessment
HPA	Hypothalamic-pituitary-adrenal axis
IUGR	Intrauterine growth retardation
INMA	INfancia y Medio Ambiente Project
IP	Intermediary phenotypes

LBW	Low birth weight of less than 2500g
LISA	Lifestyle – Immune System – Allergy Study
LTL	Leukocyte telomere length
MAF	Minor Allele Frequency
MRC	Medical Research Council
NFBC1966	Northern Finland birth cohort 1966
NFBC1986	Northern Finland birth cohort 1986
NLME	Non-linear model equation
PANIC	Physical activity and nutrition in children study
PCOS	Polycystic ovary syndrome
PHV	Peak height velocity
PWV	Peak weight velocity
RSD	Residual square deviation
SBP	Systolic blood pressure
SD	Standard deviation
SE	Standard error
SDS	Standard deviation scores
SES	Socio-economic status
SF	Symphysis-fundus
SGA	Small for gestational age
SITAR	Super Imposition by Translation and Rotation
SNP	Single nucleotide polymorphism
SWS	The Southampton Women's Survey
T2D	Type 2 Diabetes
WC	Waist-circumference
WHR	Waist-to-hip ratio

List of Tables

Table 1: The key historical and prospective cohort studies used in research on the fetal developmental origins of adult disease.

Table 2: Recent candidate gene identified for association with obesity in children [1].

Table 3: Variables used in the analyses in this thesis from NFBC 19866 and NFBC 1986.

Table 4: Characteristics of the NFBC 1986 data set.

Table 5: Univariate association of infant PHV with prenatal and size at birth variables.

Table 6: A. Association between prenatal predictors and size-at-birth factors (birth weight, birth length and birth BMI), adjusted for gestational age (GA) in NFBC 1986 Males. B. Association between prenatal predictors and size-at-birth factors (birth weight (BW), birth length (BL) and birth BMI), adjusted for gestational age (GA) in NFBC 1986 Females.

Table 7. A. The association between prenatal factors and PHV (model 1) and the mediating role of size-at-birth variables in NFBC 1986 males. Table 7. B. The association between prenatal factors and PHV (model 1) and the mediating role of size-at-birth variables in NFBC 1986 females.

Table 8: Summary characteristics of study population in Northern Finland 1986 Birth Cohort (NFBC 1986).

Table 9: Influence of the pre- and post-natal obesogenic environment on the glucose-insulin axis in adolescent males and females.

Table 10: Influence of the pre- and post-natal obesogenic environment on cardiovascular risk factors in adolescents of the NFBC 1986 birth cohort.

Table 11: Descriptive characteristics of study participants in Northern Finland Birth Cohort (NFBC 1966).

Table 12: The unadjusted and adjusted associations between leukocyte telomere length (LTL) at 31 years and early life/adult adiposity related measures by sex. The beta (95% confidence interval) is given as percentage difference. The $P < 0.05$ is considered significant.

Table 13: Summary results for top-three loci associated with head circumference (HC) for each participating cohort. For each of the 13 cohorts, the median age, total N, percentage of male, minor allele frequency (MAF), effect size (and SE), and P-value are shown as well as the latter for overall discovery stage, replication stage and meta-analysis results.

Table 14: Combined analyses of the top-three SNPs associating with head circumference (HC) with other phenotypes.

Table 15: Association of top-three head circumference SNPs with infant height and weight. Of the total of 13 study cohorts, 12 study cohorts contributed to these analyses since data on infant height and weight were not available in one cohort.

Table 16: Association of top-three SNP associating with head circumference (HC) at infancy, before and after adjustment for infant height.12 cohorts out of 13 contributed to these analyses.

Table 17. Association of 180 known adult height variants with head circumference in infancy, the table shows only significant association after correction for multiple testing at $P < 10^{-4}$.

Table 18. Mediation analysis for infant height influencing relationship between top-three SNPs and head circumference (HC).

Table 19: Characteristics of study populations and outcome phenotypes in the early growth GWAS.

Table 20. Summary results for discovery meta-analysis of top-eight SNPs from five European cohorts, and with replication in 12 European cohorts and the combined meta-analysis results in all 17 studies for early growth GWAS.

Table 21. Results of meta-analysis from mediation analyses of the five identified loci in four European population-based studies (N=9,128) in early growth GWAS.

Table 22. The five loci associating with early growth phenotypes from the discovery meta-analysis compared with the published GWAS studies (from GWAS Catalogue). The table shows the top five early growth SNPs (index SNPs) with their equivalent variant in LD with the index SNP (R2 > 0.5) from all published GWAS studies mapped to nearest gene. We provide distance of variant from the index SNP, r^2, risk allele and *P* values from early adiposity meta-analysis and GWAS catalogue, association to traits, direction of effect from early adiposity meta-analysis references. These SNPs give us good indication of the genetic susceptibility of top-five early growth phenotypes SNPs with various disease traits.

Table 23: Associations between 32 SNPs associating with adult BMI from GIANT consortium (N=249, 796) and early adiposity phenotypes, within the Stage 1 GWA cohorts (N=7,215). Results are from linear regression of z-score of early adiposity phenotypes against SNP (additive model), adjusted for gender and gestational age (GA). The gestational age was only adjusted in PWV, age and BMI at AP. The β are given in SD units. Results for PHV were removed as none of the results were significantly associated and due to lack of space.

Table 24: SNP selected for maternal smoking and BW interaction G x E study for follow-up in 12 indepei replication samples.

Table 25: Association between birth and infant size measures and testosterone levels at 15/16 years in the NFBC1986.

List of Figures

Figure 1. An overview of the developmental programming hypothesis followed in this thesis. The figure shows pathways which were investigated in this thesis to understand the role of prenatal genetic and environmental factors which may lead to in utero programming affecting smaller size at birth and accelerated childhood growth. Both of which have been associated to affect metabolic syndrome factors in later life. Furthermore, postnatal environment may also insert programming effect on accelerated postnatal growth and lead to adverse metabolic profile in adulthood.

Figure 2. The three phases of growth in height from birth to adulthood in an individual according to Tanner, 1989[19]. The black line represents the growth pattern of an individual from birth to adulthood and the grey line represents the growth variations that may occur in an individual due to nutrition changes. The figure has been adapted from Tanner, 1989[19].

Figure 3: Showing the waves of GWAS on adult obesity with different obesity related phenotypes.

Figure 4. Graphical illustration of height, weight and BMI growth patterns (or early growth phenotypes) in NFBC 1986. A) Peak Height Velocity (PHV); B) Peak Weight Velocity (PWV); C) Age and BMI at Adiposity Peak (AP) and Rebound (AR). The growth curves for males are in blue and for females in red.

Figure 5. A framework for the mechanisms in the association between the prenatal predictors and PHV.

Figure 6. Study groups according to maternal BMI in the Northern Finland Birth Cohort 1986 (NFBC 1986).
Offspring who participated in clinical examinations were stratified by the maternal BMI at the start of pregnancy and maternal BMI when offspring were 16 years of age. The maternal BMI definition of normal weight (N) and overweight and obese (O) were based on IOTF criteria[25].

Figure 7: Flowchart showing sample selection for the analysis from the NFBC 1966.
Figure 8: Study design showing the analyses performed for GWAS on head circumference (HC). The boxes highlighted represent the analyses performed for the NFBCs (1966 and 1986).

Figure 9. Manhattan plot of discovery genome wide association meta-analysis (n=10,768), with three loci selected for replication.

Figure 10. Mediation of effects of genotype on head circumference through an indirect effect on height.

Figure 11: Flowchart summarising the study design and results for the discovery, replication and combined meta-analysis of GWAS across 17 studies of European Ancestry.

Figure 12: Manhattan and Quantile-Quantile (QQ) plots of discovery stage results (N=7,215). The SNPs were filtered out if minor allele frequency <0.01, and adjusted twice by genomic controls, once within the study and then across the meta-analysis (lambda = 1.2). For each early growth phenotype, SNPs within loci with $P < 5 \times 10\text{-}5$ are displayed in green. The QQ plot of observed vs expected –log10P values are given in separate panels for each early growth phenotype. The black QQ lines represent the observed vs excepted associations for all SNPs. A. Peak height velocity (PHV), B. Peak weight velocity, C. Age at Adiposity peak (AP), D. BMI at adiposity peak (AP). D. Age at adiposity rebound (AR), E. BMI at adiposity rebound (AR).

Figure 13: Scatter plots of the effect size estimates (regression coefficients) corresponding to the 32 SNPs for each of the early growth phenotypes against those corresponding to adult BMI (in GIANT consortium study).

Figure 14. Schematic diagram showing five loci selected from combined meta-analysis of early growth phenotype meta-analysis. The figure illustrates the relationships between SNPs and early growth phenotypes and that how early growth may mediate the association between FTO, TFAP2B and GNPDA2 loci and adult BMI (Table 22). The arrows represents the association between the SNPs and phenotypes; blue and black arrows show an association with $P < 0.0001$ significance and grey arrow an association of $P<0.05$.

Table of Contents

	Page
Acknowledgements	**2**
Declaration of Originality and Copyright Declaration	**4**
Abstract	**6**
Abbreviations	**7**
List of Tables	**9**
List of Figures	**11**
Contents	
List of Publications	**16**
Overall aim, Scope and structure of thesis	**17**
1. Chapter 1: Introduction	**23**
1.1. Background	23
1.2. Growth	24
1.2.1. Stages of growth	
1.2.2. Growth modeling of longitudinal data	
1.3. Obesity	26
1.3.1. Measures of obesity	
1.3.2. Childhood obesity	
1.3.3. Adulthood obesity	
1.4. Genetic Variation	29
1.5. Genetic Epidemiology studies on obesity	30
1.5.1. Linkage studies	
1.5.2. Candidate Genes studies	
1.5.3. Genome-wide association studies (GWAS)	
1.6. Summary	33
2. Chapter 2: Review of the literature	**35**
2.1. Developmental origins of health and disease (DOHaD)	37
2.1.1. Background	
2.1.2. Thrifty phenotype hypothesis	
2.1.3. Postnatal Growth acceleration hypothesis	
2.1.4. Genetic susceptibility hypothesis	
2.1.5. Epigenetic modification	
2.1.6. Summary	
2.2. Growth patterns in early life and obesity	40

2.2.1. Intra-uterine period
2.2.2. Infancy
2.2.3. Childhood
2.2.4. Summary
2.3. Other phenotypes related to obesity 41
2.3.1. Maternal BMI
2.3.2. Summary
2.4. Genetic studies on early life phenotypes related to obesity 42
2.4.1. Background
2.4.2. Candidate-gene studies
2.4.3. Genome-wide association studies
2.5. Leukocyte Telomere length and obesity 47
2.6. Summary 47
3. Chapter 3: Objectives, hypothesis and research questions 49
3.1. Main objective
3.2. Study-specific hypothesis
3.3. Research questions for the thesis
4. Chapter 4: Methods and Materials 52
4.1. Description of Primary cohorts for analysis 52
4.1.1. Northern Finland Birth cohort 1966
4.1.1.1. Height and weight measurements in children and adolescents
4.1.1.2. Follow-up at 14 and 31 years
4.1.2. Northern Finland Birth cohort 1986
4.1.2.1. Height and weight measurements per person in children and adolescents
4.1.2.2. Follow-up at 16 years
4.2. Variable (outcome) selected from the NFBC1966 and NFBC1986 58
4.2.1. Prenatal period – Maternal and paternal
4.2.2. Postnatal period
4.3. Growth modeling: Application of Anthropometric measurements 60
4.3.1. Descriptive analysis of growth data
4.3.2. Models for infant height and weight growth in NFBC1986
4.3.3. Models for growth in body mass in NFBC1986
4.4. Genetic data for GWAS 69
4.4.1. Description of genetic data for NFBC1966) and NFBC1986
4.4.2. Description of other cohorts participating in GWAS
4.5. Data analysis strategy and statistical methods 76
4.5.1. Methods used for association analysis

4.5.2. Methods used for genetic association analysis
4.6. Software 79
4.7. Summary 80
5. Chapter 5: Results 81
5.1. Association Analysis 82
5.1.1. Association between Prenatal predictors and Postnatal Height Growth Velocity in Infancy 82
5.1.1.1. Introduction
5.1.1.2. Study-specific materials and methods
5.1.1.3. Results
5.1.1.4. Study-specific discussion
5.1.2. The effect of an obesogenic environment *in utero* and during childhood on metabolic health in adolescents 108
5.1.2.1. Introduction
5.1.2.2. Study specific material and methods
5.1.2.3. Results
5.1.2.4. Study-specific discussion
5.1.3. Association of leukocyte telomere length with adiposity-related anthropometric measurements at 31 years and change in BMI distribution from childhood to adulthood 124
5.1.3.1. Introduction
5.1.3.2. Study-specific materials and methods
5.1.3.3. Results
5.1.3.4. Study-specific discussion
5.2. Genome-wide association analysis 135
5.2.1. Genome-wide association study to find genetic variants affecting infant head circumference 135
5.2.1.1. Introduction
5.2.1.2. Study specific materials and methods
5.2.1.3. Results
5.2.1.4. Study-specific discussion
5.2.2. Association between common genetic variants and early growth phenotypes 149
5.2.2.1. Introduction
5.2.2.2. Study specific materials and methods
5.2.2.3. Results
5.2.2.4. Study-specific discussion
6. **Chapter 6: Conclusions** 173
6.1. General Conclusions 173
6.2. Methodology considerations 175

6.3. Strengths of the study 176
6.4. Limitations of the study 176
6.5. Future work 177
6.5.1. Gene-environment (G x E) interaction studies on pre-pregnancy maternal smoking and offspring birth weight (BW).
6.5.2. Association between birth and infant size measures and testosterone levels in adolescent with poly-cystic ovary syndrome (PCOS) symptoms in the NFBC1986.
6.5.3. A genome-wide association studies (GWAS) on maternal gestational weight gain to identify novel genetic variants.
6.5.4. The association between DNA methylation, early-life growth trajectories and adulthood metabolite levels.
Bibliography **187**
Appendices **203**

List of publications

In preparation

- **Das S*,** Alves C.A.[*], Taal HR[*], Sovio U, et al, Common variants at LEPR/LEPROT, FTO, TFAP2B and GNPDA2 are associated with childhood growth trajectories, linking early growth with adulthood obesity and metabolism. Manuscript to be submitted to Nature Communications, 2014.
- **Das S*,** Sovio U*, Kaakinen M, Millwood I, Laitinen J, Pouta A, Hartikainen A-L, Jarvelin M-R. Size at birth mediates association between prenatal predictors of infant height growth velocity in the Northern Finland Birth Cohort 1986 (NFBC 1986). Manuscript to be submitted to Plos One, 2014.
- Alves CA, Valcarcel B, Morin-Papunen L, **Das S**, Sebert S, Kangas JA, Soininen P, Iorio MD, Coin L, Korpela MA, Jarvelin M-R, Franks S. Metabolic profiling of polycystic ovary syndrome. Manuscript to be submitted to Plos One, 2014.
- **Das S***, Samuelsson A-M*, Rodriguez A, Kaakinen M, O'Reilly P, Poston L, Symonds M, Jarvelin M-R, Taylor P, Sebert S. Consequences of maternal overweight in pregnancy on child health and interaction with a post-natal obesogenic environment - a comparative study in the Finnish cohort and a murine model. Manuscirpt in Preparation, 2014.
- **Das S***, Pararajasingham J*, Serbert S, Rodriguz A, O'Reilly P, Price T, Jarvelin M-R. Genome-wide Interaction study on Birth Weight and Maternal Smoking, Manuscript in Preparation, 2014.

Work published

- **Das S***, Buxton J*, Rodriguez A, Sebert S, Jarvelin M-R and Blakemore A. BMI increases after early childhood and adulthood is inversely associated with mean leukocyte telomere length in the Northern Finland Birth Cohort 1966. *PloS one* 9, e99133 (2014).
- Sovio U, Kaakinen M, Tzoulaki I, **Das S**, Ruokonen A, Pouta A, Hartikainen L A, Molitor J, Jarvelin M.R. How do changes in body mass index in infancy and childhood associated with cardiometabolic profile in adulthood? Findings from the Northern Finland Birth Cohort 1966 Study. *Int J Obes* **38**, 53-59 (2014).
- Cousminer D, Berry D, Timpson N, Ang W, Thiering E, Byrne E, Huikari V, Bradfield J, **Das S**, et al. Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, childhood adiposity, timing of puberty. *Hum. Mol. Genet.* 22 (13): 2735-2747(2013).
- Morandi A, Meyre D, Lobbens S, Kleinman K, Kaakinen M, **Das S**, et al. Estimation of newborn risk of child or adolescent obesity: lessons from longitudinal birth cohorts. *PLoS one7,* e49919 (2012).
- Horikoshi M*, Yaghootkar H*, Mook-Kanamori D*, Sovio U, H, **Das S**, et al. Novel loci associated with birth weight provide biological links between intrauterine growth, postnatal growth and metabolic traits. *Nat Genet* 45, 76-82 (2013).
- Taal HR*, **Das S*,** Pourcain SB*, Thiering E*, Mook-Kanamori OD*, et al, Common genetic variants in SBNO1, and near HMGA2 and CRHR1/MAPT affect infant head circumference. *Nat Genet 44,* 532-538 (2012).
- Bradfield J, Taal R, **Das S**, et al. A genome-wide association meta-analysis reveals new childhood obesity loci. *Nat Genet **44**,* 526-531 (2012).

Overall Aim, Scope and Structure of thesis

Overall Aim of the thesis

There is convincing evidence supporting Barker's hypothesis on the 'fetal programming of adult disease' and its continuum; the Developmental origins of health and disease (DOHaD) hypothesis. As presented in this thesis, the work on Barker's hypothesis concluded that several genetic and environmental prenatal and postnatal factors may pre-program long-term health during *in utero* development. More specifically, in terms of metabolic health, it was understood that genetic, environmental and nutritional factors influence early fetal development, altering hypothalamic development and function that permanently may 'reset' to high energy efficiency to promote the rapid gain of weight in childhood. Conversely, if fetal malnutrition occurs late in gestation, when adipocyte cell division is occurring more rapidly, then accrual of body fat is inhibited by a reduced number of adipocytes.

Many complex diseases generally considered to be due to adult 'lifestyle' are now recognized to have a significant developmental origin. Developmental processes, acting in part through genetic mechanisms, can have echoes on structure and physiology of complex metabolic diseases such as obesity which can possibly be transmitted across generations. Some observations suggest that the fetal and early postnatal environment may provide a snapshot of the environment in which individuals grow and could be an amalgam for different factors affecting development of overweight and obesity in an individual. The main aims of this thesis, in brief:

1) The overall aim of this thesis was to investigate the putative role of genetic and environmental prenatal and postnatal factors associated with early growth in response to programming in *in utero* and postnatal. The work focuses on understanding, from a developmental point of view, how growth during early childhood may be associated with obesity in later life.
2) The thesis utilises novel analytic strategies and methods for growth modeling in two population-based longitudinal studies of Northern Finland to develop rapid growth phenotypes which are proxies for the growth phase. My focus has mainly been on three time periods – birth, infancy and early childhood. A more detailed comparative description of growth modeling was conducted for an initial part of the research which was presented in the MSc Dissertation by me (Appendix Note 2).

3) The thesis also aimed to understand if 'programming' during *in utero* and 'programming' during the postnatal development period exerts different risk on an individual's metabolic profile.
4) To examine the genetic variants underlying head circumference and early life growth and that how they associated with adulthood adiposity-related factors. To explore possible mechanisms underlying these associations and to understand if any of the early growth factors mediates the association between genetic variants and later health.

Scope of the thesis

The field of research based on the 'fetal development' of disease hypothesis has expanded significantly and has made major progress over the last 20 years. The studies based on this hypothesis have demonstrated a clear link between intra-uterine environment, early growth and later health outcomes. However, research has not been able to establish the connection between the effects of predictor factors, such as mother's health and environment during *in utero*, and birth size or early growth patterns.

Genome-wide studies have been very successful in uncovering several genetic variants associated with complex diseases. These studies now need to expand to explain the pathways related to these disease phenotypes. It is important that the current studies investigate not only the disease endpoints at certain ages but also trajectories of phenotypic changes to understand the longitudinal variations in the etiology of disease.

The availability of prospective population-based studies with longitudinal phenotype data has now become extremely valuable for epidemiological research. Diverse phenotypic information has been collected in the Northern Finland Birth Cohort 1966 (NFBC1966) and NFBC 1986, including repeated early growth measurements. When combined with genotype data, this gives an excellent opportunity to study the genetic background of complex traits such as obesity, however, collaboration with other studies is essential to ensure adequate statistical power to identify associations.

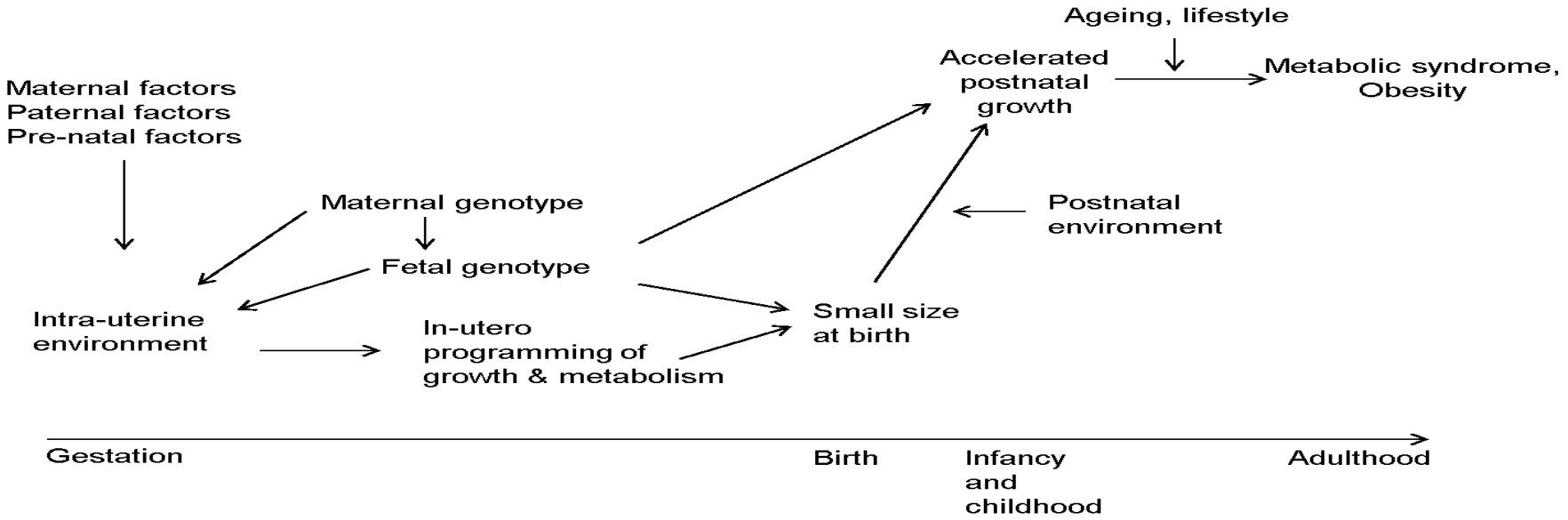
This study focuses on growth phenotypes – size-at-birth (such as birth weight, birth length and head circumference), peak height velocities (PHV) at infancy, the timing of adiposity peak (AP) and rebound (AR) in childhood. I have also investigated anthropometric and metabolic phenotypes (height, body mass index (BMI), and waist circumference, and blood pressure, lipid and glucose

measurements) in adolescence, which are influenced by growth and are also associated with the risk of obesity. **Figure 1** gives an overview of the DOHaD hypothesis pathway which the thesis addresses, highlighting the key prenatal factors which were investigated to understand the *in utero* programming and then postnatal factors which can also program the developmental stages during childhood. One way to start answering questions on genetic factors of longitudinal growth is to fit biologically motivated growth curves to longitudinal growth measurement data, then to derive interpretable and biologically meaningful growth parameters (e.g. PHV and timing of AR) from these data. These parameters can then be used as outcomes in genetic association analyses.

Results from these analyses are synthesised, interpreted and discussed in a wider context. In addition, future challenges in the field of genetic epidemiological studies, in particular in the DOHaD field, are discussed.

However, the scope of this thesis was restricted, due to research priorities, data availability, and space and time limitations. The analyses using comparisons for longitudinal growth were not included in subsequent chapters of this thesis. For easier interpretability, a concise explanation of growth models on height/weight and BMI was presented from NFBC 1986. The genome-wide studies conducted in this thesis examined common genetic variants which have an allele frequency >= 0.01. The low frequency variants were not explored as it would require new methodology that was outside the scope of this thesis. Genome studies were conducted using additive genetic models, which assumed the alleles frequency to be coded as 0, 1 and 2. This has been a standard procedure for meta-analysis in large collaborations. Other alternative approaches such as Mendelian Randomisation, gene-gene interaction and gene-environment analysis were beyond the scope of this thesis. More complex growth modeling methodologies, such as SITAR or Structural Equation Modeling (SEM), were omitted from the thesis due to time constraints.

Figure 1. An overview of the developmental programming hypothesis followed in this thesis. The figure shows pathways which were investigated in this thesis to understand the role of pre-natal genetic and environmental factors which may lead to *in utero* programming affecting smaller size at birth and accelerated childhood growth. Both of which have been associated to affect metabolic syndrome factors in later life. Furthermore, postnatal environment may also insert programming effect on accelerated postnatal growth and lead to adverse metabolic profile in adulthood.



Structure of the thesis

The thesis is divided into four main parts which correspond to `Background', `Subjects and Method', `Study-specific results ' and `Overall conclusion'.

Chapter 1 provides some background to the subject matter covered in the thesis. The general concepts of "Developmental Origins of Health and Disease (DOHaD)" hypothesis and human growth phases are described. A brief introduction of obesity, which is the central theme of the project, is also discussed. The chapters give a brief summary of genetic variations and the application of genome-wide approaches in studies of obesity, with a brief introduction to telomeres and why they need to be studied in obesity.

Chapter 2 examines the existing literature related to DOHaD, growth patterns in early life and their association with obesity, phenotypes which have been related to obesity, and genetic studies on early life phenotypes relating to obesity. In particular the literature review focuses on published studies investigating the association between prenatal, postnatal, early growth factors and childhood, adulthood obesity-related phenotypes. Additionally, current studies exploring genetic and early life factors associated with adverse adult health are also presented.

Chapter 3 presents the overall objective, hypothesis and research questions of the thesis.

Chapter 4 introduces the two main study populations; Northern Finland Birth Cohorts (NFBCs) 1966 and 1986. It also gives an overview of the statistical methods used for the epidemiological and genome-wide association analyses. It describes the growth modeling of height, weight and body mass index (BMI) during childhood, based on longitudinal growth data, used to calculate the growth phenotypes considered as outcome variables in the association analyses.

Chapter 5 presents results from association studies conducted on prenatal predictors and various outcomes at birth, during infancy, childhood. It describes an investigation into the association between prenatal predictors, size at birth and infant peak height velocity (PHV). This investigation places an emphasis on mediation by size at birth variables, on PHV in infancy in the NFBC 1986. This chapter then describes an association study between pre-pregnancy maternal body mass index (BMI) and adolescent metabolic factors. Finally describing an investigation on Leukocyte Telomere Length (LTL) and adiposity-related factors in childhood and adulthood

in NFBC1966. The chapter also presents results from genome-wide association studies (GWAS) conducted on infant head circumference (HC) and early growth phenotypes of peak height and weight velocities (PHV, PWV) during the first year of life, BMI at adiposity peak (AP) in infancy, BMI and age at adiposity rebound (AR) in childhood. NFBC (1966 and 1986) results were incorporated into meta-analyses combining several GWAS from the early growth and genetics (EGG) consortium.

Chapter 6 presents the overall summary of methodologies used, conclusions from each study performed as well as the strengths and weakness of the studies. The thesis gives a broad background by presenting studies from published literature which have investigated several prenatal, postnatal, genetic and environmental factors to understand the underlying cause of childhood and adulthood obesity. However, the literature review also shows that there is a gap which needs to be filled. Each study included in this thesis provided evidence towards the 'programming effect' in the development of obesity. This chapter also summarises further work being undertaken and the work that can be carried out in the future.

The Appendix gives additional tables and reproduces papers, manuscripts by written by myself, which are either submitted or to be submitted to journals.

Chapter 1: Introduction

This chapter gives a brief background of the epidemiological basis for the research undertaken. It gives an overview of the stages of growth, definition of obesity and genetic architecture of complex diseases. The purpose of this chapter is to provide a historical context for the research presented in this thesis.

1.1. Background

More than twenty years ago, the Clinical Epidemiologist Prof David Barker, highlighted a link[2] between geographic distribution of infant mortality and later adult mortality. His pioneering research led the way, forming the basis of the fetal origin of adult diseases (FOAD) hypothesis or Barker's hypothesis. This hypothesis states that small size at birth or in infancy is associated with an increased adult death rate from chronic diseases such as cardio-vascular disease (CVD), obesity and Type 2 diabetes (T2D). Barker's hypothesis has now been expanded to incorporate a wider framework of life course approaches relating to chronic disease epidemiology, called "The Developmental Origins of Health and Disease (DOHaD)" hypothesis. The DOHaD hypothesis states that multiple adverse *in utero* conditions programme individual health; resulting in the development of diseases in later life. It contributes significantly to the understanding of how the adverse fetal environment influences health in childhood and adulthood, including the effects of nutritional changes during gestation and infancy together with stress during pregnancy

Several epidemiological studies[3-6] have outlined that adiposity during infancy and childhood may have an intermediating role between the pre-natal environment and obesity in later life. These studies established that there are "sensitive" or "critical" periods during fetal development and in early childhood, during which physiological alterations may be linked with increased risk of childhood and adulthood obesity[7]. In the last decade, there have been numerous studies on rapid growth and obesity, although their epidemiological interpretation has been limited due to several reasons[8-10]. For example, these studies had no consensus on the quantification of growth rapidity or focused on growth measures from one time period in a cross-sectional design for measuring rapid growth, which does not take into account that there may be variations in growth patterns at different childhood stages. Similarly, few studies have considered gender differences, which is important given that the growth rate between males and females is fairly different.

With the availability of dense longitudinal data, a larger prospect for testing the DOHaD hypothesis in life-course approach has emerged. The longitudinal data benefits the development of complex statistical models to

understand growth changes during childhood and adolescence, and allows the fitting of growth models to individual growth data. These complex statistical growth models provide functional parameters that follow the biological pattern of growth, for example the rate of growth (also called growth phenotypes) and quantify infliction points during growth phases. Several studies have emphasized the significance of growth modeling and have associated growth phenotypes with the risk of metabolic diseases such as obesity[11-13] and other metabolic phenotypes later in life[14,15].

There is a vast amount of research on rapid growth, and its determinants, associated with the development of obesity in later life. These aspects are very important in order to understand the disease mechanisms as well as for preventive purposes because it has been shown that body mass tend to "track" from early childhood into adult life. However, there is little evidence about prenatal links between rapid growth in longitudinal study design and adult health. In particular, it is largely unknown how prenatal determinants affect early childhood growth, or how early childhood growth may mediate the effects of prenatal determinants on adult phenotypes.

Genes play an important role in determining the appearance and behaviour of an organism. The ultimate outcomes in an organism are mostly the combination of these genes and their interaction with environmental exposure. The availability of HapMap data[16], the Encyclopedia of DNA Elements (ENCODE) project[17] and more recent "1000 Genomes" project[18], has made it possible to associate between common forms of genetic variations, single nucleotide polymorphisms (SNPs) , and the occurrence of diseases such as obesity. However there are still a great deal of complexities to be explored, multidisciplinary research is needed to be understand the genetic and environmental factors associated with childhood growth. Certainly, longitudinal studies on early childhood growth remain very limited.

1.2. Growth

Human growth is a complex process characterised by an increase in size or mass of tissue over time, which is regulated through hormones secreted by the endocrine system. The metrics of growth can be defined as a change in any one of the anthropometric parameters. The anthropometric parameters for which measurements may be routinely recorded include head circumference, height, weight, waist/hip circumference, skinfold thickness and body composition. Growth can be an excellent indicator of nutritional and health status in individuals and is routinely measured in the clinical setting to review a child's healthy development. Any

deviation from the typical growth trajectories may indicate a particular health problem in an individual and regular follow-up could aid in early detection of the problem. Height and weight growth is generally a very regular process, although growth in weight is generally less regular than growth in height. Females are generally shorter and weigh less than males.

Monitoring growth has two purposes:

1) If there is growth retardation due to any medical, nutritional or developmental problems, routine assessments of growth patterns will ensure early identification and intervention.

2) Monitoring growth of infants and children undergoing medical care allows for the evaluation and adjustment of dietary intake or use of nutritional therapies, when required.

1.2.1 Stages of growth

Growth from the fetal period until adolescence (and adulthood) is usually separated into four different stages by differences in growth velocities and patterns: 1) intra-uterine period (very high growth velocity overall), 2) infancy (postnatal growth "spurt"), 3) childhood (decreasing growth velocity, "levelling-off" and 4) adolescence (second postnatal growth "spurt").

The *intra-uterine* growth spurt depends on nutrients plus oxygen supply for cell division; the key adaptation made by the fetus in response to a lack of supply in any of these, is a reduction in the rate of cell division in tissues. The interpretation of these trends is that relatively small babies have suffered some degree of intrauterine growth retardation (IUGR). In addition to affecting overall size, IUGR is often associated with growth asymmetries or disproportional head circumference due to the preferential allocation of limited energy resources to the growth of the brain.

Infancy is defined as the period between the second postnatal month and 2 years; it is characterised by a high growth velocity immediately after birth and rapid deceleration thereafter until about 2 years of age. This is followed by *childhood*, characterised by a period of slow deceleration of growth velocity that lasts until the onset of puberty. During childhood the spurts in growth occur in two stages; the first stage sees a slight, sudden, growth increase (peak) and the second sees a sudden growth decrease (rebound).

During puberty the *adolescent* growth spurt provides marked growth acceleration, then after a period of peak velocity, growth decelerates until it finally ceases. Adolescence is marked by an increasing growth velocity, which generally occurs earlier in girls than in boys. In terms of the magnitude of growth velocity, the growth spurt after infancy is much weaker compared to the prenatal period and infancy, further stressing the importance of the present study. This study will be able to explain some biological pathways, otherwise not possible, without extensive growth modeling.

1.2.2 Growth Modeling of Longitudinal data

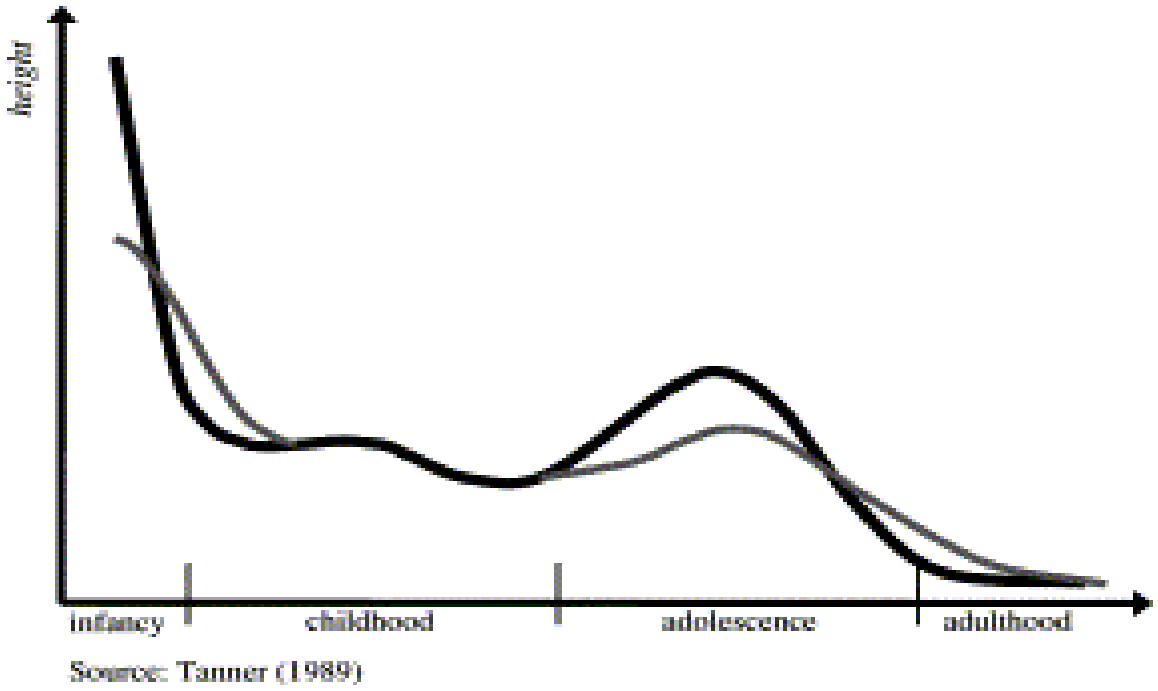
Tanner, a famous epidemiologist working on human growth, first suggested that fitting a curve to an individual's growth data is the best way to extract maximum information on an individual's health[19] (Figure 2). Indeed, epidemiologists have since attempted to find mathematical models which fit the growth measurements and thus summarise growth data. Growth models can be successful in reducing large amounts of growth data into more meaningful parameters, which can then be used to compare growth between individuals and populations. The overall aim of growth models is to fit the simplest possible model; which may provide biologically interpretable parameters. There have been several models that have been used to describe human growth; some using specific anthropometric variables over a specific age range, while others use more general statistical modeling techniques.

1.3. Obesity

Obesity is a complex multi-factorial disease, characterized by abnormal or excessive fat accumulation in adipose tissue, affecting the health of the individual. Obesity occurs when energy intake from food and drink consumption is greater than energy expenditure, resulting in the accumulation of excess fat. It arises as a result of the interplay between behavioural, environmental and genetic factors, and can reveal very different individual responses to the same nutritional intake or physical activity. Obesity has been associated with various early-onset adult cardio-vascular health complications. In last 20 years, obesity has been predicted to increase the risk of type 2 diabetes (T2D) by 44%, ischemic heart diseases by 23% and different cancers by 7-41%[20], also contributing towards substantial overall morbidity and mortality in later life[21].

Obesity has now become a major public health challenge in both developed and developing countries[22]. The World Health Organization (WHO) reported[22] that the worldwide prevalence of obesity nearly doubled between

Figure 2. The three phases of growth in height from birth to adulthood in an individual according to Tanner, 1989[19]. The black line represents the growth pattern of an individual from birth to adulthood and the grey line represents the growth variations that may occur in an individual due to nutrition changes. The figure has been adapted from Tanner, 1989[19].



1980 and 2008. In the UK, the prevalence tripled in the last 25 years, with a recent health survey showing over 30% of children and approximately 60 % of adults are either overweight or obese. This alarming rise in obesity causes various physical[23] and psychological[24] problems, increasing an individual's risk of developing a number of chronic diseases.

1.2.3 Measures of Obesity

The most common measurement to assess overweight and obesity status is body mass index (BMI), defined as weight (kg)/height (m) ^{2}. The international task force for obesity (IOTF)[25] and WHO[26] recommends BMI cut-points to classify weight status in both children and adults. Other indicators commonly used to determine aspects of obesity include skin fold thickness, waist circumference, and waist to hip ratio. The skin fold thickness refers to the measurement of subcutaneous fat located directly behind the skin and is measured by calipers, grasping a fold of skin and subcutaneous fat[27]. Skin fold thickness is mainly used to determine relative fatness and percentage body fat. The waist circumference (WC) is a measurement of the circumference of the waist with a measurement of WC in men > 94 and WC in women > 80 indicating an increased risk of health problems[28]. The waist to hip ratio (WHR) examines the fat distribution and according to guidelines from the WHO, abdominal obesity status can be identified as a WHR > 0.90 in men and > 0.85 in women[28]. More recently it has been suggested that body fat mass, measured by dual energy x-ray absorptiometry, provides a more accurate estimate of adiposity status than BMI[29].

1.2.4 Childhood Obesity

Childhood and adolescent obesity are major issues in modern societies, with an alarming rise in European countries, such as the UK. The WHO predicts that patterns of overweight/obesity in children will continue into early adulthood for over 60% of European populations[30]. The main reason for this alarming rise in childhood obesity is suggested to be the increased availability of high-fat food and reduced physical activity[31]. Efforts towards understanding contributors to the childhood obesity epidemic have become essential to both scientific and clinical communities, and more effort is being directed towards prevention strategies. It should be noted that while an internationally accepted definition (from the WHO) exists for adult obesity, there is no similar definition for childhood obesity based on BMI, as the body mass index in childhood changes with age.

Previous epidemiological studies have outlined various risk factors for childhood obesity, such as parental obesity, more than 8 hours spent watching television per day at 3 years of age, rapid growth during infancy,

height-weight z-score in childhood, large weight gain in early life, a high birth weight and short sleep duration. There is also evidence that childhood obesity is associated with the timing of puberty. Childhood obesity has been consistently found to be associated with most of the major cardio-vascular risk factors in adulthood[32].

It is well established[33] that tracking obesity patterns from childhood to adulthood may provide valuable information on overweight children, who may be at high risk of becoming overweight in early adulthood as this can be associated with adverse health problems in later life. However, it has also been observed that not all obese children became obese adults.

1.2.5 Adulthood obesity

Adult obesity has sharply increased during the 1990s and 2000s with the 2011 Health Survey in England showing that obesity, defined in BMI terms, rose between 1993 and 2011 in men from 13% to 24% and in women from 16% to 26%[34]. It is predicted that the prevalence of obesity may increase to 60% of adult men and 50% of adult women by 2050[34].

Maternal pre-pregnancy obesity and gestational weight gain is one of the top risk factors positively associated with obesity in offspring, both in childhood and in adulthood. Studies from the individuals exposed to famine *in utero* during World War II showed that mothers who suffered famine periconceptually and in the first trimester of pregnancy gave birth to babies who later had an increased risk of childhood obesity[35]. It was hypothesized by Rooij et al, 2010, that under-nutrition during the first months of gestation, when the central nervous system is formed in the fetus, generates better metabolic efficiency to compensate for the energy deficit[36]. This energy sparing system would lead to excessive fat storage when enough food is available. However, the links between the nutritional state of the mother and predisposition to obesity in children needs to be further investigated.

1.4 Genetic variation

Human DNA is a double stranded molecule that gets translated by cells to produce chains of amino acids known as genetic code. Whilst the discovery of the structure and function of DNA provided vast understanding of gene function, it did not lead to knowledge about how DNA influenced the behavior of the cells. Sequencing of the human genome in 2003[37] was therefore a major development in understanding how genetic variation relates to human health and diseases. Most complex human diseases, such as CVD and obesity, are multifactorial caused by the effect of both multiple genes (polygenic) and environmental factors. Genes are influenced to a varying

degree, by the environment (depending on the population), which is called *heritability*. The measure of the expression of a trait is relative to the environment an organism has experienced.

1.5. Genetic epidemiology studies on obesity

Obesity has been shown to result from the interplay between genetic and environment factors. Studies show that polymorphisms in various genes controlling appetite and metabolism predispose individuals to obesity under certain dietary conditions. Although genetic mutation is currently considered rare, variations in these genes may predispose individuals to common obesity. Genetic epidemiology examines many common genetic variants in different individuals to see if any variant is associated with a trait. These studies usually focus on associations between single nucleotide polymorphisms (SNPs) and traits from complex diseases, such as obesity. Genetic studies on childhood obesity started by investigating single-gene mutations in the monogenic form of obesity, which is both rare and severe. However, further studies have shown that the genetic architecture of common obesity is due to a network of gene (polygenic) and gene-gene interactions.

Studies of twins and adopted children indicate that genetic factors have a much stronger effect than environmental factors on changes in BMI of children up to the age of 18 years[38]. Longitudinal cohort studies have shown a strong correlation between parental BMI and the BMI of their offspring in both childhood and mid-adulthood[39]. Therefore, genetic epidemiological methods for the gene discovery of complex traits, such as obesity, can be divided into two broad classes: hypothesis-driven (for example, candidate gene studies) and hypothesis-free approaches (for example, genome-wide linkage studies and genome-wide association studies). Some of these are outlined below:

1.5.1 Candidate gene studies

Candidate gene studies use a hypothesis-driven approach to conduct genetic association studies focusing on associations between genetic variants (selected on known biological and physiological functions of obesity) and related traits. Candidate-gene studies include functional and positional candidates based on data derived from animal models, cellular systems, or extreme/monogenic forms of obesity. Considered together, the candidate gene studies provide suggestive evidence that multiple genes are involved in the predisposition to obesity. The discoveries from these studies have been used in physiological experiments to study the functional consequences of the identified genetic variants. However, previous candidate-gene studies were grossly underpowered and when replicated in large-scale studies using meta-analytical approaches, a definitive effect

for most of the variants could not be identified. As the results from such studies have been unreliable, there is a case for the application of the genome-wide approach to search for the genetic determinants of obesity.

1.5.2 Genome-wide association linkage studies

Genome-wide linkage studies (GWLS), typically using a 10-centimorgan (cM) (measure of genetic distance) marker density usually containing 400 microsatellite markers evenly covering the entire human genome, identifies broad intervals of several mega bases that might contain hundreds of susceptibility genes for diseases of interest. Linkage analysis has being extremely successful in identifying genes causing monogenic disorders in obesity. However when compared to common complex diseases, linkage analysis has been less powerful and its success limited. In addition the relative high cost and family data requirements make linkage analysis more restricted in practice. Linkage analysis has provided insights into strongly associated genes with large phenotypic effects identified by fine-mapping the regions around the peaks of linkage without any pre-existing knowledge of the underlining trait being studied. GWLS has identified several single-genes for the monogenic and syndromic forms of obesity, which are both rare and severe, but it had limited success in detecting polygenes. Failure to replicate associations at many loci in the studies suggested that the variants have only a small genetic effect when taken alone.

1.5.3 Genome-wide association studies

The genome-wide association studies (GWAS) approach has revolutionized the field of genetic epidemiology by screening the whole genome at a very high resolution, covering hundreds of thousands of genetic variants. GWAS offers a new approach by identifying important regulatory elements of the genome as well as the genetic variants that disrupt these. Either of these could impart a predisposition to complex disease. The GWAS approach has only been made possible due to the discovery of human genome[37] and International HapMap projects[16](a large-scale effort aimed at understanding human sequence variation). These projects[16] have enabled the development of unprecedented technology and tools to investigate the genetic basis of complex diseases. Since much of human diversity is due to single base pair variations together with variations in copy number throughout the genome, current advances in single-base extension (SBE) biochemistry and hybridization/detection of synthetic oligonucleotides now make it possible to accurately genotype and quantify allelic copy number. There is now a revolution occurring in SNP genotyping technology, with high quantity

genotyping methods allowing large volumes of SNPs to be genotyped in large cohorts of patients and controls, enabling large-scale GWAS of complex diseases.

1.5.4 Gene-environment interaction studies

It is accepted that human embryonic development is determined by both fetal genes and the intrauterine environment, which in turn depends on the maternal genes and health behaviour. Genetic and environmental factors interplay in different scenarios to cause complex diseases. This interplay may function via fetal genes to increase susceptibility to diseases or via the mother's genes causing less favourable environments for development. Different scenarios could arise, such as those based on interaction between environmental exposure and fetal or maternal genes, or interaction between fetal and maternal genes or through a combination of several scenarios. This shows that gene-environment (G x E) interaction studies are an important topic in human genetics as they can further explain the heritability of complex diseases.

G x E interaction studies are in the early stages of development but have been successful in providing evidence demonstrating that certain exposures have different effects in different populations or ethnic groups or in individuals with different genetically determined phenotypes. For example, G x E interaction studies on risk of melanoma has shown that fair skinned individuals are at higher risk than dark-skinned individuals and that there is an interaction between ultraviolet light and skin pigmentation[40].

The term 'interaction' has been very loosely used in G x E interaction studies. The biological definition of 'interaction' is the joint effect of two factors that act together in a direct physical or chemical reaction and co-participation of two or more factors in the same causal mechanism of disease development. In statistics, interaction is usually defined as a relationship between two or more independent variables modified by other variables. However, G x E studies has a few limitations. The sample size required to identify statistically significant interaction is much higher than the sample size required to identify marginal effects. The methodological requirements for G x E studies are greatly driven by the research question and thus may have in-built sample bias. A number of study designs and analytical methods have been proposed for testing G x E interactive effects, including case-control, case-only and case-parent approaches, each with its own strengths and weaknesses.

1.5.5 Telomere Length

Telomeres are non-coding functional, repeat sequences at the ends of chromosomes [5(TTAGGG) n-3]. Telomere function is to stabilize the chromosome during mitosis, prevent aberrant recombination, and protect the chromosomes from end-degrading enzymes. After each cell division, however, telomeric base pairs are lost due to the inability of DNA polymerase to fully replicate chromosome ends. This "end replication problem" would have disastrous results were genetic information to be contained in the chromosome tips. Thus, the highly repetitive telomere sequences serve to protect chromosomes from loss of essential coding sequences. The obligatory loss of telomere DNA with each cell division thus serves as a "mitotic clock" which marks the number of cell divisions that have taken place. Telomere length is heritable, and appears to differ among individuals, even at birth.

In recent times, telomere length has been associated with obesity[41]. It has been suggested that oxidative stress, associated with obesity[42], can shorten telomeres, and this can be dangerous to healthy cells. Shortening of telomere length is a known mechanism of premature cellular death and loss of longevity, however, not much is known about the role of telomere length in childhood obesity and later life factors.

1.6 Summary

In summary, obesity is a complex disease that has now become a major public health challenge globally. Both childhood as well as adult obesity is on the rise in developing as well as developed societies. Several studies have consistently indicated that obesity-mediated cardiovascular morbidity in adulthood may have its origins in childhood obesity. Therefore, research towards understanding contributors to the childhood obesity epidemic has become essential so that effort can be directed towards appropriate prevention strategies.

Based on the DOHaD hypothesis, several epidemiological studies have outlined that there are "sensitive" or "critical" periods during fetal development and during childhood, during which rapid growth may be linked to higher childhood and adulthood obesity risk factors. The studies on rapid growth and obesity, using either average measurement of growth parameters such as weight/height over the life course or cross-sectional data for measuring rapid growth have limitations in terms of epidemiological interpretation. Longitudinal studies address this key limitation by providing the dense growth data necessary to test the DOHAD hypothesis. Complex statistical models have been developed to understand growth rate in early childhood, and fitted to the dense growth data from longitudinal studies. These models provide functional parameters, such as growth velocity (growth phenotypes), which have been associated with later risk of obesity. Studies on these growth

phenotypes emphasize the importance of investigating growth velocities between childhood growth points, among other things, and their association with later metabolic phenotypes at 31 years of age.

Whilst obesity is highly heritable and arises from the interaction between genetic, environmental and behavioural factors, epidemiological research has not explored associations of growth patterns in terms of genetic and environmental factors. Recent genetic epidemiological methods for the gene discovery of complex traits, such as obesity, are either hypothesis-driven (for example, candidate gene studies) or hypothesis-free (for example, genome-wide association studies). Lately, telomere length has also been associated with obesity and it has been suggested that the oxidative stress associated with obesity can shorten telomeres, and this can be dangerous to healthy cells. However, not much is known about the role of telomere length in childhood obesity and later life factors.

Chapter 2: Review of the literature

This chapter describes the literature review undertaken on the developmental origins of health and disease (DOHaD) paradigm and several hypotheses related to it . Over the years, 'Barker's fetal origin' hypothesis has evolved and developed further by several epidemiological studies. In this literature review, I focus on published studies which form the basis for the research performed in this thesis, and give an overview of the current knowledge in the field. This literature review establishes a theoretical framework and the methodological basis to explain the importance of the work conducted in this study.

2.1. The Developmental origins of health and disease (DOHaD) approach

The DOHaD approach evolved from the 'fetal origin' hypothesis, which suggested that infant growth was associated with adult mortality[43]. Several reviews present the summary of the genesis of the developmental origin hypothesis[44], presenting a strong link between the fetus and adult diseases. Barker's hypothesis stimulated great worldwide interest in developmental plasticity, although with advance in research it was felt that a broader scope of developmental cues were required. Gillman et al, 2007[45], summarised a report on the DOHaD 'paradigm', suggesting that the 'fetal origin' hypothesis can be expanded to incorporate a multidisciplinary enterprise that can look beyond the oocyte and development into other areas of epidemiological research to answer questions.

As observed during the course of this literature review, there is an overwhelming publication on the DOHaD paradigm. In this literature review, I focus on giving an overview of the most heavily cited ideas to develop good background information for my thesis. I present studies which have been previously published showing links between prenatal factors with intra-uterine environment, rapid growth with adult adiposity factors but not all together. This chapter will give a brief background on the hypothesis which the thesis has utilised for conducting analyses.

2.1.1. Background

One of the initial epidemiological study on the 'fetal origins' of CVD was described by Kermack et al, 1934[46]. Kermack et al, summarised using demographic information from England and Wales from 1845, Scotland from 1860, and Sweden from 1751 that the death rates of adolescents and adults depended on the individual's body composition that one acquired during the first 15 years of life. This gave epidemiologists a historical perspective, associating mortality with geographic location. In 1962, Neel et al[47], first proposed the idea of the "thrifty gene hypothesis", suggesting that individuals are genetically predisposed to diabetes[48], discussed in

detail in section 2.1.2. In the same year, McCance et al, 1962, introduced the concept that nutrition in infancy may have a long-term influence on, or "program" as later coined, adiposity in rats[49]. In 1976, Ravelli et al[50], proposed a landmark article, using the Dutch Hunger Winter Study, demonstrated that association between nutritional restriction *in utero* and development of obesity in later life[50]. This was followed by a paper by Forsdahl et al, 1977[51], which reported that the incidence of adult atherosclerotic heart disease is correlated with the infant mortality rate. Lucas et al,1991, introduced the term "programming", suggesting that early-life experiences exert long-term effects on later phenotypes[52].

Against a backdrop of these epidemiological developments, Barker et al, 1993, developed the "fetal origin" hypothesis, which states that coronary heart disease originates through adaptations that the fetus makes during *in utero* nutritional deprivations[2,53]. In particular, an inverse association between birth weight (BW) and adult death rate from ischemic heart disease was identified. A possible reason for this inverse association was that the organ system was compromised by deprivation during fetal development, and therefore is unable to successfully respond to high-calorific diets during their life-time, leading to higher risk of adult disease. Several other epidemiological studies supported the association between BW and CVD. Further on, studies associated low BW with high adulthood BMI and higher risk of CVD, indicating a possible combined effect of early-life and later-life exposures[54-56]. Huxley et al, 2000, estimated that with each kg increase in BW the blood pressure ddecreases by 2-5 mm Hg[57], suggesting that impaired fetal growth may lead to raised adult blood pressure. **Table 1,** presents the historical and prospective cohorts which initiated the research on 'fetal origin' hypothesis.

These historical life-course approaches provided a good base to understand the effect of fetal life on the development of chronic diseases, however, these studies only emphasized association at two time points, one at birth and one in later life, which was not enough to infer the intrauterine and postnatal influence on CVD. Overall, it was not clear exactly which periods are important or even whether increase or decrease of rapid growth is harmful. There were also discrepancies in the definition of catch-up growth; the standard interpretation was referred to only first 6-12 months after birth, which did not cover the postnatal environment factors influencing growth. Therefore, it was suggested that more focus should be given to serial growth data to determine which periods of growth contribute most to the adverse risk of CVD. In summary, three possibility were proposed which may have an independent effect on increased CVD disease risk, first to determine rapid postnatal growth, second was to understand the factors related to fetal growth restriction and third a combined effect of the two. Several hypotheses were proposed based on these new focuses; relevant hypotheses for this thesis are discussed below.

Table 1. The key historical and prospective cohort studies used in research on the fetal developmental origins of adult disease.

Birth Cohorts	Country	Period Of Births	*n*	Reference
Historical birth cohorts				
Uppsala Academic Hospital cohort	Sweden	1915-29	14611	Leon *et al*
Caerphilly cohort	Wales	1920-35	2512	Elwood *et al*
Hertfordshire birth cohort	England	1931-9	3000	Syddall et al
Helsinki birth cohort	Finland	1934-44	15846	Eriksson *et al*
Boyd Orr cohort	UK	1918-39	4999	Martin *et al*
Dutch Hunger Winter cohort	The Netherlands	1943-7	3307	Lumey *et al*
Aberdeen children of the 1950's cohort	Scotland	1950-6	12150	Leon *et al*
Human Capital Study*	Gautemala	1962-72	2393	Grejeda *et al*
Prospective birth cohorts				
1946 National birth cohort	UK	1946	5362	Wadsworth *et al*
1958 British birth cohort	UK	1958	17416	Power & Elliott
New Delhi birth cohort	India	1969-72	8181	Bhargava *et al*
Vellore birth cohort	India	1969-73	10670	Antonisamy *et al*
1970 British birth cohort	UK	1970	16571	Elliott & Shepherd
Pelotas 1982 birth cohort	Brazil	1982	5914	Victoria & Barros
Cebu Longitudinal Health Nutrition Survey	The Phillipines	1983-4	3080	Tudor-Locke
Birth to twenty birth cohort	South Africa	1990	3273	Richter *et al*
Avon Longitudinal Study of Parents and Children	England	1991-2	Approx 14000	Golding *et al*
Pelotas 1993 birth cohort	Brazil	1993	5249	Victoria *et al*

*Not all those in the cohort were recruited at birth.

2.1.2. The 'thrifty' gene and phenotype hypothesis

Neel et al, 1967 was the first to suggest the 'thrifty' gene hypothesis, explaining that a fetus stores fat when there is sufficient availability of nutrients and allows re-utilisation when there is starvation[48]. Adding to this, twin studies from 1980s suggesting the role of genetics in T2D, with monozygotic twins showing 100% concordance[58]. Hales and Barker, further added to the idea called 'thrifty' phenotype, which emphasises the importance of epidemiological association between poor fetal nutrition and infant growth and the subsequent development of adverse metabolic health through the mechanisms of developmental plasticity[59]. It was suggested that poor fetal nutrition produced permanent changes in glucose-insulin metabolism, which when combined with the effects of obesity, ageing and physical inactivity, are the most important factors in determining an adverse metabolic profile in adult life. A recent study by Jones et al, 2009, show that although the underlying mechanisms remain unexplained, the observations supporting the 'thrifty' phenotype hypothesis can be reproduced in both animal and human models [60]. However, there are critical differences between the 'thrifty genotype' and 'thrifty phenotype' hypotheses. The 'thrifty' genotype hypothesis suggests that genes react differently in response to its environmental stimuli while the 'thrifty' phenotype hypothesis suggests that genes undergo adaptation due to continued environmental changes.

2.1.3. Postnatal growth acceleration hypothesis

The concept of fetal programming suggests the idea that there are critical periods during fetal development, where a 'stimulus or insults' can have long-term effects on the health of an organism. There has been much interest in the so called critical window during which 'programming' effects might be exerted[61]. Experimental studies indicate that infants who were fed with growth-promoting enriched diets, had greater risk of developing an adverse metabolic profile[62]. Further research on preterm and full-term infants suggested that growth acceleration in infants is associated with later insulin resistance and blood pressure[63-65]. These findings lead to a proposal by Singhal et al, 2004, of the 'postnatal growth acceleration hypothesis', states that an adverse long-term effects of early programming is cardio-vascular disease[63]. Singhal explains that growth acceleration is a continuous process from *in utero* to adolescence, and restriction during fetal growth can be relative to genetic growth potential. A fetus which undergoes growth restrictions during *in utero* may have postnatal growth acceleration by a nutrient enriched diet which can have adverse long-term effects. This hypothesis was supported by consistent findings from animal data showing that early postnatal nutrition and faster growth led to

higher propensity of metabolic syndrome[63]. Similarly, comprehensive reviews on studies in humans also ddemonstrated support for this hypothesis[63,66-68].

2.1.4. Genetic susceptibility hypothesis

The genetic susceptibility hypothesis is a recent hypothesis based on underlying genetic links with birth size and metabolic diseases in adulthood. It uses the developmental plasticity concept, proposing that genetic changes occurring during critical time windows of development in the fetus can lead to irreversible morphological and functional changes and higher susceptibility to disease later in life. The 'fetal insulin hypothesis' states that paternal insulin resistance results in impaired insulin-mediated growth in the fetus as well as insulin resistance in adult life suggesting it may be genetically determined[69]. This hypothesis has been supported by studies on rare monogenic variants that were associated with both low BW and altered insulin secretion or resistance later in life[70,71]. These findings are interesting as the factors for parental diabetes do not directly influence the intra-uterine environment but there is some evidence from life-course studies that after the birth of offspring there is an inverse association between offspring weight and maternal insulin resistance[72] and diabetes[70]. However, the evidence supporting this hypothesis remains weak. For example, twin studies examining this hypothesis have been inconclusive overall and population-based genetic association studies conducted before the genome-wide era have produced conflicting results[64]. Recent studies have also indicated that at least some of the association between the BW of individuals and their later risk of diabetes may be genetic, and therefore may not susceptible to modification by interventions that influence intra-uterine development[73]. Furthermore, these inferences are now increasing; pointing towards the epigenetic inheritance of disease risk.

2.1.5. Epigenetic modification

In recent years, human and animal studies indicate that disturbances during critical periods of prenatal and postnatal development, due to nutritional and other environmental stimuli influence developmental pathways, which induces permanent changes in disease susceptibility[74]. Currently genetic and environmental factors that influence disease susceptibility, due to disruptions of energy homeostasis and metabolic regulation, are explored. Researchers are now increasingly interested in exploring the epigenome, which may be acting as the interface between gene-environment interactions and changes in gene expression. There is an increasing body of evidence suggesting that the brief environment influences during early-life development can cause permanent changes in epigenetic modification leading to induction of human diseases [75].

2.2. Growth patterns in early life and obesity

Over the last decade, life-course epidemiological research has been focused on identifying determinants of growth patterns and CVD, such as, obesity and T2D. Previous studies suggest the concept of 'critical' or 'sensitive' periods of growth during early life, which can alter functional changes in an individual's health. In this review, I collect evidence from studies conducted on growth patterns and later outcomes into four growth periods 1) the intra-uterine period, 2) infancy, 3) mid-childhood and 4) adolescence. However, growth patterns during adolescence are not discussed in this review due to time restrictions and will be at looked in the future.

2.2.1 Intra-uterine period

The structure and timing of the critical periods are considered important factors in predicting long-term health. Observations from previous epidemiological studies show that growth restriction during fetal growth can alter infant growth and have been shown to have different metabolic consequences in adult life[76]. Important studies indicating links between intra-uterine periods has been discussed above under the 'fetal origin' hypothesis.

2.2.2 Infancy

Infancy has been characterised as the postnatal period during which rapid growth occurs. It commences after the first neonatal weeks (first 2 weeks). Various definitions of infancy have been proposed in studies using infant data. For consistency of results, this literature review uses the period between birth and 2 years as the definition for 'infancy'. Several studies refer to postnatal weight gain or height gain as 'catch-up growth'; however, in this thesis 'catch-up' has been used in context of an individual's growth in height or weight to reach its genetic potential. Prader et al, 1963, was the first to introduce the term, "catch-up", in infants who were exposed to stress during in-uterine growth, leading to deviations in growth trajectory[77]. Further studies confirmed this link between infancy rapid growth and intra-uterine constraint[78-81]. Studies on appropriate for gestational age (AGA) children in both developed and developing countries, supported associations between rapid infant growths with metabolic diseases[82]. Other research have highlighted links between AGA infant exhibiting rapid growth and adult hypertension [57,83], coronary heart disease [84], and T2D or noninsulin- dependent diabetes (NIDDM) [85]. An overall review on size-at-birth and postnatal rapid growth associates higher blood pressures in low birth weight (LBW) and high infant growth[57]. In summary, all evidence indicated that infancy forms an important critical

period during which a combination of low BW and high growth centile can be indicative of a greater risk for future obesity and perhaps of other associated morbidities.

2.2.3 Childhood

Childhood is defined as the period of slow deceleration of growth velocity that lasts until onset of puberty. There have been several studies conducted on growth during childhood and later health factors. However, for this thesis, I was only interested in exploring the role of adiposity peak (AP), the time when BMI reaches its peak, and adiposity rebound (AR), the time when BMI reaches its nadir. Both AP and AR have been suggested to have associations with adiposity in later life. Rolland-Cachera et al, 1984, was first to describe the term "adiposity rebound (AR)", suggesting that there is an association between timing during which the BMI reaches its nadir (AR) and adiposity in adolescence and young adulthood [86]. This inverse relationship between the timing of AR and adiposity in later life has been confirmed in a number of other longitudinal studies[87-89], which remains even after adjusting for parental obesity [3,90]. Some studies also show links between faster infant height with AR, suggesting that obese children tend to be taller as well as fatter and heavier[91-93]. Thus, whilst there is clear evidence that early childhood AR predicts later obesity, it is not clear that this association is entirely independent of parental and early-life BMI, which are themselves related to the risk of obesity. At this point the evidence does not clearly suggest if adiposity rebound can be considered as a critical period for the development of obesity.

2.3 Other phenotypes related to obesity

2.3.2 Maternal BMI

Maternal BMI has been reported as the most significant marker for offspring obesity by several epidemiological studies. Studies have shown maternal obesity affects both mother and offspring, which in the offspring, the implications could be either intermediate or long-term, such as congenital anomalies, small for gestational age, fetal overgrowth, preterm birth and neonatal mortality [94-96]. It is now widely acknowledged that maternal BMI has a significant impact on infant body weight, although the mechanisms underlying the biological pathways are still unclear. Reviews on various epidemiological studies have shown that maternal body composition is strongly associated with visceral adipose tissue in offspring, which is associated with insulin resistance, dyslipidemia, systemic inflammation, diabetes, hypertension, myocardial infarction and all-cause mortality.

Children of obese mothers are at increased risk of developing metabolic syndrome regardless of birth weight [97]. It has been shown that maternal obesity directly links to higher metabolic syndrome factors, such as non-fasting glucose levels and insulin resistance and leptin concentration, in young adults (particularly male offspring)[98-101]. Data from animal studies indicates that maternal obesity is related to increased appetite in the offspring. Whilst it could be predicted that the increased appetite may be the causal effect of obesity in the later life of the offspring, the pathway for the increased birth weight due maternal obesity still cannot be explained.

2.4 Genetic Studies on early life phenotypes related to obesity

2.4.1 Background

The first evidence linking genetics to common obesity came from Feinleib et al, 1977, indicating a possibility that the observed familial aggregation for obesity could be genetic rather than environment [102]. Subsequently, studies from family, twins and adopted children showed that there was a 40-70% inter-individual variation in the BMI, particularly fat mass distribution, within the same population that shares the same environment which had a heritability value[103-107], although the heritability estimation varied within different population according to methodology and sample size.

With increasing evidence of the genetic heritability of the obesity trait, many researchers started to study the genetics of obesity as a polygenic association. In this review, I focus on two genetic association approaches which have been used across the field of complex human genetics; the candidate-gene and genome-wide association studies.

2.4.2 Candidate gene studies

Table 2 presents an update on all the candidate genes which have been shown to be associated with obesity-traits in childhood. In 2006, an update from the human obesity gene map by Rankinen et al, 2006, reported 127 candidate genes were associated with obesity-related traits, of which only eight of the candidate genes were identified in children[108]. Furthermore, Walley et al in 2004 and 2009, also reviewed the published candidate-genes associations in children[109,110].

Although, results on candidate genes studies in children has been equivocal as most of the studies studying obesity-related outcomes in children focused on replicating candidate genes in adult studies. For example, two

Table 2: Recent candidate genes identified for association with obesity in children[1].

Chromosome	Location	Regional Gene(s)	Gene name	SNP type	phenotype measured	References	year
1	1p31	*LEPR*	Leptin receptor	Coding	BMI, fat mass, overweight	Clement[111]	1998
4	4q31	*NPY2R*	Neuropeptide Y receptor Y2	Coding	Obesity	Siddiq[112]	2007
5	5q15	*PCSK1*	proprotein convertase subtilisin/kexin type 1	Coding	Obesity	Benzinou[113]	2008
6	6q22	*ENPP1*	ectonucleotide pyrophosphatase/phosphodiesterase 1	Coding	Obesity	Meyre[114]	2005
6	6q14	*CNR1*	cannabinoid receptor 1 (brain)	Coding	Obesity	Benzinou[113]	2008
7	7q22.3	*NAMPT*	nicotinamide phosphoridbosyltransferease	Coding	Severe Obesity	Blakemore[115]	2009
8	8p12	*ADRB3*	Adrenoceptor beta 3	Coding	BMI, body weight, WHR, Fat mass	Farooqi [116]	2003
18	18q22	*MC4R*	Melanocortin 4 receptor	Coding	BMI, body fat, fat mass, WHR	Hinney[117]	2003
16	16q12.2	*FTO*	Fat mass and obesity associated	Coding	Early on-set obesity	Meyre[118]	2009
16	16q22-q23	*MAF*	v-maf avian musculoaponeurotic fibrosarcoma oncogene	Coding	Early on-set obesity	Meyre[118]	2009
10	10p12	*PTER*	Phosphotriesterase related	Coding	Early on-set obesity	Meyre[119]	2009
18	18q11.2	*NPC1*	Niemann-Pick disease, type C1	Coding	Early on-set obesity	Meyre[119]	2009

studies - one by Haworth et al, 2008, found no evidence of 10 candidate genes at various ages in children while Podolsky et al, 2007, reported multiple associations in 11 candidate genes from adult studies[120,121]. Therefore, it was acknowledged that more comprehensive studies that examined the association of all common variations in a gene were required[1].

2.4.3 Genome-wide association studies (GWAS)

The foundation of understanding the role of genes involved in obesity was started by Zhang et al, 1994 and co-workers. They cloned and characterized the *ob* genes in *ob/ob* and *db/db* strains of mice highlighting the role of leptin in control of food intake[122]. The first GWAS was performed on 86, 604 SNPs were analysed in 694 participants from family studies, and one SNP rs7566605 near *INSIG2* was suggested to associated with obesity, which was validated in the independent replication stage[123].

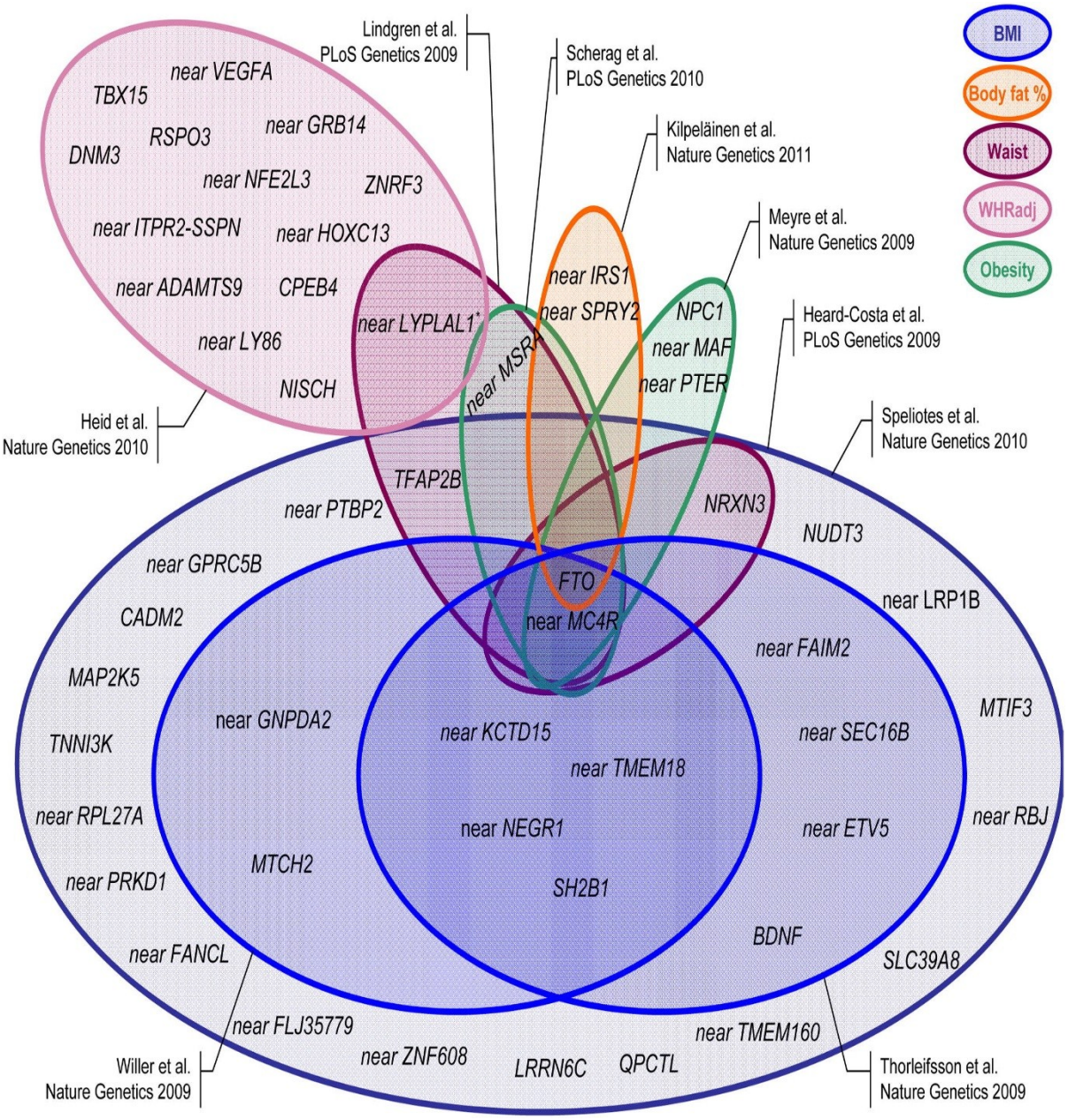
The true GWAS era was introduced in 2007, and so far four waves of large-scale high-density GWAS for BMI, a marker of overall adiposity, have been performed, identifying 32 loci that reached genome-wide significance. Most of these have been studied on white Europeans and adults only. The first obesity GWAS wave resulted in the suggestion of four susceptibility loci. *FTO* was originally highlighted in a GWAS of type 2 diabetes[124], however , adjustment for BMI revealed that the association was mediated through obesity in two different studies[125,126]. Furthermore, *FTO* locus has also been found to be associated with obesity-related traits in white European children and adolescent[118,127-129], and other ethnic decent[130-132]. In the second GWAS wave, the GIANT (Genetic investigation of Anthropometric Traits) consortium performed meta-analysis of ~17000 Caucasian individuals and identified variants in or near *MC4R* associating with measures of obesity and waist circumference[133,134].

The third wave included studies identifying variants in or near *TMEM18*, *SH2B1*, *KCTD15*, *NEGR1*, *GNPDA2*, *MTCH2* , *BDNF*, *SEC16B*, *FAIM2*, *ETV5* and *MAF*[119,135,136] genome-wide significantly associated with BMI. Another four novel loci were identified in a parallel GWAS strategy, focused on measures of fat distribution using waist circumference and WHR adjusted for BMI: *LYPLAL1*, *TFAP2B*, *MSRA* and *NRXN3* [137,138]. The fourth obesity GWAS was dominated by two meta-analysis by GIANT on BMI [139] and WHR[140]. These identified 18 and 13 novel loci, shown **Figure 3**. Overall, WHR variants had a stronger association in women than in men, confirming the sex difference in distribution of fat. Further studies identified three new loci for

extreme obesity in children and adolescents, *SDCCAG8* and *TNKS*[141], one loci *KCNMA1* found in an adult population[142]. In total, 43 loci were proposed which may predispose overall adiposity and 18 loci to visceral fat accumulation[143]. Of these, 32 BMI and 14 waist/WHR variants were genome-wide significant, and one variant (*MAF*) associating with morbid obesity[143].

In recent years, after successful identification of adult adiposity variants, the focus has now shifted to identify common childhood obesity, which is not widely studied. In one of the latest effort, Bradfield et al, 2010, performed a GWAS meta-analysis of totally 5530 cases and 8318 controls, using age- and gender-matched measures of BMI, identifying two loci, *OLFM4* and *HOXB5*, in common childhood obesity[144]. However, in understanding common variants of childhood obesity, studying the extremes of the BMI distribution might be just one way to further unravel obesity genetics. More functional studies are needed to fully characterise the functions of genes in relation to childhood obesity. This thesis fills the gap of understanding on genetic variants relating to early life 'programming' and provides further evidence towards genetic backgrounds of complex diseases.

Figure 3: Showing the waves of GWAS on adult obesity with different obesity related phenotypes.



2.5 Leukocyte telomere length (LTL) and obesity

Leukocyte Telomere length (LTL) has been suggested as a biomarker of human aging. LTL is supposed to be associated cellular senescence which has been predicted to be associated with onset and development of cardio-vascular diseases (CVD) such as atherosclerosis [145]. Although conflicting results had been published on LTL, two researchers clarified with twin same-sex models that co-twins with shorter LTL were more likely to die first [146 147] supporting the role of LTL as a biomarker. Further studies supported association between telomere lengths, independently of age, and cardio-vascular diseases[148,149]. Although shortening of telomere length is not the cause of the onset of a CVD, it has been shown that there is a significant relationship between telomere length and leukocyte and vascular tissues that may affect the tissue related to CVD[150,151].

Shorter telomere in peripheral blood has also been suggested to associate with obesity, and oxidative stress and inflammation could be the underlying mechanisms for this association[152-156]. This has been supported from longitudinal studies as well, reflecting that weight loss is associated with longer telomere length and rectal mucosa in obese men [157]. In children, initial small-scale studies of telomere length and childhood obesity presented contradictory results in comparisons to adults[158,159]. Lee et al, 2011, showed an inverse association between adiposity and telomere length, however, they also stressed on utilizing longitudinal studies to establish a causal association between early life adiposity and biological aging[160]. Later, Sidorov et al, 2008, using models for weight-gain charts from birth to 20 years showed that LTL shortening happened in an age-dependent manner[161]. More recently, Buxton et al, 2011, demonstrated that obese girls and boys have significantly shorter leukocyte telomere compared to their non-obese counterparts, emphasizing the adverse risk of childhood obesity on adulthood health.

2.6 Summary

The concept of “critical” or “sensitive” periods of growth, during which the antecedents of chronic degenerative diseases may be initiated, has been investigated in hundreds of publications over the last decade. The approach that has emerged within this review combines classic axiological and epidemiological methods within a life-span perspective that identifies accumulating and interacting risks that manifest from prenatal life onwards.

The current study proposes to form the basis of a fuller understanding of patterns of growth from birth to adolescence and the identification of factors that exert their influence on obesity in later life. The main goal is to bridge the gap by developing statistical models to study human growth from birth until adolescence and integrate both genetic and environmental factors that underlie growth trajectories leading to obesity.

In summary, the above literature provides ample evidence that exposures in genetic and environmental factors can have long-term effects on the development of diseases such as obesity in the future. However, only few investigators examined the critical interplay between genes, environment and early pre- and postnatal events in the development of obesity. Even fewer studies have utilized human only data to model this effect of genetic and environmental factors together, as most of these have been done on animals. These studies emphasize the importance of identifying factors within the pre- and postnatal period that promote and those that inhibit the development of obesity.

However, recent work on telomere attrition provides a new area of investigation that may eventually delineate causal connections between early growth and later disease.

To resolve the effects of rapid growth and obesity, dense growth measurements throughout the postnatal life are needed, as well as complex statistical models to take into account the dependency between consecutive growth measurements in each individual [64]. Sovio et al, 2009, demonstrated this through their study by deriving peak height velocity from parametric growth curves fitted to longitudinal height growth data to test their association with known genetic height variants[12]. Recent work on infant growth velocities and their association with later risk of obesity [14,61] has shown that determinants of height and weight growth during these early life periods need to be studied. A recent example by Tzoulaki et al, 2010, has shown the importance of studying postnatal infant growth and associating it with later metabolic profile at 31 years of age [8]. Such studies form the basis for this thesis to further add to the understanding of various prenatal and postnatal factors, which may influence postnatal infant growth, and may have an indirect effect on later health.

Chapter 3: Objective, hypotheses and research questions

The overall aim of this work was to gain new insights into the genetic and environmental factors leading to obesity risk. These insights may potentially help in developing new means of effective prevention and treatment of obesity in the future. The focus of the thesis was to understand the association between prenatal predictors, size-at-birth, early growth factors and their genetic determinants.

3.1 Main objective

1) To understand the role of size-at-birth in the association, testing mediation by size-at-birth, between prenatal predictors and peak height velocity (PHV).

2) To identify the role of the obesogenic environment *in utero* and during childhood on metabolic health of the offspring. It was also to distinguish the causal effect of '*in utero* programming' by pre-pregnancy maternal weight, and 'postnatal programming' by obesogenic environment during childhood and adolescence.

3) To examine leukocyte telomere length (LTL) at 31 years and its associations with childhood and adulthood adiposity-related measurements.

4) To identify the role of genetic variants on head circumference (HC) in a GWAS design, and to check if infant head circumference mediates the association between genetic variants and adult height.

5) To identify the role of genetic variants for early adiposity phenotypes in GWAS design, and if early adiposity phenotypes mediates the association between genetic variants and adult BMI.

3.2 Study-specific hypotheses

The hypotheses tested in this thesis were:

1) Several prenatal predictors associate with size at birth and infant height growth. Also, the size at birth mediates the association between prenatal predictors and infant height growth.

2) Maternal weight exerts a 'programming' effect on fetal growth, size-at-birth as its marker, and later metabolic phenotypes in adolescents. The postnatal obesogenic environment during childhood adds to this adverse risk.

3) Leukocyte Telomere length (LTL), a biomarker for age-related diseases, is associated with childhood and adulthood adiposity related measures. Longitudinal change in BMI over childhood to adulthood may have the effect of shortening telomere length.

4) The common genetic variants associated with infant HC variations. These variants result in smaller head size which may be associated with adulthood stature and related risk factors.

5) Common genetic variants influence the variation in early growth phenotypes, such as peak weight and height, age/BMI at adiposity peak (AP) and age/BMI at adiposity rebound (AR). The early growth phenotypes also mediate the association between these genetic variants and adulthood BMI, leading to risk of diseases such as obesity.

3.3 Research questions for the thesis

1) What are the maternal, paternal and environmental determinants of early growth velocities? What is the evidence for an association between prenatal predictors, size at birth and infant growth? Is infant growth velocity associated with later metabolic health in adults?

2) Are these associations with prenatal predictors a reflection of the intra-uterine environment affecting size at birth? More specifically, does size-at-birth act as mediator between the association of prenatal predictors and infant growth phenotypes?

3) What is the role of the postnatal environment in generating the association between infant growth and subsequent adult health?

4) Is there a difference between adolescent metabolic phenotypes born to overweight/obese mothers during pre-pregnancy compared to adolescents of overweight/obese mothers after child's birth? Is there is a programming effect or postnatal environmental effect on the offspring's overweight/obese mothers?

5) What is the association between telomere length (a biomarker for age-related diseases) and childhood or adulthood adiposity related measures? Is there any evidence that change in BMI over childhood to adulthood may have an effect on the shortening of telomere length?

6) What are the genetic factors associated with infant head circumference? Do these genetic variants also associate with adulthood phenotypes such as height, brain size? Is there a mediating link between infant head circumference and adulthood phenotypes?

7) What are the genetic factors associated with early growth phenotypes? Do these genetic variants also associate with adult adiposity phenotypes, such as BMI? What are the other disease phenotypes associated with these genetic variants? Is there an inter-link between these disease phenotypes and obesity? What is the correlation between published adult BMI-associating SNPs and early growth phenotypes? Do these early growth phenotypes mediate the association between genetic variants and adult adiposity phenotypes such as BMI?

Chapter 4: Methods and Materials

This chapter introduces the study populations and statistical methods used for the epidemiological and genome-wide association analyses in Chapter 5. Section 4.1.1 provides information on the two primary study populations used in this thesis, the Northern Finland Birth Cohort 1966 and 1986 (NFBC1966 and NFBC1986). Section 4.1.2 describes the growth modeling of height, weight and body mass index (BMI) during childhood, based on longitudinal growth data, used to calculate the growth phenotypes considered as outcome variables in the association analyses in Chapter 5. Section 4.1.3 provides details of the other phenotypes used in the analyses in Chapter 5 on the NFBC1966 and NFBC1986 data, and confounding variables. Section 4.4 describes the statistical analysis conducted on association analysis and preliminary stages of genome-wide association studies (GWAS).

4.1 Description of primary cohorts for analysis

The core data used in this thesis is comprised mainly of individuals born in Northern Finland in the 1960s and 1980s that form part of the Northern Finland Birth Cohort (NFBC) studies. The NFBC study is an epidemiological and longitudinal research program initiated by Professor Paula Rantakallio, University of Oulu, in 1965. The aim of the study was to investigate biological, social and behavioral risk factors for preterm birth, low birth weight and determinants of later development, diseases and disorders. The present data derives from two cohorts of mothers who lived in Oulu and Lapland, the northernmost provinces of Finland, with an expected date of delivery between 1st January to 31st December 1966, known as NFBC1966, and between July 1st 1985 and June 30th 1986 known as NFBC1986. In both studies, mothers and children have been followed up; since mothers enrolled at their first antenatal clinic visit (in NFBC1966 on average on the 16th gestational week, in NFBC1986 from the 10th gestational week onwards and on average on the 12 the gestational week)[63,66,162]. The NFBCs have one of the world's most dense early growth data collections, with measurements from birth until adulthood. These cohorts include vast amount of data related to subjects' living circumstances and well-being, diseases, genetic, metabolic and other biological measures from early pregnancy[66,163]. The Ethics committees of the University of Oulu and Northern Osthrobothnia Hospital District have approved the study in accordance with the Declaration of Helsinki.

4.1.1 Northern Finland Birth Cohort 1966 (NFBC 1966)

The Northern Finland Birth Cohort 1966 (NFBC1966) was recruited through maternity health centres and data was collected from women living, during pregnancy, in Finland's two northernmost provinces, Oulu and Lapland, with expected deliveries between the 1st of January to the 31st of December 1966 (n=12,055 mothers). A high proportion of women (approximately 80%) first attended the clinics by the 16th gestational week; from this point onwards the measurements related to the course of pregnancy are available (by midwives). The questionnaires on baseline socio-demographic, health behavioural (e.g. smoking) and other background information were administered between the 24th and 28th gestational weeks. Data on the course of the pregnancies and birth were obtained from maternity cards filled in and updated by midwives at each antenatal visit and from hospital records.

In NFBC1966, a total of 12,231 babies were born, of which 12,058 alive and 173 stillborn. The study population has a high coverage of 96.3% of all eligible births in the area during the catchment period. All cohort members are Finns (white Caucasians), and less than 1% of these are Gypsies or Lapps. Birth outcomes were collected at delivery by trained medical staff and entered in their medical records. This data was immediately transferred to the study specific forms and returned to the research centre in Oulu. The children were then followed up, using questionnaires, clinical examinations, child welfare cards from birth to ages 1 and 14 and clinical examination at 31 years, covering information on growth, development, health, measures of function (such as lung function, physical performance, blood pressure), lifestyle and socio-economic indicators[164]. The written informed consents were requested at the age of 31 for the follow-up in 1997, where they consented to the use of current and previously collected data for scientific research.

The NFBC1966 also provides rich sources of data from blood samples and DNA of its 5,753 subjects, which is described in more detail, together with description of the genetic data of the NFBC1986, in section 4.3. Data from the NFBC1966 cohort was analysed throughout the current work and described in Chapter 5.

4.1.1.1 Height and weight measurements in children and adolescents

Data on the height and weight growth of each child was collected from records obtained from communal health clinics and using questionnaires at the age of one year when nurses provided the information on child growth and development as well as at the age of 14 years. On average, 7 height measurements were obtained per person from birth until age 2 years in NFBC1966 and 15 measurements per person from 2 years until adulthood (~ 16 years). This growth data was not available for all individuals in the NFBC data base (although available on paper files), as growth measurements was entered for those with DNA available at 31 years (maximum N=5,753) because of plans for genetic studies of growth and development. The copies of original child health clinic records were requested from the health care centres, institutions and archives , but around one quarter of the records were missing over the years, with the final number of records for those with DNA samples amounting to N=4,311. The number of measurements of height and weight varied between the individuals which were taken into account in the analyses. Growth curves were fitted to derive clinically meaningful growth phenotypes, such as growth velocities for the genetic and other association analyses.

4.1.1.2 Follow-up at 14 and 31 years

All individuals alive at 14 years were contacted by postal questionnaire to collect information on their growth and health, health behaviour (smoking, drinking, and physical activity), school performance and family's social circumstances. The target population, all alive, was 11,637. Of them 11,541 adolescents were traced with their recorded address and questionnaires were sent. If the adolescent did not respond, then a questionnaire (N=754) was sent to the parents of the adolescents. To get all possible information, a similar questionnaire was sent to the regional school offices and to the school nurse if the children or their parents did not respond. After these efforts the final response rate was 97%.

In 1997, individuals alive at 31 years of age; and that could be traced, were invited for follow-up by a postal questionnaire on health, lifestyle and socio-economic status (N=11,541), and all individuals who were resident in Northern Finland or had moved to the capital city Helsinki, area (N=8,463), were invited to a clinical examination. The response rate of the postal questionnaire was 75% (N=8,600) and for the clinical examination 71% (N=6,007). Trained nurses performed the clinical examination, collecting anthropometric and other clinical data as well as samples for metabolic

phenotypes including, height, body mass index (BMI), waist-circumference (WC), high density lipid (HDL), cholesterol, triglyceride, glucose, systolic blood pressure (SBP) and diastolic blood pressure (DBP), as described below.

1) *Anthropometric measures at 31 years*

BMI was calculated from weight and height measurements, using the formula of weight (kg)/height (m)[2165]. WC was measured at the level midway between the lowest rib margin and the iliac crest. SBP and DBP were measured by trained nurses using a mercury sphygmomanometer following a standardized procedure and on-going quality control. Two readings of SBP and DBP were recorded and the average of the two measurements was used for the association analyses[66].

2*) Blood sampling*

The blood samples were taken after overnight fasting and drawn between 8.00-11.00 AM and stored at -70°C until analysed. The glucose analysis was performed by radioimmunoassay with commercial reagents (Pharmacia Diagnostics, Uppsala, Sweden) within 7 days. The other samples were analysed within 24 hours in Oulu University Hospital laboratory by using on-going internal/external quality control. Enzymatic assays of fasting serum glucose, insulin, triglycerides (TG), total cholesterol (TC), and high density lipoprotein cholesterol (HDL-C)were carried out using Hitachi 911 automatic analyser and commercial reagents (Boerhringer Mannheim, Mannheim, Germany) in the accredited laboratory of Oulu University Hospital (Oulu, Finland).

The fasting plasma concentration of low density lipoprotein cholesterol (LDL-C) was measured by the Friedewald equation: [LDL-C] = [Total C] - [HDL-C] - ([TG]/2.2)[166]. Insulin resistance was assessed using the homeostasis model assessment (HOMA-IR) and calculated using the following formula: HOMA-IR (mmol/L x μU/ml) = fasting glucose (mmol/L) x fasting insulin (μU/ml)/22.5. Pancreatic ß cell function was determined using the homeostasis model assessment of β-cells (HOMA-B) and calculated by the following formula: 20 × fasting insulin (μIU/ml)/fasting glucose (mmol/ml) − 3.5 (http://www.dtu.ox.ac.uk/homacalculator/).

4.1.2 Northern Finland Birth Cohort 1986 (NFBC1986)

The younger cohort, Northern Finland Birth Cohort 1986 (NFBC1986), was also recruited from the same region, Oulu and Lapland, with an expected date of delivery between the 1st July 1985 and the 30th June 1986. Information on health and socio-economic status was collected through questionnaires handed out at the first antenatal visit, on average in the 12th gestational week and returned by the 24th gestational week if still pregnant. Overall, data collection in this younger cohort followed that of the twenty years older NFBC1966 with some modifications and availability of new technology such as ultrasound examinations during pregnancy.

The recruited women had 9,362 deliveries, and information was collected through maternity health centers with 99% coverage of births in the area[65]. In total, 9,479 children were born of which 9,432 were live-born and 47 still born. Obstetric and birth data was obtained from hospital records and maternity cards were filled in and updated by midwives at each antenatal visit. Data concerning maternal medical characteristics, including pre-pregnancy body mass index (BMI; kg/m2), were collected via antenatal records at the time of inclusion and were either self-reported or measured when mothers attended the clinics. Birth outcomes were collected at delivery by trained medical staff, entered into the medical records and into the study-specific forms which were returned to the research center in Oulu. The cohort was followed-up by postal questionnaires, hospital records and register data at the ages of 7 and 8 (questionnaires to the parents and teachers) as well as the ages of 15 and 16 (questionnaires to the parent and adolescents and a clinical examination). At the ages of 15 and 16, the cohort members and their parents gave written informed consent. The University of Oulu Ethics Committee and the Ethical Committee of Northern Ostrobothnia Hospital District have approved the study.

4.1.2.1 Height and weight measurements per person in children and adolescents

The height and weight growth data was entered from scanning or from the paper copies of original records obtained from health clinics. The data in the original records was filled in by the study nurse, following procedure described in the protocol. On average there were 11 height measurements and 13 weight measurements in infancy during first year of life and 9 measurements from childhood to adolescent. After combining all the data, those with missing values, double entries and unclear values were set to be missing. Similar to NFBC1966, growth curves were fitted

to derive clinically meaningful growth phenotypes, e.g. growth velocities for the genetic and other association analyses.

4.1.2.2 Follow-up at 16 years

The cohort was followed-up by postal questionnaire when offspring were 15 to 16 years old, between May 2001 and June 2002 (80% responded), and then by invitation for clinical examination between August 2001 and June 2002 (74% attended). Target population, those alive with known address (N=9,215), were contacted by postal questionnaire to collect information on their growth and health, health behaviour (smoking, drinking, physical activity), school performance and family's social circumstances. 7,344 replies on questionnaires from adolescents and 6,985 replies on questionnaire from parents were received. The target population, all alive and known address (N=9,215), were invited for clinical examination; 6,798 responded, covering 74% of target population. Trained nurses performed the clinical examination, collecting anthropometric and other clinical data as well as samples for metabolic phenotypes including, height, body mass index (BMI), waist-circumference (WC), high density lipid (HDL), cholesterol, triglyceride, glucose, systolic blood pressure (SBP) and diastolic blood pressure (DBP), as described below.

Of the NFBC1986 individuals who provided blood samples at the 16 year of age follow-up DNA was extracted for 6,266 individuals and genotyping for the present study was performed with Genetics group at Imperial College London, UK, in Pasteur Institute Laboratories, Lille, France or in KBiosciences, UK, depending on the study. Details are given below in section 4.3.

Anthropometric measures

The clinical examinations were conducted by teams of trained nurses. The following information was recorded: weight, height, waist and hip circumferences, sitting height, spirometry, blood pressure, pulse rate, skin prick test, physical activity and questions about puberty, nutrition, smoking and use of alcohol. Adolescent BMI (kg/m^2) and waist-to-hip ratio (WHR) were calculated from the height and weight recorded at the clinical examinations. The stages of puberty were assessed via the Tanner scale and the age of menarche was assessed using a questionnaire in the clinical examination. In the NFBC 1986 seated blood pressure (BP) was measured in the

adolescents (after 15 min rest) by using an oscillometric blood pressure meter Omron 705C, (or a mercury sphygmomanometer if previous failed). Two readings of SBP and DBP were recorded and the mean of the two measurements was used for the association analyses.

5 Blood sample

The blood sample was taken at the age of 16 years, and the serum samples were stored at -70C. Enzymatic assays of fasting serum glucose, insulin, triglycerides (TG), total cholesterol (TC), high density lipoprotein cholesterol (HDL-C), apolipoprotein A1 and apolipoprotein B were carried out using Hitachi 911 automatic analyzer and commercial reagents (Boerhringer Mannheim, Mannheim, Germany) in the accredited laboratory of Oulu University Hospital (Oulu, Finland). The fasting plasma concentration of low density lipoprotein cholesterol (LDL-C) was measured by the Friedewald equation: [LDL-C] = [Total C] - [HDL-C] - ([TG]/2.2).

In NFBC1986, as for NFBC1966, insulin resistance was assessed using the homeostasis model assessment (HOMA-IR) and calculated using the following formula: HOMA-IR (mmol/L x μU/ml) = fasting glucose (mmol/L) x fasting insulin (μU/ml)/22.5. Pancreatic ß cell function was determined using the homeostasis model assessment of β-cells (HOMA-B) and calculated by the following formula: 20 × fasting insulin (μIU/ml)/fasting glucose (mmol/ml) − 3.5 (http://www.dtu.ox.ac.uk/homacalculator/).

4.2 Variables (or exposure) selected from the NFBC1966 and NFBC1986

Phenotype and other variable selections in this study was based on a prior knowledge from previous published studies [68] and on expert knowledge of my supervisors (Prof Marjo-Riitta and Dr Paul O'Reilly). Phenotypes selection were study-specific, and are described below, based on if the data was collected prenatally or postnatally. The prenatal period includes the maternal and paternal variables of the individual included in the study, while postnatal period includes data on individual's birth, infancy, childhood and adolescent. One of the studies also utilised the data on maternal and paternal variables when the adolescent was 16 years old and has been described under the postnatal period. **Table 3**, describes each variable in detail per study. A more extensive data exploration was done for the MSc dissertation (given in Appendix Note 2) for selection of variation (or exposures).

4.2.1 Prenatal period – Maternal and paternal

Variables on prenatal period were obtained from questionnaires administered by mid-wives when pregnant mothers attended the maternity clinics or from medical records (maternity cards). The questionnaire, medical records and population registry provided data on maternal age, height, weight, parity, hypertensive disorders, and smoking during pregnancy, marital status and SES. Maternal pre-pregnancy and paternal BMI (kg/m^2) was calculated by weight in kilograms divided by height in meters squared. Maternal SES and hypertensive complications during pregnancy were re-classified to make them more interpretable and comparable between the NFBCs (details of the reclassification are given in Appendix Note 1).

4.2.2 Postnatal period

Birth

Data on birth was collected from the hospital records soon after delivery. Individual's birth weight (BW), birth length (BL), head circumference (HC), and gestational age (GA) was obtained from birth records prospectively at the time of birth. The BW was given in grams (converted to kg where appropriate) and gestational age (in weeks according to the mother's last menstrual period, for all in NFBC1966 and in 16% of the sample in NFBC1986; by ultrasound examination in 84% of the NFBC1986 sample) were clinically measured according to standard procedures at the time of delivery.

Infancy, childhood and adolescent

Data on postnatal height and weight growth in infancy and childhood has been described above for each of NFBCs in section 4.1.1.1.

Adolescent variables were obtained from questionnaires completed at 14 years in NFBC1966 and at 16 years in NFBC1986. Adolescent BMI (kg/m^2) and waist-to-hip ratio (WHR) were calculated from height and weight recorded at clinical examinations. The stage of puberty was assessed via the Tanner scale and the age of menarche was recorded for the females. The Tanner scale is a physical measurement for sexual development in puberty. In girls, the Tanner scale is based on their breast size, shape and pubic hair distribution. In boys the Tanner scale is based on the size and shape of the penis and scrotum, and on their pubic hair development.

Data on maternal and paternal weight, height and SES were also collected via the postal questionnaire at 14 years in NFBC1966 and at 16 years in NFBC1986.

Adulthood

Adulthood height and weight was measured at the clinical examination at 31 years in NFBC 1966. Adult BMI was calculated by weight in kilograms divided by height in meters squared and presented in units' kg/m^2. Data on smoking, parity and SES was collected by the postal questionnaire at 31 years preceding the clinical examination.

4.3 Growth modeling: application of parametric methods

4.3.1 Descriptive analysis of growth data

The description of the data and genome-wide association analyses of growth data are presented in Chapter 5, section 5.1 and 5.2. For the purpose of modeling and to reflect the literature that defines the three stages of growth[167], the data was divided into three categories: 0-2 years for infancy, 2 to 13 years for childhood and 13 to 16 years for adolescence. This thesis only covers the growth data for infancy and childhood. The adolescence data was used for training purposes but not described in this thesis. Individuals with at least three data points available within these time periods were selected for analysis. Mean and standard deviation (SD) of height, weight and BMI was calculated at each age interval. The analysis was sex-stratified, as males have a higher body size than females at most ages, and growth velocities differ, particularly during puberty, as shown previously[168]. The differences between males and females were tested by the Student t-test in univariate analyses, with $P < 0.05$ considered as significant.

From a statistical perspective fitting growth curves has various advantages compared to the analysis of serial measurements over time in a cross-sectional manner. It allows the estimation of rate of change, for example height growth velocity over time in each individual, and the estimation of biologically meaningful parameters such as the timing of peak growth. Curve fitting summarises a large amount of raw anthropometric data into a few constants or function parameters and thus serves as a data reduction technique, which is useful in defining biologically meaningful phenotypes for the association analyses. The resulting fitted curve helps to describe typical growth in the sample. The growth curve evens out the fluctuations in the data caused by measurement errors and

Table 3. Variables used in the analyses in this thesis from NFBC 1966 and NFBC 1986.

Variables	Unit/categories	Study and notes
Prenatal period		
Maternal		
height	cm	Height information was collected from questionnaire when mother first visited maternity clinic.
weight	kg	Weight information was collected from questionnaire when mothers first visited the maternity clinic.
BMI	kg/m^2	Calculated from reported height and weight. Divided into a)underweight b) normal c) overweight d) obese using IOTF classification
age	years	Age was collected from the questionnaire when the women first visited the maternity clinic. Continuous variable
parity	a) 0; b)1; c)2; d)3; e)>=4	Categorical variable
marital status at birth	a) married; b)unmarried/cohabiting; c) divorced/widow	Unmarried and cohabiting were grouped together due to small sample size
SES	a)white collar; b)blue collar; c)farmer	If mother's SES was missing then father's occupation. I used three categories in my analyses. A detail on how the variable was re-classified is given in Appendix.
education	a)primary; b)vocational; c)secondary; d)university	
smoking at 2nd month of pregnancy	a) non-smoker; b) smokers	Maternal smoking at 2nd month of pregnancy was collected from questionnaire, and the mother smoking was categorised as a non-smoker (<1 cigarettes per day) or smokers (>=1 cigarette per day).
hypertensive complications during pregnancy	a) normotensive; b) gestational hypertension; c) pre-eclampsia; d) chronic hypertension; e) super-imposed pre-eclampsia;	See Appendix Note 3.

	f) proteinuria	
weight gain during pregnancy	kg	Collected during ante-natal checks by maternal nurses
Paternal		
weight	kg	Collected when pregnant mothers' visited
height	cm	
BMI	kg/m^2	Divided into a)underweight b) normal c) overweight d) obese using IOTF classification
SES	a) white collar; b) blue collar; c) farmer	
Birth		
Birth weight	kg	
Birth length	cm	
BMI	kg/m^2	Divided into a)underweight b) normal c) overweight d) obese using IOTF classification
Gestational age	weeks	
Head circumference	cm	
Infancy		
height	cm	
weight	kg	
peak height velocity	cm/year	Calculated from growth models, as presented below Section 4.3.
peak weight velocity	kg/year	Calculated from growth models, as presented below section 4.3.
Childhood		
height	cm	
weight	kg	
Age and BMI at adiposity peak (AP)	years, kg/m^2	AP=adiposity peak derived from fitted growth curves
Age and BMI at adiposity rebound (AR)	years, kg/m^2	AR=adiposity rebound derived from fitted growth curves
Adolescent		
BMI	kg/m^2	
age at menarche	years	
Fasting glucose	mmol/L	
Fasting insulin	mmol/L	

Triglyceride	mmol/L	
Total cholesterol	mmol/L	
HDL and LDL	mmol/L	
Apoprotein AI	mmol/L	
Apoprotein B	mmol/L	
SBP/DBP	mmHg	SBP and DBP were measured twice with a mercury sphygmomanometer in a sitting position from the right arm after 15 minutes of rest. The average of two measurements was used. For those on medication for hypertension, 15 mmHg and 10mmHg were added to the recorded SBP and DBP values, respectively
Adulthood		
BMI	kg/m^2	Calculated as weight/height ^2
Body adiposity index (BAI)	% fatness	Calculated as hip circumferences (cm)/height(m) ^1.5 and subtracted by 18, as described by Bergman et al, 2011.
smoking	a) light (<10 cigarettes/day); b) heavy smokers (>= 10 cigarettes/day)	The adulthood smoking was used in the study for association between LTL and childhood and adulthood measurements. Non-smokers were combined with light smokers, as there was no association between LTL and non-smokers.
SES	a) Farmer; b) self-employed; c) white collar; d) blue collar	
Parity	a)0; b) 1; c) 2 or more	Number of live births.

other sources of variation. The individual growth curve gives an estimate of the variable of interest (e.g. height) at each time point between observed measurements, and in effect interpolates a value for all the 'missing' measurements[169].

The modeling of growth has been an extensive task. Consequently this task was shared between team members; Dr. Ulla Sovio and I. In this thesis, in terms of growth modeling, I have mainly focused on and consequently described here the growth modeling for infants and young children in the NFBC1986. However, the analyses in this thesis utilize the growth modeling from NFBC1966, which was conducted by Dr. Ulla Sovio and other team members. The modeling of later childhood and adolescent growth parameters has also been performed in NFBC1986 but not included in the study, as it was not part of the analyses for this thesis.

4.3.2 **Inclusions, exclusions and data included in growth modeling**

The growth modeling analyses were restricted to singletons, as multiple births are associated with smaller size at birth and higher growth velocities in early life[170]. The growth data were checked and outliers removed by Dr Marika Kaakinen and Dr Ulla Sovio in the project centers in Oulu and in London. Any measurements that indicated downward growth in height or that did not fall within the age interval were omitted from the descriptive analyses. Individuals with at least three height measurements between birth and two years of age were selected for growth modeling, since growth velocity estimation is unreliable with a low number of measurements.

4.3.3 **Models for infant height and weight growth in the Northern Finland Birth Cohort 1986 (NFBC1986)**

The infant growth modeling was first performed within NFBC1966 longitudinal data. A detailed comparison of the different models available, and those most suitable for longitudinal data for NFBC1966, was undertaken both by me and by Dr Ulla Sovio as described previously[11] . A detailed report on the comparison of models in NFBC1986 was performed by me for the MSc Dissertation, as presented in **Appendix Note 2**. Parametric modeling was applied, as it can reflect what is known biologically, and is easy to interpret. Utilising Dr Sovio's approach, I initially compared two

parametric growth models relevant for modeling of infancy and early childhood weight and height measurements in NFBC1986. The first was a three-parameter Count model and the second a four-parameter Reed1 model, as described by Simondon et al, 1992 154, and these were compared in terms of the residual standard deviation (RSD) and Akaike Information Criteria (AIC). Both were fitted to height and weight growth data from 0-2 years in the NFBC1986155. The four-parameter Reed1 model 156 gave a better fit than the three-parameter Count152in both girls and boys of the NFBC1986. This result was consistent with the modeling approach used in NFBC1966 analyses and a previous study from a population in Simondon,1992[171] .

The four-parameter Reed1 model was used for infant height and weight modeling. Reed1 growth curves were fitted for each sex using nonlinear random effects models, allowing both fixed and random effects for each function parameter 156. The nlme function in R program (version 2.4.1) was used to fit the models. As Reed1 model is not defined at birth, therefore a modified version which includes an additional parameter to account for changes in height/weight growth after birth, as described in Simondon et al 154, was used. The model was defined as:

$$Y = A + Bt + C \ln(t+1) + D/(t+1)$$

where, t is the postnatal age at the time of modeling, Y is the height reached at age t and A, B, C and D are the function parameters. The function parameter A relates to the baseline height at birth, B to the linear component of the growth velocity, C to the decrease in the growth velocity over time, and D to the inflection point that allows growth velocity to peak after birth rather than exactly at birth. The best-fitting curves for each individual were estimated by the maximum likelihood method 12. The growth velocity curves were estimated as first derivatives of the height or weight growth curves. The growth phenotypes were derived from the velocity curves for height at the maximum velocity, that is, peak height velocity in infancy (PHV). Similarly the estimate of peak weight velocity in infancy (PWV) was defined as the maximum value of the weight velocity curve (**Figure 4 A and B**).

4.3.4. Models for growth in body mass index (BMI) in the Northern Finland Birth cohort 1986 (NFBC1986)

The initial data preparation for model selection on body mass index (BMI) in infancy and childhood was previously considered in detail by Dr Ulla Sovio at Imperial College London, Dr. Nicholas Timpson at Bristol University and Ms. Nicole Warrington at the University of Western Australia, Perth. A detailed description on the selection of BMI growth modeling in NFBC1966 is given in the doctoral thesis of Sovio, 2011[11]. The comparative statistical models used for longitudinal childhood BMI growth was recently published by Ms Nicole Warrington, University of Western Australia[13]. For work described in this thesis, I have only utilised age and BMI growth modeling in NFBC1986 infant and childhood BMI growth model. However, the work in thesis also utilises BMI growth models from NFBC1966 as part of the research in Chapter 5, section 5.1.3 and 5.2.2, which were calculated by Dr. Ulla Sovio, details for which are given in her doctoral thesis[11].

In summary, due to the complexities in modeling BMI longitudinally, the data was divided into two age windows: 2 weeks to 18 months (infancy) and 18 months to 13 years (childhood). These cut-offs points were chosen based on biological phases of growth and from the growth curves in the NFBC1966[172]. Measurements taken in the first two weeks of life were excluded due to potential decrease in BMI especially during the first week of life and to avoid the error in the growth curve modeling that this may have introduced.

For infancy growth model (0 – 1.5years), the mid-point for age was 0.75 years (about 9 months), which was close to the average age when BMI reaches its peak. For childhood growth model (1.5 – 13 years) the mid-point for age was 7.25 years, which was on average the age shortly after that at which the BMI reaches its minimum. BMI measurements for both infancy and childhood were right skewed and hence log transformed to improve the statistical model in infancy and childhood.

Linear mixed effects (LME)[173] models were fitted using sex, gestational age and their interactions with age as covariates, with baseline BMI as the random effects and change in BMI over time as slope. In addition to a linear age effect, quadratic and cubic terms for age were included to account for non-linearity of change in BMI over time.

The model specification by age window:

1) Infancy. For calculations of age and BMI at adiposity peak (AP) in infancy, the analysis was restricted to singletons with at least three BMI measures from two weeks to 18 months of age. The model is as follows:

$$\log(\text{BMI (kg/m2)}) = \beta0 + \beta1\ \text{Age} + \beta2\ \text{Age2} + \beta3\ \text{Age3} + \beta4\ \text{Sex} + u0 + u1\ (\text{Age}) + \varepsilon$$

where β0, β1, β2, β3, β4 are the fixed effects terms, u0 and u1 are the individual level random effects and ε is the residual error. Age at AP was calculated from the model as the age at maximum BMI between 0.25 and 1.25 years.

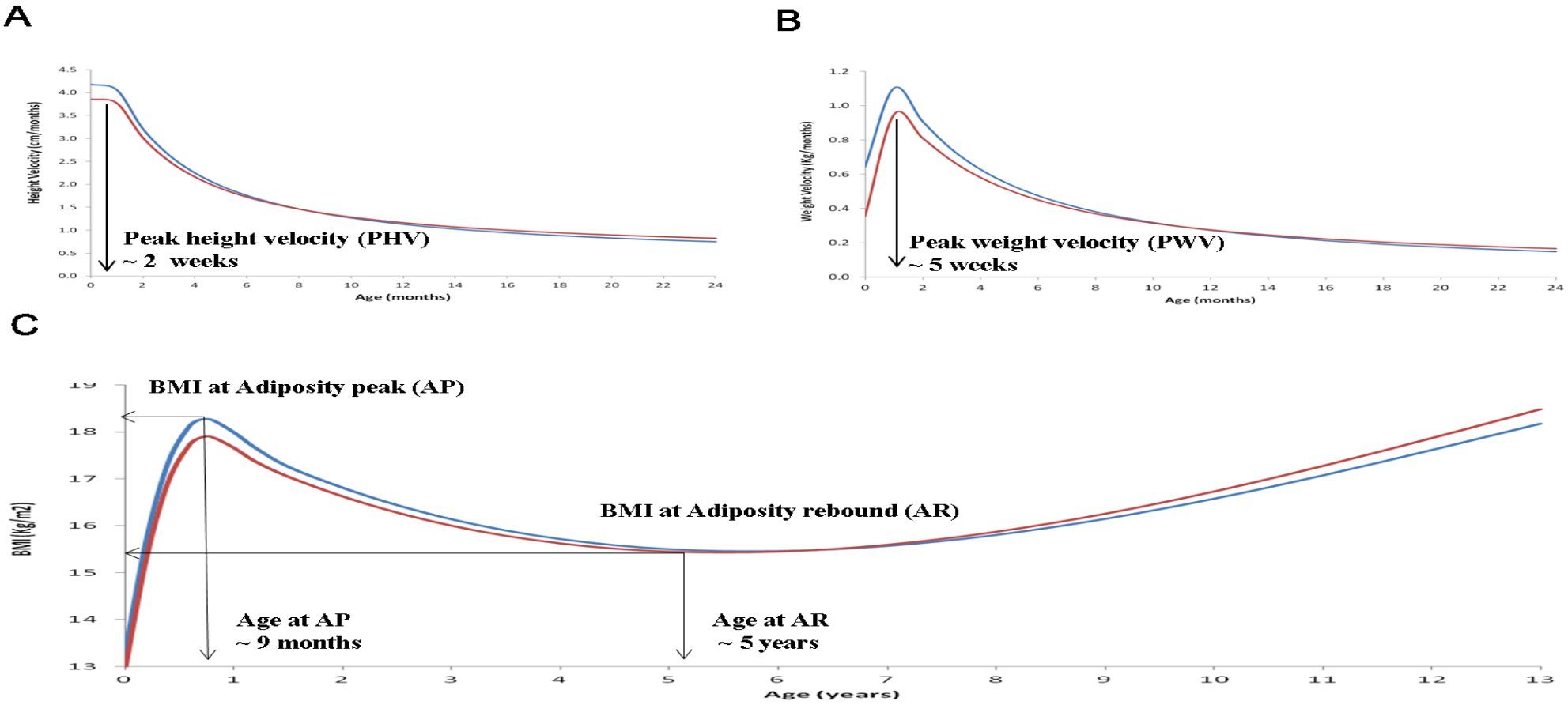
2) Childhood. For modeling of age and BMI at adiposity rebound (AR) in childhood, the analysis was restricted to singletons with BMI measurements, from 18 months to 13 years (**Figure 4C**). The model is as follows:

$$\log(\text{BMI (kg/m2)}) = \beta0 + \beta1\ \text{Age} + \beta2\ \text{Age2} + \beta3\ \text{Age3} + \beta4\ \text{Sex} + \beta5\ \text{Age} * \text{Sex} + \beta6\ \text{Age2} * \text{Sex} + u0 + u1\ (\text{Age}) + \varepsilon$$

where β0, β1, β2, β3, β4, β5 and β6 are the fixed effects, u0 and u1 are the individual level random effects and ε is the residual error. Age at AR was calculated as the age at minimum BMI between 2.5 and 8.5 years (**Figure 4 C**).

The work from the BMI growth modeling in infancy and childhood in NFBC1986, contributed to the research in Chapter 5, section 6.2 for genome-wide association analysis on early growth phenotypes. For the same study, I also provided support and conducted growth modeling in other cohorts which included, ALSPAC, COPSAC, INSI and LISA, SWS, details of which are given in Chapter 5, section 5.2.2.

Figure 4. Graphical illustration of height, weight and BMI growth patterns (or early growth phenotypes) in NFBC 1986. A) Peak Height Velocity (PHV); B) Peak Weight Velocity (PWV); C) Age and BMI at Adiposity Peak (AP) and Rebound (AR). The growth curves for males are in blue and for females in red.



4.4 Genetic data for GWAS

4.4.1 Description of genetic data for Northern Finland Birth cohort 1966 (NFBC1966) and 1986 (NFBC1986)

The NFBC 1966 individuals who provided blood at 31 years and gave consent for use of data were eligible for genotyping (N=5,753). IIlumina's HumanCNV370-Duo Analysis BeadChip was used to obtain genome-wide data, which contains tag SNPs derived from the HapMap European-derived (CEU) sample[174]. The genome-wide data consisted of over 318,000 SNPs and about 52,000 markers to target approximately 14,000 copy number variants (CNV) regions, giving a total of over 370,000 markers. The Beadstudio algorithm was used for calling the genotypes. Sample and SNP quality control (QC) was completed by Drs. Susan Service and Chiara Sabatti at the University of California, USA, Dr. Inga Prokopenko at the University of Oxford, United Kingdom, and Marika Kaakinen at the University of Oulu, Finland[163]. Imputation was performed using IMPUTE software version 3.1, applying information call threshold of >0.4 and MAF threshold of >1%. In total, over 2.5 million SNPs were analysed[163]. In addition, several SNPs were separately genotyped for candidate gene studies and for replication purposes using TaqMan SNP genotyping assay (Applied Biosystems, Warrington, UK) according to the manufacture's protocol. After quality control (QC) and imputation of the data (April 2008), the final data was for 5,402.

In NFBC1986, DNA was extracted for 6,266 individuals who provided blood samples at 16 years of age follow-up. Eleven SNPs were directly genotyped which were then used in the replication stage of genetic association studies (of section 5.2.1 and section 5.2.2) in this thesis. Genotyping was performed with Genetics group at Imperial College London, UK or in Pasteur Institute Laboratories, Lille, France. Some genotyping for replication in genetic association studies was performed by KBiosciences Ltd (Hoddesdon, UK) using their own novel system of fluorescence-based competitive allele-specific PCR (KASPar). Genotyping success rate for all SNPs was >95.0% and none of the SNPs deviated from Hardy-Weinberg equilibrium (all *P*-values>0.0001). The genetic association studies were performed for this thesis as part of a large consortium, early growth genetics (EGG), which involved contributions from other datasets to participate in the study. This contribution increased the statistical power for genome-wide analysis and made the results robust. A short summary of other datasets used in the thesis are described below.

4.4.2 Description of other cohorts participating in Genome –wide association studies (GWAS)

Avon Longitudinal Study of Parents and Children (ALSPAC)

ALSPAC is a prospective study, which recruited pregnant women with expected delivery dates between April 1991 and December 1992 from Bristol UK. The ALSPAC data was collected from four clinical exams at 6 weeks, 9 months, 18 months and pre-school and follow at the ages of 7, 8, 9, 10, 11, ~12 and ~14 years. Height and weight measurements were taken at each of these time points when individuals visited clinics. Height was measured to the last complete mm using the Harpenden Stadiometer. Weight and bioelectrical impedance were measured using the Tanita Body Fat Analyser (Model TBF 305). Weight was measured to the nearest 50g. All proposals for new data collection and policies for data handling and analysis were approved by ALSPAC's own Law and ethics and by Local Research Ethics Committees (LRECs). Ethical approval for the study was obtained from the ALSPAC Law and Ethics Committee and the Local Research Ethics Committees.

For the GWA analysis in Chapter 5, section 5.2, n=3,714 participants (discovery sample) were genotyped using either the Illumina 317K or 610K genome-wide SNP genotyping platforms by the Wellcome Trust Sanger Institute, Cambridge, UK and the Centre National de Génotypage, Evry, France. A common set of SNPs (present in both genotyping platforms) were extracted and the resulting raw genome-wide data was subjected to standard quality control methods. Individuals were excluded on the basis of having incorrect sex assignments; minimal (0.34) or excessive (0.36) heterozygosity; disproportionate levels of individual missingness (>3%) and evidence of cryptic relatedness (PI HAT >0.11). The remaining individuals were assessed for evidence of population stratification by multidimensional scaling (MDS) analysis, using CEU, YRI, JPT and CHB individuals from the HapMap as reference ethnic groups. Only those clustering with the CEU individuals were included. The underlying population stratification was thereafter controlled by using the first two EIGENSTRAT derived ancestry informative covariates (EIG-PC1 and EIG-PC2) as well as any additional EIGENSTRAT derived ancestry informative covariates that were correlated with the phenotype (EIG-PC3 for g-men; EIG PC9 for exr.XY; EIG-PC4 for prn-alL; EIG-PC10 for ls-pg; EIG-PC5 for PC3; EIG-PC5 for PC11; EIG-PC9 for PC9).

SNPs with a minor allele frequency of <0.5% and call rate of <97% were removed. Furthermore, only SNPs which passed an exact test of Hardy-Weinberg equilibrium ($P>5x10^{-7}$) were considered for analysis. The resulting dataset consisted of 3,233 individuals and 285,531 SNPs. We then conducted imputation with MACH 1.0 Markov Chain Haplotyping software17; using CEU individuals from phase 2 of the HapMap

project as a reference set (release 22). The final imputed dataset consisted of 3,233 subjects, each with 2,543,887 imputed autosomal markers (http://www.bristol.ac.uk/alspac/).

Cambridge Baby Growth Study (CBGS)

The CBGS cohort recruited pregnant mothers attending a single antenatal centre in Cambridge UK, representative of the South Cambridgeshire population. A trained paediatric research nurse approached and recruited mothers' at their first antenatal clinic appointment during early pregnancy. In total N=1,662 babies born between the years of 2001 to 2009 were selected. A paediatric research nurse measured the offspring at birth, and months 3, 12, 18 and 24. The study was approved by the local Cambridge research ethics committee and all mothers gave informed written consent. The data used in this study were based on 556 infants where DNA samples were available from cord blood samples collected at birth. Individual SNP genotyping was performed by TaqMan (Applied Biosystems) at the MRC Epidemiology Unit, Cambridge UK. For all SNPs, call rates were >99% and Hardy Weinberg Equilibrium P-values >0.05 (http://www.medschl.cam.ac.uk/paediatrics/pages/growthstudy.htm).

Children's Hospital of Philadelphia (CHOP)

CHOP recruited infants born from 2006 to 2010 at the Children's Hospital of Philadelphia, in the Greater Philadelphia area. The cohort consisted of 1,445 obese children and 2,802 lean children of European ancestry. All of the participants had their blood drawn in to an 8ml EDTA blood collection tube and were subsequently DNA extracted for genotyping. All subjects were biologically unrelated and were aged between 2 and 18 years old. This study was approved by the Institutional Review Board of the Children's Hospital of Philadelphia. Parental informed consent was given for each study participant for both the blood collection and subsequent genotyping. Self-reported ethnicity was confirmed by multidimensional scaling methodologies. The high throughput genome-wide SNP genotyping was performed using the Illumina Infinium™ II HumanHap550 BeadChip technology (Illumina, San Diego), at the Center for Applied Genomics at CHOP. We used 750ng of genomic DNA to genotype each sample, according to the manufacturer's guidelines (http://www.chop.edu/).

Copenhagen Study on Asthma in Children (COPSAC)

The COPSAC cohort recruited mothers with a history of asthma and the cohort total was 411 babies. The newborns were enrolled at the age of 1 month, the recruitment of which was previously described in detail. Each newborn was assessed at the COPSAC clinical research unit at six monthly intervals; additional visits were arranged immediately upon the onset of symptoms. All growth parameters were measured and obtained by the COPSAC doctors and history was obtained using structured questions and closed response

categories. The study was approved by the Ethics Committee for Copenhagen (KF 01-289/96) and The Danish Data Protection Agency (2008-41-1754) and informed consent was obtained from both parents. High throughput genome-wide SNP genotyping were performed using the Illumina Infinium™ II HumanHap550 v1, v3 or quad BeadChip platform (Illumina, San Diego), at the Children's Hospital of Philadelphia's Center for Applied Genomics. Further details can be found on www.copsac.com.

Danish National Birth Cohort (DNBC)

The DNBC were collected for 92,274 pregnant women recruited between 1996 and 2002, and their children were followed through childhood and adolescence. The BMI data was collected when the infant was 18 months old. The mothers were asked to report height and weight from the child exams at age 5 and 13 months of age. For the purpose of research in this thesis, the observation closest to age 9 months was used in the analysis of growth modeling and anlaysis. Genotype data were performed with the Illumina 660 Quad chip, and genotypes for all SNPs from the HapMap phase 2 CEU reference set were imputed with MACH 1.0. The analysis was based on 861 children (447 boys, 414 girls). The local Ethics Committee approved the study. Details for this data collection are described at http://www.dnbc.dk.

Étude des Déterminants pré et postnatals du développement et de la santé de l'ENfant (EDEN)

The EDEN cohort is an ongoing study of pre- and post-natal determinants of children's growth and development. The cohort enrolled 2,002 mother-child pairs at two centers in France, Nancy and Poitiers, and continued follow-ups with the children until their fifth birthday. The study was approved by the ethics committee (CCPPRB) of Kremlin Bicêtre (France) and by the Data Protection Authority "Comission Nationale de l'Informatique et des Libertés" (CNIL). DNA was extracted from cord blood samples collected in 1,367 of the 1,907 births. Individual SNP genotyping was performed by TaqMan (Applied Biosystems) at the MRC Epidemiology Unit, Cambridge UK. For all SNPs, call rates were >99% and Hardy Weinberg Equilibrium P-values >0.05 (eden.vjf.inserm.fr).

The Exeter Family Study of Childhood Health (EFSOCH)

The EFSOCH is a prospective study, set up to test the fetal insulin hypothesis, and to identify genetic polymorphisms that play a role in determining birth weight and early postnatal growth. The study recruited 1,017 families from a postcode-defined area in central Exeter, with inclusion criteria were; a homogeneous, non-diabetic, UK Caucasian cohort. Detailed anthropometric measurements were taken from both parents at 28 weeks' gestation, and from their children at birth, 12 weeks, 1 year and 2 years of age. Insulin and other biochemical analysis were measured in fasting parental samples and an umbilical cord blood sample taken at

delivery. Parental and offspring DNA were extracted to allow molecular genetic analysis of candidate genes implicated in fetal growth. Local Ethics Committee approved the study.

The Generation R Study (GenR)

The GenR Study is a population-based prospective cohort study, with children born between April 2002 and January 2006 recruited to participate in the study. This study is designed to identify early environmental and genetic determinants of growth, development and health from fetal life until young adulthood and has been described previously in detail[175,176]. Detailed measurements of height and weight were performed using ultrasound and physical examinations, biological samples and advanced imaging techniques. Analysis was restricted to Caucasian individuals with genome-wide data and sufficient growth measurements available for growth modeling (N=2,227). Length was measured to the nearest millimeter and weight was recorded by well-trained staff in community health centers using standardized procedures. The study has been approved by the Medical Ethics Committee of the Erasmus Medical Center, Rotterdam. Written informed consent was obtained from all participants or their parent(s). Cord blood for DNA isolation was available in 59% of all live-born participating children. Individuals with sex-mismatch or missing cord blood samples were excluded. Individual genotype data were extracted from the genome-wide Illumina 610 Quad Array51(http://www.generationr.nl).

Helsinki Birth Cohort Study (HBCS)

The HBCS is a prospective birth cohort which includes individuals 4,630 boys and 4,130 girls born at the Helsinki University Central Hospital between the years 1934 and 1944. Serial measurements of height and weight until age 12 were extracted from child welfare clinics and school health clinic records, with an average of 10 measurements between birth and 2 years, and 8 measurements between 2 and 11 years of age. A representative set of 928 males and 1075 females (59 and 70 years old) returned for clinical examinations, between 2000 and 2002, when adult stature was measured and blood was taken for DNA extraction. HBCS has been approved by local ethics committees. Genotyping was performed on the Illumina 670 Quad chip and calling was done with the Illuminus algorithm at the Sanger Centre on a total of 1566 samples. Pre-imputation quality control filters included: call rate 95%, MAF > 0.01, HWE 1x10-6. Closely related individuals were screened by examining pair-wise IBD estimates. >0.2 IBD-sharing was used as cut-off. Population stratification was examined using multidimensional scaling analyses, but there were no outliers. Imputation was performed with MACH, and the analysis software was MACH2QTL. Post-imputation quality control filters were R2HAT >0.3 and MAF >0.01

Lifestyle – Immune System – Allergy Study (LISA)

LISA is a population-based prospective birth cohort study on lifestyle-related factors on the immune system and the development of allergies in childhood. Individuals born healthy, mature (gestational age over 37 weeks) and with a birth weight over 2500g were recruited between December 1997 and January 1999 (N=3,097 in Munich, Leipzig, Wesel, and Bad Honnef, Germany to participate in the study. Weight, length, and head circumferences were measured highly time-resolute by the family physician during the first two years of life. Head circumference measurements were scaled to the nearest centimeter and obtained via medical records. DNA was collected at the age 6 & 10 years and analysed using the Affymetrix Human SNP Array 5.0 for each individual. After quality control (MAF>1%, HWE>0.01, call rate per SNP and person >95%) genome-wide data was imputed using BRLMM-P algorithm in IMPUTE [177]. Local Ethics Committee approved the study.

The Raine Study

The Raine Study was a randomized controlled trial to evaluate the effects of repeated ultrasounds in pregnant women in Perth, Western Australia. In total, 2,900 pregnant women prior their 18th week gestation were recruited between 1989 and 1991 at the King Edward Memorial Hospital (Perth, Western Australia). Women were randomized to repeat ultrasound measurements at 18, 24, 28, 34 and 38 weeks gestation or to a single ultrasound assessment at 18-weeks. Children have been assessed at average ages of 1, 2, 3, 5, 8, 10, 14 and 17. DNA was collected at the 14 and 17 year follow-ups. The study was conducted with appropriate institutional ethics approval, and written informed consent was obtained from mothers at all follow-ups and participants at the 17 year follow-up. DNA was collected using standardized procedures from 74% of all adolescents who attended the 14 year follow-up on and a further 5% from the 17 year follow-up measurements. Genotype data was extracted from the genome-wide Illumina 660 Quad Array for each individual. Analysis was restricted to individuals with genome-wide data and BMI (http://www.rainestudy.org.au/). Local Ethics Committee approved the study.

INfancia y Medio Ambiente [Environment and Childhood] Project (INMA)

INMA is a population-based birth study; cohorts were established in several regions of Spain following a common protocol. This analysis uses the INMA cohorts of Valencia and Sabadell (Catalonia) established between 2003 and 2006. This project aims to study the associations between pre- and postnatal environmental exposures and growth, health, and development from early fetal life until adolescence. Pregnant women were recruited for the project during the 1st trimester of pregnancy at public primary health care centers or public hospitals. Detailed measurements were performed using ultrasounds, physical examinations and biological samples. Informed consent was obtained from all participants and the study was

approved by the Hospital Ethics Committees in each participating region. Cord blood for DNA isolation was available for 909/1409 (64.5%) of all live-born participating children and 752 of these samples were selected for the genome-wide genotyping using the HumanOmni1-Quad Beadchip (Illumina). Sex-mismatch rate between genome based sex and midwife recorded sex was 0, indicating that possible contamination of maternal DNA was extremely low (http://www.proyectoinma.org/presentacion-inma/en_index.html).

The Physical Activity and Nutrition in Children (PANIC) Study

The PANIC Study is a long-term controlled exercise and diet intervention study in a population sample of 512 primary school children conducted by the University of Eastern Finland. Baseline examinations at the ages of 6-8 years were performed in 2007-2009. Half of the children were allocated into the exercise and diet intervention group and the other half to the control group. The 2-year follow-up examinations at 8-10 years of age (participation rate 86%) were done in 2009-2011. Physical activity and inactivity, cardiorespiratory and musculoskeletal fitness, nutrition, overweight, obesity, body composition, fetal and early childhood growth and development, metabolic, cardiovascular, musculoskeletal and cognitive functions were assessed comprehensively at baseline and the 2-year follow-up. After the 2-year examinations, the intervention group will continue with a less intensive exercise and diet intervention until adulthood. The children will be re-examined in adolescence and early adulthood. DNA was collected at the baseline and 2-year follow-up and analyzed using the Metabochip array. The PANIC Study protocol was approved by the Research Ethics Committee of the Hospital District of Northern Savo in 2006. Both children and their parents gave their written informed consent (http://www.lastenliikuntajaravitsemus.fi/5).

The Southampton Women's Survey (SWS)

The SWS study enrolled female patients aged 20-34 years between 1998 and 2002 by general practitioners in Southampton. Of those women recruited into the study, N=12, 583 (75%) agreed to be interviewed by a research nurse, and all interviews were conducted between April 1998 and October 2002. Women who subsequently become pregnant were recruited into the pregnancy phase of the SWS. They visited the SWS Ultrasound Unit at the Princess Anne Hospital in Southampton at 11, 19 and 34 weeks gestation where fetal and placental size and blood flows were measured using ultrasound. At 11 and 34 weeks, the women were interviewed by research nurses, and anthropometric measurements of the baby were made within 48 hours of birth. By the end of 2007, when the pregnancy follow-up stopped, N= 3,159 babies had been born to SWS women. Cord blood samples were obtained at birth from which DNA has been extracted. The children were followed up at home at 6 months, 1, 2, and 3 years of age. Thereafter a subsample was seen at 4 years and further subsamples at 6, 8 and 10-12 years are currently being assessed. At all visits body composition is

assessed by anthropometry and/or DXA. Buccal swabs are being taken at the 6, 8 and 10-12 year visits from which further DNA is being extracted. The Southampton Women's Survey has received approval for all stages of the study from Southampton and South West Hampshire Local Research Ethics Committee http://www.leu.soton.ac.uk/sws/). DNA was extracted from Kbiosciences, Hoddesdon, Herts, UK.

4.5 Data analysis and statistical methods

4.5.1 Association analysis

An analysis plan was prepared for each association study and it included a detail description of the study design and the planned data analysis (Appendix Note 4).

4.5.1.1 Descriptive analysis

The selected variables from NFBC 1966 and NFBC 1986 were examined for distribution to check for any possible errors. Histograms were developed to look for any outliers and extreme values were removed to prevent a skewing of results. The variables were transformed accordingly by looking at their distribution plots. SAS procedure PROC PLOT was used for producing scatter plots for independent and dependent variables. The continuous variable mean and standard deviation (SD) were presented and percentage distributions for each category in the categorical variables were given. Missing records for variables in the sample data set were also presented as percentage of total. SAS procedure PROC MEANS and PROC FREQ in version 9.2 (SAS Institute Inc. 2010) were used to perform this analysis. Possible errors were checked against original records and if there were errors in data then a missing value was inserted (for example an impossible height). Once the data was cleaned, distribution tables were noted. This gave the characteristics of the study population. For each analysis, an outcome and exposure variable/variables were specified and confounders were selected by *a priori* or by expert advice.

4.5.1.2 Regression analyses

The representativeness of each sample selected was examined in detail to reveal any potential bias due to dropout (that is, unit non-response) for each study sample. The item non-response were analysed according to the study to make sure the data was not biased and there were no differences between complete datasets and incomplete data sets (those with missing values). For univariate and multivariate regression analysis, the complete case approaches were considered, missing values from each were removed. Detail on representativeness and item-nonresponse are explained within individual sections of Chapter 5, as per the requirements of the analysis. Correlations were calculated using PROC CORR procedure to check for multicolinearity problems and Pearson product-moment correlation coefficient for variables were calculated.

Assumptions of the linear regression model were tested by evaluating plots of residual vs. predicted values, reviewing scatter plot matrix showing all dependent against all independent variables. A two-sided $P < 0.05$ were used to denote statistical significance. The Benjamin-Hochberg procedure was considered for false discovery rate at $P < 0.05$ which was statistically significant [178]. Bonferroni corrections for multiple comparisons were calculated, with $P < 0.05$ considered as statistically significant.

4.5.1.3 Multiple imputation

Multiple imputations were considered to recover the statistical power and reduce selection bias due to item-nonresponse in the full model. The number of imputations (m=20) were based on the recommendations by Graham et al [179] for a data set with the fraction of missing information of 10-30% and a low (<1%) tolerance for power fall-off. Imputed data sets were analysed using multiple regression with SAS PROC REG, and analysis results were pooled together for inference with SAS PROC MIANALYZE. Earlier work done in a simulation study in the NFBC 1966 demonstrated how MI can reduce bias and increase statistical power when data is missing at random, and even when data is not missing at random [180]. The association analyses of the NFBC1986 data did not show any difference between the results based on imputed dataset compared to those based on the non-imputed dataset. Therefore, I do not use the imputed data for my analyses.

4.5.1.4 Mediation

Mediation analyses were conducted using investigative methods described by Baron and Kenny[181] as well as formal significance tests of the indirect effect by the Sobel test[182]. According to Baron and Kenny, 1986, there is a set of three criteria which need to be fulfilled to establish mediation: 1) the exposure significantly associated with the outcome, 2) the outcome significantly associated with the mediator, and 3) that the mediator substantially reduces the estimate coefficient for the association between outcome and exposure. If the effect size is reduced by >10% then the partial mediation is considered, however, if mediator-adjusted coefficient of the exposure approaches null (0) then complete mediation is considered.

Sobel test was used to test the mediation effect in the models. Sobel, et al, 1981, hypothesised that the association between independent variables and dependent variables is in-directly via the mediator. When the mediator is added to the model, between the independent variables and dependent variables, the effect of association is decreased showing the mediation. An online webpage for Sobel test were used to assess if the mediation effects were significant (http://quantpsy.org/sobel/sobel.htm).

Over the years, more sophisticated analytical methods have been developed for testing mediation, and Baron and Kenny's (with Sobel's test) casual steps approach has been criticized on multiple grounds[183].

Importantly, it has been shown from stimulation studies that testing for intervening variable effects reduces the power and that the quantification of casual steps is not based on the intervening effect. More recently, there have been proposals for several new analytical opportunities for testing mediation, for example, bootstrapping[184] and *M*-test[185]. Although these methods have been more robust in calculating mediation, it was beyond the scope of this thesis to compare results from Baron and Kenny's method against these new methodologies.

4.5.2 Genetic association analysis

A detailed description of how genome-wide association studies (GWAS) has been previously described by Bush et al, 2012[186]. A detailed analysis plan was prepared for each genetic association study including the motivation for the analysis and the planned study design (see Appendix Note 4). Briefly, two genetic association analyses on head circumference (HC) and early growth phenotypes were conducted. These analyses were divided into three stages: i) Stage 1 or discovery stage consisting of meta-analysis of studies with genome-wide data available and ii) Stage 2, replication or follow-up, consisting of a meta-analysis on leading SNPs (selected from discovery stage), and iii) Stage 3 or combined stage consisting of the procedure when discovery stage and replication is meta-analysed together.

At discovery stage, each study applies a Hardy-Weinberg Equilibrium (HWE) cut-off within the study and an overall HWE $P > 1 \times 10^{-4}$ cut-off for overall discovery, meta-analysed data was also applied. SNPs with poor imputation quality were excluded if minor allele frequency (MAF) < 5%, Call rate < 98%, $r^2 < 0.3$ for MACH and proper info<0.4 for IMPUTE). The additive association analysis was run using bioinformatics software such as PLINK[187] or SNPTEST[188], which supports various statistical alternatives for analysis. The discovery study beta estimates with their standard errors(SE) were meta-analysed using the inverse-variance weighting method in the METAL software[189]. Double genomic control[190] was applied to the individual study results in Stage 1 to adjust the statistics generated within each cohort and then adjust the overall Stage 1 meta-analysis statistics. A Manhattan plot is created to identify top SNP *P*-values from the discovery meta-analysis. This is a scatter plot with X-axis showing the genomic co-ordinates' and Y-axis showing the negative logarithm (denoted by $\log_{-10}(P)$) of the associated SNP *P*-value for each SNP[191]. I have also used regional association plots to show individual SNPs in the region of interest. A regional association plot is similar to a Manhattan plot but uses a limited section of genome with X-axis that shows the position of the SNP on the chromosome and the Y-axis shows the negative logarithm for the associated SNP *P*-value[192]. Study heterogeneity, within each study, were quantified to determine the degree to which studies differ. For study heterogeneity, the *Q* statistics and the I^2 index, which represents the approximate proportion of the

variability in coefficient, were calculated. The I^2 values were categorized as, low (<25), medium (>25 and <75), and high (>75) heterogeneity. If high study heterogeneity were observed, then associations for overall directional consistency between all the studies were checked. Furthermore, GWAS analyses removing studies contributing to wide confidence interval were run to see if results were driven only by them. Finally, both fixed and random effect models were tested to check the difference in results.

In Stage 2, SNPs are followed up in individuals from independent dataset with *in silico* or *de novo* genotyped data available for replication. All Stage 2 studies provided beta estimates and standard errors for association. The effect estimates were oriented to the forward strand of the NCBI build 36 reference sequence of the human genome, consistently using the alphabetically higher allele as the coded allele. Effect estimates and standard errors were combined across the stage 2 studies using an inverse-variance weighting method in the METAL software[189]. To address the problem of winner's curse, the replication sample size was kept to a sufficient size to detect the effect of susceptibility allele. Same GWAS design was applied as closely as possible in the replication sample.

In the combined stage, a meta-analysis of Stage 1 and 2 results was undertaken using inverse-variance weighting method using two software: METAL software[189] and usually replicated using another software, such as R software program (version 2.14.2; http://www.r-project.org/), to remove any errors or bias due to software. Associations with $P < 5 \times 10^{-8}$ were considered as genome-wide significant and $P< 5 \times 10^{-6}$ were considered as suggestive association. Informed consent was obtained from all Stage 1 and Stage 2 study participants (or parental consent, as appropriate). Data exchange facilities were provided by the AIMS server[193]. All study protocols are approved by the local ethics committees as appropriate.

4.6 Software

For the purpose of growth modeling, I used R software (http://www.r-project.org/). R software is a free software environment for statistical computing and graphics. It runs on the UNIX platform or in Windows platform. The quantile-quantile plots, Manhattan plots and forest plots for genetic association studies were developed using packages from R software. For regional plot, I used the LocusZoom tool. LocusZoom is a tool to plot regional association results from GWAS. For other statistical analysis, for example, descriptive data analysis or association analysis, I used SAS version 9.1, software (SAS Institute, Inc. Cary, North Carolina).

In genetic association studies, SNPTEST[188] version2.5 program was used for genome-wide linear additive models[188]. The METAL software was utilised for meta-analysis of large datasets[189].

4.7 Summary

This chapter describes the two main datasets used for analysis, the Northern Finland Birth Cohorts (NFBCs) 1966 and 1986, referred to as NFBC1966 and NFBC1986. It also describes the NFBC1986 growth modeling and the restrictions applied to the number of individuals selected for growth modeling. The availability of dense longitudinal data from the NFBCs 1966 and 1986 is one of the main strengths of this thesis. It provided the platform necessary to model growth data and derives biologically meaningful parameters for further analysis.

As part of my training for this doctoral work, I undertook extensive research to understand the various types of models and comparisons for both NFBC datasets. However, only relevant information has been included in this thesis. It was most useful to have Dr Ulla Sovio's expertise and guidance during the early days of growth modeling as my initial work used her NFBC1966 methodology for fitting models. The growth modeling was an essential part of this thesis and this chapter gives a brief outline of the models. Extensive details were already published as part of my master's dissertation[194] and also as part of Dr. Ulla Sovio's doctoral thesis[11,172].

A short description of genetic data from NFBCs (1966 and 1986) is given in this chapter. The imputation and quality control of overall genetic data for both NFBCs was performed by other team members at Imperial College London and elsewhere; hence, only relevant information about study-specific quality control is described in this chapter. Contributions of other cohorts from EGG consortium for genetic association analysis are described in the chapter. I provided advice and contributed to the development of growth models for several cohorts participating in genetic analysis studies.

The statistical methods described in this chapter relate to the general approach taken for this thesis. Study-specific statistical approaches have been described in results section as appropriate.

Chapter 5: Study-specific results and discussions

This chapter presents results from association and genetic association studies conducted on prenatal predictors and various outcomes at birth and during infancy and childhood. The section (5.1.1) describes an association analysis between various prenatal predictors, size-at-birth and peak height velocity (PHV), with an emphasis on mediation by size-at-birth variables on PHV in infancy in NFBC 1986. Section 5.1.2 describes an investigation into the association study between pre-pregnancy maternal body mass index (BMI) and adolescent metabolic factors. In this section, I describe work performed in collaboration with researchers from King's College London. This work was designed as a translational study, for which I (as part of the team from Imperial college, London) performed analysis on human data, discussed in this chapter, and the team from King's College, London, performed similar analyses on murine data (not discussed here).Section 5.1.3 describes an investigation on leukocyte telomere length (LTL) and adiposity-related factors in childhood and adulthood in NFBC1966. I describe an association analysis performed on mean LTL and adiposity related measures in childhood and adulthood. This work was performed in collaboration with Dr Jess Buxton, Imperial College London, to be submitted as a scientific paper to PlosONE in 2014. The genetics association studies (section 5.2) present the genetic association for head circumference and early growth phenotypes in various populations.

5. 1 Association analyses

5.1.1. Association between prenatal predictors and postnatal height growth velocity in infancy is mediated by size-at-birth

5.1.1.1 Introduction

Fetal life is an important stage of development for future adult health[2]. Adverse factors during fetal life can lead to smaller size-at-birth and accelerated postnatal growth in infancy[96,195]. Both of these associate with an increased risk of cardiovascular disease (CVD), obesity and type 2 diabetes in adulthood[66,196]. Several epidemiological studies have reported prenatal risk factors for small size-at-birth such as lower maternal socio-economic status (SES), maternal smoking, deprived living conditions, under-nutrition and infections during pregnancy[64,189,197]. However, none have considered the relationships between these prenatal risk factors and postnatal growth in infancy with size-at-birth variables as potential mediators, in order to disentangle the potential causal pathway.

This study hypothesises that multiple prenatal predictors are associated with accelerated height growth in infancy, and may have their effect mediated through intra-uterine environment. The size-at-birth variables;

birth weight (BW), birth length (BL) and birth body mass index (BMI), were used as a proxy for the intra-uterine environment[198]. Several epidemiological studies associate size-at-birth (BW, BL and birth BMI) with various adverse risk factors in mothers[199] and with the long term health of their offspring[200]. The study uses the NFBC1986 data to evaluate the association between prenatal predictors, size-at-birth and accelerated growth in infancy, quantified as peak height velocity (PHV). A detailed description on how the NFBC 1986 data was used to model PHV is described in Section 4.3.3. The main objective of this study is to understand the pathway(s) that links prenatal factors with accelerated growth in infancy. Previous studies show that accelerated postnatal growth associates with adverse cardio-vascular health, therefore it would be an important risk factor to study[201]. Consequently, the aim of the study was to investigate the relationship between multiple prenatal predictors and PHV, and PHV's association with size-at-birth variables. Additionally, the study also investigates if the association is indirect via size-at-birth with size-at-birth acting as a mediator.

5.1.1.2. Study specific material and methods

Details on the collection of variables are given in Chapter 4. The variables were selected based on information from previously published epidemiological studies and expert knowledge. The study sample (n=5,577) was representative of the full cohort, as shown in **Representativeness**. More detailed information on missingness and the analysis on item non-response in both cohorts are in the **Item non-response**.

Representativeness

The attributes of the NFBC 1986 individuals included in the analyses (N=5,577) were evaluated by comparing the included group with the remainder of the cohort traced and alive at the time of invitation to the clinical examination at age 15/16 years (N=3,171). The proportion of children of single mothers (16% vs. 21%), mothers who had completed only primary education (25% vs. 29%) and smoking mothers (20% vs. 26%) was lower in the included group than in the remainder. The proportion of mothers in manual work (blue collar; 33% vs. 38%) and fathers in manual work (14% vs. 18%) was also lower in the included group. There was a trend towards a higher number of previous deliveries in the mothers of the included children compared to the remainder of the cohort while no significant trend was observed with respect to maternal BMI at $P < 0.05$. Maternal height and hypertensive complications were distributed equally and did not differ in the inclusion group and the remainder (P=0.30). Mothers of included children were on average eight months older than the rest of the mothers. The proportion of girls was higher in the included group than in the remainder of the cohort (51% vs. 45%). The child's BL, birth BMI and GA did not differ between the included and remaining group in the analyses (P=0.64). The overall results suggested that the included

individuals did not differ significantly from the remainder of the cohort, showing a good sample of the population.

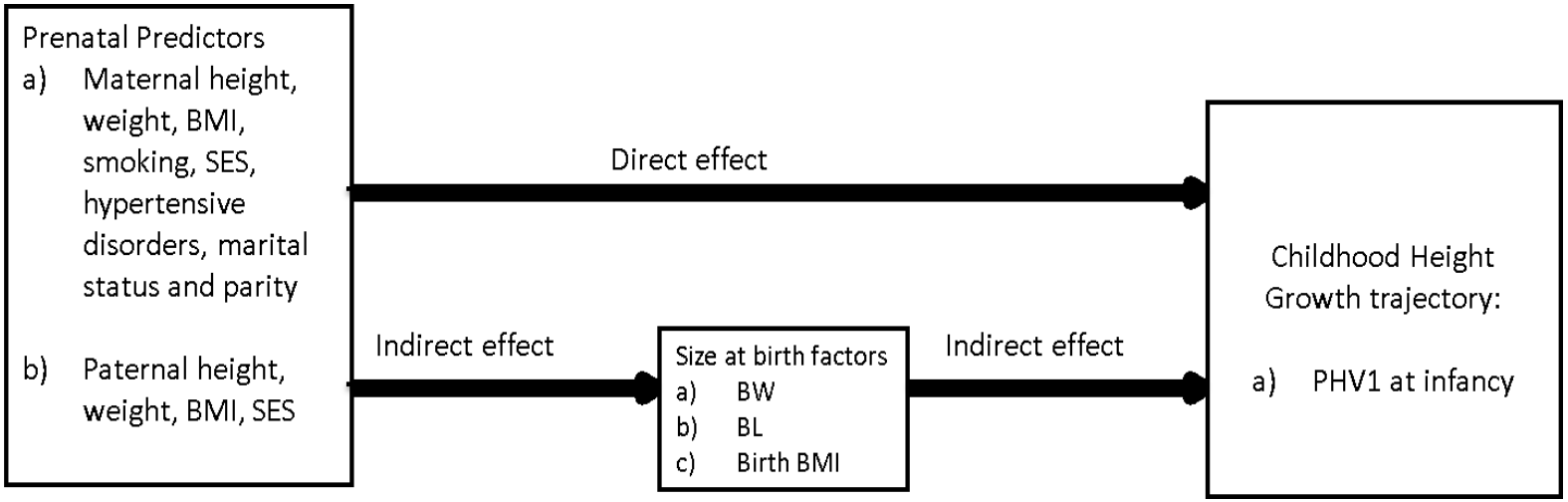
Item non-responses or missing data

The item non-response of the NFBC 1986 individuals with complete dataset (N=5,577) was evaluated by comparing individuals who had complete data on all variables selected for analysis (complete cases) with those who had missing information on one or more variables (incomplete cases). As shown in Table 4, mother's age at pregnancy, weight, height, pre-pregnancy BMI had a variable range for missing data (between 0.9% to 2.3%), however overall the mean for complete cases and incomplete cases did not differ ($P < 0.05$).The proportion of children of married mothers (85% vs. 74%) and mothers with secondary education (55% vs. 40%), and the proportion of children of farmer mothers (6% vs. 2%) and fathers (8% vs. 3%) was higher in complete cases than in incomplete cases. The mothers of children who had complete data were less often nulliparous (32% vs. 37%) and more often primiparous (34% vs. 29%) than the mothers of children who had incomplete data. Maternal smoking was lower in complete cases than in incomplete cases (19% vs. 23%). Children with complete data were on average 0.2 cm longer and they were born 0.2 weeks later than children with incomplete data.

Statistical analysis

Descriptive statistics are presented as the mean and standard deviations (SD) for continuous variables and percentages of categorical variables are in the available dataset (**Table 4**). PHV was log-transformed for the association analyses since it had a right-skewed distribution, therefore, results were presented as the percentage change in PHV per one unit change in the predictor variable. To account for the missing measurements in the derived growth parameters, the regression models were weighted by the number of measurements per person assuming a linear relationship between the number of measurements and the accuracy of estimated PHV. A weighted linear regression allows the parameter estimates to determine the contribution of each observation to the final parameter estimates, hence accounting for the assumptions made during growth modeling. All the analyses were run separately for males and females due to their different growth trajectories (Appendix Note 2). A two-sided $P < 0.05$ was used for statistical significance in all association tests.

Figure 5. A framework for the mechanisms in the association between the prenatal predictors and PHV.



For tests of association, a multivariate linear regression model was used. The first step was to examine the direct association between the outcome with potential predictors and mediator (**Figure 5**). The first set of models used the Univariable model between PHV with prenatal predictors and size-at-birth variables (**Table 5**). The second set was multivariable models between prenatal predictors and size-at-birth variables, adjusted for gestational age (GA), presented in **Table 6**.

Finally for testing mediation, Baron and Kenny's method of mediation[181] was applied in the analysis. Details on Baron and Kenny's method have been described in Section 4.2.3.4. The first two steps of association analysis (**Table 5 and 6**) supported the first two steps of mediation by Baron and Kenny. For the third step, four separate models were fitted to examine mediation roles by size-at-birth variables between prenatal predictors and PHV (see **Figure 5, Table 7**). Model 1 included PHV association with the prenatal predictor adjusted for gestational age (GA); Model 2 included the model 1 with BW as mediator and adjusted for GA; Model 3 included the model 1 with BL as mediators and adjusted for GA; and finally model 4 included model 1 with birth BMI as mediator and adjusted for GA. The size-at-birth variables were fitted separately to distinguish the roles of BW, BL and birth BMI separately and adjusted for GA as confounder for each size-at-birth variable. Reviews of previous studies show that GA has been associated with both prenatal predictors and size-at-birth variables, hence it was included in the analyses as a confounder[202]. Sobel's test was used to test the significance of mediation association with $P_{sobel}<0.05$ considered as statistically significant.

Table 4. Characteristics of the NFBC 1986 data set.

NFBC 1986 (n=5,577)					
	All (n=5577)	**Males[a] (n=2741)**	**Females[a] (n=2836)**	**Missing (%)**	***P*[b]**
Maternal					
Age at pregnancy, years	28.0(5.5)	28.0(5.4)	28.0(5.5)		0.8
Weight, kg	59.5(9.6)	59.3(9.5)	59.6(9.7)	1.8	0.26
Height, cm	163.2(5.5)	163.3(5.4)	163.1(5.5)	0.9	0.3
Pre-pregnancy body mass index, kg/m^2 [c]	22.3 (3.4)	22.2(3.3)	22.4(3.4)	2.3	0.11
% Underweight (<18.5)	5	7	7		
% Ideal weight (18.5-24.9)	79	75	74		
% Overweight (>24.9)	9	13	13		
% Obese (>30)	5	3	4		
% Marital Status				0.2	0.44
Married	83	83	82		
Unmarried/cohabiting	16	16	17		
Divorced/widow	1	1	1		
% Socio-economic status				3	0.15
White collar	26	27	28		
Blue collar	66	66	64		
Farmer	5	5	5		
% Education				11.9	0.32
Primary education	8	8	9		
Vocational education	28	25	25		
Secondary education	53	47	47		
University	9	7	8		
% Smoking at 2 months of pregnancy				0.5	0.54
No	80	80	80		
Smoker	20	19	20		
% Previous Deliveries				0.4	0.28
No previous deliveries	33	33	33		
1 previous delivery	33	32	34		
2 previous deliveries	18	19	18		
3 previous deliveries	6	6	6		
More than 3 previous deliveries	9	9	9		
% Hypertensive complications in pregnancy				1.8	0.25
Normotensive	78	78	78		
Gestational hypertension	5	5	5		
Pre-eclampsia	3	3	3		
Chronic hypertension	5	5	5		
Superimposed pre-eclampsia	2	1	2		
Proteinuria	5	5	6		

Paternal					
Weight, kg	75.1(9.6)	75.1(9.4)	75.1(9.8)		0.85
Height, cm	176.8(6.2)	176.9(6.1)	176.6(6.3)		0.06
Body mass index, kg/m^2 [c]	24.0(2.7)	24.0(2.7)	24.1(2.7)		0.39
% Underweight (<18.5)	16	17	16		
% Ideal weight (18.5-24.9)	58	57	58		
% Overweight (>24.9)	24	24	24		
% Obese (>30)	2	2	3		
% Socio-economic status				5.4	0.98
White collar	29	32	32		
Blue collar	59	56	56		
Farmer	7	6	7		
Infant					
Birth weight, kg	3.6(0.5)	3.6(0.5)	3.5(0.5)		<0.0001
Birth length, cm	50.6(2.2)	51.0(2.1)	50.2(2.1)	0.7	<0.0001
Body mass index, kg/m^2 [c]	14.0 (1.3)	14.0(1.3)	14.0(1.3)	0.7	0.38
% Underweight (<10th percentile)	6	5	6		
% Ideal weight (10th-80th percentile)	80	79	80		
% Overweight (>80th percentile)	10	10	10		
% Obese (>90th percentile)	5	5	5		
Gestation age, weeks	39.5(1.5)	39.4(1.5)	39.5(1.5)	0	0.32
PHV, cm/year	50.6 (2.2)	55.3 (6.7)	51.2 (7.1)	0	<0.0001

[a] Values for continuous variables are given as means and standard deviations (SD) and categorical variable as percentage.
[b] *P* for comparison between the sexes.
[c] Calculated from the IOTF on BMI classification[203]

5.1.1.3 Results

5.1.1.3.1 Subject Characteristics

Table 4 shows the details on prenatal predictors, size-at-birth and PHV characteristics for an overall study sample and then by males and females in the NFBC 1986. The mothers' average age at pregnancy was 28.0 years in the overall sample. Pre-pregnancy weights and height in mother were similar in both male and female offspring (*P*=0.26; *P*=0.30; respectively). Mothers of both sexes were, on average, 163.2 cm tall and weighed 59.5 kg before pregnancy. The average pre-pregnancy BMI was 22.3 kg/m2 overall, and no difference between mother's BMI of males and females was observed (*P*=0.11). 79% mothers were of normal weight, 5% were underweight, 9% were overweight and 5% were obese (using WHO cut-offs). High proportions of mothers were married (83%), with 16% of mothers single or cohabiting and 1% divorced or widowed. In the socio-economic status (SES), a high proportion of mothers (66%) were classified as blue collar workers, 26% were white collar and 5% of the mothers were farmers. Most mothers were educated to secondary level (53%), 8% had at least primary education, 28% attained vocational education and 9% had been to university. The proportion of mothers who smoked during the 2nd month of pregnancy were similar in both males and females (*P*=0.54) with only 20% of mothers smoking while 80% were non-smokers.

Both nulliparous (no previous deliveries) and primiparous (one previous delivery) mothers were equally distributed (33%). About 18% of mothers had 2 previous deliveries, with 6% of mothers having 3 previous deliveries and 9% of mothers with more than 3 previous deliveries. No difference in the distributions of the groups was observed between sex (*P*=0.28). The hypertensive complications during pregnancy were similarly distributed in both males and females (*P*=0.25) with approximately 78% of mothers normotensive, 5% of mothers had gestational hypertension, 3% had pre-eclampsia, 5% had chronic hypertension, 2% had superimposed pre-eclampsia and 5% had proteinuria.

Father's weight, height and BMI were similar in both males and females. Fathers weighed an average of 75.1 kg, had an average height of 176.8 cm and an average BMI of 24.0 kg/m^2. 58% of fathers were classified as normal weight, 16% as underweight, 24% as overweight and 2% as obese, with the distribution similar in both sexes (*P*=0.39). Fathers' SES was distributed as: 59% blue collar, 29% white collar and 7% farmers. Their education and smoking status were not recorded in NFBC 1986, therefore unavailable for analysis.

The mean birth weight (BW) was 3.6 kg for all the participants from the study, and BW in males and females differed by 100 g (P <0.0001). The mean birth length (BL) was 50.6 cm, with males being 8 mm taller than females (P < 0.0001). The average birth body mass index (BMI) in the overall study sample was 14.0 kg/m2. The birth BMI was not significantly different in males and females (P=0.38), with up to 80% children being normal weight, 5% underweight, 10% overweight and 5% obese, the cut-offs were selected according to IOTF guidelines[203]. The average gestational age (GA) was 39.5 weeks in all samples. The mean PHV was 50.6 cm/year in overall study population, which differed between sexes, with an average of 55.3 cm/year in males and 51.2 cm/year in females (P < 0.0001).

Table 5. Univariate association of infant PHV with prenatal and size at birth variables.

NFBC 1986											
	Males						Females				
	B[#]	se	95% CI[#]		*P*[$]		B[#]	se	95% CI[#]		*P*[$]
Maternal											
Age, years	0.00	0.04	0.00	0.00	0.9876		-0.15	0.05	-0.24	-0.06	<.0001
Weight, kg	-0.01	0.01	-0.03	-0.05	0.4274		-0.03	0.03	-0.08	0.02	0.2727
Height, cm	0.32	0.04	0.24	0.50	<.0001		0.32	0.05	0.23	0.41	<.0001
Body mass index, kg/m^2 (ref = normal weight)											
Underweight	0.13	0.90	-1.63	1.90	0.8827		2.93	1.00	0.97	4.90	0.0037
Overweight	-1.94	0.69	-3.30	-5.77	0.0046		-0.28	0.77	-1.79	1.24	0.7156
Obese	-4.89	1.33	-7.50	-13.37	0.0002		-2.67	1.34	-5.29	-0.05	0.0417
Marital Status (ref = married)											
Unmarried/cohabiting	2.04	0.62	0.83	2.24	0.0011		2.00	0.67	0.68	3.32	0.0031
Divorced/widow	-0.11	2.61	-5.21	-7.61	0.9664		1.77	2.43	-2.99	6.53	0.4643
SES (ref = white collar)											
Blue collar	0.88	0.53	-0.15	0.23	0.0949		-0.24	0.58	-1.39	0.90	0.6753
Farmer	-1.52	1.19	-3.86	-6.37	0.1953		-2.29	1.28	-4.79	0.22	0.0683
Education (ref=primary)											
Vocational education	-0.53	0.97	-2.43	-3.79	0.5808		-1.51	1.01	-3.49	0.46	0.1279
Secondary education	0.92	0.92	-0.87	-0.79	0.3129		-0.90	0.94	-2.75	0.94	0.3323
University	0.03	1.21	-2.34	-3.38	0.9795		-2.06	1.26	-4.54	0.42	0.0976
Smoking at 2 months of pregnancy (ref=no)											
Smoker	0.94	0.25	0.45	1.12	0.0002		2.95	0.64	1.71	4.20	<.0001
Previous Deliveries (ref=no)											
1 previous delivery	-2.11	0.56	-3.20	-5.71	0.0001		-3.73	0.61	-4.92	-2.54	<.0001
2 previous deliveries	-2.10	0.65	-3.38	-5.98	0.0011		-4.04	0.74	-5.49	-2.60	<.0001
3 previous deliveries	-4.74	1.00	-6.70	-12.14	<.0001		-5.91	1.12	-8.11	-3.71	<.0001
More than 3 previous deliveries	-4.03	0.93	-5.86	-10.56	<.0001		-7.31	1.01	-9.29	-5.33	<.0001
Hypertensive complications in pregnancy (ref=normotensive)											
Gestational hypertension	1.61	1.31	-0.96	4.19	0.2195		0.78	1.58	-2.30	3.87	0.6177
Pre-eclampsia	8.62	1.58	5.52	12.40	<.0001		12.48	1.79	8.98	15.99	<.0001

Chronic hypertension	3.34	1.52	0.37	2.24	0.0292		3.19	1.79	-0.32	6.70	0.0772
Superimposed pre-eclampsia	2.02	2.79	-3.45	-3.97	0.4676		0.81	2.23	-3.56	5.17	0.7157
Proteinuria	-0.67	1.09	-2.81	-4.41	0.5380		0.22	1.09	-1.92	2.35	0.8409
Paternal											
Weight, kg	0.03	0.01	0.01	0.03	0.0034		0.36	0.04	0.27	0.44	<.0001
Height, cm	0.13	0.02	0.09	0.20	<.0001		0.13	0.03	0.07	0.18	<.0001
Body mass index, kg/m^2 (ref=normal)											
Underweight	1.13	0.73	-0.30	0.14	0.1223		1.13	0.73	-0.30	2.56	0.1223
Overweight	-0.55	0.63	-1.78	-2.87	0.3729		-0.55	0.63	-1.78	0.67	0.3729
Obese	-1.31	1.67	-4.58	-7.32	0.4253		-1.31	1.67	-4.58	1.96	0.4253
SES (ref=white collar)											
Blue collar	0.07	0.51	-0.93	-1.31	0.8930		0.50	0.57	-0.62	1.62	0.3826
Farmer	-1.46	1.01	-3.44	-5.74	0.1453		-0.41	1.11	-2.59	1.76	0.7076
Infant											
Birth weight (BW), kg	-4.12	0.17	-4.46	-8.57	<.0001		-10.52	0.45	-11.41	-9.63	<.0001
Birth length (BL), cm	-1.15	0.04	-1.23	-2.37	<.0001		-3.05	0.00	-3.05	-3.05	<.0001
Birth body mass index (BMI), Kg/m2	-1.05	0.08	-1.20	-2.28	<.0001		-2.41	0.00	-2.41	-2.40	<.0001

[#]Beta, se and 95% CI were presented in % change as PHV was log-transformed to reduce right skewness.

[$]Significance was considered at $P<0.05$.

5.1.2.2 Sex stratified association between prenatal factors, size-at-birth variables and PHV (Table 5)

Table 5 shows that an increase in mother's age associated with 0.15% lower PHV in NFBC 1986 females ($P < 0.0001$), but did not associate in males ($P=0.98$). Mother's weight did not associate with PHV in both sexes ($P=0.42$ and $P=0.27$, respectively). However, mother's height showed a strong positive association with PHV in both sexes ($P<0.0001$ for both), confirming previous findings that mother's height may contribute via the intrauterine environment or via genetic influence to the development of infant height growth[204]. Mother's higher pre-pregnancy BMI groups showed an inverse association with PHV in both set of offspring, using the normal group as reference. Males born to overweight or obese mothers had lower PHV than offspring of normal weight mothers (β=-1.94%, CI: (-3.30, -5.77) and β=-4.89%, CI: (-7.50, -13.37); respectively), but offspring of underweight mothers did not associate with PHV ($P=0.88$). Females born to underweight mothers had higher PHV (B=2.93%, CI: (0.97, 4.90)) than offspring of normal mothers, while female offspring of obese mothers had 2.67% lower PHV than offspring of normal mothers, supporting maternal weight hypothesis[205]. Mother's marital status associated with PHV, and children of single or cohabiting mothers had 2% higher PHV than children of married mothers in both sexes ($P=0.001$ and $P=0.003$, respectively). Their SES or education did not significantly associate with PHV in either of the sexes (**Table 5**).

Mothers' smoking during the initial 2 months of pregnancy strongly associated with PHV in both sexes, male offspring of mothers who smoked had 0.94% higher PHV than offspring of non-smoker mothers ($P=0.0002$). The strength of the association was much stronger in females, where offspring of smokers had 2.95% higher PHV than offspring of non-smokers ($P<0.0001$). This is the first study which associates maternal smoking with PHV, suggesting a link between smoking during pregnancy, intrauterine growth retardation (IUGR) and accelerated growth after birth to catch-up to their genetic potential. This study confirms that male and female offspring of smoking mothers had different associations[206]. There was a clear trend towards a lower PHV by the increasing number of previous deliveries in both ($P_{trend}<0.0001$ for both males and females), this fits well with the previous results that bigger babies grow more slowly[207]. The hypertensive complications in pregnancy associated with PHV, both maternal pre-eclampsia and chronic hypertension predicted a higher PHV in males ($P<0.0001$ and $P=0.03$, respectively), but only pre-eclampsia associated with higher PHV in females ($P<0.0001$).

As expected, father's higher weight and height were both associated with higher PHV in both sexes, as height and weight are both genetically determined phenotypes[204]. Although, both father's BMI and SES did not associate with PHV in the sexes at $P<0.05$ significance.

All three size-at-birth variables (BW, BL and birth BMI) had a stronger inverse association with PHV in both sexes (**Table 5**). For each 1kg increase in BW, PHV was on average 4.12% lower in males (β = -4.15%, CI: (-4.46, -3.78), $P < 0.0001$) and 10.52% lower in females (β = -10.52%, 95% CI: (-11.41, -9.63), $P < 0.0001$). Similarly, for each 1cm increase in BL, PHV was reduced by 1.15% in males (β= -1.15, 95%CI: (-1.23, -1.07), $P < 0.0001$) and 3.05% for females (β=-3.05, 95% CI: (-3.26, -2.84), $P < 0.0001$). Birth BMI was also inversely associated with PHV in both sexes, with every 1 kg/m2 increase in Birth BMI, PHV reduced by 1.05% in males (β=-1.05%, 95% CI: (-1.20, -0.90) , and $P < 0.0001$ and 2.41% (β=-2.41%, 95% CI: (-2.78, -2.03) , $P < 0.0001$) in females (**Table 5**).

5.1.2.3 Sex stratified association between prenatal predictors and size-at-birth factors adjusted for gestational age (Table 6, A and B)

In both sexes, maternal age was associated with higher BW, BL and birth BMI ($P < 0.0001$, $P = 0.03$ and $P<0.0001$, respectively), consistent with previous findings suggesting that with age mothers gain weight and give birth to bigger babies[208]. Maternal weight, height and pre-pregnancy BMI were positively associated with size-at-birth variables in both sexes ($P <0.05$ for all) (**Table 6**), except the association between offspring of obese mothers did not associate with BL in males ($P=0.19$) and female offspring of overweight mothers did not associate with BL ($P=0.07$). Mother's marital status associated with all size-at-birth variables significantly ($P<0.0001$ in both sexes), with results showing that unmarried or cohabiting mothers gave birth to smaller size offspring in comparison to married mothers in both sexes (**Table 6A**). The maternal SES in both male and female offspring, associated negatively with BW ($P=0.03$ in males and $P=0.01$ in females) and with BL ($P=0.04$ in males and $P=0.0009$ in females) but did not associate with birth BMI ($P=0.12$ in males, P=0.66 in females), especially blue collar mothers who had smaller BW and BL in their offspring compared to the white collar mothers in both sexes (**Table 6A**). These results are in line with previous studies suggesting a link between

mother's SES and smaller size-at-birth[209]. In female offspring, mother's education associated positively with BL (P=0.009) but negatively with birth BMI (P=0.04) in female offspring, while no association was observed in male offspring (P>0.05 for both). No association between BW and maternal education was observed in both sexes (P>0.05) (**Table 6A and B**).

Maternal smoking status during the 2nd month of pregnancy was significantly associated with all three size-at-birth variables in both sexes (P<0.0001 for BW and BL in both sexes, P=0.0002 for birth BMI in males and P<0.0001 in females), however the strength of the association was stronger in females than males (**Table 6A and B**). Mother's parity (previous delivery) significantly associated with all size-at-birth variables in both sexes (P_{trend} for all < 0.001), except for male offspring of mothers who had 3 or more pregnancies, which did not associate with BL (P=0.12). Mother's hypertensive complications during pregnancy associated with size-at-birth factors, but the association varied positively and negatively for each category separately between males and females. For example, mother's gestational hypertension associated with 0.1 kg smaller BW (P=0.003) and 0.35 kg/m2 lower birth BMI (P=0.0006) in males but not in females (P=0.36 and P=0.54). Mother's pre-eclampsia strongly associated with smaller size-at-birth variables in both sexes (**Table 6A and B**), which shows consistency with previous findings[57]. In male offspring chronic hypertension associated with smaller BW (P<0.0001), while in female offspring proteinuria in mothers were associated with BW (P=0.030) and birth BMI (P=0.007).

Similarly to mothers, the fathers weight and height was associated with all size-at-birth variables in both sexes except for father's height which did not associate with birth BMI in both sexes (P=0.27 in males and P=0.36 in females). Paternal BMI associated with all size-at-birth variables in females (P<0.05) but only with BW and birth BMI in males (see **Table 6A and B**). Male offspring of overweight fathers had higher BW and birth BMI, while female offspring of overweight fathers was associated positively with all size-at-birth variables at P<0.05 significance. Similarly, male offspring of obese fathers had 0.13 kg higher BW than offspring of normal weight fathers (P=0.03). Father's SES associated with smaller BW in females (P=0.03) but no association with males was observed, suggesting that females from lower SES may be at a more adverse risk than males.

Table 6: A. Association between prenatal predictors and size-at-birth factors (birth weight (BW), birth length (BL) and birth BMI), adjusted for gestational age (GA) in NFBC 1986 Males.

A	**NFBC 1986 Males (N=2741)**								
	Birth weight (Kg)			**Birth Length (m)**			**Birth BMI (kg/m2)**		
	β	**95% CI**	***P***	**β**	**95% CI**	***P***	**β**	**95% CI**	***P***
Maternal									
Age, years	0.01	(0.01, 0.01)	<.0001	0.02	(0.00, 0.03)	0.0314	0.02	(0.01, 0.03)	0.0002
Weight, kg	0.01	(0.01,0.01)	<.0001	0.03	(0.02, 0.04)	<.0001	0.02	(0.02, 0.03)	<.0001
Height, cm	0.01	(0.01, 0.02)	<.0001	0.07	(0.06, 0.08)	<.0001	0.01	(0.01, 0.02)	0.0009
Body mass index, kg/m2 (ref= normal weight)									
Underweight	-0.13	(-0.20, -0.07)	<.0001	-0.29	(-0.55, -0.02)	0.0358	-0.32	(-0.50, -0.15)	0.0003
Overweight	0.10	(0.05, 0.15)	<.0001	0.28	(0.07, 0.48)	0.0087	0.25	(0.11, 0.38)	0.0003
Obese	0.19	(0.09, 0.29)	0.0001	0.26	(-0.14, 0.66)	0.1995	0.60	(0.34, 0.85)	<.0001
Marital Status (ref=married)									
Unmarried/cohabiting	-0.15	(-0.20, -0.11)	<.0001	-0.53	(-0.71, -0.34)	<.0001	-0.31	(-0.43, -0.19)	<.0001
Divorced/widow	-0.14	(-0.32, 0.05)	0.1562	-0.11	(-0.87, 0.65)	0.7812	-0.52	(-1.01, -0.02)	0.0410
SES (ref=white collar)									
Blue collar	-0.04	(-0.08, -0.01)	0.0265	-0.16	(-0.31, 0.00)	0.0491	-0.08	(-0.18, 0.02)	0.1186
Farmer	0.05	(-0.03, 0.14)	0.2220	0.01	(-0.34, 0.36)	0.9536	0.21	(-0.01, 0.44)	0.0670
Education (ref=primary education)									
Vocational education	-0.06	(-0.13, 0.01)	0.0805	-0.14	(-0.43, 0.14)	0.3219	-0.01	(-0.18, 0.16)	0.9009
Secondary education	-0.05	(-0.12, 0.02)	0.1394	-0.04	(-0.31, 0.23)	0.7899	-0.02	(-0.19, 0.14)	0.7630
University	-0.02	(-0.11, 0.07)	0.6421	0.14	(-0.22, 0.50)	0.4380	-0.05	(-0.27, 0.17)	0.6624
Smoking at 2 months of pregnancy (ref= non-smoker)									
Smoker	-0.14	(-0.18, -0.10)	<.0001	-0.62	(-0.79, -0.45)	<.0001	-0.21	(-0.32, -0.10)	0.0002
Previous Deliveries (ref=no previous delivery)									
1 previous delivery	0.13	(0.09, 0.17)	<.0001	0.29	(0.12, 0.45)	0.0008	0.34	(0.23, 0.44)	<.0001
2 previous deliveries	0.19	(0.15, 0.24)	<.0001	0.40	(0.20, 0.59)	<.0001	0.51	(0.38, 0.63)	<.0001
3 previous deliveries	0.14	(0.06, 0.21)	0.0003	0.23	(-0.06, 0.53)	0.1233	0.38	(0.19, 0.58)	<.0001

More than 3 previous deliveries	0.24	(0.18, 0.31)	0.0345		0.47	(0.19, 0.75)	0.0010		0.69	(0.51, 0.87)	<.0001
Hypertensive complications in pregnancy (ref=normotensive)											
Gestational hypertension (GH)	-0.11	(-0.19, -0.04)	0.0032		-0.25	(-0.56, 0.06)	0.1108		-0.35	(-0.55, -0.15)	0.0006
Pre-eclampsia (PE)	-0.23	(-0.33, -0.14)	<.0001		-0.81	(-1.20, -0.43)	<.0001		-0.55	(-0.80, -0.30)	<.0001
Chronic hypertension (CH)	-0.23	(-0.33, -0.14)	<.0001		0.11	(-0.21, 0.43)	0.5038		0.08	(-0.12, 0.29)	0.4320
Superimposed pre-eclampsia (SP)	-0.02	(-0.16, 0.12)	0.8062		0.17	(-0.40, 0.74)	0.5559		-0.20	(-0.57, 0.17)	0.2883
Proteinuria (PR)	0.02	(-0.06, 0.10)	0.5564		0.00	(-0.32, 0.33)	0.9845		0.09	(-0.12, 0.30)	0.4043
Paternal											
Weight, kg	0.01	(0.01, 0.01)	<.0001		0.03	(0.03, 0.04)	<.0001		0.01	(0.01, 0.02)	<.0001
Height, cm	0.01	(0.01, 0.01)	<.0001		0.05	(0.04, 0.06)	<.0001		0.00	(0.00, 0.01)	0.2679
Body mass index, kg/m2 (ref= normal weight)											
Underweight	-0.01	(-0.05, 0.04)	0.7420		-0.17	(-0.36, 0.02)	0.0769		0.05	(-0.08, 0.17)	0.4466
Overweight	0.07	(0.03, 0.12)	0.0004		0.16	(0.00, 0.33)	0.0554		0.20	(0.09, 0.31)	0.0004
Obese	0.13	(0.01, 0.24)	0.0308		0.38	(-0.09, 0.85)	0.1134		0.30	(-0.01, 0.60)	0.0556
SES (ref = White collar)											
Blue collar	-0.02	(-0.06, 0.02)	0.2992		-0.10	(-0.25, 0.05)	0.1806		-0.02	(-0.12, 0.08)	0.7055
Farmer	0.02	(-0.05, 0.10)	0.5403		-0.07	(-0.37, 0.24)	0.6717		0.14	(-0.05, 0.34)	0.1470

[#]Beta, se and 95% CI were presented in % change as PHV was log-transformed to reduce right skewness.

[$]Significance was considered at $P<0.05$.

Table 6. B. Association between prenatal predictors and size-at-birth factors (birth weight (BW), birth length (BL) and birth BMI), adjusted for gestational age (GA) in NFBC 1986 Females.

B.	**NFBC 1986 females (N=2836)**											
	B[#]	**95 % CI[#]**		***P***	**β[#]**	**95 % CI[#]**		***P***	**β[#]**	**95 % CI[#]**		***P***
Maternal												
Age, years	0.01	0.01	0.01	<.0001	0.02	0.02	0.02	0.0024	0.02	0.02	0.02	<.0001
Weight, kg	0.01	0.01	0.01	<.0001	0.04	0.03	0.05	<.0001	0.02	0.02	0.03	<.0001
Height, cm	0.01	0.01	0.02	<.0001	0.08	0.07	0.09	<.0001	0.01	0.01	0.02	0.0003
Body mass index, kg/m2 (ref= normal weight)												
Underweight	-0.24	-0.30	-0.18	<.0001	-0.55	-0.81	-0.29	<.0001	-0.34	-0.52	-0.17	0.0001
Overweight	0.09	0.04	0.13	0.0005	0.18	-0.02	0.38	0.0747	0.19	0.05	0.32	0.0057
Obese	0.25	0.16	0.33	<.0001	0.65	0.30	0.99	0.0002	0.59	0.36	0.82	<.0001
Marital Status (ref=married)												
Unmarried/cohabiting	-0.12	-0.16	-0.08	<.0001	-0.40	-0.58	-0.23	<.0001	-0.24	-0.36	-0.12	<.0001
Divorced/widow	-0.10	-0.25	0.05	0.2002	-0.43	-1.05	0.19	0.1752	-0.19	-0.60	0.23	0.3818
SES (ref=white collar)												
Blue collar	-0.05	-0.08	-0.01	0.0144	-0.25	-0.40	-0.10	0.0009	-0.02	-0.12	0.08	0.6686
Farmer	0.02	-0.06	0.10	0.6548	-0.05	-0.38	0.28	0.7628	0.09	-0.13	0.31	0.439
Education (ref=primary education)												
Vocational education	-0.01	-0.07	0.05	0.739	0.02	-0.02	0.07	0.8669	-0.17	-0.36	0.02	0.0793
Secondary education	0.02	-0.04	0.08	0.5569	0.23	-0.01	0.47	0.0631	-0.18	-0.36	-0.01	0.0438
University	0.05	-0.03	0.13	0.2155	0.43	0.11	0.75	0.0094	-0.16	-0.39	0.08	0.19
Smoking at 2 months of pregnancy (ref= non-smoker)												
Smoker	-0.17	-0.21	-0.13	<0.0001	-0.72	-0.89	-0.56	<0.0001	-0.27	-0.38	-0.16	<0.0001
Previous Deliveries (ref=no previous delivery)												
1 previous delivery	0.15	0.11	0.18	<0.0001	0.31	0.15	0.47	0.0002	0.44	0.33	0.55	<0.0001
2 previous deliveries	0.17	0.13	0.22	<0.0001	0.39	0.20	0.58	<0.0001	0.46	0.33	0.59	<0.0001
3 previous deliveries	0.21	0.14	0.28	<0.0001	0.42	0.12	0.71	0.0054	0.58	0.39	0.78	<0.0001
More than 3 previous deliveries	0.26	0.20	0.33	<0.0001	0.62	0.35	0.88	<0.0001	0.70	0.53	0.88	<0.0001
Hypertensive complications in pregnancy (ref=normotensive)												

Gestational hypertension (GH)	-0.04	-0.11	0.04	0.3598		-0.16	-0.47	0.16	0.3269		-0.07	-0.28	0.15	0.54
Pre-eclampsia (PE)	-0.12	-0.22	-0.02	0.0162		-0.59	-0.99	-0.19	0.0038		-0.36	-0.63	-0.09	0.0083
Chronic hypertension (CH)	0.02	-0.06	0.09	0.6716		0.20	-0.10	0.51	0.194		-0.06	-0.27	0.14	0.559
Superimposed pre-eclampsia (SP)	0.02	-0.09	0.14	0.7079		0.19	-0.29	0.66	0.4344		0.05	-0.27	0.37	0.7768
Proteinuria (PR)	0.08	0.01	0.14	0.033		0.15	-0.14	0.43	0.3096		0.26	0.07	0.46	0.0072
Paternal														
Weight, kg	0.01	0.00	0.01	<0.0001		0.03	0.02	0.03	<.0001		0.01	0.00	0.01	0.0022
Height, cm	0.01	0.00	0.01	<0.0001		0.04	0.03	0.06	<.0001		0.00	0.00	0.01	0.366
Body mass index, kg/m2 (ref= normal weight)														
Underweight	0.00	-0.05	0.05	0.9827		-0.06	-0.25	0.13	0.548		0.03	-0.10	0.16	0.6666
Overweight	0.06	0.02	0.10	0.0033		0.21	0.04	0.37	0.0132		0.12	0.01	0.23	0.026
Obese	0.04	-0.06	0.15	0.45		-0.14	-0.57	0.29	0.5274		0.23	-0.06	0.52	0.1221
SES (ref = White collar)														
Blue collar	-0.04	-0.08	0.00	0.0328		-0.14	-0.29	0.00	0.0557		-0.06	-0.16	0.04	0.2059
Farmer	-0.01	-0.08	0.06	0.8211		-0.03	-0.32	0.25	0.8286		-0.04	-0.23	0.15	0.7033

[#]Beta, se and 95% CI were presented in % change as PHV was log-transformed to reduce right skewness.

[$]Significance was considered at $P<0.05$.

5.1.1.2.4 Sex-stratified association showing the potential mediation roles by size-at-birth between prenatal factors and PHV (Table 7 A and B)

Tables 7A and 7B show the four models compared to understand the role of mediation by size-at-birth variables in the relationship between prenatal predictors and PHV. Comparing effect size for the four models, it was observed that mother's age in female offspring showed an indirect effect via all size-at-birth variables but not in male offspring (**Table 7A, 7B; model 1 vs 2, 3 and 4**). In female offspring, after adjustments with individual size-at-birth variables in the models, the effect size for each of the association between maternal age and PHV was reduced by > 10% suggesting mediation by size-at-birth variables which were confirmed by Sobel's test ($P_{sobel} < 0.05$ for all size-at-birth variables). Maternal weight in Model 1 did not associate with PHV in either sex, however, after adjustments with size-at-birth variables the association turned significant indicating mediation by these factors (**Table 7A,B; model 1 vs 2,3 and 4,** $P_{sobel} < 0.05$ for all). Maternal height did not show significant association after adjustments of size of birth variables in each model for both sexes, suggesting no mediation by any of these variables ($P_{sobel} > 0.05$). Maternal BMI showed mediation with PHV by BW and birth BMI in males ($P_{sobel} <0.05$) but not in females.

The association between male PHV and overweight mothers was mediated by birth BMI ($P_{sobel} <0.05$**)**, but was not mediated in females ($P_{sobel}>0.05$**)**. Similarly association between PHV and obese mothers, suggested mediation by BW and birth BMI ($P_{sobel} <0.05$) but not by BL ($P_{sobel} >0.05$) and no significant association was observed in female offspring ($P_{sobel} > 0.05$**)**. The association between female offspring of single/cohabiting mothers and PHV in female offspring mediated by all size-at-birth (**model 1 vs 2,3 and 4**), whereas boys born to single/cohabiting mothers had their association mediated only by birth BMI ($P_{sobel}<0.05$). Mother's SES and education did not show any mediation by size-at-birth variable after Sobel's test ($P_{sobel} >0.05$).

Association between offspring of mothers smoking during the initial 2 months of pregnancy and PHV were mediated by birth BMI in males ($P_{sobel} <0.05$), and in females the association were mediated by both BW and birth BMI ($P_{sobel} <0.05$). Maternal parity showed mediation by all size-at-birth factors in both sexes except mothers with 2 previous deliveries did not show mediation by BL (**model 1 vs 2, 3 and 4**), which could be due to small sample size for this category. For the hypertensive complication during

Table 7. A. The association between prenatal factors and PHV (model 1) and the mediating role of size-at-birth variables in NFBC 1986 males

NFBC 1986 Males (N=2741)																							
A	**Model1a**						**Model 2a**						**Model 3a**						**Model 4a**					
	Beta	**95% CI**					**Beta**	**95% CI**					**beta**	**95% CI**					**beta**	**95% CI**				
Maternal																								
Age, years	-0.03	(	-0.06	0.00	)	0.0859	0.00	(	-0.03	0.03	)		0.00	(	-0.03	0.03	)		-0.02	(	-0.05	0.02	)	
Weight, kg	-0.01	(	-0.02	0.01	)	0.6017	0.03	(	0.01	0.05	)	[b]	0.02	(	0.00	0.04	)	[b]	0.02	(	0.00	0.04	)	[b]
Height, cm	0.16	(	0.12	0.19	)	<.0001	0.21	(	0.17	0.24	)		0.23	(	0.20	0.26	)		0.16	(	0.13	0.19	)	
BMI, kg/m2 (ref= normal weight)																								
Underweight	-0.02	(	-0.73	0.68	)	0.9447	-0.41	(	-1.08	0.27	)		-0.29	(	-0.95	0.37	)		-0.19	(	-0.89	0.51	)	
Overweight	-0.83	(	-1.37	-0.29	)	0.0027	-0.53	(	-1.05	-0.01	)		-0.59	(	-1.10	-0.08	)		-0.72	(	-1.25	-0.18	)	[b]
Obese	-2.24	(	-3.25	-1.21	)	<.0001	-1.70	(	-2.68	-0.71	)	[b]	-2.11	(	-3.07	-1.14	)		-2.05	(	-3.06	-1.02	)	[b]
Marital (ref=married)																								
Unmarried/cohabiting	0.77	(	0.29	1.26	)	0.0019	0.32	(	-0.15	0.79	)		0.30	(	-0.16	0.76	)		0.61	(	0.13	1.10	)	[b]
Divorced/widow	0.09	(	-1.91	2.13	)	0.9283	-0.31	(	-2.22	1.64	)		-0.01	(	-1.89	1.91	)		-0.19	(	-2.17	1.84	)	
SES (ref=white collar)																								
Blue collar	0.41	(	0.00	0.83	)	0.0492	0.28	(	-0.11	0.68	)		0.27	(	-0.12	0.66	)		0.38	(	-0.03	0.79	)	
Farmer	-0.79	(	-1.71	0.14	)	0.0944	-0.63	(	-1.51	0.26	)		-0.78	(	-1.64	0.09	)		-0.68	(	-1.59	0.25	)	
Education (ref=primary)																								
Vocational education	0.06	(	-0.70	0.82	)	0.8835	-0.14	(	-0.86	0.60	)		-0.07	(	-0.78	0.65	)		-0.03	(	-0.79	0.74	)	
Secondary education	0.69	(	-0.03	1.42	)	0.0604	0.54	(	-0.16	1.24	)		0.65	(	-0.03	1.33	)		0.58	(	-0.14	1.31	)	
University	-0.03	(	-0.98	0.93	)	0.9499	-0.09	(	-1.00	0.82	)		0.12	(	-0.77	1.02	)		-0.11	(	-1.05	0.85	)	
Smoking (ref= no)																								
Smoker	0.84	(	0.38	1.29	)	0.0003	0.42	(	-0.02	0.86	)		0.26	(	-0.17	0.69	)		0.71	(	0.26	1.17	)	[b]
Previous Deliveries (ref=no)																								
1 previous delivery	-0.92	(	-1.35	-0.49	)	<.0001	-0.56	(	-0.97	-0.14	)	[b]	-0.65	(	-1.06	-0.24	)	[b]	-0.74	(	-1.18	-0.31	)	[b]
2 previous deliveries	-1.01	(	-1.51	-0.50	)	<.0001	-0.46	(	-0.96	0.03	)		-0.64	(	-1.12	-0.16	)	[b]	-0.76	(	-1.27	-0.25	)	[b]
3 previous deliveries	-2.15	(	-2.91	-1.38	)	<.0001	-1.77	(	-2.51	-1.02	)	[b]	-1.97	(	-2.69	-1.25	)	[b]	-2.00	(	-2.76	-1.23	)	[b]
More than 3 previous deliveries	-2.08	(	-2.79	-1.37	)	<.0001	-1.40	(	-2.09	-0.70	)	[b]	-1.71	(	-2.39	-1.03	)	[b]	-1.79	(	-2.52	-1.07	)	[b]
Pregnancy complications (ref=normotensive)[c]																								
GH	0.18	(	-0.62	0.99	)	0.6612	-0.15	(	-0.92	0.64	)		-0.07	(	-0.84	0.70	)		-0.03	(	-0.84	0.79	)	
PE	1.82	(	0.82	2.84	)	0.0004	1.14	(	0.17	2.12	)	[b]	1.05	(	0.08	2.02	)		1.50	(	0.48	2.53	)	[b]

CH	0.23	(	-0.61	1.07	)	0.5944	0.34	(	-0.47	1.15	)		0.33	(	-0.47	1.13	)		0.27	(	-0.57	1.11	)	
SP	0.65	(	-0.85	2.18	)	0.3957	0.60	(	-0.84	2.07	)		0.80	(	-0.62	2.24	)		0.54	(	-0.95	2.06	)	
PR	-0.17	(	-1.02	0.68	)	0.6941	-0.10	(	-0.92	0.72	)		-0.16	(	-0.96	0.65	)		-0.12	(	-0.97	0.73	)	
Paternal																								
Weight, kg	0.04	(	0.02	0.06	)	0.0004	0.16	(	0.13	0.19	)		0.07	(	0.05	0.09	)	[b]	0.04	(	0.02	0.07	)	[b]
Height, cm	0.13	(	0.10	0.16	)	<.0001	0.06	(	0.04	0.08	)	[b]	0.18	(	0.15	0.21	)	[b]	0.13	(	0.10	0.16	)	
BMI, kg/m2 (ref= normal weight)																								
Underweight	0.50	(	0.00	1.01	)	0.0516	0.48	(	-0.01	0.96	)		0.33	(	-0.15	0.81	)		0.33	(	-0.15	0.81	)	
Overweight	0.10	(	-0.34	0.54	)	0.6597	0.32	(	-0.10	0.75	)		0.23	(	-0.19	0.65	)		0.23	(	-0.19	0.65	)	
Obese	-0.85	(	-2.08	0.38	)	0.1762	-0.47	(	-1.65	0.72	)		-0.52	(	-1.67	0.65	)		-0.52	(	-1.67	0.65	)	
SES (ref = White collar)																								
Blue collar	-0.09	(	-0.49	0.31	)	0.6535	-0.15	(	-0.53	0.23	)		-0.18	(	-0.55	0.20	)		-0.09	(	-0.49	0.30	)	
Farmer	-0.70	(	-1.48	0.09	)	0.0828	-0.63	(	-1.38	0.13	)		-0.78	(	-1.52	-0.04	)		-0.65	(	-1.43	0.14	)	

[a] Model 1 = PHV~ prenatal predictors (PP) + GA, Model 2 =PHV~ PP + BW + GA, Model 3=PHV~ PP + BL+GA, Model 4=PHV~ PP+Birth BMI +GA;

[b] Sobel test $P < 0.05$

[c]GH = gestational hypertension, PE= pre-eclampsia, CH = chronic hypertension, SP = superimposed pre-eclampsia, PR = proteinuria

Table 7. B. The association between prenatal factors and PHV (model 1) and the mediating role of size-at-birth variables in NFBC 1986 females.

	NFBC 1986 Females (N=2836)																							
B.	**Model1a**						**Model 2a**						**Model 3a**						**Model 4a**					
	Beta	**95% CI**					**Beta**	**95% CI**					**beta**	**95% CI**					**beta**	**95% CI**				
Maternal																								
Age, years	-0.22	(	-0.31	-0.14	)	<.0001	-0.12	(	-0.21	-0.04	)	[b]	-0.12	(	-0.20	-0.04	)	[b]	-0.20	(	-0.28	-0.11	)	[b]
Weight, kg	-0.02	(	-0.07	0.03	)	0.4693	0.08	(	0.03	0.13	)	[b]	0.09	(	0.04	0.14	)	[b]	0.09	(	0.04	0.14	)	[b]
Height, cm	0.38	(	0.29	0.46	)	<.0001	0.52	(	0.44	0.60	)		0.63	(	0.55	0.71	)		0.40	(	0.32	0.49	)	
BMI, kg/m2 (ref= normal)																								
Underweight	1.62	(	-0.22	3.50	)	0.0845	0.20	(	-1.56	1.98	)		0.20	(	-1.50	1.93	)		1.24	(	-0.60	3.12	)	
Overweight	-1.13	(	-2.52	0.27	)	0.1135	-0.44	(	-1.78	0.93	)		-0.74	(	-2.04	0.57	)		-1.00	(	-2.38	0.41	)	
Obese	-2.90	(	-5.23	-0.51	)	0.0178	-0.92	(	-3.22	1.43	)		-1.17	(	-3.38	1.10	)		-2.15	(	-4.50	0.25	)	
Marital (ref=married)																								
Unmarried/cohabiting	2.95	(	1.70	4.22	)	<.0001	1.93	(	0.74	3.15	)	[b]	1.71	(	0.55	2.89	)	[b]	2.48	(	1.23	3.75	)	[b]
Divorced/widow	2.69	(	-1.71	7.30	)	0.2343	1.86	(	-2.33	6.24	)		1.56	(	-2.49	5.78	)		2.46	(	-1.90	7.02	)	
SES (ref=white collar)																								
Blue collar	0.07	(	-0.99	1.14	)	0.9002	-0.31	(	-1.32	0.71	)		-0.58	(	-1.56	0.41	)		0.06	(	-0.99	1.12	)	
Farmer	-2.94	(	-5.16	-0.66	)	0.0117	-2.79	(	-4.92	-0.61	)		-2.85	(	-4.93	-0.72	)		-2.62	(	-4.85	-0.33	)	
Education (ref=primary)																								
Vocational	-0.52	(	-2.32	1.32	)	0.5775	-0.61	(	-2.33	1.14	)		-0.46	(	-2.14	1.24	)		-0.54	(	-2.33	1.29	)	
Secondary	-0.13	(	-1.82	1.60	)	0.8848	0.03	(	-1.59	1.67	)		0.54	(	-1.04	2.15	)		-0.11	(	-1.80	1.60	)	
University	-1.83	(	-4.05	0.45	)	0.1153	-1.40	(	-3.53	0.78	)		-0.76	(	-2.85	1.38	)		-1.98	(	-4.18	0.28	)	
Smoking (ref= no)																								
Smoker	2.94	(	1.75	4.14	)	<.0001	1.55	(	0.41	2.70	)	[b]	0.89	(	-0.22	2.01	)		2.45	(	1.26	3.65	)	[b]
Previous Deliveries (ref=no)																								
1 previous delivery	-3.97	(	-5.03	-2.91	)	<.0001	-2.92	(	-3.96	-1.87	)	[b]	-3.25	(	-4.25	-2.25	)	[b]	-3.61	(	-4.68	-2.53	)	[b]
2 previous deliveries	-4.45	(	-5.72	-3.16	)	<.0001	-3.21	(	-4.46	-1.93	)	[b]	-3.48	(	-4.69	-2.27	)	[b]	-4.02	(	-5.31	-2.72	)	[b]
3 previous deliveries	-6.81	(	-8.68	-4.89	)	<.0001	-5.32	(	-7.17	-3.43	)	[b]	-5.81	(	-7.58	-4.00	)	[b]	-6.29	(	-8.18	-4.36	)	[b]
More than 3 previous deliveries	-7.86	(	-9.53	-6.16	)	<.0001	-6.04	(	-7.70	-4.34	)	[b]	-6.38	(	-7.97	-4.76	)	[b]	-7.22	(	-8.92	-5.50	)	[b]
Pregnancy complications (ref=normotensive)																								
GH	1.33	(	-0.90	3.62	)	0.2431	1.03	(	-1.10	3.21	)		0.94	(	-1.13	3.04	)		1.27	(	-0.94	3.53	)	

PE	5.88	(	2.94	8.89	)	<.0001	4.83	(	2.04	7.69	)	[b]	3.98	(	1.28	6.76	)	[b]	5.12	(	2.20	8.12	)	[b]
CH	0.63	(	-1.51	2.82	)	0.5661	0.77	(	-1.29	2.86	)		1.18	(	-0.82	3.23	)		0.57	(	-1.55	2.74	)	
SP	-3.25	(	-6.36	-0.03	)	0.0482	-3.07	(	-6.06	0.01	)		-3.11	(	-6.08	-0.05	)		-3.53	(	-6.69	-0.27	)	
PR	0.04	(	-1.93	2.05	)	0.9667	0.67	(	-1.24	2.61	)		0.28	(	-1.58	2.18	)		0.23	(	-1.75	2.26	)	
Paternal																								
Weight, kg	0.12	(	0.06	0.17	)	<.0001	0.17	(	0.12	0.22	)	[b]	0.49	(	0.42	0.57	)	[b]	0.37	(	0.29	0.44	)	[b]
Height, cm	0.36	(	0.28	0.44	)	<.0001	0.43	(	0.35	0.50	)	[b]	0.19	(	0.14	0.24	)	[b]	0.13	(	0.08	0.18	)	[b]
BMI, kg/m2 (ref= normal)																								
Underweight	1.13	(	-0.21	2.48	)	0.0996	1.13	(	-0.15	2.43	)		0.89	(	-0.35	2.15	)		0.89	(	-0.35	2.15	)	
Overweight	-0.58	(	-1.70	0.56	)	0.3211	-0.08	(	-1.17	1.02	)		-0.05	(	-1.11	1.02	)		-0.05	(	-1.11	1.02	)	
Obese	-0.56	(	-3.53	2.49	)	0.7143	-0.23	(	-3.08	2.71	)		-0.90	(	-3.65	1.92	)		-0.90	(	-3.65	1.92	)	
SES (ref = White collar)																								
Blue collar	0.78	(	-0.26	1.84	)	0.1430	0.47	(	-0.54	1.48	)		0.39	(	-0.58	1.38	)		0.69	(	-0.35	1.74	)	
Farmer	-0.49	(	-2.48	1.54	)	0.6341	-0.55	(	-2.46	1.39	)		-0.39	(	-2.25	1.51	)		-0.35	(	-2.34	1.67	)	

[a] Model 1 = PHV~ prenatal predictors (PP) + GA, Model 2 =PHV~ PP + BW + GA, Model 3=PHV~ PP + BL+GA, Model 4=PHV~ PP+Birth BMI +GA;

[b] Sobel test $P < 0.05$

[c]GH = gestational hypertension, PE= pre-eclampsia, CH = chronic hypertension, SP = superimposed pre-eclampsia, PR = proteinuria

pregnancy, mother's with pre-eclampsia showed mediation by all size-at-birth variables in female offspring (P_{sobel}<0.05**)**, but only with BW and birth BMI in male offspring (P_{sobel} <0.05). Father's weight and height were mediated by all size-at-birth variables in female offspring as tested by Sobel's test (P_{sobel} <0.05), however, in males mediation was only observed by BL and Birth BMI (P_{sobel} <0.05). Father's BMI and SES did not show any mediation by any of the size-at-birth variables after testing for Sobel's test (P_{sobel} >0.05), which may indicate that fathers BMI may account for high muscular mass rather than fat mass.

5.1.1.4 Discussion

This study examines a wide selection of prenatal predictors which associate with size-at-birth and PHV. Moreover, mediation by size-at-birth which was assumed to influence association between prenatal predictors and PHV was also investigated. This is a novel study as it uses longitudinal growth models to determine individual growth velocities in NFBC1986. The NFBCs are an ideal source for such growth modeling purposes, as they provide large numbers of serial height measurements, and a rich set of prenatal factors, which enables a detailed study of the prenatal and postnatal predictors of early growth velocity as shown by Tzoulaki et al, 2010[14].

One of the biggest strengths of this study was the use of longitudinal data to measure height growth acceleration. As discussed previously, fitting appropriate growth models has added advantages as they provide biologically meaningful parameters and summarise a large amount of raw anthropometric data into functional parameters at time periods of interest[210]. Moreover, this is the first study which uses a comprehensive set of prenatal factors to understand the mechanisms behind interrelationships between prenatal factor, size-at-birth (proxy for fetal growth) and postnatal immediate outcome. The study shows that due to adverse *in utero* environment a fetus may not be able to reach its genetic growth potential, and this "damage" is compensated by increased postnatal growth. This study demonstrates how important these prenatal factors are for early growth and how they may later impact wellbeing and health.

The results of this study are consistent with previously reviewed literature, confirming an inverse association between size-at-birth variables on infant rapid postnatal growth, also called *catch up* growth[67,211,212]. However, results from most previous studies have been ambivalent. With studies mostly focusing on the associations between size-at-birth variables, rapid *weight gain* and adverse health outcomes in adulthood, such as obesity or metabolic syndrome[57,95], none tried to trace the causal pathway connecting prenatal predictors with rapid growth. Moreover these studies use an average of height/weight growth over two or

more time periods as an indicator of rapid postnatal growth[61], which makes direct comparison between studies difficult. There was no previous study that had tried such longitudinal modeling before, using PHV. With different definitions for rapid growth, one cannot assess the relative contribution of each potential growth phase during which the individuals catch up or drop below their genetic growth potential.

The study showed an inverse association between size-at-birth variables and PHV, which was much stronger in females than males. The result is in line with the *growth acceleration hypothesis* that smaller babies experience faster infant postnatal growth, and also suggests a possible temporal effect in terms of the strength of the association[201]. Results indicated that maternal age, weight, height, pre-pregnancy BMI, marital status, smoking during the initial 2 months of pregnancy, parity, pre-eclampsia, chronic hypertension, father's weight and height were significantly associated with PHV, independently of size-at-birth variables. This means that adjustment for size-at-birth factors did not completely remove the associations although there was partial mediation observed for many of the prenatal determinants of postnatal growth velocity. Consistent with published literature[65,68], the study also found prenatal predictors associated with size-at-birth variables.

Size at birth mediated the relationship between various prenatal predictors and PHV, with males and females showing varying degree of mediation for the prenatal predictor. Maternal age associated with lower PHV and mediated by all size-at-birth variables but only in females. Maternal age has been associated with larger babies and it is known that larger babies grow slowly, hence the results from this study is consistent with observed process. However, maternal age has been also associated with favorable height growth in children[213] as well as childhood hypertension[214]. Keeping this in mind, the result of this study can therefore indirectly indicate that association between maternal age and low PHV may affect later health, although further studies will be required to understand this association in depth.

Maternal weight did not associate with PHV in either sex which could be due to some exogenous factors[215], however, once adjusted for size-at-birth variables, the association turned to significant mediation shown by size-at-birth variables in both sex. This result fits well with the *maternal weight hypothesis*, which states that maternal pre-pregnancy weight affects the intra-uterine environment which indirectly affects infant growth. It was also observed that paternal weight and height in the female offspring showed a clear trend of mediation by size-at-birth variables, although there was no such trend for males. It is a well-known fact that parental anthropometric measurements are genetically inherited by offspring[204,216], however results from this

study indicate that there may be a sex interaction between inheritances of anthropometric measures, which needs to be analysed further in future.

Higher maternal pre-pregnancy BMI associated with a higher PHV. There was a clear trend that maternal BMI groups had a negative or positive association with PHV in both sexes. For example, infants of underweight mothers were born small and therefore had higher PHV in comparison to infants of lean mothers suggesting that these infants *catch-up* to compensate for a small size-at-birth to reach their genetic potential for growth. Similarly, infants of overweight and obese mothers are born large and therefore have slower postnatal growth than infants of normal mothers. These individuals *catch-down* to reach their genetic potential. Interestingly, it was observed that mediation between maternal BMI and PHV only associated in males but not in females.

Maternal marital status associated with higher infant height growth. Female infants showed mediation by all size-at-birth variables, but in males the association was only mediated by birth BMI. This result is in line with a previous study, which showed that mother's SES and marital status may affect the intra-uterine nutritional environment during fetal development, making these individuals small at birth[217]. Similarly, mothers smoking during pregnancy also associated with higher infant height growth. The association in female infants shows mediation by BW and birth BMI but in male infants this association was mediated only by birth BMI. Although, this result was in agreement with a previous study [218], this study accounts for the longitudinal variation in height also showing a sex interaction.

Maternal parity had a consistent, strong association with PHV, in all groups, even after adjustment for other size-at-birth variables showing partial mediation by them. Ong *et. al,* 2002, have already previously reported that maternal factors such as weight gain and smoking during pregnancy and parity are associated with the *catch-up* growth [218], and this study confirms the findings. Studies have also shown that firstborn babies are born small and hence gain weight more rapidly during the first year of life than other siblings [66,219]. The current study shows a similar trend, with slower infant height growth as number of children increases.

Pre-eclampsia was found to be mediated by all size-at-birth variables in both sexes except for BL in males. This is similar result to that by Srinivas et al, 2009, who demonstrated that intrauterine growth retardation, occurs in pre-eclampsia followed by rapid postnatal growth[220].

One of the limitations of this study was that it accounted for the error in PHV estimations by weighting the number of measurements per person. This process assumes that the reliability of the growth data has a proportional relationship with the measurement frequency. These are broadly reasonable assumptions, but they are difficult to verify using this data alone[12]. Moreover, although the study considered an extensive set of prenatal factors which may influence infant height growth, no adjustment for variables such as maternal infections during pregnancy, postnatal infections and other diseases as well as infant feeding could be addressed in this study. There was no information about mother's diet, lifestyle, breastfeeding or overall child's nutrition in NFBC 1986 therefore these could not be analysed in this study.

To conclude, this study of a large population-based Finnish birth cohort supports the hypothesis that babies with smaller birth sizes have faster infant growth and mediates the relationship between prenatal predictors and postnatal height growth. Several maternal and paternal factors were directly associated with faster postnatal height growth, some of which had their association mediated by size-at-birth variables. The clinical importance of this effect should be evaluated as it may inform public health policy aimed at improving the pre-pregnancy environment and also at monitoring infant's growth during the first few years.

5.1.2. The effect of an obesogenic environment *in utero* and during childhood on metabolic health in adolescents.

5.1.2.1. Introduction

There has been a dramatic increase in the prevalence of obesity in women of reproductive ages in the last 20 years[221], with the WHO predicting that one in five women are obese at the time of their antenatal check in the UK[222]. Maternal obesity has been identified as an independent risk factor for adverse health consequences in both mother and child during the prenatal and postnatal period. Overweight and obese mothers have been found to have a higher risk of pre-eclampsia and gestational diabetes mellitus[223] during pregnancy, poor mental health and several chronic diseases, such as T2D and cardiovascular diseases in later life[224].

Babies born to overweight or obese mothers are usually large for gestational age (LGA)[225]. These mothers have high incidences of obstetric complications during birth, and a higher incidence of neonatal deaths compared to newborns of lean mothers[226]. Recent reviews suggests that high pre-pregnancy maternal weight associates with an altered offspring metabolic profile including higher blood pressure and adverse lipid profile[222], although these studies are underpowered and consider the pre-pregnancy weight. Studies have found evidence of correlation between parental obesogenic environment and weight status of children at various ages, suggesting an association between parental obesogenic environment and adiposity in children[227].

Childhood obesity increases the probability of adverse psychological, social and metabolic health consequences during childhood and later life, which could reduce the quality of life of an individual [228]. Maternal overweight and obesity, both at conception and during childhood, has been independently associated with offspring adiposity and insulin resistance [97,229-231]. One of the limitations for these studies is that they have not considered the role of postnatal lifestyle influences on this association, hence making it difficult to prove the causal link between pre-pregnancy weights with offspring metabolic profile.

Animal studies support the 'developmental programming hypothesis' here, which proposes that maternal nutrition during fetal development exerts a 'programming' effect on the offspring's obesity, metabolic and behavioral abilities later in life[232-235]. In humans, similar studies to study the 'development programming' effect are challenging. The complex nature of the potential interaction between genetic, behavioral and life-style factors in pre- and post- environments, makes it difficult to segregate the contribution of each of these independently[236]. Most previous studies have focused on comparing either the anthropometric or metabolic

phenotypes for overweight /obese mothers vs. lean mothers, or birth outcomes of their newborn. No previous study has investigated the long-term effects of maternal overweight/obesity during *in utero* and during childhood on adolescent metabolic health[233,234,237]. Such studies are important, as they will be able give public health clinicians the opportunity to identify pregnant mothers who are at risk for prenatal programming of obesity[238].

Considering limitations from other studies on maternal obesity effect on offspring, the current study tests the hypothesis that both the prenatal and postnatal 'obesogenic' environment predisposes offspring of O/O mothers to adverse metabolic health, and that these environments effect, either in combination, or independently the offspring's metabolic phenotypes. The main objective was to use an experimental study design by Samuelsson et al, 2008, [232] and apply on human data using longitudinal data from the NFBC 1986. The aim was to support 'developmental programming' of obesity in offspring within the four groups of mothers by comparing insulin-resistance, lipid and blood pressure outcomes of adolescents.

The overall study was conducted as a trans-disciplinary approach to establish the consequences of maternal overweight and obesity during pregnancy in both humans and mice. The murine model is not discussed here, however, comparisons between human and mice models is an ideal research in translational studies since various animal studies show that models in mice have been able to mimic a wide range of human disease and health problems. In addition, it is technically possible to conduct experiments in mice model, which may not be possible in humans.

5.1.2.2. Study-specific methods and material

NFBC 1986 has been an ideal dataset for this study, as its dense longitudinal data from the start of pregnancy to adolescent health at 16 years is ideal for replicating the experimental murine model for this study. The details on data collection of anthropometric and metabolic variables (used for analyses in this study) are described in Chapter 4, Section 4.2. The comparison between groups and results were presented as two groups based on whether the metabolic function related to glucose-insulin axis or to cardiovascular risk factors. The glucose-insulin axis included fasting glucose, fasting insulin, HOMA-IR and HOMA-B. The metabolic phenotypes related to cardio-vascular risk factors included Triglyceride, total cholesterol, HDL, LDL, Apoprotein, Aproprotien B, SBP, DBP. Details on how these variables were calculated are given in Chapter 4, section 4.2. The description of study specific inclusion, exclusion and methods is given below.

Inclusion and exclusions

The flowchart showing the division of the four study groups for mothers of N=5,739 offspring with fasting blood and maternal BMI data available are shown in **Figure 6**. The offspring sample was restricted to singletons that had fasting glucose data available at 16 years of age. Underweight mothers (BMI < 18.5 kg/m2), mothers who refuse the use of data, and those with missing weight or height data were excluded from the analysis. Using the International Obesity Task Force (IOTF) cut-off, a BMI of between 18.5 kg/m2 and 25 kg/m2 was classified as 'normal' (N) and a BMI of over 25 kg/m2 was classified as 'overweight or obese' (O)[25]. The overweight and obese mothers were grouped as one, as health outcomes of children born to both overweight and obese mothers are prone to a similar direction and grouping gives us a larger sample size[239].

First, the offspring were divided into 'N' and 'O' groups according to mother's BMI at the start of pregnancy. Next, the N and O groups were further divided into four groups based on mother's BMI when offspring were 16 years, resulting in four groups: Normal-Normal (N/N), Normal-overweight or obese (N/O), overweight or obese-Normal (O/N) and overweight or obese-overweight or obese (O/O) groups. In this study, maternal BMI represents a proxy for the obesogenic environment that affects offspring's health since the nutritional information during pregnancy and during childhood was not available for analysis in the NFBC1986.

Statistical methods

Summary statistics of the adolescence metabolic phenotypes were reported as means with 95 % confidence interval (CI) for continuous variables and percentages for each group for the categorical variable. The main tests compared the mean of each group O/O, O/N, N/O with the reference group (N/N) to find if any of the metabolic outcomes were elevated within the groups and if there was a significant difference between the groups, indicating the effect of an obesogenic environment *in utero* or during childhood. Analysis of Variance (ANOVA) was used to test high variance between different maternal groups, which provided an overall P_{anova} value for variability within the groups. Tukey posthoc tests were performed to check pair-wise comparisons between the groups. The ANOVA requirements of normality and equal variances were tested for each group, Dunnett's t-test was performed. All statistical tests of hypotheses were two-sided, applying a threshold for significance of $P < 0.05$ after Bonferroni corrections.

The initial analyses were adjusted for maternal parity, weight gain during pregnancy and socio-economic status (SES), paternal BMI and SES, gestational age in weeks (GA), Tanner stage (scale to represent pubertal development) and age at menarche as potential confounding factors. All of these potential

confounders have been shown to be associated with maternal BMI. Adjustments for maternal parity, weight gain during pregnancy and socio-economic status (SES), paternal BMI and SES did not modify the associations; therefore these covariates were not adjusted to maintain high sample size. The final analyses were adjusted for three potential cofounders: GA, Tanner stage and age at menarche (females only), as they had a significant ($P<0.05$) modification of the association between outcome and explanatory variables. The analyses were sex-stratified, as previous studies confirmed that metabolic phenotypes differ between adolescent males and females[240]. Moreover, one of the aims of this study was also to investigate if there is any interaction between maternal BMI group and sex.

5.1.2.3. Results

Descriptive statistics

Table 8 shows the characteristics of the sample population by maternal BMI groups according to offspring's sex. At the start of the pregnancy, 75% of mothers were normal ("N") and 16% of mothers were overweight or obese ("O"), while the remaining 7.3% underweight mothers were removed from the analysis. Within the normal BMI group of mothers, 65% remained normal 16 years later (N/N) while 35% of became O/O (N/O). In the O group of mothers, 8% had normal BMI 16 years later (O/N) and 92% remained O/O (O/O). The remaining 1% mothers, who were underweight when their offspring were 16 years old, were again removed from the analysis. The underweight mothers were removed as they had a smaller sample size compared to other groups, and the offspring of the underweight mothers have other threats (or concerns), which may have deviated from the hypotheses in this study. The N/N groups were considered as the reference group in both sexes.

Male offspring

a) Maternal characteristics

The O/O mothers of male offspring were older (31 years) compared to the other three other groups, with an average maternal age of approximately 28 years in the study sample. The average BMI at the start of pregnancy in mothers was 20.9 kg/m2 in N/N mothers, 22.5 kg/m2 in N/O mothers, 27 kg/m2 in O/N

mothers and 27.9 kg/m2 in O/O mothers. The maternal weight gain during pregnancy supported our assumption that BMI at the start of pregnancy is a proxy for nutritional environment *in utero*, with N/O mothers gaining the most weight during pregnancy in comparison to the other three groups, an average of 14.9 kg. The O/O mothers had the highest parity among the groups, with an average of 2.4 children born to these mothers compared to 1.5 born to both N/O and O/N and 1.1 born to N/N mothers (Panova < 0.0001).

The maternal BMI when offspring were 16 years old was 22.5 kg/m2 in N/N mothers, 27.3 kg/m2 in N/O mothers, 24 kg/m2 in O/N mothers and 30.9 kg/m2 in O/O mothers.

b) Paternal characteristics

The paternal BMI at birth observed the same trend as maternal BMI, with an average BMI in husbands of N/N mothers of 23.9 kg/m2, 24.2 kg/m2 in N/O mothers, 24.4 kg/m2 in O/N mothers and 24.8 kg/m2 in husbands of O/O mothers. Father's BMI when offspring were 16 years old showed a slightly different trend to their BMI at offspring's birth, with husbands of N/O mothers and husbands of O/O mothers showing a similar average BMI of 27 kg/m2.

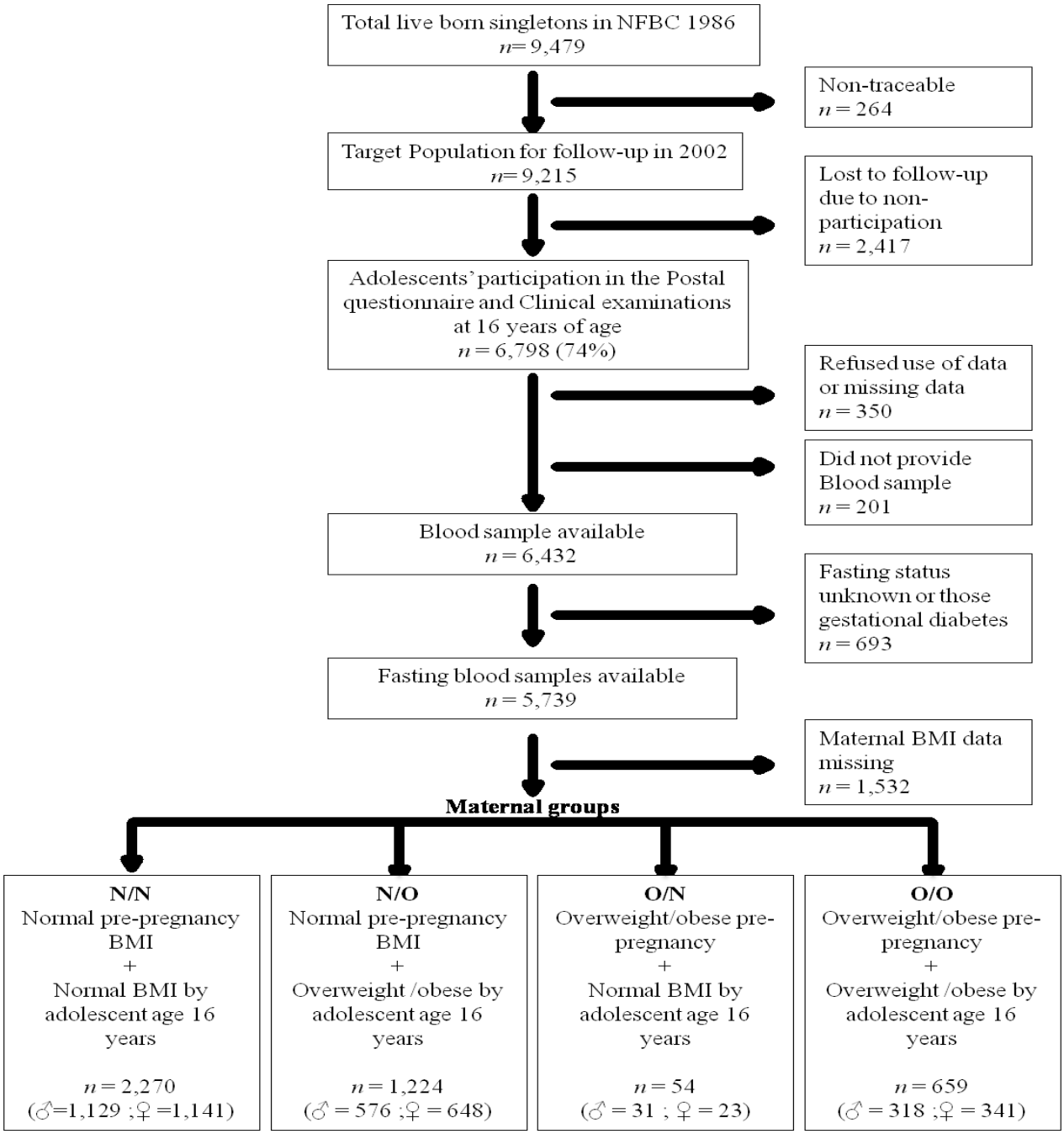
c) Newborn characteristics

Within all groups the average gestational age (GA) was approximately 39 weeks. The BW of newborns varied significantly between groups, with newborns of N/N being the lightest with an average of 3.6 kg and newborn of N/O, O/N and O/O mothers heaviest with an average BW of 3.7 kg (**Table 9**).

d) Adolescent at 16 years

The adolescent BMI at 16 years and WHR followed the same trend as the mothers' and fathers' BMI, with children of O/O mothers having an average ~ 2 kg/m2 higher BMI at 16 years and 0.02 higher WHR than children born to N/N mothers. Offspring of N/O mothers were significantly different to offspring of O/O and N/N mothers for both BMI and WHR, indicating that both environments, *in utero* and postnatal, may influence anthropometric outcomes. The Tanner scale was an average of 4.2 for the male offspring sample, and did not differ significantly between the groups (Panova =0.7488)

Figure 6. Study groups according to maternal BMI in the Northern Finland Birth Cohort 1986 (NFBC 1986). Offspring who participated in clinical examinations were stratified according to the maternal BMI at the start of pregnancy and maternal BMI when offspring were 16 years of age. The maternal BMI definition of N (normal weight) and O (overweight and obese) were based on IOTF criteria[25].



Females

a) Maternal characteristics

Similarly to males, the average maternal age was 28 years and the O/O mothers were the oldest compared to the three other groups. The average BMI at the start of pregnancy was 20.9 kg/m2 in N/N mothers, 22.5 kg/m2 in N/O mothers, 26 kg/m2 in O/N mothers and 28.1 kg/m2 in O/O mothers. Following the pattern in male offspring, the N/O mothers of females gained the most weight during pregnancy in comparison to the other three groups, an average of 14.3 kg, justifying the assumption that maternal BMI acts as a proxy for the obesogenic environment *in utero* (Panova P<0.0001). The maternal BMI when offspring were 16 years followed the same pattern as BMI at the start of pregnancy, with an average BMI of 22.6 kg/m2 in N/N mothers, 27.5 kg/m2 in N/O mothers, 23.8 kg/m2 in O/N mothers and 31.4 kg/m2 in O/O mothers.

b) Paternal characteristics

The paternal BMI showed the same trend as maternal BMI, with an average BMI in husbands of N/N mothers of 23.9 kg/m2, 24.2 kg/m2 in N/O mothers, 24.4 kg/m2 in O/N mothers and 24.8 kg/m2 in husbands of O/O mothers. The paternal BMI when the offspring were 16 years, showed a slightly different trend to their BMI at birth but a similar trend to male offspring, with husbands of N/O mothers showing the highest average BMI of 26.9 kg/m2 in comparison to all other groups, representing the postnatal obesogenic environment.

c) Newborn characteristics

The average GA was approximately 39 weeks in all groups. The BW of newborns varied significantly between groups (Panova P=0.0021), with newborns of N/N mothers lightest and newborns of O/O mothers the heaviest (**Table 9**). It was observed that the average BW for newborns of N/N and O/N mothers were 3.5 kg, whereas BW for newborn of N/O and O/O were 3.6 kg, suggesting that newborns of mothers with the same postnatal obesogenic environment show similar birth characteristics, which could be due to some influence of *in utero* obesogenic environment on BW.

d) Adolescent at 16 years

The BMI at 16 years in females varied slightly from their male offspring, as offspring of O/N mothers had the highest BMI of 22.6 kg/m2, although this estimate is less accurate due to low numbers in the group. The O/N mothers have the smallest sample size and thus may not be representative of the total population of O/N

Table 8: Summary characteristics of study population in Northern Finland 1986 Birth Cohort (NFBC 1986).

\	N/N (n=2,311)		N/O (n=1,243)		O/N (n=54)		O/O (n=691)		P_{anova}*
	Mean	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**	
%N	53%		30%		1%		16%		
Males									
Maternal									
Maternal Age (years)	27.6	(27.3, 27.9)	28.4	(28.0, 28.9)	28.3	(26.4, 30.2)	31	(30.4, 31.6)	<0.0001
BMI at start of pregnancy (kg/m^2)	20.9	(20.8, 21.0)	22.5	(22.3, 22.6)	27	(26.1, 28.0)	27.9	(27.6, 28.2)	<0.0001
Weight gain during pregnancy	13.5	(13.2, 13.7)	14.9	(14.5, 15.3)	12.3	(10.3, 14.4)	12.2	(11.6, 12.8)	<0.0001
Parity	1.1	(1.1, 1.2)	1.5	(1,3, 1.6)	1.5	(1,3, 1.6)	2.4	(2.2, 2.7)	<0.0001
BMI when offspring 16y (kg/m^2)	22.5	(22.5, 22.6)	27.3	(27.1, 27.5)	24	(23.7, 24.3)	30.9	(30.4, 31.3)	<0.0001
Paternal									
BMI at birth (kg/m^2)	23.9	(23.7, 24.1)	24	(23.8, 24.3)	24.1	(23.7, 24.3)	24.9	(24.6, 25.2)	<0.0001
BMI when offspring 16y (kg/m^2)	26	(25.8, 26.2)	26.9	(26.3, 27.6)	24.1	(24.1, 27.1)	26.9	(26.5, 27.4)	0.0011
Newborn									
Gestational age (weeks)	39.4	(39.4, 39.5)	39.5	(39.4, 39.6)	39.2	(38.4, 39.9)	39.4	(39.3, 39.6)	0.6149
Birth weight (kg)	3.6	(3.6, 3.7)	3.7	(3.7, 3.7)	3.7	(3.4, 4.0)	3.7	(3.7, 3.8)	0.0044
Adolescent at 16									
Body mass index (BMI) (kg/m2)	20.8	(20.6, 20.9)	21.5	(21.2, 21.8)	21.6	(20.7, 22.4)	22.6	(22.1, 23.1)	<0.0001
Waist hip ratio (WHR)	0.8	(0.8, 0.8)	0.8	(0.8, 0.8)	0.8	(0.8, 0.8)	0.8	(0.8, 0.8)	<0.0001
Tanner scores	4.2	(4.2, 4.3)	4.3	(4.2, 4.3)	4.2	(3.8, 4.5)	4.3	(4.2, 4.4)	0.7488

Females									
Maternal									
Maternal Age (years)	27.6	(27.3, 27.9)	28.4	(28.0, 28.8)	28.0	(25.4, 30.6)	29.9	(29.3, 30.6)	<0.0001
BMI at start of pregnancy (kg/m^2)	20.9	(20.9, 21.0)	22.5	(22.4, 22.6)	26.0	(25.6, 26.4)	28.1	(27.8, 28.4)	<0.0001
Weight gain during pregnancy	13.4	(13.2, 13.6)	14.3	(13.9, 14.7)	13.6	(11.7, 15.5)	11.9	(11.3, 12.5)	<0.0001
Parity	1.2	(1.1, 1.3)	1.5	(1.3, 1.5)	2	(1.1, 2.9)	2.2	(2.0, 2.5)	<0.0001
BMI when offspring 16y (kg/m^2)	22.6	(22.5, 22.7)	27.5	(27.3, 27.7)	23.8	(23.2, 24.4)	31.4	(30.9, 31.9)	<0.0001
Paternal									
BMI at birth (kg/m^2)	23.8	(23.7, 24.0)	24.3	(24.0, 24.5)	24.8	(23.0, 26.6)	24.7	(24.3, 25.0)	<0.0001
BMI when offspring 16y (kg/m^2)	25.8	(25.7, 26.1)	26.9	(26.5, 27.2)	26.8	(24.7, 28.9)	26.4	(26.4, 27.4)	<0.0001
Newborn									
Gestational age (weeks)	39.5	(39.5, 39.6)	39.5	(39.4, 39.6)	39.1	(37.8, 40.4)	39.3	(39.1, 39.5)	0.0897
Birth weight (kg)	3.5	(3.5, 3.5)	3.6	(3.5, 3.6)	3.5	(3.2, 3.8)	3.6	(3.5, 3.7)	0.0021
Adolescent at 16									
Body mass index (BMI) (kg/m2)	20.7	(20.6, 20.8)	21.7	(21.4, 22.0)	22.6	(20.8, 24.4)	22.5	(22.1, 22.9)	<0.0001
Waist hip ratio (WHR)	0.8	(0.7, 0.8)	0.8	(0.8, 0.8)	0.8	(0.8, 0.8)	0.8	(0.7, 0.8)	<0.0001
Tanner scores	3.7	(3.7, 3.8)	3.8	(3.8, 3.9)	3.9	(3.4, 4.4)	3.8	(3.7, 3.8)	0.4986
Age at menarche	12.6	(12.5,12.6)	12.4	(12.3, 12.5)	12.5	(12.0, 12.9)	12.4	(12.3, 12.5)	0.0043

BMI, body mass index; CI, confidence interval

*P value for the difference between groups

mothers. The BMI of offspring from N/O mothers, an average of 21.7 kg/m2, significantly differed from the BMI of offspring of both O/O mothers, 22.5 kg/m2, (Panova P=0.0006) and N/N mothers, 20.7 kg/m2, (Panova P<0.0001), suggesting that both *in utero* and childhood obesogenic environment have an added effect in females. The average WHR was similarly distributed as male offspring, with offspring of O/O mothers having the highest WHR of 0.78, although WHR between offspring of N/O differed with N/N mothers (P=0.011) and O/O mothers differed from the offspring of N/N mothers (P<0.0001). The Tanner scale was similar across the groups, with an average of 3.7 in all female offspring showing similar pubertal development within all groups (P=0.4986). Age at menarche occur at around 12.5 years in the overall sample, with offspring of N/O mothers being a significantly earlier age at menarche than offspring of N/N mothers (P=0.0085), suggesting that the childhood obesogenic environment may exert an effect on the age at menarche[241].

e) Association between obesogenic environment, glucose metabolism and insulin resistance

Males

The average fasting glucose in male offspring was 5.1mmol/L in all the groups, and was unaffected by the obesogenic environments during *in utero* and childhood (*P*anova =0.8970) (**Table 9**). The fasting serum insulin concentrations were elevated in all groups relative to offspring of N/N mothers (*P*anova=0.0015), with the highest observed in offspring of O/O mothers, which was significantly greater than both the N/O (P = 0.0007) and N/N mothers (*P*=0.0002). This variation in fasting insulin concentration indicates that the obesogenic environment *in utero* has an impact on insulin resistance. The early changes are reflected first in insulin whilst the glucose levels stay normal, O/N group differed in insulin but this may be insignificant due to low numbers.

In addition, the two HOMA indices followed the same trend as fasting insulin, with offspring of O/O mothers having higher levels than offspring of N/O (PHOMA-IR=0.0001 and PHOMA-B=0.0001 respectively) and N/N mothers (PHOMA-IR=0.0005 and PHOMA-B=0.0007 respectively), confirming the findings of serum insulin concentrations and suggesting that the maternal nutritional status during pregnancy 'programs' insulin resistance and β-cell alteration in offspring.

Table 9: Influence of the pre- and post-natal obesogenic environment on the glucose-insulin axis in adolescent males and females

Outcome variables	**N / N**	**N / O**	**O / N**	**O / O**	**Global ANOVA**	**N/N vs. O/O**	**N/O vs. O/O**
	Mean (95% CI)	**Mean (95% CI)**	**Mean (95% CI)**	**Mean (95% CI)**	**P**	**P**	**P**
Males							
Fasting glucose (mmol/L)	5.1 (5.1, 5.1)	5.1 (5.1, 5.2)	5.1 (5.1, 5.2)	5.1 (5.1, 5.2)	0.8970	1	1
Fasting insulin (mmol/L)	9.9 (9.5, 10.3)	9.9 (9.3, 10.5)	11.3 (8.5, 14.1)	12.0 (11.2, 12.8)	0.0015	0.0002	0.0007
HOMA-IR index	1.3 (1.2, 1.3)	1.3 (1.2, 1.3)	1.5 (1.1, 1.8)	1.6 (1.5, 1.7)	0.0012	0.0001	0.0005
HOMA-B index	99.3 (96.4, 102.1)	99.7 (95.7, 103.7)	112.4 (94.4, 130.4)	113.5 (109.0, 120.1)	0.0006	<0.0001	0.0007
Females							
Fasting glucose (mmol/L)	4.9 (4., 5.0)	5 (4.9, 5.1)	4.7 (4,1, 5.3)	4.9 (4.7, 5.0)	0.4285	1	0.6300
Fasting insulin (mmol/L)	10 (9.6, 10.5)	10.5 (9.9, 11.1)	8.6 (5.4, 11.8)	11.2 (10.4, 12.0)	0.0753	0.0417	0.8300
HOMA-IR index	1.3 (1.2, 1.4)	1.4 (1.3, 1.5)	1.1 (0.6, 1.6)	1.4 (1.3, 1.5)	0.2705	0.2998	1
HOMA-B index	109.9 (106.9, 112.8)	113.9 (109.7, 118.1)	104.4 (82.6, 126.2)	121.7 (116.2, 127.2)	0.0006	0.0009	0.1046

Legend: Adjusted mean and 95% confidence intervals by gestational age and tanner scale. Only statistically significant differences at P <0.05 between groups after Bonferroni correction are presented.

Females

In females, the mean fasting glucose was an average of 5 mmol/L but there was no significant variability between groups (*P*anova=0.4285), indicating that fasting glucose was unaffected by the obesogenic environment either *in utero* or during childhood (**Table 9**). The fasting serum insulin concentrations were elevated in offspring of O/O mothers, with an average of 11.2 mmol/l, which was significantly higher than offspring of N/N mothers 10 mmol/l (P=0.0417); although, unlike in males, insulin concentrations for offspring of O/O mothers did not differ from that of the offspring of N/O mothers (P=0.8300). Females did not show a substantial change either in glucose or serum insulin concentrations and were unaffected by the obesogenic environment *in utero*.

HOMA-IR levels were highest for offspring of O/O mothers, although there was no statistically significant variability between any of the groups (Panova=0.2705). The HOMA-B index was again highest in offspring of O/O mothers, with a significant difference between offspring of O/O mothers and N/N mothers (*P*=0.0009), but not between O/O and N/O mothers (*P*=0.1046), suggesting that there are other environmental factors that are responsible for the insulin-resistance difference in females.

f) Association between obesogenic environment, adolescent lipids and blood pressure

Males

Fasting triglycerides (TG) concentration varied between the maternal BMI groups (*P*anova=0.0046) and, as expected, offspring of O/O mothers had the most elevated TG of the groups which significantly differed from offspring of N/N mothers (P=0.0036). The total cholesterol (TC) varied across groups (*P*anova=0.0008), being elevated in the offspring of N/O mothers and significantly different to the offspring of N/N mothers (*P* = 0.0088), suggesting that childhood obesogenic environment adversely affects TC levels in males. The overall high-density lipoprotein (HDL-C) varied between groups (*P*anova<0.0001), and was lowest in the offspring of O/O mothers and significantly different to that in the offspring of N/O mothers (P =0.0017), suggesting an impact via the *in utero* obesogenic environment. The alteration in low-density lipoprotein (LDL-C) concentrations was mirrored, similar to HDL-C variations, within all groups (*P*anova=0.03) but no significant differences between groups was observed. The serum apoprotein A-I concentrations varied significantly between groups (*P*anova< 0.0001) but there was no in-group differences. Similarly, apoprotein B concentration did not vary overall (*P*anova=0.07) and no significant difference

between the offspring of N/N, N/O, O/N and O/O mothers were observed for apoprotein A-I and B. The systolic blood pressure (SBP) and diastolic blood pressure (DBP) showed variability across all groups (*P*anova<0.0001 and *P*anova=0.0006, respectively) and were highest in the offspring of O/O mothers. However, none of the between-group comparisons were significant for SBP or DBP, indicating roles of other factors responsible for this variation which cannot be explained completely by the obesogenic environment (**Table 10**).

Females

The TG was similar across all groups in female offspring, with no significant variability across the groups (Panova=0.2521), unlike in males. The TC level was elevated in offspring of O/O mothers in comparison to N/N mothers, however no significant difference between O/O and N/N (P=0.7714) or for the overall groups (P=0.2096) was observed. Offspring of N/O mothers differed significantly in TC level from offspring of N/N mothers (P=0.04), which may be explained by the influence of the childhood obesogenic environment. The HDL-C did not show significant variability across the groups (Panova=0.0913). There was a statistically significant variability, although it was small magnitude-wise, in LDL concentration between all groups (Panova=0.0368) with N/O mothers showing significant differences between both offspring of O/O mothers (P=0.0295) and N/N mothers (P=0.0277), showing that the obesogenic environment *in utero* and childhood has an effect on LDL concentrations in females. The serum apoprotein (Apo) A-I did not show any variability between groups (Panova=0.2934) but apoprotein B concentrations vary between groups (Panova=0.0384), although none of the groups differed significantly from each other. Both SBP and DBP did not differ between groups (Panova=0.1271 and Panova=0.0677 respectively) in contrast to males, and no significant difference between groups was observed, indicating no effect of the obesogenic environment on blood pressure in female offspring in this sample (**Table 10**).

Discussion

This study was done as a part of study designed by Samuelsson et al, 2008, who aimed to look for a programming effect on mice and this study replicates the experimental design using human data[232]. This is the first study which looks at these factors in a more longitudinal way than previously published studies. As it is difficult to develop a diet-induced experimental model in human, the study uses the NFBC dataset which is very dense and has several prenatal and postnatal data points available to test this hypothesis.

There is increasing evidence for a trans-generational transmissibility of a risk for chronic diseases associated with common pre-gestational obesity. The nature of such an association between maternal

metabolic health during the childbearing period and the health of her offspring later in life is poorly understood owing to important genetic, social and behavioural contributing factors[242]. This study specifically intended to assess the influence of both pre-gestational and postnatal overweight and obese mothers and to study the potential sexual dimorphism into the response of metabolic phenotypes. It was previously reported that pre-pregnancy BMI in overweight women predicts their risk of obesity 20 years later[243] exposing their child to an environment that may favor obesity. However the study observed that both the pre- and the postnatal exposure to maternal overweight may have a differential impact on an offspring's BMI. The extent to which the pre-natal exposure to pre-gestational overweight carries a specific risk for the later development of T2D and/or CVD is yet to be identified.

The fasting insulin, HOMA-IR, HOMA-B, triglyceride and Systolic blood pressure (SBP) differed significantly between offspring of mothers who remained lean (N/N) and those who were O/O throughout, but the difference was sex-stratified[244,245]. The mothers who were lean pre-pregnancy gained weight during pregnancy (N/O) and turned O/O showing an adverse profile within their adolescents. The characteristics of these mothers resonated with O/O mothers.

The most important result this study shows is that, in male offspring, mothers who were lean pre-pregnancy but became obese in later life gained the most weight, and differed in the fasting insulin, HOMA-IR and HOMA-B from the O/O group[246]. This result distinctly demonstrates that it is not merely obesity which has a programming effect but that the weight gained during pregnancy in combination with obesogenic environment is the cause of this elevation. Using the same dataset, Vaarasmaki et al, 2009, showed that adolescents of gestational diabetic mothers (GDM) mothers had symptoms of metabolic syndrome, with high WC and insulin level and homeostatic model assessment insulin sensitivity was lower in the offspring of women with GDM (OGDM) group[247]. Although this study has now been able to provide further evidence which supports that such changes not only occur in adolescents of GDM but in all O/O mothers, therefore, it is an indication that programming happens much earlier than the gestational stage. HDL and apoliprotein AI differs between N/O groups differed from O/O groups, suggesting that the characteristics of this group replicated very much that of O/O mothers and there is a hidden obesogenic environment which influences this group.

Table 10: Influence of the pre- and post-natal obesogenic environment on cardiovascular risk factors in adolescents of the NFBC 1986 birth cohort.

Outcome Variables	N/N	N/O	O/N	O/O	Overall ANOVA	N/N vs. O/O	N/O vs. O/O	N/O vs. N/N
	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	P	P	P	P
Males								
Triglycerides (mmol/L)	0.9 (0.8, 0.9)	0.9 (0.8, 1.0)	0.7 (0.4, 1.0)	1.0 (0.9, 1.1)	0.0046	0.0036	0.0929	1
Total cholesterol (mmol/L)	4 (3.9, 4.1)	4.2 (4.1, 4.3)	3.7 (3.4, 4.1)	4.1 (4.0, 4.2)	0.0008	1	1	0.0088
HDL (mmol/L)	1.3 (1.3, 1.3)	1.4 (1.3, 1.4)	1.3 (1.2, 1.5)	1.2 (1.2, 1.3)	<0.0001	0.0734	0.0017	0.3350
LDL (mmol/L)	2.4 (2.3, 2.4)	2.5 (2.4, 2.6)	2.2 (1.8, 2.5)	2.5 (2.4, 2.6)	0.0300	0.1858	1	0.0664
Apoprotein AI (mmol/L)	1.3 (1.3, 1.3)	1.3 (1.3, 1.3)	1.2 (1.1, 1.3)	1.3 (1.2, 1.3)	<0.0001	0.1489	0.0427	1
Apoprotein B (mmol/L)	0.6 (0.6, 0.6)	0.7 (0.6, 0.7)	0.6 (0.5, 0.6)	0.7 (0.6,0.7)	0.0742	0.8817	1	0.2915
SBP (mmHg)	119.7 (118.7, 120.7)	121 (119.6, 122.4)	120.7 (114.2, 127.1)	121.8 (119.9, 123.8)	<0.0001	0.3815	1	1
DBP (mmHg)	67.9 (67.3, 68.9)	67.9 (67.1, 68.8)	68.2 (64.1, 72.4)	68.7 (67.4, 69.9)	0.0006	1	1	1
Females								
Triglycerides (mmol/L)	0.9 (0.8, 0.9)	0.9 (0.8, 0.9)	0.9 (0.6, 1.2)	0.9 (0.8, 0.9)	0.2521	1	1	1
Total cholesterol (mmol/L)	4.4 (4.3, 4.5)	4.3 (4.2, 4.4)	4.2 (3.7,4.7)	4.4 (4.3, 4.5)	0.2096	1	0.7714	0.0369
HDL (mmol/L)	1.5 (1.5, 1.5)	1.5 (1.4, 1.5)	1.5 (1.2, 1.6)	1.5 (1.4, 1.5)	0.0913	0.3283	0.3787	1
LDL (mmol/L)	2.6 (2.5, 2.6)	2.5 (2.4, 2.5)	2.4 (1.9, 2.8)	2.6 (2.5, 2.7)	0.0368	1	0.0295	0.0277
Apoprotein AI (mmol/L)	1.4 (1.4, 1.4)	1.4 (1.4, 1.4)	1.4 (1.3, 1.5)	1.4 (1.3, 1.4)	0.2934	1	1	1
Apoprotein B (mmol/L)	0.7 (0.7, 0.7)	0.7 (0.6, 0.7)	0.7 (0.5, 0.7)	0.7 (0.6, 0.7)	0.0384	1	0.1419	0.0579
SBP (mmHg)	109.1 (108.2, 110.1)	111.1 (109.7, 112.3)	109 (101.8, 116.2)	111.2 (109.4, 112.9)	0.1271	0.4808	1	0.1382
DBP (mmHg)	66.5 (65.9, 67.1)	67.8 (66.9, 68.6)	64.7 (60.0, 69.4)	67.5 (66.4, 68.6)	0.0677	1	1	0.3096

Legend: Adjusted mean and 95% confidence intervals by gestational age and tanner scale. Only statistically significant differences at P <0.05 between groups after Bonferroni correction are presented.

The maternal pre-gestational obesity predisposes the offspring to a supplementary risk of being obese later in life. Critically onset-obesity induced by the obesogenic environment, such as diet after weaning, did not overcome the programmed changes induced by pre-gestational obesity. The study observed that some risk factors for T2D and CVD show specificity for maternal pre-gestational obesity and for sex. Suggestive additive effects are notably observed for fasting glucose and cholesterol. Most strikingly, for fasting insulin, the male offspring showed very atypical responses and despite a greater body fat mass in O/O offspring compared to N/O, those male offspring with a double exposure had higher insulin sensitivity but not within fasting glucose.

Whilst evidence establishing the programming of later health by maternal under-nutrition and fetal restriction is clear, the programming impact of fetal "overnutrition" on later health is questionable. The question of whether maternal gestational obesity relates to a problem of over nutrition is currently highly debatable and may be misleading. Our study excludes gestational diabetic individuals as they may confound our study results. Pre-gestational overweight mothers in human generally gained less weight than normal weight mothers and did give systematically gave birth to macrocosmic babies.

The study has yet to clearly identify the metabolic impact of common pre-gestational overweight or mild obesity which exposed the developing offspring to a mixture of contributing factors with potential programming effects. Pre-gestational overweight mothers are at a high risk of developing pre-diabetic states[243]. The fetus is likely to be exposed to higher concentrations of insulin and growth factors compared to those of normal weight mothers with potential subsequent epigenetic and trophic alterations in various developing organs. One key issue will be to understand how this trans-generational transmissibility of diabetogenic and cardiovascular risks is related to changes in pancreatic, liver and cardiovascular functions.

To our knowledge, this is the first study to demonstrate the difference between 'programming' due to maternal overweight/obesity during *in utero* and postnatal environment. Although the study accounts for wide metabolic anthropometric phenotypes and confounding factors, it was not possible to quantify for postnatal obesogenic environment due to lack of information on dietary and lifestyle factors from birth to adolescence in NFBC 1986. Although it is reasonable to assume that the mothers' weight status may indicate an obesogenic environment during child's growth; there could be other environmental factors which may have an effect that were not included. Therefore, further studies are required which can provide further information about nutritional and lifestyle status of mothers to justify such studies. In conclusion, this study shows the importance of weight monitoring during pregnancy and also during developmental stages of offspring.

5.1.3. Association of leukocyte telomere length at 31 years with adiposity-related anthropometric measurements at 31 years and change in BMI distribution from childhood to adulthood.

5.1.3.1. Introduction

Studies show that accelerated leukocyte telomere length (LTL) attrition is associated with CVD factors such as obesity, smoking and sedentary life-style factors[248]. Evidence from epidemiological studies suggest that obesity linking pathways, such as systemic inflammation and oxidative stress, may also influence LTL length and function[249]. Most previous studies report an inconsistent association of LTL and adiposity measures in adulthood. An inverse association between LTL and body mass index (BMI), waist-to-hip ratio (WHR) and visceral fat in adults is shown in some studies[152,154,156,160], while others suggest no association[153,250]. Studies indicate that this inverse association between adulthood LTL and adiposity measures may be more strongly associated in younger women (<30 years) compared to older women (>60 years)[160]. Additionally, one study suggests that women with fluctuating body weight between 50-55 years [251] may have a higher risk for shortening of LTL.

More recent epidemiological studies show an association between childhood obesity with adulthood LTL[252], however the number of studies investigating childhood obesity links with LTL are very limited, with results being equivocal[158,159,253]. Two cross-sectional studies, one in a European cohort aged 2-17 years [253] and another in Arab youth aged 5-12 years, have shown that severely obese children have a shorter mean LTL than those of their normal weight peers[159]. Another study suggests that this effect can only be observed in adults but not in children[158]. Moreover, none of the studies address the key issue, which is whether childhood obesity has an effect on adulthood LTL measures, given that it is an established fact that both child and adult adiposity-related phenotypes are associated with increased risk of age-related CVD[82,254].

In summary, there is a gap in understanding the longitudinal change in weight affecting LTL at 31 years. This study hypothesises that LTL at 31 years associates with childhood and adulthood adiposity measures, and that upward change in BMI from childhood to adulthood is inversely associated with LTL at 31 years. The main aim of this study was i) to understand the association between LTL at 31 years and childhood adiposity measures, ii) to confirm findings on LTL at 31 years association with adulthood adiposity-related phenotypes, and iii) to investigate association between LTL at 31 years and change in BMI from childhood to adulthood. The objective of the study was to understand the relationship between LTL at 31 years with longitudinal variation in adiposity-related factors in childhood and adulthood. The study compared

unadjusted and adjusted models, to investigate how these childhood and adulthood factors associated with LTL and if adjusting for confounders such as maternal parity and socio-economic status (SES) or individual's smoking, and SES affects the association. The literature search on this topic suggested that there are no longitudinal studies such as this which had prospective longitudinal growth data with multiple measurements as well as prenatal and adult contributing or confounding factors.

5.1.3.2. Study specific methods

Subjects and Study design

The description of the study population, NFBC1966, is given in detail in Chapter 4, Section 1. Briefly, of the 6,007 individuals who attended the clinical examination at 31 years n=5,753 individuals with DNA data were available at 31 years for the analysis in the current study (**Figure 7**). These individuals were combined with their available anthropometric data at 31 years, LTL measurements (n=5,620) and their childhood growth data (n=4,121). Participants who had missing LTL measurements or had missing adult BMI information (n=193) or less than 3 child growth measurements for growth modeling (n=190) were excluded from the analysis. The study utilised three datasets for analysis to maximise sample size and for each analyses increase the statistical power. The first dataset was restricted to only those individuals who had had data available for childhood BMI at AR (5.5 years) and LTL measurements at 31 years (n=3,903). The second dataset consisted of those who had data available for LTL measurements and adult anthropometric measurements at 31 years (n=5,404). The third dataset was restricted to individuals with LTL measurements at 31 years and longitudinal variation in BMI (n=3,703). The selected sample was representative of the wider NFBC 1966 subjects and no differences between the complete dataset and incomplete dataset were observed when tested at $P < 0.05$.

Variable selection and confounders

The independent variables selected for analysis were divided into childhood (Age and BMI at adiposity rebound (AR)) and adulthood adiposity measures (WHR, BMI, BAI and BMI *z-score* difference). Several confounders, such as mother's parity, socio-economic status (SES) and individual's age at menarche, smoking and SES at 31years were selected based on the *a priori* selection method and on previous knowledge of the field. Details on collection of these variables from questionnaire and hospital records are given in Chapter 4. Measurements for childhood BMI at 5 years were limited in sample size therefore the study uses the BMI at adiposity rebound (AR) as a proxy for BMI at 5 years. The growth modeling method for age and BMI at AR is presented in Chapter 4, section 2.1. Many previous studies associated age and

BMI at AR with adolescent adiposity [86] and adulthood metabolic health [14,255] therefore, these parameters can be considered good markers for childhood obesity.

The body mass index (BMI) was calculated as body weight in kilograms divided by height in meters squared (kg/m^2). Waist-hip ratio (WHR) was calculated as waist circumference (cm) divided by height (cm). Body adiposity index (BAI) was calculated as (hip circumference (cm)/height (m)) $^{1.5}$ -18, as described by Bergman et al, 2011[256]. The longitudinal variations in BMI from childhood and adulthood was calculated as the difference between the *z-score* of BMI at AR, around 5.5 years and *z-score* of BMI at 31 years, standardized to zero mean and 1 standard deviation (SD), calculated separately in males and females.

The mean relative LTL was measured by Dr Jess Buxton, Imperial College as described previously[253]. In summary, a multiplex quantitative real-time PCR (qPCR) method was used to measure the mean relative LTL in the genomic DNA samples prepared from peripheral blood[257]. The multiplex qPCR method provides a 'T/S ratio' for each DNA sample, which is the relative measure of the amplication of the telomeric DNA sequence (T) compared to that of a single copy of gene(S) in each test sample, normalized using a common reference DNA sample. The overall mean coefficient of variation (CV) for T/S values of duplicate test samples on the same plate was 5%, and the mean inter-run CV for selected samples was 6.2%.

Statistical analyses

The LTL measurements were selected as the main outcome. LTL at 31 years were right skewed and hence it was log transformed to achieve normal distribution and to improve interpretability. Descriptive characteristics of the study samples are described in the overall study sample, and sex-stratified, with means and SD calculated for continuous variables and percentages calculated for categorical variables. The Kolmogorov-Smirnov test was performed to test normal distribution of continuous variables. Pearson's coefficient was then used to test for correlation between covariates; Student's t-test for unpaired data was used to test the mean difference between sexes and the Chi square test for the difference in percentage distribution in categorical data with a P_{diff} value of < 0.05 was considered statistically significant. The linear regression model was used to estimate effect size of association between independent variables and the outcome.

Two models were tested, 1) Unadjusted model, testing association between log-transformed LTL and adiposity-related measurements from childhood and adulthood and 2) Adjusted model, testing association

between log-transformed LTL and adiposity-related measurements from childhood and adulthood controlling for maternal parity, SES, age at menarche (only females), SES, smoking and parity in offspring at 31 years. The estimated effects were reported as a percentage change in LTL by multiplying the exponential regression coefficients by 100. All statistical tests of hypotheses were two-sided with a significant association threshold of $P < 0.05$. The Benjamin and Hochberg method was used to address the issue of multiple testing, with a $P_{corrected}$ value of < 0.05 considered statistically significant[258].

5.1.3.3. Results

At birth measurements

Table 11 shows the descriptive statistics for all participants, and by sex. Mother's parity collected at birth was on average 3 children overall, and no difference between sexes was observed (P_{diff}=0.7122). Mother's socio-economic status (SES) collected at birth had similar distributions in both sexes with no significant difference observed (P_{diff}=0.1522). 7% of children were born to mothers who were housewives, 17% belonged to white collar, 33% to blue collar and 43% were born to a manual worker or a farmer.

Childhood measurements

The mean age and BMI at adiposity rebound (AR) for all individuals were 5.7 years and 15.37kg/m2 in the sample population, however, there was a sex difference in the mean age and BMI at AR (P_{diff}<0.0001 and P_{diff}=0.0008, respectively). Males on average had 15.42 kg/m^2 BMI at AR while females had lower BMI at AR, 15.33 kg/m^2 (P_{diff}=0.0008 for gender difference), and age at AR was earlier in females (5.6 years) than in males (5.8 years) (P_{diff}<0.0001). The BMI at AR were divided into categories according to IOTF guidelines for children[25] with 5% of children being underweight, 80% normal weight, 10% overweight and 5% were obese. No gender difference was observed.

Adulthood measurements

The mean BMI at 31 years was, overall, 24.7 kg/m^2, and it was observed that males had a 1.04 kg/m2 higher mean BMI than females (P_{diff}<0.0001). Overall, 2% of individuals were classified as underweight, 58% normal weight, 31% overweight and 8% obese in both sexes; however adult males had a bigger proportion of overweight individuals than females. The mean WHR at 31 years in males (0.91) was higher than WHR of females (0.81) (P_{diff}<0.0001), this result is consistent with previous findings showing that females have lower WHR than males[259]. BAI indicated a significantly higher percentage of fatness in females (28%) than in males (23%) (P_{diff}<0.0001), this result was also consistent with another study suggesting that women carry more body fat than men [260].

Figure 7: Flowchart showing sample selection for the analysis from the NFBC 1966.

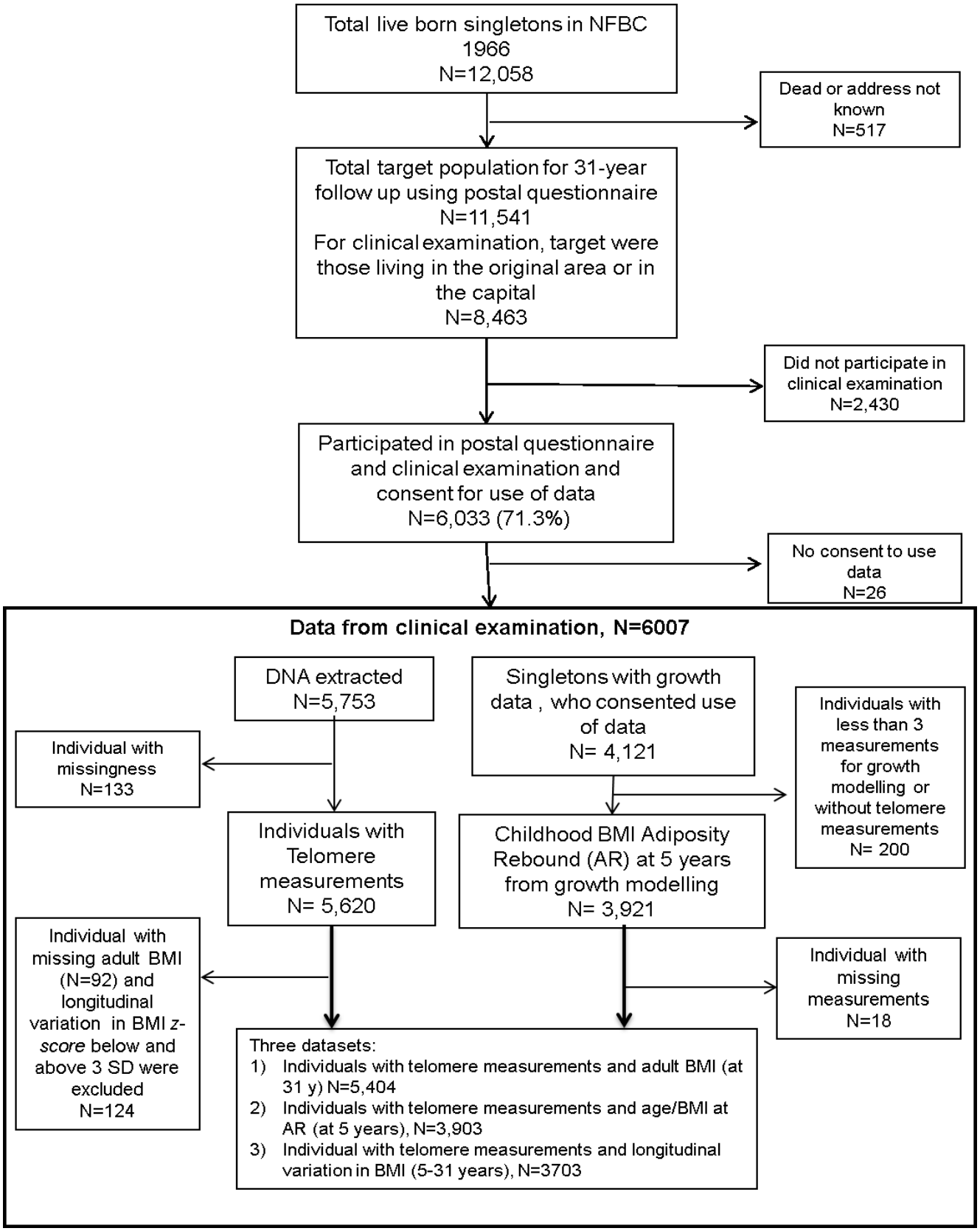


Table 11 shows that males were more often heavy smokers (>10 cigarettes) than females (30% compared to 12%) (P_{diff} <0.0001). The SES distribution between males and females' differed, with higher proportions of females being white and blue collared (P_{diff} <0.0001) in comparison to men. Parity at 31 years was, on average, 1.26 children per individual in the overall sample and there was a significant difference between males and females observed (P <0.0001). Higher proportion of males (46%) than females had no children while females had more often 2 or more children (49%). The mean leukocyte telomere length (LTL) ratio was 1.12 in males and 1.17 in females, with females showing significantly longer LTL than males (P_{diff} <0.0001). This association result also fitted with as previously suggested by Barrett et al in 2011[261]. The mean longitudinal variations in BMI *z-score* was -0.02 in males and -0.01 in females, both being within the normal distribution of *z-score*.

*Association between childhood adiposity measures and LTL at 31years (**Table12**)*

There was no association between age at AR with LTL at 31 years in either males (P = 0.0923) or females (P = 0.88) (**Table 12**). The adjustment for potential confounding factors did not substantially change the association ($P_{corrected}$ =0.26 in males; $P_{corrected}$ =0.71 in females). BMI at AR associated positively with LTL at 31 years in females (P=0.0074), but not in males (P=0.79) (**Table 12**). After adjustments the effect size reduced (β=1.71%, 95% CI: (0.26% to 3.18%)) but the association remained significant even after multiple adjustments ($P_{corrected}$ = 0.0406). The results indicate that girls with higher BMI at AR at 5 years have longer LTL at 31 years. In males, no association was observed even after adjustments for potential confounding factors ($P_{corrected}$ = 0.68).

Association between adulthood adiposity-related phenotypes and LTL at age 31

The unadjusted model showed that LTL at 31 years associated with lower WHR at 31 years in both sexes (P = 0.0138 in males, P=0.0003 in females), demonstrated consistently in other studies[153,251]. However, after adjustments with confounders, the association remained significant in both males ($P_{corrected}$ =0.0490) and females ($P_{corrected}$ =0.0078), suggesting the independent (of accounted factors) association between LTL and WHR at 31 years. BMI also showed a negative association with LTL in males (P=0.0017) and females (P=0.0048), with the association being stronger in males. Adjustments for confounders reduced the strength of association in males ($P_{corrected}$ =0.0246), but no difference for females was observed ($P_{corrected}$ =0.0286). The percentage of fatness measured by BAI at age 31, was also inversely

Table 11: Descriptive characteristics of study participants in Northern Finland Birth Cohort (NFBC 1966).

	All			Males			Females			
	N	Mean		N	Mean		N	Mean		Pdiff
At birth										
Maternal parity	5462	2.95 (2.28)		2639	2.93(2.29)		2823	2.97(2.26)		0.7122
No previous pregnancy (%)	136	2		70	3		66	2		
1 previous pregnancy (%)	1731	31		843	31		888	31		
2 or more previous pregnancies (%)	3731	67		1796	66		1935	67		
Socio-economic status at birth	5438			2626			2812			0.1522
Stay at home (%)	373	7		194	7		179	6		
White collar (%)	903	17		457	17		446	16		
Blue collar (%)	1812	33		865	33		947	34		
Farmer (%)	2350	43		1110	42		1240	44		
Childhood (5y)										
Age at Adiposity Rebound (AR) (years)	3903	5.71(0.92)		1971	5.81(0.89)		1932	5.61(0.95)		<0.0001
BMI at Adiposity Rebound (AR) (Kg/m2)	3903	15.37(1.07)		1971	15.42(1.00)		1932	15.31(1.13)		0.0008
Underweight (%)	193	5		98	5		95	5		
Normal (%)	3138	80		1581	80		1557	80		
Overweight (%)	392	10		197	10		195	10		
Obese (%)	198	5		99	5		99	5		
Adulthood (31y)										
Age at menarche (years)	2851						2851	12.92		
Waist-Hip Ratio (WHR)	5398	0.86 (0.08)		2679	0.91(0.06)		2719	0.81(0.07)		<0.0001
Body mass index (BMI) (kg/m2)	5404	24.70(4.24)		2662	25.23(3.63)		2742	24.19(4.69)		<0.0001
Underweight (%)	117	2		28	1		89	3		
Normal (%)	3069	58		1309	50		1760	65		
Overweight (%)	1647	31		1049	40		598	22		
Obese (%)	486	9		223	9		263	10		
Body adiposity index (BAI) (% fatness)	5277	25.60(4.58)		2615	23.01(2.90)		2662	28.15(4.50)		<0.0001
Adult Smoking (distribution among smokers)	5252			2528			2724			<0.0001
Light (%)	4174	79		1775	70		2399	88		
Heavy (%)	1078	21		753	30		325	12		
Socio-economic status	5534			2669			2865			<0.0001
Farmer (%)	208	4		130	5		78	3		
Self-employed (%)	1312	24		731	27		581	20		
White collar (%)	3128	57		1441	54		1687	59		
Blue collar (%)	886	16.00		367	14.00		519	18.00		
Parity at 31years	5502	1.26 (1.29)		2646	1.03 (1.21)		2856	1.47(1.33)		<0.0001
No previous child (%)	2184	39		1297	47		887	30		
1 previous child (%)	1117	20		506	19		611	21		
2 or more previous children (%)	2297	41		906	34		1391	49		
Leukocyte Telomere Length (LTL)	5598	1.14 (1.40)		2709	1.12(1.40)		2889	1.71(1.41)		<0.0001
Longitudinal BMI difference (5 to 31 year) [#]	3683			1896	-0.02		1787	-0.01		

Sample sizes and means (standard deviation (SD)) are given for all continuous variables and sample sizes and category percentage are given for categorical variables.
AR, adiposity rebound; LTL, leukocyte telomere length. BMI, body mass index - calculated as weight (kg)/ height $(m)^2$. BAI, body adiposity index (% fat) - calculated as hip circumference (m) / $height(m)^{1.5} - 18$.
Student's t-test was performed for calculating gender differences, P_{diff}.

associated with LTL at 31 years in females (P=0.004) but not in males (P=0.16). Every 1 unit increase in % fatness, the LTL decreased by 0.48 % in females although after adjustment the association attenuated but remained significant ($P_{corrected}$ =0.02).

Association between increase in BMI from childhood to adulthood and LTL at 31

The longitudinal change in BMI *z-score* from 5 years to 31 years was inversely associated with LTL at 31 years in females (P=0.0004) but not in males (P=0.2381). One unit increase in BMI *z-score* associated with about 3.08% decrease in LTL at 31 years in females, which strengthened to about a 4.7% decrease in LTL at 31 years after adjustments with confounding factors, suggesting independent association in females ($P_{corrected}$ =0.0078). In males, the association between LTL at 31 years and BMI *z-score* did not change even after adjustments for confounder ($P_{corrected}$ =0.2681).

5.1.3.4. Discussion

LTL has been proposed as a biomarker for risk of disease development in later life[262], although this evidence is still debatable. However, as shown by reviews on telomere length and CVD risks factors, not many studies have shown consistent results regarding LTL association with CVD disease risk factors[262], while most are underpowered. This is the first study to present such large samples for LTL measurements and it is unique as it uses growth modeling phenotypes to investigate the relationship between childhood adiposity-related phenotypes and LTL in adulthood. The advantage of growth modeling has been discussed in detail previously.

The major strengths of this study include: the uniform age and ethnicity of the participants, the even sex ratio and the consistency of the method used to obtain LTL data for a large number of samples. Although doubts have been expressed over the reliability of the qPCR method [263,264], the telomere measurements used here show the expected sex difference, and have previously been used in the identification of robust genetic associations for LTL [265].

Table 12. The unadjusted and adjusted associations between leukocyte telomere length (LTL) at 31 years and early life/adult adiposity related measures by sex. The beta (95% confidence interval) is given as percentage difference. The $P < 0.05$ is considered significant.

	Unadjusted model					*Adjusted model*[d]				
	% difference	**95 % CI**		***P***		**% difference**	**95 % CI**		$P_{Adjusted}$	$P_{corrected}$
MEN										
Childhood (~5 years)										
Age at AR	1.47	-0.24	3.20	0.0923		1.22	-0.57	3.05	0.1836	0.2681
BMI at AR	0.21	-1.28	1.73	0.7853		0.40	-1.20	2.02	0.6252	0.6820
Adulthood (31 years)										
Waist-hip ratio (WHR)	-0.44	-0.79	-0.09	0.0138		-0.42	-0.79	-0.04	0.0286	0.0490
BMI (Kg/m2)	-29.51	-43.31	-12.35	0.0017		-26.86	-41.97	-7.80	0.0082	0.0246
Body adiposity index (BAI), % fatness	-0.25	-0.76	0.13	0.1640		-0.29	-0.76	0.18	0.2234	0.2681
BMI trajectory *z-score*	-0.98	-2.58	0.65	0.2381		-1.08	-2.78	0.64	0.2179	0.2681
WOMEN										
Childhood (~5 years)										
age at AR	-0.12	-1.69	1.48	0.8818		-0.33	-2.07	1.44	0.7123	0.7123
BMI at AR	1.84	0.49	3.21	0.0074		1.71	0.26	3.18	0.0203	0.0406
Adulthood (31 years)										
Waist-hip ratio (WHR)	-0.51	-0.78	-0.23	0.0003		-0.50	-0.79	-0.21	0.0008	0.0078
BMI (Kg/m2)	-23.04	-35.82	-7.71	0.0048		-22.03	-35.76	-5.37	0.0119	0.0286
Body adiposity index (BAI), % fatness	-0.48	-0.71	-0.14	0.0035		-0.44	-0.74	-0.13	0.0053	0.0212
BMI increase *z-score* (5-31y)	-3.08	-4.74	-1.40	0.0004		-2.91	-4.65	-1.15	0.0013	0.0078

AR, adiposity rebound

BMI, body mass index, calculated as weight (kg)/height(m)2.

BAI, body adiposity index, calculated as hip circumference (m)/height (m)1.5 -18.

BMI trajectories z-score were calculated from a difference between z-scores of BMI at AR and BMI at 31 years.

[d]Linear regression model adjusted for maternal parity, SES at birth, SES at 31 years, smoking at 31 years, SES at 31 years, parity at 31 years and qPCR plate. Additionally, in females the model was also adjusted for age at menarche in addition to other confounding factors (given above).

The study shows that anthropometric measures from early childhood to adulthood associate with a shorter mean relative LTL in women at age 31, regardless of confounding factors. This study also confirms the inverse association of adulthood adiposity measures such as BMI and WHR at 31 years with concurrent LTL in both adult men and women. The BAI was also inversely associated with LTL at 31 years but only in females, which remained associated after adjustment for confounders. There was a clear indication that these adiposity-related measures showed stronger association in female than males.

The inverse association between BMI, BAI and LTL at 31 years was consistent with several previous reports of an inverse relationship between measures of adiposity and concurrent telomere length[160,252,264,266]. The BAI is a new method for measuring body fat, although it remains inconclusive if this measure is a better indicator of adiposity than BMI[256,267]. However, results from this study show that females may have a disproportionate fat distribution which could be linked with shortening of LTL, and is a marker for higher risk of CVD than in men[268]. Further work needs to be conducted to completely understand these findings. The study also identified a significant inverse association between WHR and LTL in adult males and females, which were consistent with a previous study that found WHR to be a significant predictor of a faster telomere shortening rate in a largely male cohort (>80%) [153]. The results from this study suggest that both males and females with higher adiposity-related factors eventually drive themselves towards earlier aging and a greater risk for CVD.

There was no evidence for an association between age at adiposity rebound (AR) and LTL at age 31 in either sex in this cohort, despite previous studies showing that earlier age at AR predicts poor metabolic health in adulthood [255]. Interestingly, the study identified a *positive* association between BMI at AR and adult telomere length in women, which remained significant after correcting for maternal parity and SES at birth, age at menarche, adult SES, smoking status and parity at 31 years. It was evident that only a small proportion of children were overweight (10%) and obese (5%) in the NFBC1966 compared to normal group (80%), suggesting that being within a normal range of BMI may be beneficial for telomere length in adulthood. These results need further work to detangle the causal mechanisms behind this positive association between BMI at AR and LTL measurements.

The main finding of this study was that a greater increase in a standardised BMI in women spanning from childhood to adulthood is associated with shorter telomeres at age 31. This result supports a previously published report showing that maintaining a healthy body weight helps maintain the LTL in women[251]. However, this study adds to this knowledge and suggests that women who maintain their longitudinal

variation of body weight from childhood (5 years) to adulthood (31 years) are in lower risk groups compared to women with fluctuating body weight.

However, one of the limitations of the study was that as DNA samples were available for only one time-point, it was not possible to study the association between longitudinal changes in BMI and telomere shortening rate over the same period. I was also unable to study the potential effects of interim changes in BMI between childhood and adulthood, and thus the influence of fluctuating adiposity levels during this time. Finally, it would have been interesting to investigate the effects of underweight (BMI < 18.5) on LTL, but I did not have sufficient numbers of such individuals available in this cohort.

Despite these limitations, the study adds to the growing evidence supporting the hypothesis that factors associated with obesity may have similar effects on lifespan and cellular function to aging [269]. Specifically, the results suggest that high levels of overall body fat in women and central adiposity in men are associated with accelerated ageing at the cellular level. The biological mechanisms that may underlie associations between LTL and increased BMI are currently unknown, although obesity is recognized as a state of both increased oxidative stress and low-grade inflammation; both processes that are believed to accelerate leukocyte telomere attrition [264,270,271].

The findings from this study only provide an indication towards identification of risk factors. Further investigation should be carried out into the potential role of telomere length as a biomarker to assess the effectiveness of interventions to combat obesity, such as maintain healthy weight. Longitudinal studies are now required to further assess the long-term effects of weight gain after childhood on LTL, healthy aging and cardio-metabolic disease risk and to investigate the sex differences observed here.

5. 2 Genome-wide association analyses

This section presents results from genome-wide association studies (GWAS) conducted on infant head circumference (HC) and early growth phenotypes of peak height and weight velocities (PHV, PWV) during the first year of life, BMI at adiposity peak (AP) in infancy, and BMI and age at adiposity rebound (AR) in childhood. NFBC (1966 and 1986) results were incorporated into meta-analyses combining several GWAS from the early growth and genetics (EGG) consortium. The first section (5.2.1) describes the GWAS conducted on infant HC in the NFBCs (1966 and 1986). This EGG project was led by a team from Erasmus Medical College, The Netherlands (Dr. Taal as the lead analyst and myself as joint first author) and published in Nature Genetics[176]. Section 5.2.2 presents the GWAS conducted on early growth phenotypes in the NFBCs (1966 and 1986). This EGG project was led by a team from Imperial College, London, UK (with myself as lead analyst) and a manuscript on the work has been submitted to Nature Communications. Details of my role in both studies are described in sections 5.2.1 and 5.2.2.

5.2.1. Genome-wide association study to find genetic variants affecting infant head circumference (HC)

5.2.1.1 Introduction

Most human brain development occurs in the first two years of life. It is thought that brain growth increases the pressure on the cranium, leading to growth of the skull and an increase in head circumference (HC). This makes HC an indirect measure for the development of the brain[272]. Previous studies associated measurements of HC with intrauterine growth restriction (IUGR) in newborns[273], linking HC to smaller birth weight[274], cognitive[275] and behavioral development[276]. Galton (1869) was the first to propose the association between HC and general mental ability (GMA)[277], suggesting that head size could be an important factor in the relationship between intelligence and learning. Further studies confirmed a strong positive correlation between HC and both brain volume and intelligence [276,278-280], although one study did not[281]. Among these studies, the most reliable findings may be from the within-family study [279] because it controls for most confounders, such as social class, parenting style, and general nutrition, giving a more accurate estimate of the relationship. Studies have also associated HC with mother's education[282], socio-economic status, and other variables known to be associated with variations in intellectual ability[283].

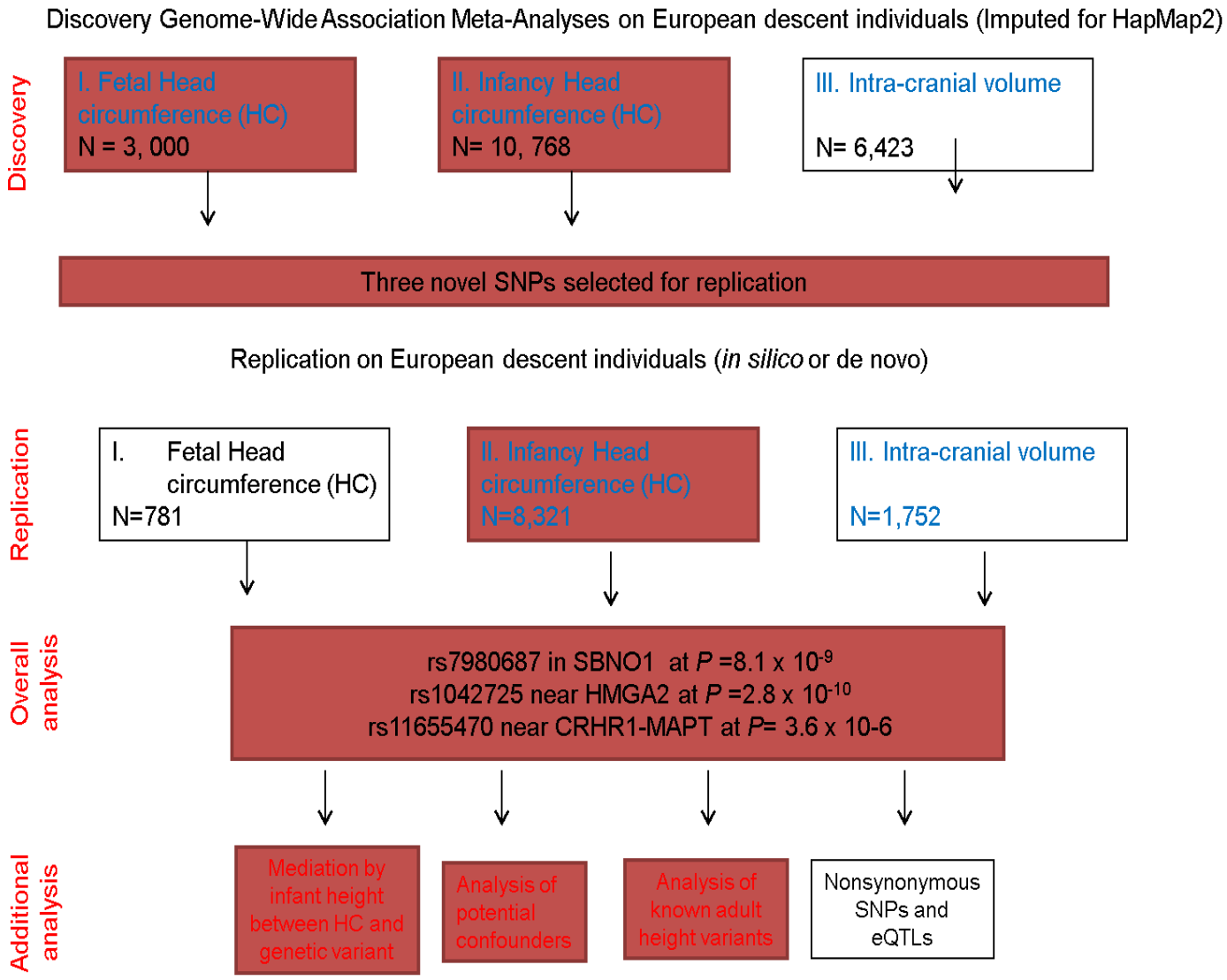
HC is a complex trait with high estimated heritability ($0.78 > R^2 > 0.98$), based on Smit et al, 2010[284], which investigated Dutch and Australian twin families. Based on current evidence and knowledge of neurodevelopment it is likely that both genetic and environmental factors affect brain development and also head size[285]. Several rare genetic mutations have been identified for HC, which has been linked to diseases, such as, microcephaly, autism and intelligence disability[286,287]. None of these studies have investigated normal variation in HC.

The aim of the present study is to identify common genetic variants that influence the population variation in infant HC and test if infant height mediates the association between any identified genetic variants and infant HC. The objective is to provide further insight into the molecular mechanisms that may be responsible for brain development. The hypotheses tested are: (1) there are genetic variants that determine HC at birth and in infancy, and (2) the associations between genetic variants and infant HC may have an indirect effect via infant height.

The overall project was led by the team from Erasmus Medical College, The Netherlands (Dr. Taal as a main analyst). I was responsible for analysing data from the NFBC1966 and NFBC1986. My role in this study was to conduct 1) a discovery stage GWAS on NFBC1966 data, and 2) perform follow-up association analyses of the leading SNPs in the NFBC1986. I also conducted a mediation analysis in the NFBCs (1966 and 1986) as part of the project. In this chapter, I explain the discovery and replication stages and finally the combined analysis of all 13 participating cohorts. The mediation meta-analysis performed on the NFBCs (1966 and 1986) and 10 other cohorts are also described. This chapter covers the topics directly relevant to this thesis and presents overall results which involved my participation.

The Nature Genetics publication [288] includes further details, such as association analyses between genetic variants and both fetal head circumference and intra-cranial volume, however these have not been included in this chapter since the NFBC1966 and NFBC1986 did not contribute to those analyses. Further details on these association analyses can be found in Taal et al, 2010[288]. On the initiation of the study it was agreed that this EGG-consortium related work, where I and the Imperial College London team had a substantial role, would form part of my thesis (the NFBC data represented about 1/3 of the whole study population and 40% of the discovery sample).

Figure 8: Study design showing the analyses performed for GWAS on head circumference (HC). The boxes highlighted represent the analyses performed for the NFBCs (1966 and 1986).



5.2.1.2. Study specific materials and methods

The genotyping details of the NFBCs (1966 and 1986) and other cohorts have been described in detail in Chapter 4, section 4. In summary, with birth and infant head circumference (HC) the primary outcome was transformed into sex and age-adjusted standard deviation scores (SDS) using Growth Analyser software (https://www.growthanalyser.org/; Dutch Growth Research Foundation, Rotterdam, The Netherlands), which used reference growth curves from the Netherlands, 1997. These sex and age-adjusted SDs were calculated, as the relationship between head circumference and age during infancy is non-linear, and the variance increases by age.

As a general strategy, HC closest to 18 months (range 6-30 months) was selected for analysis if the cohort had multiple measurements. In the NFBCs (1966 and 1986), the closest measurements were available at 12 months. One child of multiple birth (first alive born) and premature subjects (NFBC 1966, N=358; NFBC 1986, N=496) (born before 37 weeks of gestation) were also included in the study to increase the sample size. Non-caucasians were excluded from the study.

Statistical analysis

Figure 8 shows all the analyses performed for the study, with boxes highlighted in red representing those performed in the NFBCs (1966 and 1986). In summary, approximately 2.5 million directly genotyped and imputed SNPs from infants of European descent from seven discovery GWA studies (N=10,768) were meta-analysed (4, 287 subjects from NFBC1966). The associations were validated in six replication cohorts (N=8,321 of whom 2,533 from NFBC1986). The final meta-analyses were performed by inverse-variance method (N=19,089).

As shown in **Figure 8**, further analyses were performed to verify that the association between SNPs and infant HC were not confounded by other covariates which have previously associated to HC. Therefore, covariates such as height (cm), weight (kg) and age at measurement (years), breastfeeding (months), maternal educational level and sex were associated separately with HC and genetic variants, where data were available, using linear or logistic regression models.

Taal et al, 2010, also conducted the analyses for the association of the previously identified 180 adult height loci, published in Allen et al, 2010[289] and HC in infancy, using a linear regression model on discovery meta-analysis results. Results of *P*-value lower than 2.8 x 10^{-4} (0.05/180) was considered significant (Bonferroni correction).

To investigate possible mediation effects of the genetic variants on HC through height, the study conducted linear regression analysis with, and without, adjustment for the height SD score. The mediation framework (seemingly unrelated regression analysis (STATA)[290]or a simple pathway analysis (MPLUS)[290]) using an additive model was performed to assess the direct and indirect SNP effects (mediated through height) on HC for each of the identified genetic variant using univariate and multiple linear regression models. Furthermore, to test whether the associations between the SNPs and infant HC were similar in males and females, the analyses were repeated by sex. Where possible, the results were meta-analyzed to assess overall evidence of association.

5.2.1.3 Results

Study sample

Table 13 shows the selected characteristics of study samples in infant HC GWAS; and results from discovery and replication analyses by the studies and overall. A more detailed table describing the cohorts, genotyping information and quality control cut-offs applied by each study is presented in the supplementary information of the published paper[288]. The range of age for individuals in participating cohorts varies from 11.8 to 18.1 months. All subjects were of European descent.

Summary of overall discovery stage meta-analyses

Overall, the discovery stage meta-analyses were performed on ~ 2.5 million directly-genotyped and imputed SNPs (**Table 13**). From the meta-analyses of seven studies (ALSPAC, CHOP, COPSAC, LISA (D), NFBC 1966 and RAINE), top-three lead SNPs at $P < 1.5 \times 10^{-6}$ were selected for follow-up (**Appendix Table 1**). Two loci were on chromosome 12 and one on chromosome 17 (**Appendix Table 2**). Nine other loci associated with infant HC at $P < 1 \times 10^{-5}$ but were not taken forward for replication (**Appendix Table 3**).

Table 13: Summary results for top-three loci associated with head circumference (HC) for each participating cohort. For each of the 13 cohorts, the median age, total N, percentage of male, minor allele frequency (MAF), effect size (and SE), and P-value are shown as well as the latter for overall discovery stage, replication stage and meta-analysis results.

Study Type	Study	Year(s) of birth	Median age (months)	Total N	% male	SBNO1 (rs7980687)				nearest gene HMGA2 (rs1042725)				near gene CRHR1/MAPT (rs11655470)			
						MAF	B (SD score)	SE	*P*-value	MAF	B (SD score)	SE	*P*-value	MAF	B (SD score)	SE	*P*-value
Discovery	ALSPAC(D)	1991-2	18.9	1,748	53	0.19	0.105	0.038	$6x10^{-3}$	0.47	-0.071	0.031	0.02	0.41	0.114	0.031	$3x10^{-4}$
	CHOP	2006-10	18.5	1,008	59	0.20	0.041	0.058	0.48	0.48	-0.017	0.046	0.72	0.39	0.036	0.048	0.45
	COPSAC	1998-2001	18.1	369	49	0.19	0.083	0.086	0.33	0.47	-0.026	0.065	0.69	0.42	0.159	0.063	0.01
	Generation R	2002-6	13.1	2,240	52	0.21	0.064	0.031	0.04	0.49	-0.059	0.026	0.02	0.39	0.060	0.026	0.02
	LISA(D)	1998-9	11.8	357	56	0.21	-0.045	0.077	0.56	0.48	-0.059	0.060	0.33	0.49	0.068	0.061	0.26
	NFBC 1966	1966	12.3	4,287	49	0.20	0.181	0.041	$1x10^{-5}$	0.49	-0.074	0.033	0.02	0.41	0.068	0.033	0.04
	RAINE	1989-91	13.1	759	53	0.19	0.108	0.058	0.06	0.50	-0.179	0.043	$4x10^{-5}$		-0.001	0.044	0.09
Discovery meta-analysis				10,768			0.091	0.018	$3.3x10^{-7}$		-0.072	0.014	$6.6x10^{-7}$		0.070	0.015	$1.4x10^{-6}$
Replication	ALSPAC	1991-2	18.9	3,163	51	0.20	0.042	0.030	0.16	0.49	-0.088	0.024	$3x10^{-4}$	0.40	0.044	0.024	$6x10^{-4}$
	DNBC	1996-2002	12.1	531	54	0.20	0.120	0.070	0.09	0.45	-0.049	0.058	0.40	0.45	0.060	0.058	0.30
	EFSOCH	2000-4	12.1	703	52	0.20	0.054	0.061	0.37	0.50	-0.019	0.046	0.67	0.41	0.027	0.046	0.56
	INMA	2004-7	13.9	693	53	0.16	0.020	0.062	0.75	0.44	-0.029	0.045	0.52	0.36	0.022	0.046	0.64
	GINI + LISA	1995-9	11.8	698	51	0.21	0.020	0.060	0.74	0.50	-0.092	0.049	0.06	0.40	-0.070	0.050	0.16
	NFBC 1986	1985-6	12.0	2,533	48	0.22	0.082	0.035	0.02	0.49	-0.034	0.029	0.25	0.50	0.019	0.287	0.51
Replication meta-analysis				8,321			0.055	0.018	$2.5x10^{-3}$		-0.058	0.015	$8.3x10^{-5}$		0.025	0.015	0.093
Overall meta-analysis				19,089			0.074	0.013	$8.1x10^{-9}$		-0.065	0.010	$2.8x10^{-10}$		0.048	0.010	$3.6x10^{-6}$

MAF; Minor allele frequency, SE; standard error; β is the difference in head circumference SD score per minor allele (additive model).

P-value is obtained from linear regression of the SNP against infant head circumference (HC) SD score (additive model). All study samples were of European descent.

Key to study names: ALSPAC (D), Avon Longitudinal Study of Parents and Children Discovery subset; CHOP, Children's Hospital Of Philadelphia; COPSAC, Copenhagen Prospective Study on Asthma in Childhood; Generation R, the Generation R Study; LISA (D), Lifestyle – Immune System – Allergy Discovery subset; NFBC1966, Northern Finland Birth Cohort 1966; RAINE, The Western Australian Pregnancy Study; ALSPAC (R), Avon Longitudinal Study of Parents and Children Replication subset; DNBC, Danish National Birth Cohort; EFSOCH, Exeter Family Study Of Childhood Health; INMA, INfancia y Medio Ambiente [Environment and Childhood] Project; GINI+LISA (R), German Infant Study on the influence of Nutrition Intervention Munich + Lifestyle – Immune System – Allergy Replication subset; NFBC1986, Northern Finland Birth Cohort 1986.

Table 14: Combined analyses of the top-three SNPs associating with head circumference (HC) with other phenotypes.

Marker	Nearest gene	Head circumference in third trimester of pregnancy (SD score)				Head circumference at birth (SD score)				Intra cranial volume (ml)[a]				
		Total N	Beta	Se	P-value	Total N	Beta	Se	P-value	Total N	Mean age at measurement (years)	Beta	Se	P-value
rs7980687_A	*SBNO1*	3,781	0.089	0.029	$1.9x10^{-3}$	17,330	0.05	0.012	$5.2x10^{-5}$	8,175	67.5	0.72	2.03	0.72
rs1042725_T	*HBGA2*	3,781	-0.075	0.023	$9.9x10^{-4}$	17,074	-0.031	0.1	$1.9x10^{-3}$	8,175	67.5	-7.18	1.61	$8.8x10^{-6}$
rs11655470_T	*CRHR1*	3,781	0.049	0.024	0.037	17,695	0.03	0.1	$2.0x10^{-3}$	8,175	67.5	3.54	1.69	.036[#]

MAF; minor allele frequency, Se; standard error, *B* is the difference in head circumference SD score per minor allele. P value is obtained from linear regression: the SNP and sex against the head circumference SD score in fetal life (additive model); SNP, sex and gestational age against birth head circumference SD score at birth (additive model); and SNP, age and sex against intracranial volume[291] (additive model). All study samples were of European descent.

[a]The intra cranial volume additive model were run using data from Ikram et al, 2012[291].

[#] A variant further downstream (rs9915547; r^2=0.22 with HapMap CEU) showed association at genome-wide significance (P=1.5 x 10-12) with adult intracranial volume[291].

Table 15: Association of top-three head circumference SNPs with infant height and weight. Of the total of 13 study cohorts, 12 study cohorts contributed to these analyses since data on infant height and weight were not available in one cohort.

Index SNP	Location	Nearest gene	Association between SNPs and infant height (SD score)				Association between SNPs and infant weight (SD score)		
			N in the meta-analysis	Effect size (95% CI)	P value		N in the meta-analysis	Effect size (95% CI)	P value
rs7980687	12q14.31	SBNO1	16,664	0.034 (0.009, 0.059)	7.3×10^{-3}		16,532	0.038(0.013, 0.064)	3.3×10^{-3}
rs1042725	12q15	*HBGA2*	16,557	-0.042 (-0.063, -0.020)	1.5×10^{-4}		16,457	-0.038(-0.059, -0.018)	$2.9x 10^{-4}$
rs11655470	17q21.1	*CRHR1*	16,512	-0.009(-0.030, 0.012)	0.38		16,392	0.005(-0.016, 0.025)	0.67

95% CI; 95% Confidence interval. Effect estimates reflect the difference in head circumference SD score per minor allele.

The *P* value is obtained from a linear regression of the SNP against the head circumference SD score (additive model).

The *P* value is obtained from a linear regression of the SNP + the current height SD score against the head circumference SD score (additive model).

All study samples were of European descent. Of the total of 13 study cohorts, 12 study cohorts contributed to these analyses since data on height and weight were not available in the CHOP (Children's Hospital Of Philadelphia) study

Figure 9 shows the Manhattan plot for the association at overall discovery stage. The locus at SNP rs7980687 on chromosome 12 located in protein strawberry notch homo log 1 gene (*SBNO1*, *P*=3.3 x 10^{-7}), associated with higher HC by 0.09 SDS per A allele (**Table 13**). The locus at SNP rs1042725 on chromosome 12 near high mobility group AT-hook 2 gene (*HMGA2*, *P*=6.6x10-7) associated with a decrease in HC by 0.07 SDS per T allele of the SNP. Finally, locus at SNP rs11655470 at chromosome 17 near corticotropin releasing hormone receptor1 gene (*CRHR1*, *P*=1.4x 10-6) associated with higher HC by 0.07 SDS per T allele of SNP (**Table 13**). Genomic control correction was applied to adjust for population stratification within each cohort (lambda value ranging from 1.007-1.054).

Summary of overall replication stage meta-analyses

The three top-SNPs or closely related proxy (HapMap R^2), selected from discovery meta-analysis, were followed up in infants from six independent (ALSPAC, DNBC, EFSOCH, INMA, GINI + LISA, NFBC 1986) replication samples of European descent (N=8,321). The genetic data was generated either via direct genotyping of these top SNP or computed from *in silico* genome-wide data, as presented in published work[288]. The results at chromosome 12, for rs7980687 (*P*=0.003) and rs1042725 (*P*=8.1 x 10^{-5}), gave consistent association, however SNP rs11655470 only had marginal evidence of association in the replication sample (*P*=0.09).

Combined results

Results from overall discovery and replication were combined, using inverse variance fixed effect meta-analysis. The combined results from discovery and replication samples (N=19,089), showed that SNPs at the rs7980687 (*P*=8.1 x 10^{-9}) and rs1042725 (*P*=2.8x 10^{-10}) associated with infant head circumference (HC) at genome-wide significance. SNP rs11655470 did not reach the genome-wide significance but

Table 16: Association of top-three SNP associating with head circumference (HC) at infancy, before and after adjustment for infant height.12 cohorts out of 13 contributed to these analyses.

Index SNP	Nearest gene	Model 1. Association between SNP and head circumference (SD score)				Model 2. Association between SNP and head circumference (SD score) adjusted for infant height		
		N in the meta-analysis	Effect size (95% CI)	P value		N in the meta-analysis	Effect size (95% CI)	P value
rs7980687	SBNO1	16,664	0.065 (0.042, 0.088)	3.8×10^{-8}		16,664	0.057(0.035, 0.080)	3.8×10^{-7}
rs1042725	*HBGA2*	16,557	-0.063 (-0.082, -0.045)	3.5×10^{-11}		16,557	-0.048(-0.066, -0.030)	1.3×10^{-7}
rs11655470	*CRHR1*	16,512	0.045 (0.026, 0.064)	3.6×10^{-6}		16,512	0.048(0.030, 0.066)	1.4×10^{-7}

95% CI; 95% Confidence interval. Effect estimates reflect the difference in head circumference SD score per minor allele.
The P value is obtained from a linear regression of the SNP against the head circumference SD score (additive model).
The P value is obtained from a linear regression of the SNP + the current height SD score against the head circumference SD score (additive model).

showed suggestive association (P=3.8 x10^{-6}). Each A allele of rs7980687 in *SBNO1* was associated with a 0.07 SD larger head circumference (95% CI: 0.049, 0.099); P=8.1 x 10^{-9}) and each T allele of rs1042725 near *HMGA2* with a 0.065 SD smaller head circumference (95% CI: -0.085, -0.045; P=2.8x 10^{-10}). In terms of raw values, these alleles changed HC by around 1.2 mm and 1.0 mm per allele, respectively. The T allele of rs1165540 near *CRHR1* did not reach genome-wide significance in the combined analysis, but was associated with 0.05 SD larger head circumference (95% CI: 0.028, 0.068, P=3.8 x10^{-6}). A low heterogeneity was observed between the studies for all three SNPs (P>0.1, I2=5-33%).

The associations between the top three SNPs and infant height and weight were also tested (**Table 16**). The signals in *SBNO1* (P= 7.3 x 10^{-3}) and near *HMGA2 (*P=1.5 x 10^{-4}*)*, associated with infant height (but not weight), were collected at the same time as the HC measurements. The variant near *CRHR1 (*P= 0.38*)* did not associate with infant height or weight.

Association of adult height associated polymorphisms with head circumference

The top two SNPs associated with head circumference (HC), rs7980687 in *SBNO,* rs1042725 near *HMGA2,* have also been associated with adult height[289]; therefore the study investigated the association between 180 previously associated adult height variants with HC. The effect size estimates for the association between the SNP at *HMGA2* and infant HC (0.065 SD, P= 2.8x10^{-10}) were similar to that of adult height (0.060 SD, P= 9.0x10^{-6}) however the effect size between the SNP at *SBNO1* was stronger for HC (0.074 SD, P= 8.1x10^{-9}) than for adult height (0.035 SD, P= 5.9x10^{-6}). Besides the SNPs at *SBNO1* and *HMGA2,* 23 other height associated SNPs were nominally associated with HC in infancy. After applying the Bonferroni correction for multiple testing, tested at $P < 2.8 \times 10^{-4}$ for significance, besides *SBNO1* and *HMGA2*, three signals at *ZNFX1* (P=6.1 x 10^{-6}), *OR2J3* (P=1.8 x 10^{-5}) and *ZBTB38* (P=1.8 x 10^{-4}) showed significant association (**Table 17**).

Mediation analysis

To examine the role of mediation by infant height in the relationship between three top SNPs and HC, the direct and indirect (via height) SNP and HC associations were tested using a mediation analysis framework (seemingly unrelated regression analysis (STATA)[290] or a simple pathway analysis (MPLUS)[290]) (**Figure 10**). Twelve studies out of 13 participated in the analyses; as one cohort (CHOP) did not have height measurements.

The adjustment for height reduced the strength of the effect size between HC and two of the three SNPs, suggesting that height may be mediating this relationship (**Table 18**). For example, the effect size for SNP rs7980687 reduced from 0.065 SD (95% CI: 0.042, 0.088) to 0.057 SD (95% CI: 0.035, 0.080), showing that the effect was partially mediated by infant height (12% variance explained). The effect size for SNP rs1042725 reduced from -0.063 SD (95% CI: -0.082, 0.045) to -0.048 (-0.066, -0.030), demonstrating that the effect was partially mediated by infant height (24% variance explained). Moreover, the association at *CRHR1* did not show significant association, indicating that there may be only a direct association between this locus and HC. The models were further adjusted with several potential confounding factors, which have been previously associated with HC or with height. First, the model was adjusted for the first two principal components to account for population stratification, but the associations remained unchanged. Then other confounders, such as breastfeeding, socio-economic status and education level, were also included in the model, but again these did not change the associations. Moreover, there was no evidence of interaction between these variants and sex or breastfeeding after Bonferroni correction at $P < 0.05$.

5.2.1.4. Study Specific Discussion

This is the first GWAS on HC in infancy. The SNP rs7980687 and SNP rs1042725 were associated with infant HC at genome-wide level. These SNPs have previously been associated with adult height[289]. The SNPs were located in/near *SBNO1* and *HMGA2,* respectively. The *SBNO1* gene is a downstream component of the Notch signaling pathway that is known to function in the oncogenic process in humans[292] and is shown to have an important role during embryogenesis and oogenesis in *Drosophila*[293,294]. The Notch signaling is a molecular signaling pathway that plays an important role in the determination of cell differentiation[295]. Studies on *Drosophila* show that knocking this gene led to mal-development of the central nervous system[293]. This suggests that the gene may have an important role in affecting brain growth leading to skull growth and hence affecting HC. The SNP rs1042725 near *HMGA2* was the one of the first variants to be associated with adult height[140]. The *HMGA2* gene codes the high-mobility group AT-hook2 protein, which functions as an architectural factor, it is also an essential component of the enhanceosome and acts as a transcriptional regulating factor. Previous studies on mice, show that a complete knock-out of this gene associates with diet-induced obesity[296]. The clinical implication of this gene is that it associates with a variety of cancers in humans[297,298]. Animal studies shows that a mutant *HMGA2* in mice were unusually small, suggesting a role of

this gene in stature as well[299]. Previous studies also associate head size with total body size[300], it could also be possible that these two genes may function in regulation of skeletal growth and bone development, although the underlying mechanisms are largely unknown.

The SNP near a promotor region of *CRHR1-MAPT* associated with infant HC in discovery stage but did not associate with adult height. This SNP did not reach genome-wide significance in the final combined meta-analysis. The SNP was included in further analysis, despite marginal association with infant HC, as the SNP has shown to be associated with brain development and bone minerals. The CRHR1/MAPT gene encodes a G-protein that binds with corticotropin neuropeptides releasing hormones for regulating the hypothalamic-pituitary-adrenal pathway[301]. This SNP lies within a 900-kb region which corresponds to a conversion of four genes, such as CRHR1, MAPT, STH and GRN. All these SNPs have been associated to neurodegenerative diseases such as Parkinson's, Alzheimer's and fronto-temporal degeneration.

The study identifies two novel genetic variants in or near *SBNO1* and *HMGA2* genes associated with fetal and infant HC measurements and adult height, another variant near *CRHR1-MAPT* showed marginal association with fetal and infant HC. All these genes are shown to be functional in regulations of skull development, skeleton growth and bone density, indicating that these variants could play an important role in adulthood stature. HC is a highly heritable trait and association of HC-associating genes to several neurological diseases, emphasise the importance of further work on these variants. Moreover, the role of mediation by HC between genetic variants and adult height also indicates that HC could be a significant marker for earlier disease traits. However, more detailed studies, and multi-ethnic cohort(s) participation, are needed to evaluate the effect of these SNPs at a wider population level.

Figure 9. Manhattan plot of discovery genome wide association meta-analysis (n=10,768), with three loci selected for replication.

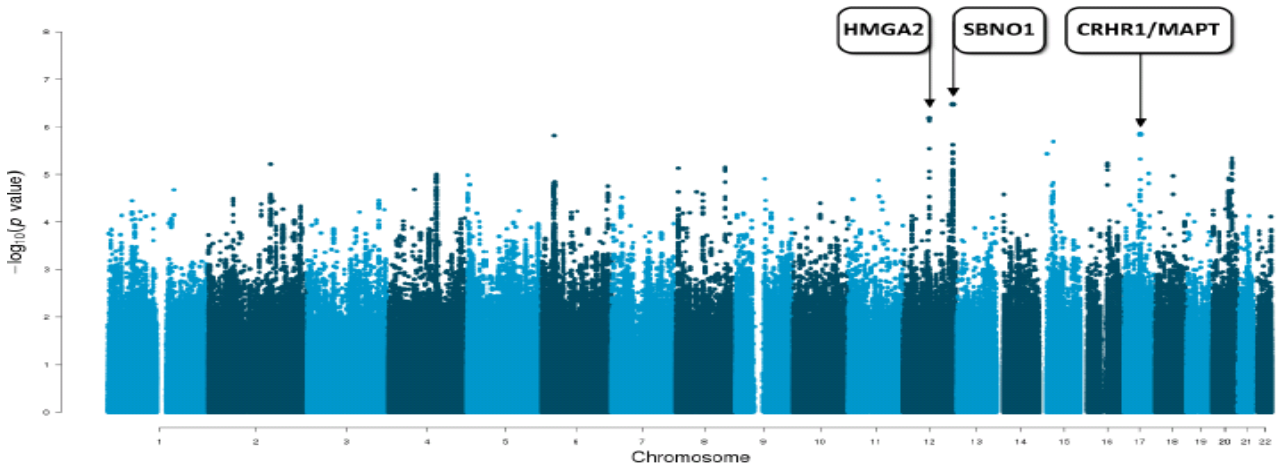


Figure 10. Mediation of effects of genotype on head circumference through an indirect effect on height.

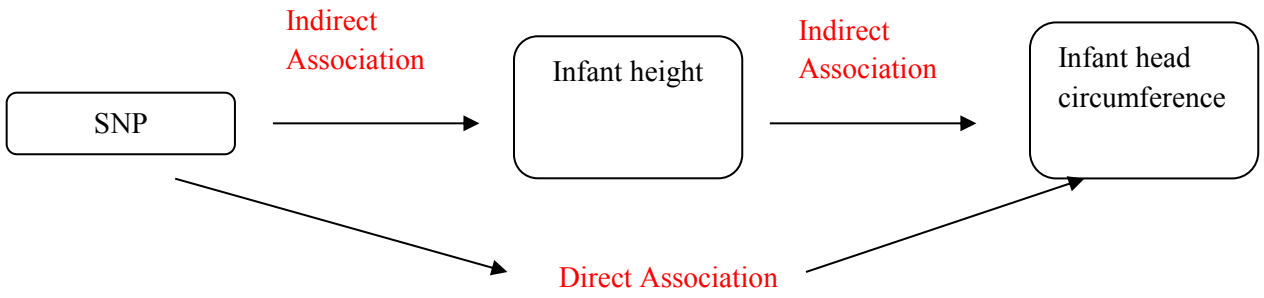


Table 17. Association of 180 known adult height variants with head circumference in infancy, the table shows only significant association after correction for multiple testing at $P < 10^{-4}$.

Chr	Position	SNP	Effect Height (SD)	Effect head circumference (SD)	Standard Error	*P*-Value	Gene	EFFECT_OTHER
3	142588260	rs724016[#]	-0.07	-0.054	0.015	$1.8x10^{-4}$	ZBTB38	A/G
6	29192211	rs3129109[#]	-0.032	-0.112	0.026	$1.8x10^{-5}$	OR2J3	T/C
12	122389499	rs11830103	-0.035	-0.08	0.018	$5.9x10^{-6}$	SBNO1	A/G
12	64638093	rs1351394	0.06	0.064	0.014	$9.0x10^{-6}$	HMGA2	T/C
20	47336426	rs237743	0.041	0.078	0.017	$6.1x10^{-6}$	ZNFX1	A/G

#Results reported for very closely related SNP (rs3117141 for rs 3129109, HapMap R2=1.000; rs6440006 for rs7224016, HapMap R2=0.967)
These results are statistically significant after applying Bonferroni correction. P-value threshold for significance: 0.05/180=2.8x10-4

Table 18. Mediation analysis for infant height influencing relationship between top-three SNPs and head circumference (HC).

Index SNP	N in meta-analysis	Total effect on head circumference (SD score)		Indirect effect (through height) on head circumference (SD score)		Direct effect on head circumference (SD score)	
		effect size (95% CI)	*P* value	effect size (95% CI)	*P* value	effect size (95% CI)	*P* value
rs7980687_A on 12q24	16,664	0.065 (0.042, 0.088)	$3.8x10^{-8}$	0.088 (0.002, 0.014)	0.011	0.057 (0.035, 0.079)	$3.8x10^{-7}$
rs1042725_T on 12q15	16,557	-0.063 (-0.082,-0.045)	$3.5x10^{-11}$	-0.015 (-0.021,-0.009)	$5.6x10^{-7}$	-0.048 (-0.066,-0.030)	$1.3x10^{-7}$
rs11655470_T on 17q21	16,512	0.045 (0.026, 0.064)	$3.6x10^{-6}$	-0.009 (-0.030, 0.012)	0.38	0.048 (0.030, 0.066)	$1.4x10^{-7}$

SD; standard deviation, 95% Confidence interval. Effect estimates reflect the difference in had circumference SD score per allele.
The P value is obtained from a mediation analysis framework seemingly unrelated regression analysis (STATA) or a simple pathway analysis (MPLUS) using an additive model.
All study were of European descent. Of the total of 13 study cohorts, 12 study cohorts contributed to these analyses since data on height and weight were not available in the CHOP (Children's Hospital Of Philadelphia) study.

5.2.2. Association between common genetic variants and early growth phenotypes, linking childhood growth with adulthood obesity

5.2.2.1 Introduction

There are critical stages in an individual's growth after birth that have been hypothesised as risk factors for several CVD diseases in adulthood [302]. These critical stages during the growth trajectory include a period of rapid gain in height and weight in infancy (at around 2 weeks for height and 5 weeks for weight), the period at which body mass index (BMI) reaches its peak in early infancy (at around 9 months) called adiposity peak (AP), and the period when BMI reaches its nadir (between 5 and 7 years) called adiposity rebound (AR) [7,10]. These early growth indicators have been quantified as Peak Weight Velocity (PWV), Peak Height Velocity (PHV), age and BMI at AP and at AR, using well-established growth models, described in Chapter 4, section 4.2. Each of these early growth parameters have been previously associated with the risk of impaired metabolic health in both adolescence [303] and adulthood [8,14,212]. However, little is known about the interplay between the common genetic variants affecting growth patterns during infancy and childhood; and metabolic homeostasis, the state of which influences cardio-metabolic disease in later life.

Most previous studies into childhood BMI or obesity have focused on testing associations with genetic variants discovered in genome-wide association studies (GWAS) on adults [38,129], with one study having used a GWAS approach that investigated extreme obesity (> 95% BMI levels) only. No studies, so far, have considered investigating the longitudinal effect of common genetic variants on early growth phenotypes in a GWAS design. To our knowledge, only one candidate gene study investigated the association of one variant at the FTO gene and early growth phenotypes with one variant[11].

The current study aims to identify common genetic variants associated with early growth phenotypes (PHV, PWV, age and BMI at AP and AR) by meta-analyzing five GWAS and twelve replication studies of European ancestry. The study also aims to investigate the nature of the relationship between genetic variants associating with early growth phenotype and adult BMI, and determining if early growth phenotypes mediate the association between the identified genetic variants and adult BMI. This study was performed as part of collaborative work with the early growth and genetics (EGG) consortium. The consortium members provided GWAS and replication data from various cohorts, details of which are given in Chapter 4, section 4.3.

This EGG project was led by me from Imperial College. I was the main analyst, responsible for developing the analytical protocol, collecting analyses results for each cohort, providing troubleshooting assistance with any issues arising during the analyses process and conducting meta-analysis on all results. My role as part of the Imperial College team was to conduct 1) a discovery stage GWAS on NFBC1966 data, and 2) perform follow-up association analyses of the leading SNPs in the NFBC1986. I conducted the combined meta-analysis for all three stages of the project; discovery, replication and joint. I provided input for analysing the growth parameters for five cohorts (ALSPAC, COPSAC, INSI, LISA and SWS). I performed the genetic susceptibility analysis using the discovery stage meta-analysis results. I also did a mediation analysis in the NFBCs (1966 and 1986) and then meta-analysed data from two other cohorts as part of the project. My work on this project, including contribution on pathway analysis from other team members of Imperial College, has been presented together in a manuscript to be submitted to Nature Genetics. As described earlier, it was agreed with members of EGG-consortium that work related to this project would form part of my thesis.

In this chapter, I explain the discovery and replication stages and finally the combined analysis of all 17 participating cohorts. It also explains the genetic susceptibility analysis performed using discovery stage meta-analysis results. The mediation meta-analysis performed on the NFBCs (1966 and 1986), and two other cohorts are described as well.

5.2.2.2. Methods and materials

Details of the NFBC1966, NFBC1986, and other relevant EGG cohort, and genotype data provided are described in Chapter 4, section 4.1 and 4.3. In summary, singletons with available gestational age and at least three height and weight measurements were selected for the study. Early growth phenotypes were derived from sex-specific individual growth curves using mixed effects models of height, weight and BMI measurements collected longitudinally from birth to 13 years (Chapter 4, section 4.2). Early adiposity phenotypes were standardized to *z* scores within each study and adjusted for sex and gestational age (only applicable for peak height, peak weight and adiposity peak models).

Statistical analysis

As shown in **Figure 11**, six separate analyses were performed in the study. After genome-wide association studies were performed, in each of the discovery cohort data sets across the six early growth phenotypes separately, the results from each were combined in meta-analyses and the SNPs corresponding to the most

significant association genome-wide were selected for replication. The next step involved association analyses at the 'replication SNPs' among the replication cohort data sets, followed by meta-analyses of the replication stage results. The final step was to run combined meta-analyses of both the discovery and replication meta-analyses results and to assess whether any SNP have association results that exceed the genome-wide significance level. Details on how meta-analysis genome-wide association studies (GWAS) are performed are given in Chapter 4, section 4.4.

As well as the discovery of genetic variants associated with these early growth phenotypes, this study aims to provide insights into the relationship between the genetic determinants of early growth, adulthood BMI and other disease phenotypes by conducting the following three investigations: 1) examine the association of the identified SNPs with other disease phenotypes by interrogating the GWAS catalogue[304], 2) assess the associations relating to the early growth phenotypes at the 32 loci found to be associated with adult BMI from the GIANT consortium[139], 3) perform mediation analyses to investigate if early growth phenotypes influence the association between genetic variants and adulthood BMI.

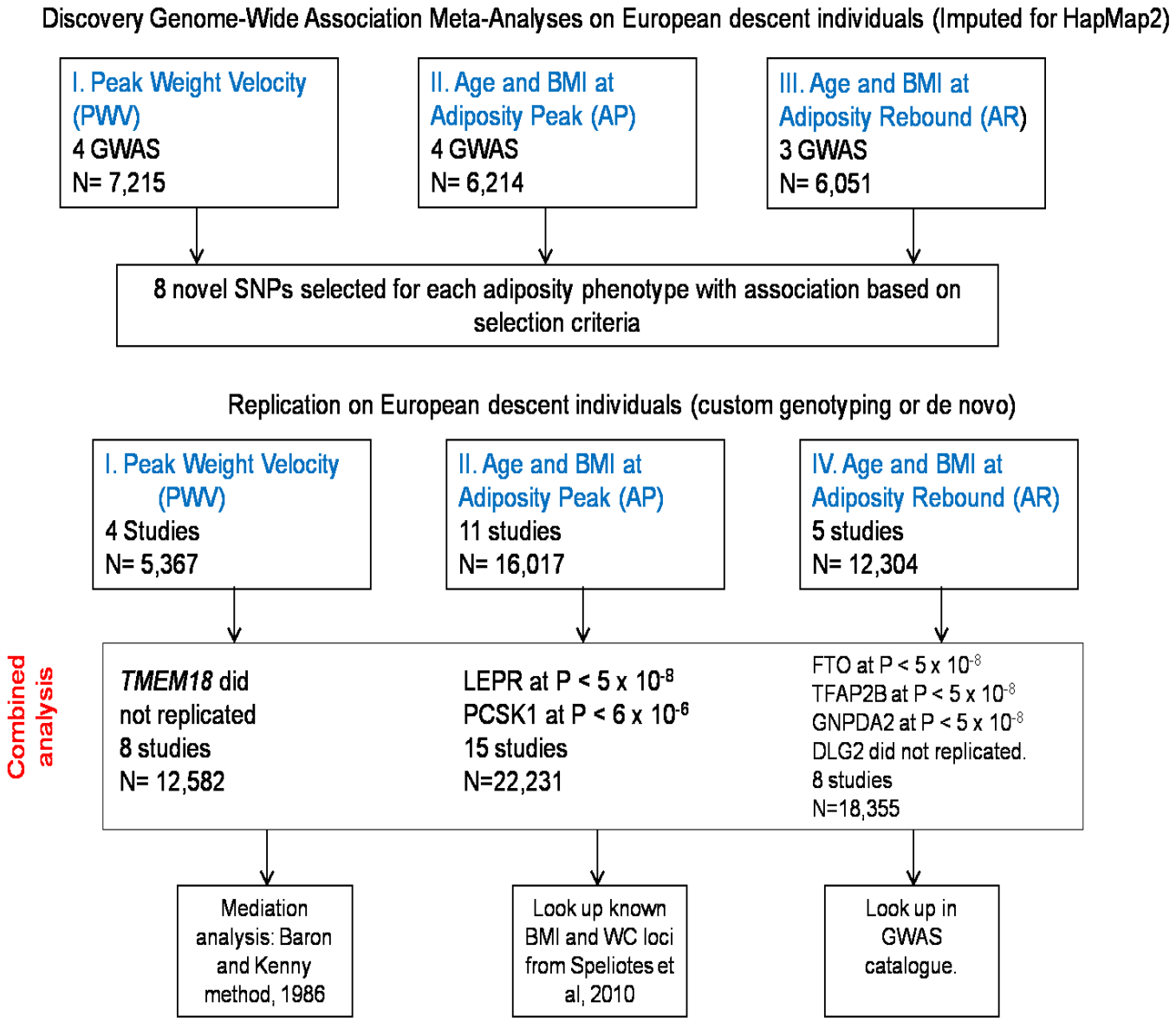
For each locus forest plots were developed, which provided the effect size for each individual study and overall, along with 2-sided 95% confidence interval. A test of heterogeneity assessed whether there was a difference in the effect size among cohorts.

5.2.2.3 Results

Study samples

Table 19 shows the study samples used for these analyses**.** Characteristics of the study samples describing the mean early growth phenotypes, genotyping information and quality control cut-offs applied by each study were similar in all cohorts (data not shown). The study cohort-specific median age range selected for growth modeling from each of the cohorts was between 2 weeks to 13 years. The mean observed early growth phenotypes ranged from 40.9 cm/year to 52.6 cm/year for PHV, 11.6 kg/year to 12.9 kg/year for PWV, 0.5 years to 1.0 years for age at AP in infancy, 16.5 kg/m^2 to 18.1 kg/m^2 for BMI at AP, 4.7 years to 5.7 years for age at AR in childhood and 15.4 kg/m^2 to 15.7 kg/m^2 for BMI at AR.

Figure 11: Flowchart summarising the study design and results for the discovery, replication and combined meta-analysis of GWAS across 17 studies of European Ancestry.



Discovery stage analyses

In the discovery stage, the association statistics of five population-based studies (HBCS, NFBC1966, RAINE, GENR and LISA) were meta-analyzed across 2.5 million directly genotyped and imputed SNPs in singletons from up to 7,215 children of European ancestry (**Table 19**). Seven loci associated with BMI at AP, age at AR and/or BMI at AR in the overall discovery meta-analyses were considered for follow-up using the following selection criteria: four SNPs were selected based on having association results with $P < 5 \times 10^{-7}$, three SNPs were selected based on a combination of association results with $P < 5 \times 10^{-6}$, biological relevance, and quality of genotyping (**Table 20)**. One further SNP for peak weight velocity (PWV) was included for follow-up, despite only a suggestive association result ($P = 5.91 \times 10^{-5}$), due to its strong biological candidacy for BMI in a previous published study[135].

Figure 12 shows the Manhattan and quantile-quantile (QQ) plots for the genome-wide association results, on a –log10 scale, for the discovery stage for each of the early growth phenotypes. The Manhattan plot represented the *P* values of entire GWAS on genomic scales, with any P value below 5×10^{-5} colored in green. The QQ plot was the graphical representation of the number and magnitude of observed association (black) between genotyped SNPs and the early growth phenotypes, compared to assuming no association under the null hypothesis (red). There was not much deviation from the null hypothesis line (red) for PHV, PWV, age and BMI at AP however the QQ plot for age and BMI at AR showed a sharp deviation from the null hypothesis (red line) suggesting that the observed *P* values were clearly more significant than those expected under the null hypothesis.

Two loci at SNP rs9436303 on chromosome 1 located near the leptin receptor gene (LEPR/LEPROT, $P = 4.15 \times 10^{-9}$) and at SNP rs10515235 on chromosome 5 near the proprotein convertase subtilisin type 1 gene (PCSK1, $P = 1.2 \times 10^{-6}$), associated with BMI at AP in infancy (**Table 20**). 31 SNPs, in the well-studied obesity-related *FTO* gene on chromosome 16, showed the strongest associations with age at AR. SNP rs1421085 showed the strongest association of the 31 ($P = 6.14 \times 10^{-8}$) in the discovery stage and was thus taken forward to the replication stage rather than the most commonly studied FTO SNP rs9939609 which was also associated with age at AR ($P = 1.6 \times 10^{-7}$). The rs9939609 was already genotyped in many studies and consequently was used as a proxy as it is in high LD ($r^2 = 0.96$) with rs1421085.

Two further SNPs associated with age at AR: one SNP rs2956578 on chromosome 5, in an inter-genic region between RAN binding protein 3-like (RANBP3L) and solute carrier family 1, member 3 (SLC1A3) ($P = 6.7 \times 10^{-8}$), and one SNP rs2817419 at transcription factor AP-2 beta (TFAP2B) on chromosome 6 ($P = 3.0 \times 10^{-6}$), which has been found to be associated with waist circumference (WC) in adults[137]. For BMI at AR, one SNP

rs2055816 on chromosome 11 in the discs, large homolog 2 (DLG2) gene ($P = 1.4x10^{-7}$) and one SNP rs10938397 on chromosome 4 near glucosamine-6-phosphate deaminase 2 (GNPDA2) gene ($P= 3.1x10^{-6}$) were selected for follow-up. One SNP rs2860323 on chromosome 2, near the trans-membrane protein 18 gene (*TMEM18*, $P=5.9 \times 10^{-5}$), associated with PWV. This SNP rs2860323 was in high linkage disequilibrium (LD) (r2 = 0.85), with the previously published SNP rs6548238, reported as associated with adult BMI [135], and was thus selected for follow-up. No SNPs were associated with PHV or age at AP, even at $P < 1 \times 10^{-4}$, and hence were not included in the replication stage.

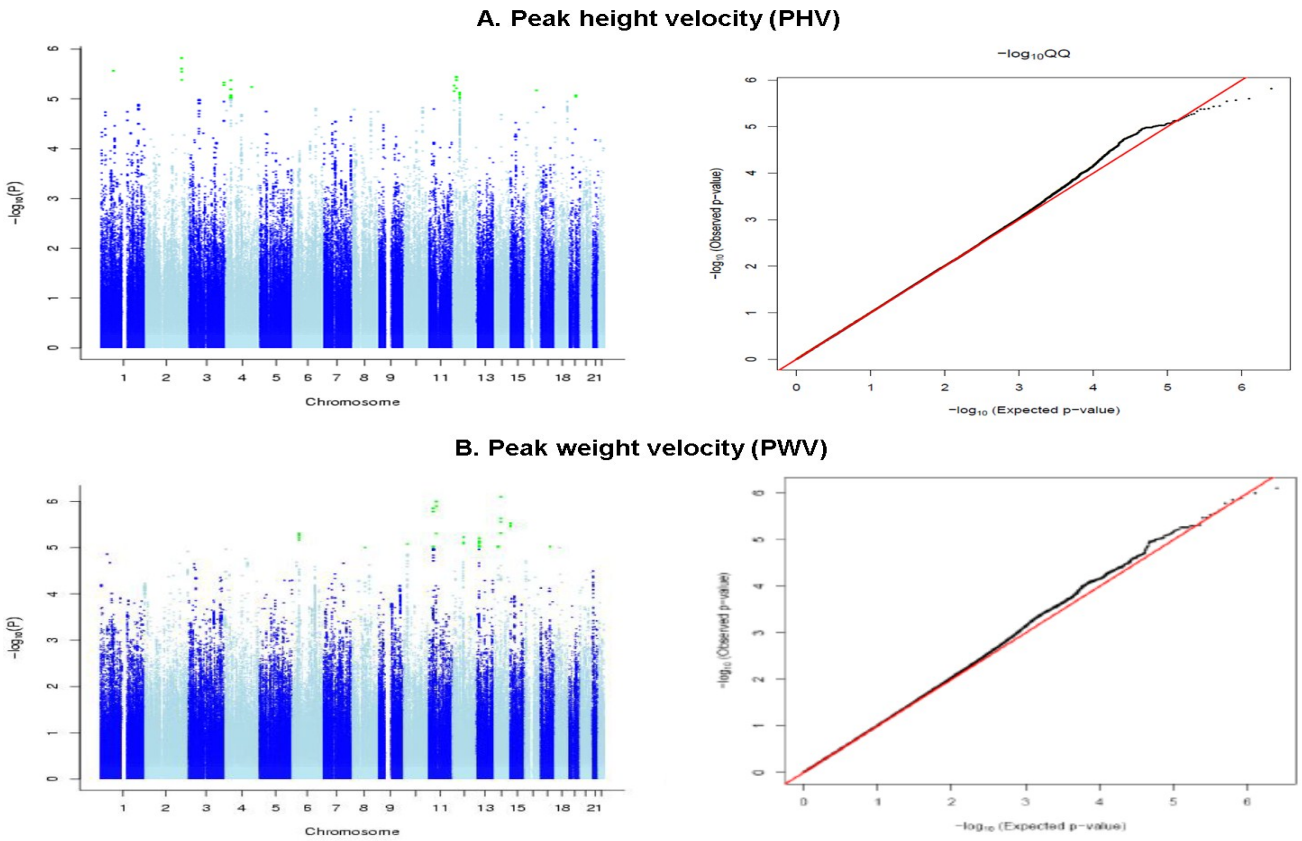
Replication stage analyses

The eight most strongly associated SNPs from each locus, or closely related proxy from HapMap Release 21 (HapMap $r^2 = 0.8$-0.96), were selected for replication (follow-up) in another twelve independent European population-based studies. These twelve European population-based studies used a mixture of *in-silico* data and *de novo* genotyped data in up to n=16,017 children (**Table 19**) for the analyses. Five top-SNPs show evidence of significant association in the replication samples, tested at $P < 0.05$ (**Table 20**). Two loci, one at *LEPR/LEPROT* replicated at $P < 6.66 \times 10^{-4}$ and, one near *PCSK1* replicated at $P < 0.01$ with BMI at AP in the replication sample. Two loci at *FTO* and at *TFAP2B*, remained significantly associated at $P < 7.10 \times 10^{-24}$ and at $P < 1.83 \times 10^{-6}$ with age at AR, in the replication sample. The locus in the inter-genic region between *RANBP3L* and *SLC1A,* did not replicate ($P = 0.83$). The locus at *GNPDA2* associated with higher BMI at AR at (P=3.15 x 10-4), while the locus at *DLG2* did not associate with BMI at AR ($P = 0.176$) in the replication sample. The SNP at locus *TMEM18* found to be associated with PWV in the discovery stage also failed to replicate in the second stage ($P = 0.468$).

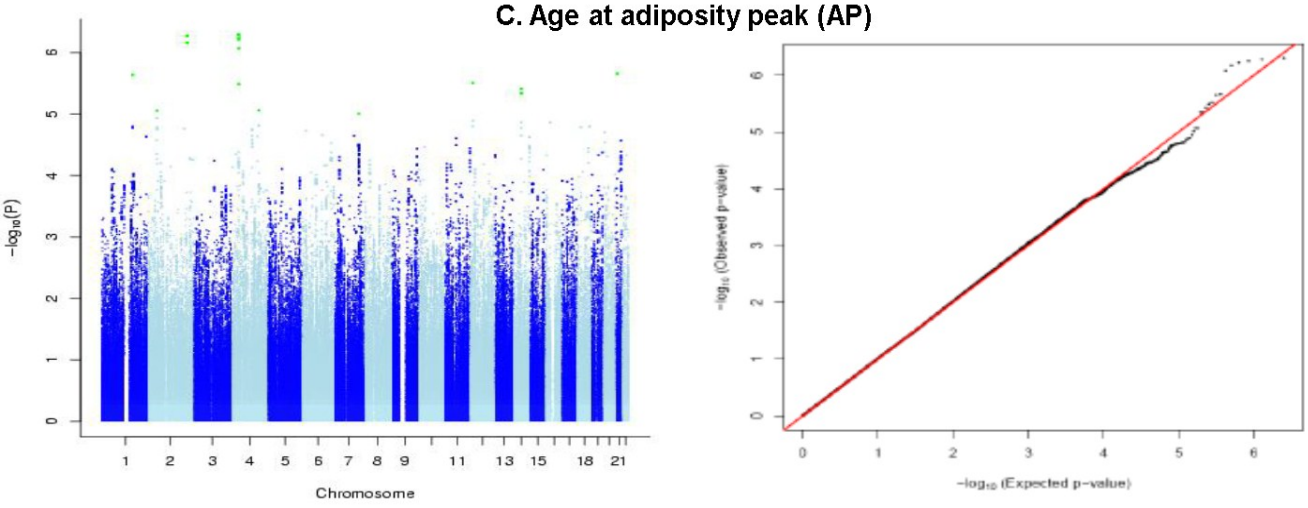
Table 19: Characteristics of study populations and outcome phenotypes in the early growth GWAS.

TYPE	Study	Country	Birth years	PHV/ PWV in Infancy	BMI/ Age at AP	BMI/Age at AR	PHV	PWV	Age at AP (years)	BMI at AP kg/m2	Age at AR (years)	BMI at AR kg/m2
Discovery	HBCS	Finland	1934-44	1449	1169	1555						
	NFBC1966	Finland	1966	3149	2585	3459	52.63 (3.99)	12.92 (2.34)	0.74 (0.05)	18.05 (1.08)	5.71 (0.93)	15.37 (1.09)
	RAINE	Australia	1989-91	-	-	1037	-	-	-	-	4.81 (1.27)	15.49 (1.05)
	LISA (D)	Germany	1998-9	390	372	-	40.9 (4.4)	11.6 (2.0)	0.70 (0.04)	16.5 (1.0)	-	-
	GENR	Netherland	2002-6	2227	2088	-	48.40 (8.03)	11.94 (1.96)	0.71 (0.04)	17.55 (0.75)	-	-
meta-analysis				7215	6214	6051						
Replication	ALSPAC	UK	1991-2	-	6357	6704	-	-	0.77 (0.05)	17.55 (1.78)	-	-
	CBGS	UK		-	557	-	-	-	1.03 (0.04)	17.00 (1.36)	-	-
	CHOP	US	2006-10	-	233	-	-	-	0.84 (0.15)	17.92 (4.12)	-	-
	COPSAC	Denmark	1998-1	364	309	319	-	-	0.61 (0.18)	17.54 (0.89)	4.7 (1.3)	15.44 (0.96)
	DNBC	Denmark	1996-2	-	861	-	-	-	-	-	-	-
	EDEN	France		-	1127	-	-	-	1.02 (0.08)	17.5 (1.5)	-	-
	EFSOCH	UK	2000-4	-	600	-	-	-	-	17.48 (1.47)	-	-
	INMA	Spain	2004-7	733	396	-	-	-	-	16.69 (1.43)	-	-
	LISA (R)	Germany	1998-9	702	619	-	-	11.6 (2.1)	-	16.7 (0.6)	-	-
	NFBC1986	Finland	1985-6	3568	4206	4020	53.25 (7.18)	12.66 (2.88)	0.69 (0.02)	17.53 (0.73)	4.9 (1.04)	15.65 (1.12)
	PANIC	Finland		-	-	433	-	-	-	-	4.7(1.5)	15.7(1.3)
	SWS	UK		-	1285	828	-	-	0.52 (0.04)	17.56 (1.53)	5.30 (1.04)	15.73 (1.24)
meta-analysis				5367	16550	12304						
Total				**12582**	**22764**	**18355**						

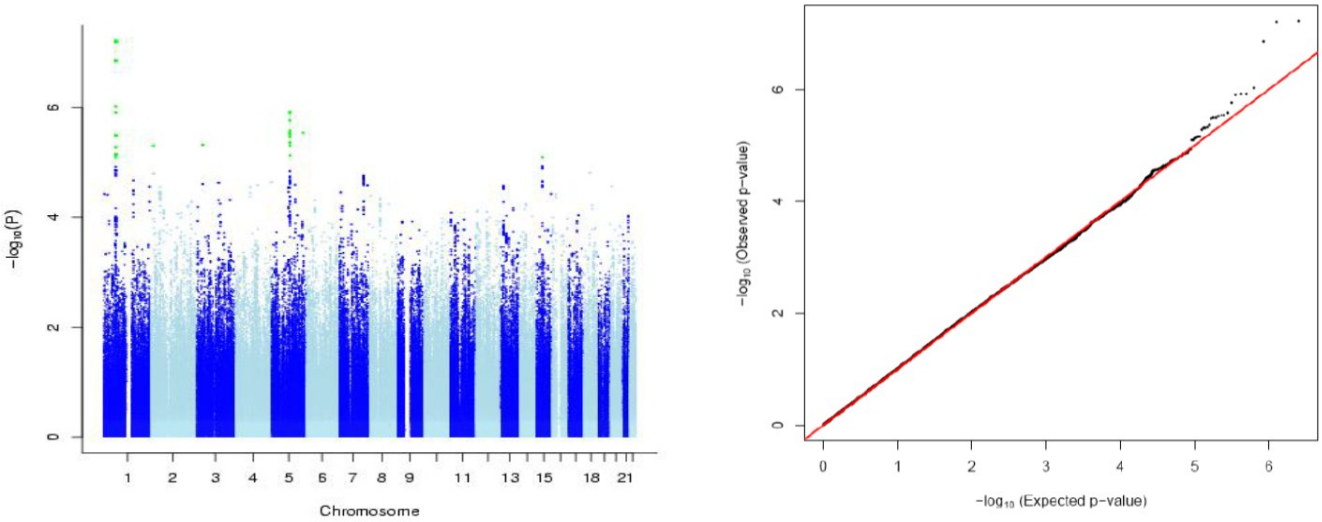
Figure 12: Manhattan and Quantile-Quantile (QQ) plots of discovery stage results (N=7,215). The SNPs were filtered out if minor allele frequency <0.01, and adjusted twice by genomic controls, once within the study and then across the meta-analysis (lambda = 1.2). For each early growth phenotype, SNPs within loci with $P < 5 \times 10^{-5}$ are displayed in green. The QQ plot of observed *vs* expected –log10P values are given in separate panels for each early growth phenotype. The black QQ lines represent the observed *vs* excepted associations for all SNPs. A. Peak height velocity (PHV), B. Peak weight velocity, C. Age at Adiposity peak (AP), D. BMI at adiposity peak (AP). D. Age at adiposity rebound (AR), E. BMI at adiposity rebound (AR).



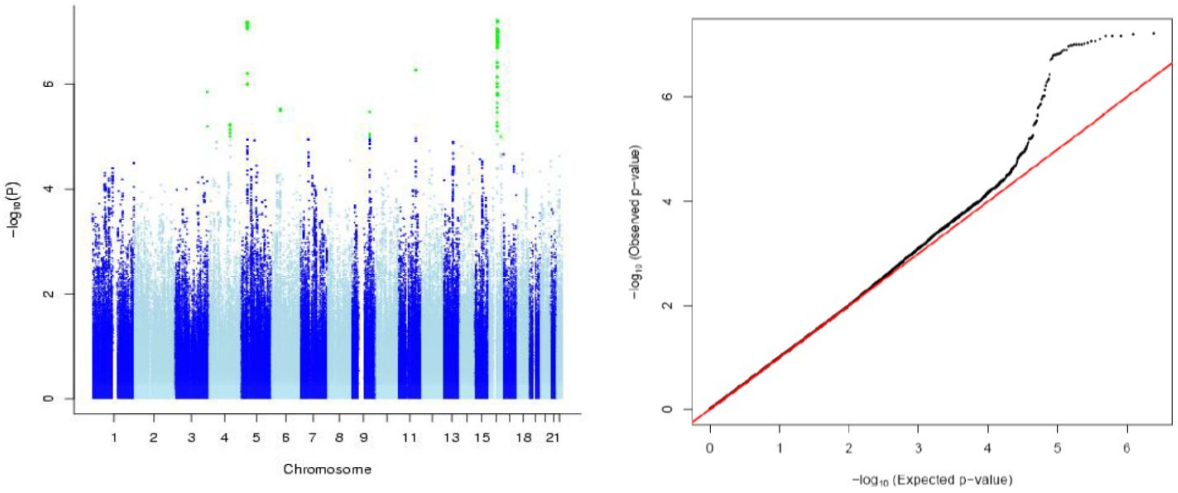
C. Age at adiposity peak (AP)



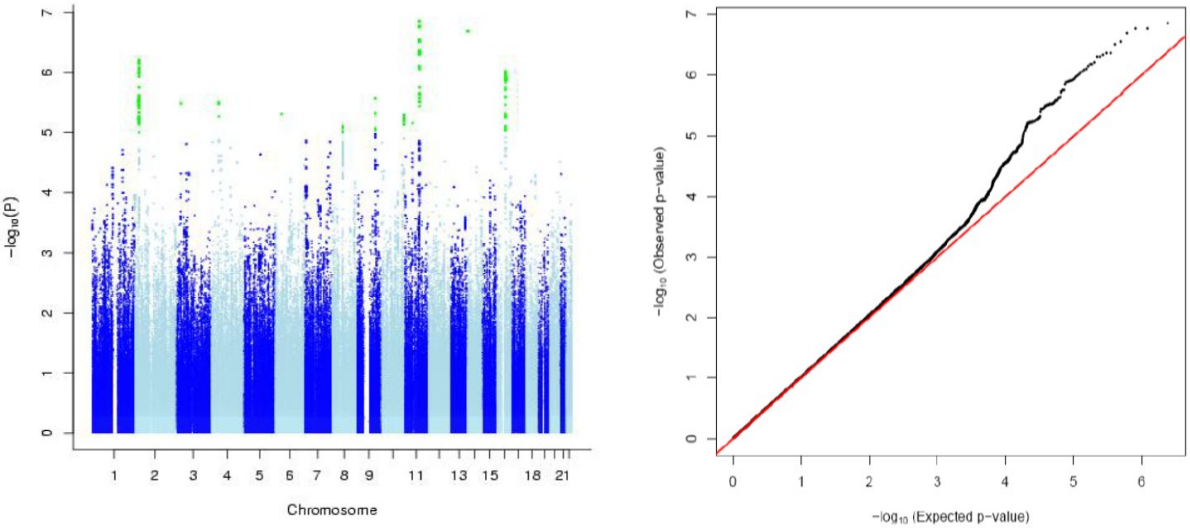
D. BMI at adiposity peak (AP)



E. Age at adiposity rebound (AR)



F. BMI at adiposity rebound (AR)



Combined discovery and replication stage meta-analysis

In the final combined meta-analysis of discovery and replication cohorts (n = 21,859), SNPs at (or near) *LEPR/LEPROT*, *FTO*, *TFAP2B* and *GNPDA2* reached genome-wide significance ($P < 5 \times 10^{-8}$), while one SNP at *PCSK1* showed association at $P < 5 \times 10^{-6}$ (**Table 20**). Each additional copy of **G** allele of SNP rs9436303 within *LEPR/ LEPROT* (minor allele frequency (MAF) = 17-28%) were associated with a *0.07 SD* (95% CI 0.04, 0.09, $P_{combined} = 2.59 \times 10^{-8}$) higher BMI at AP in the final meta-analysis. These effect sizes equate to approximate differences equivalent to an increase in BMI at AP of ~ 0.07 kg m^{-2} (95% CI: 0.04, 0.09) per allele. The **C** allele of SNP rs1421085 at *FTO* was associated with earlier age at AR by *0.12 SD* [~ 1.44 months (95% CI (1.42, 1.46)), $P_{combined} = 3.12 \times 10^{-30}$], which is directionally consistent with a cross-sectional study on BMI variation between infancy and childhood [11]. Similarly, each additional copy of **A** allele of SNP rs2817419 at *TFAP2B* was associated with earlier age of AR by 0.08 *SD* [~ 0.96 months (95% CI (0.94, 0.98)), $P_{combined} = 4.39 \times 10^{-11}$]. An additional **G** allele at SNP rs10938397 in *GNPDA2* was associated with higher BMI at AR by 0.06 SD (95% CI: 0.04, 0.07) [~ 0.06 kg m^{-2}, $P_{combined} = 3.15 \times 10^{-8}$]. The SNP at PCSK1, which associated with BMI at AP at $P = 5.11 \times 10^{-6}$, increased BMI at AP by 0.05 SD [~ 0.05 kg m^{-2} (95% CI (0.03, 0.07)] per additional A allele of SNP rs10515235.

Three loci, near *TMEM18*, in *DLG2* and in inter-genic region between *RANBP3L* and *SLC1A3*, were not significantly statistically associated with early growth phenotypes in the final analysis.

The four SNPs which reached genome-wide significance, and one SNP which showed suggestive association (combined as top-five SNP), were taken forward for further analyses described in the next sections.

There was no evidence of heterogeneity for four locus (*FTO*, *TFAPB2*, *GNPDA2* and *PCSK1*) between discovery and replication stage studies ($P < 0.05$, $I^2 = 0\%$), except at one locus, *LEPR/LEPROT* ($I^2 = 67.3\%$, $P < 0.0001$). Forest plots for each locus and justification on model selection are presented in Appendix Table 6, Appendix Figure 2. The heterogeneity at this locus could be attributed to differences in the distributions of the number of BMI at AP measurements over time between cohorts, or may be due to unmeasured environmental factors influencing growth [305]. However, applying the random effects model to the lead SNP at this locus did not change the significance of the association ($P = 0.0003$), and the overall results for this locus showed directional consistency across studies.

Genetic susceptibility of the identified SNPs to other diseases

To examine the role of shared genetic susceptibility with other disease phenotypes, the associations between the top-five SNPs and disease phenotypes reported in other studies was examined by searching the catalogue of published GWAS (http://www.genome.gov). Results for all genotype-phenotype associations where the reported SNPs were in linkage disequilibrium (LD) ($r2 > 0.5$), with lead SNPs identified here, were presented in **Table 21**. **Table 21** shows that the genetic variants, rs1421085 in *FTO* and rs10938397 near *GNPDA2* associated with age at menarche in adolescent girls, as well as adulthood adiposity related phenotypes, such as weight, BMI, obesity, T2D, metabolic syndrome and osteoarthritis. Previous studies suggest that age at menarche of adolescent girls is influenced by both birth weight and postnatal accelerated growth, while results from this study confirm that genetic variants affecting early growth phenotypes could also be associated with age at menarche[303]. The genetic variants in *LEPR/LEPROT*, *TFAP2B* and near *PCSK1,* in LD ($r^2 >0.5$ cut-off) with the lead SNP, did not associate with other later-in-life adiposity-related phenotypes.

The association between adult BMI SNPs and early growth phenotypes

The associations between early growth phenotypes and the lead SNPs at 32 loci reported for adult BMI by the Genetic Investigation of ANthropometric Traits (GIANT) consortium [139] were investigated using data from our discovery meta-analysis results. The aim of this analysis was to understand if identified adult BMI-associated SNPs also have an association with early growth phenotypes. **Table 23** shows that nine out of the 32 loci associated with one or more of the early growth phenotypes at significance of $P < 1.6 \times 10^{-3}$ (Bonferroni corrected *P* –value: 0.05/32) . For PWV, one locus near *TMEM18* associated at $P < 1.6 \times 10^{-3}$. None of the SNPs associated with age at AP or PHV at $P < 1.6 \times 10^{-3}$. One locus near *RBJ-ADCY3-POMC* associated with BMI at AP at a $P < 1.6 \times 10^{-3}$. Eight SNPs in/near *FTO, TMEM18, MC4R, GNPDA2, SEC16B, TFAP2B, FAIM2 and TNNI3K* associated with age at AR at $P < 1.6 \times 10^{-3}$. The results were consistent with a previous study on *FTO* [11], showing that alleles that increased adult BMI were associated with earlier age at AR. There were associations with BMI at AR at seven loci in/near *FTO, TMEM18, GNPDA2, SEC16B, RBJ, FAIM2* and *TNNI3K,* with the effects directionally consistent with adult BMI. **Figure 13** shows scatter plots of the regression coefficients corresponding to the 32 SNPs for each of the early growth phenotypes against those corresponding to adult BMI from the GIANT consortium study. The figure shows that adult BMI is correlated with age and BMI at AR but not for the other four growth phenotypes earlier in life, consistent with the correlations of the phenotypes reported previously [11].

Figure 13: Scatter plots of the effect size estimates (regression coefficients) corresponding to the 32 SNPs for each of the early growth phenotypes against those corresponding to adult BMI (in GIANT consortium study).

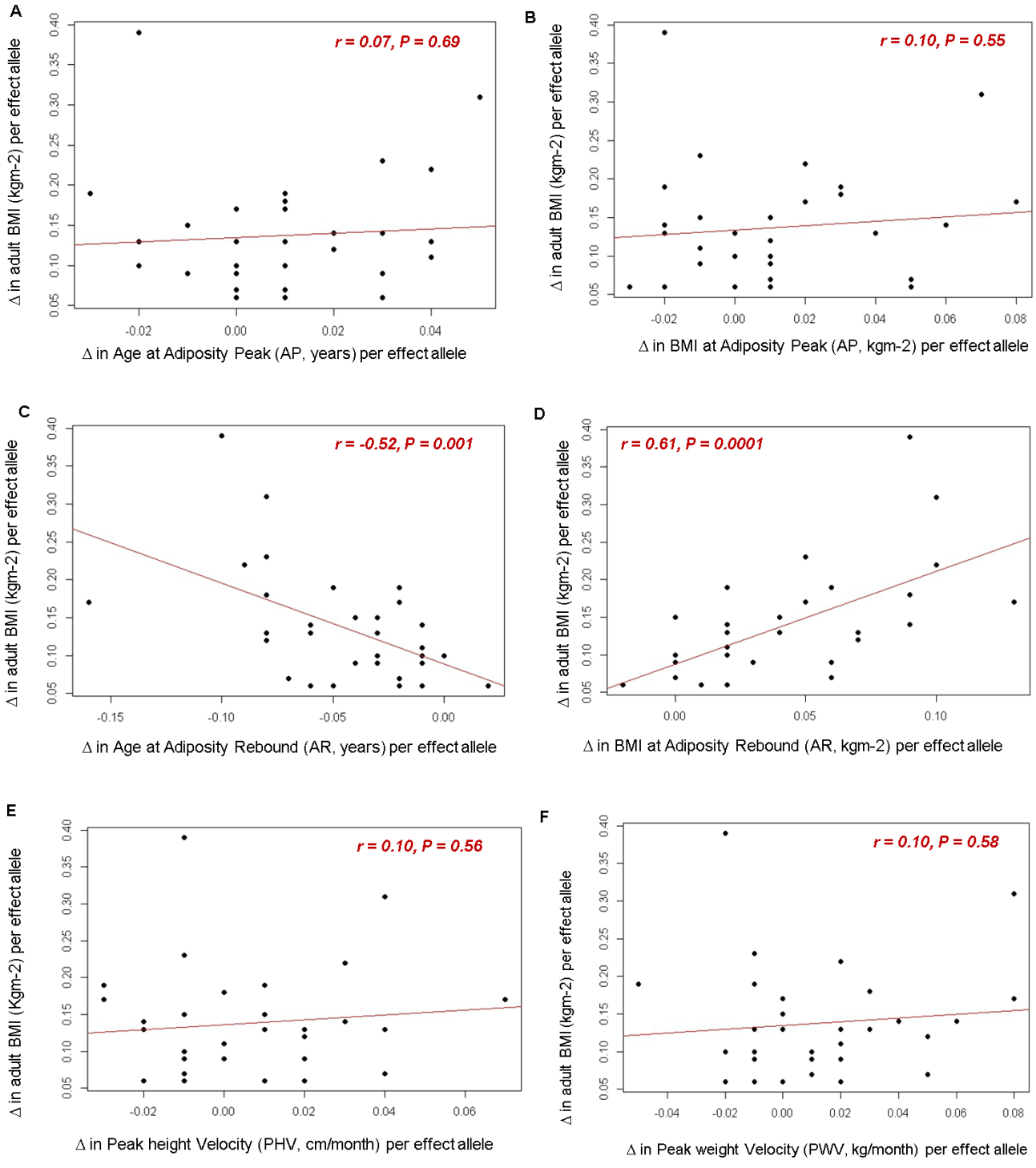


Table 20. Summary results from discovery meta-analysis of top-eight SNPs from five European cohorts, and with replication in 12 European cohorts and the combined meta-analysis results in all 17 studies for early growth GWAS.

						Discovery			Replication			Combined Stage		
SNP	**Location**	**Nearest gene**	**Coded allele/other allele**	**Coded allele freq.**	**Measure**	**B (SE) in SD units**	***P***	**N**	**B (SE) in SD units**	***P***	**N**	**B (SE) in SD units**	***P***	**N**
rs2860323	2p25.3	*TMEM18*	G/A	0.12	PWV	0.09 (0.02)	5.91E-05	7215	0.02 (0.02)	4.68E-01	5367	0.06 (0.02)	3.88E-04	12582
rs9436303	1p31	LEPR/LEPROT	G/A	0.22	BMI at AP	0.14 (0.02)	4.15E-09	5842	0.05 (0.01)	6.66E-04	16550	0.07 (0.01)	5.77E-09	22392
rs10515235	5q15	PCSK1	A/G	0.21	BMI at AP	0.09(0.02)	1.23E-06	6210	0.03 (0.01)	1.50E-02	13652	0.05 (0.01)	2.41E-06	19862
rs1421085	16q12.2	FTO	C/T	0.25	AGE at AR	-0.10 (0.02)	6.14E-08	6051	-0.13 (0.01)	7.10E-24	12122	-0.12 (0.01)	3.12E-30	18173
rs2956578	5p13	*intergenic region**	A/G	0.69	AGE at AR	-0.11 (0.02)	6.73E-08	6051	-0.01(0.01)	8.30E-01	12304	-0.04 (0.01)	1.16E-03	18355
rs2817419	6p12	TFAP2B	A/G	0.76	AGE at AR	-0.10 (0.02)	2.96E-06	6051	-0.07 (0.01)	1.83E-06	12256	-0.08 (0.01)	4.39E-11	18307
rs10938397	4p12	GNPDA2	G/A	0.35	BMI at AR	0.09 (0.02)	5.41E-06	6051	0.05 (0.01)	3.15E-04	12192	0.06 (0.01)	2.92E-08	18243
rs2055816	11q14.1	DLG2	T/C	0.78	BMI at AR	0.13 (0.02)	1.41E-07	6051	0.03 (0.02)	1.76E-01	11815	0.07 (0.02)	5.11E-06	17866

* intergenic region between RANBP3L and SLC1A3

Mediation analysis

To examine the possible mediating role of early growth phenotypes in the relationship between the top five SNPs and adult BMI, we tested the direct and indirect associations using the Baron and Kenny procedure [181]. These analyses were first conducted on the NFBC1966 and NFBC1986 datasets and then results were validated using two other European population-based cohorts (HBCS and RAINE). All the studies had available measures of adolescent and adult BMI at ages16, 18, 20 and 31 years in a sub-sample of up to 9,128 individuals. The final results, meta-analyzed among all four studies, are presented in **Table 22** and a schematic summary is shown in **Figure 14**. Using the methodological strategy described in Shrout et al, 2002 [215], a brief summary of which is given in Section 4.5, three models (direct, indirect) relevant to each locus were analysed, using the early growth phenotypes as potential mediators between the SNP and adult BMI in each model.

From the direct effect model for SNP on adult BMI, two loci at *LEPR/LEPROT* and *PCSK1* did not associate with adult BMI in this sample (P=0.79, P=0.86; respectively). Although, contrary to this result, recent studies show polymorphism at *LEPR* [306-308] and *PCSK1*[309] associating with adulthood obesity-related factors and one of the reasons for non-significant association in this study could be small sample size. In the indirect model, adjusting for BMI at AP in association between SNP and adult BMI, did not change the SNP- adult BMI association (P=0.24, P=0.35; respectively), indicating no mediation by BMI at AP for the two loci, which was confirmed by Sobel test (P_{sobel} = 0.142, P_{sobel}=0.327; respectively).

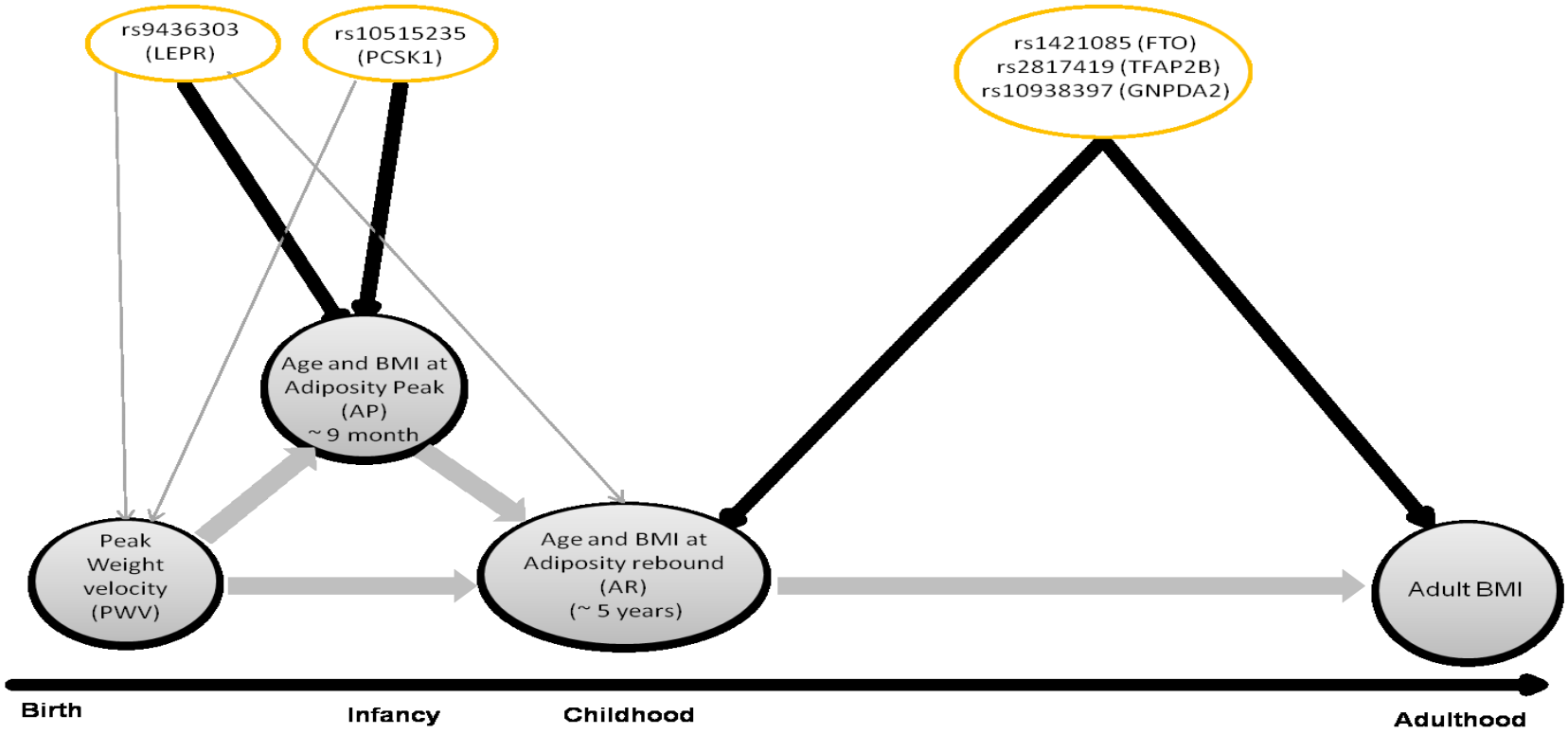
Two loci, rs1421085 and rs2817419, at FTO and TFAP2B associating with age at AR, significantly associated with adult BMI (P=1.03 x 10^{-12}, P=6.52 x 10^{-8}; respectively), confirming results from previous genome-wide studies[135,137,139] (**Figure 14**). In the indirect model, adjusting for age at AR in the model between SNPs and adult BMI, the effect size reduced from 0.10SD to 0.03SD in FTO (P_{soble} =0.01) and from 0.10SD to 0.05 SD in TFAP2B (P_{soble} =1.92 x 10^{-4}), suggesting that age at AR may have mediated this relationship. Locus at GNPDA2, associating with BMI at AR, associated with adult BMI in this sample (P= 1.19 x 10^{-5}) in the direct association. The strength of the association decreased after adjusting for BMI at AR from 0.06SD to 0.03SD (P_{soble}=0.02), suggesting that the BMI at AR may be mediating this association.

Table 21. Results of meta-analysis from mediation analyses of the five identified loci in four European population-based studies (N=9,128) in early growth GWAS.

Effect						**Sobel Test**[a]	**GIANT Consortium**[#]
	B in SD units	**SE**	**95% CI**		***P***	*P*	*P*
Direct: *LEPR* on adult BMI	-0.006	0.022	-0.037	0.048	7.917E-01	0.142	-
Direct: *LEPR* on BMI at AP	0.070	0.010	0.050	0.090	2.592E-08		
Indirect: *LEPR* on adult BMI given BMI at AP*	-0.025	0.022	-0.017	0.068	2.420E-01		
Direct: *PCSK1* on adult BMI	0.003	0.015	-0.027	0.033	8.561E-01	0.327	-
Direct: *PCSK1* on BMI at AP	0.050	0.010	0.030	0.070	1.437E-06		
Indirect: *PCSK1* on adult BMI given BMI at AP	-0.014	0.015	-0.042	0.015	3.525E-01		
Direct: *FTO* on adult BMI	0.100	0.014	-0.127	-0.072	1.034E-12	0.008	4.75E-120
Direct: *FTO* on Age at AR	-0.120	0.010	-0.140	-0.100	3.115E-30		
Indirect: *FTO* on adult BMI given Age at AR	0.027	0.011	-0.048	-0.006	1.297E-02		
Direct: *TFAP2B* on adult BMI	0.097	0.018	0.062	0.132	6.523E-08	0.001	2.90E-20
Direct:*TFAP2B* on Age at AR	-0.080	0.010	-0.100	-0.060	4.388E-11		
Indirect: *TFAP2B* on adult BMI given Age at AR	0.053	0.014	0.025	0.081	1.952E-04		
Direct: *GNPDA2* on adult BMI	0.064	0.015	0.093	0.035	1.190E-05	0.018	3.78E-31
Direct: *GNPDA2* on BMI at AR	0.060	0.010	0.040	0.080	2.920E-08		
Indirect: *GNPDA2* on adult BMI given BMI at AR	0.031	0.012	0.054	0.008	8.107E-03		

*Model for BMI at AP is adjusted for gender and gestational age.
Estimates for association between *LEPR* and *PCSK1*and adult BMI were not available in the GIANT paper.
[a] Sobel (1982) test provides statistical method to assess the significance of the mediator in relation to independent variables and dependent variable.

Figure 14. Schematic diagram showing five loci selected from combined meta-analysis of early growth phenotype meta-analysis. The figure illustrates the relationships between SNPs and early growth phenotypes and that how early growth may mediate the association between *FTO, TFAP2B* and *GNPDA2* loci and adult BMI (Table 22). The arrows represents the association between the SNPs and phenotypes; blue and black arrows show an association with *P* < 0.0001 significance and grey arrow an association of *P*<0.05.



5.2.2.4. Discussion

This is the first genome-wide meta-analysis association study on early growth trajectories, with a sample of up to 21,859 individuals of European descent. The early growth phenotypes are selected as a proxy for the childhood adiposity variable; the advantages of using early growth phenotypes are that they summaries longitudinal growth data into a few functional parameters which presents the biological growth phase. This finding adds to the understanding of the underlying molecular mechanisms linking early growth phenotypes with increased risk of higher adult BMI, which is a risk factor for diseases such as obesity and T2D.

The study finds four SNPs, in and near *LEPR/LEPROT*, *FTO*, *TFAP2B* and *GNPDA2*, associating with early growth phenotypes, such as BMI at adiposity peak (AP), age and BMI at adiposity rebound (AR), at the genome-wide significance level (at $P < 5 \times 10^{-8}$). One SNP near *PCSK1* showed suggestive evidence (at $P=2.41 \times 10^{-6}$) for an association with infant BMI at AP. This SNP did not reach the genome-wide significance but it was included in further analysis, since *PCSK1* has implications in the regulation of fat metabolism and insulin biosynthesis[113,310] and has demonstrated a role in the development of obesity[113,311]. Three of the SNP, in/near *FTO*, *TFAP2B* and *GNPDA2*, have previously been associated with adult BMI[139], and in this study we show that the association between these variants and adult BMI may be mediated by age of adiposity rebound (AR) (**Figure 14**). Figure 14 gives a schematic diagram showing the association of five loci with early adiposity phenotypes.

The SNP at *LEPR/LEPROT* associating with a higher BMI at AP, were located in an overlapping region of two genes, *LEPR* and *LEPROT*. The genes belong to the gp130 family of cytokine receptors that are known to stimulate gene transcription via activation of cytosolic STAT proteins. The function of these two genes, *LEPR* and *LEPROT,* is to encode the cellular receptor for leptin, a hormone involved in regulating appetite and energy balance[122]. Mutations at this gene has previously shown to be associated with C-reactive protein[312], skeleton growth[313], slower pubertal growth[314], monogenic obesity [111], polygenic obesity and related metabolic complications[163,315-317]. Leptin has been primarily expressed in central and peripheral nervous regions, and functions to regulate food intake [318]. The longitudinal effect of SNP in LEPR/LEPROT on obesity is poorly understood[313,319], the result from current studies suggest that genetic variants in LEPR/LEPROT may be involved in body mass development during early infancy. Although, no direct association between SNP in LEPR/LEPROT and adult BMI was found in either GIANT consortium[139] or in this study at nominal

significance, it could not be confirmed if this SNP had a continued effect from infancy to adulthood or if SNP only affects infancy and early childhood. Further, and more expansive, GWAS are required to understand the role of this gene.

The results show that variation in *FTO* and *TFAP2B* is associated with earlier age at AR, and that variants near GNPDA2 are associated with higher BMI at AR. The FTO [125], TFAP2B [320] and GNPDA2 [321] genes are expressed in the hypothalamus, suggesting a potential role for the hypothalamic-pituitary-adrenal axis (HPA) in body weight regulation. All three loci have been previously associated with childhood and adult BMI [38], but none have investigated their role in longitudinal variation in childhood BMI. One previous candidate-gene study on childhood growth patterns from within the same team at Imperial College London, demonstrated FTO's association with earlier AR[11] at genome-wide level and this current study confirms the association. The *TFAP2B* gene has been explored in terms of WC [137] and type-2-diabetes [322], and this study adds its association with earlier age at AR. According to a recent study by Mumby et al,201, a higher BMI at 20 years associates with earlier pubertal timing and adverse metabolic profiles in adults[323] for FTO and GNPDA2 genes, emphasizing the importance of identifying risk indicators for high childhood BMI and adverse growth trajectories. Our search on the GWAS catalogue confirms that FTO and GNPDA2 are potentially involved, not only in the growth trajectory, but also in phenotypes related to metabolic health and disease, reproductive biology in women and bone disease.

The role of these three genes on longitudinal variations of BMI is still unclear, and none examine the age at which these gene starts their development of body mass. A study by Haworth et al, 2008[120], tested a variety of adult BMI associating candidate genes loci with childhood BMI at 4, 7 and 10 years but the results were inconclusive. This is the first study which presents three adult-BMI associating SNPs at FTO, TFAP2B and GNPDA2 also associated with child BMI and the effects of these three SNPs on adult BMI are partly mediated via age and BMI at AR at 5 years. The results suggest that these early growth phenotypes can be important early indicators in predicting later life obesity.

The study also observed suggestive associations at genetic variants near *PCSK1* with BMI at AP ($P=2.41x10^{-6}$). Despite fundamental insights into the role of this gene in monogenic obesity[113,311], there are no previous analyses of its role in growth trajectories representing longitudinal growth patterns . Because PCSK1 encodes an enzyme that has a key role in the biosynthesis of insulin and other hormones involved in regulating energy

metabolism, our finding may be important in linking infant growth with later obesity and so further investigation is warranted.

Nine of the 32 adult BMI-loci reported by the GIANT consortium[139] show evidence of association with one or more early adiposity phenotypes in our study. The other twenty three loci may not show association, either due to a difference in the design and power of the studies, or may only impact BMI later in life[324]. The scatter plots of effect size estimates further showed that while there are strong and significant correlations between the effect sizes of the early growth phenotypes at AR and adult BMI, consistent with the associations with the phenotypes themselves[325], there are no significant correlations between the effect sizes of the other early growth phenotypes (PHV, PWV, age and BMI at AP) and those of adult BMI. These results suggest that genetic variants affecting adult BMI may often begin having an effect on adiposity related phenotypes by AR (~ 5yrs) but not as early as in the first year of life, a finding that agrees with previous reports[11,120,129,326].

One of the limitations of our study is that, despite being one of the largest GWAS in children, sample size was restricted by the availability of longitudinal growth data within the participating cohorts; however the study still had sufficient power to identify four loci at the genome-wide significance level. GWAS with even larger sample sizes will be required to identify other small-effect variants associated with early adiposity phenotypes. Our study might also be useful as a guide to similar studies in other ethnic groups.

In conclusion, we have identified four genetic loci associated with early adiposity phenotypes at a genome-wide significance and one suggestive locus that did not reach genome-wide significance. Three of these loci associate with adulthood BMI in the present study, with effects partly mediated by early growth phenotypes. New insights into the biology of growth in childhood are valuable as they can guide the development and targeting of future therapies and interventions for weight management. Using childhood adiposity phenotypes derived from longitudinal growth models, we provide novel insights into the mechanisms linking early growth patterns with later metabolic health, and establish further evidence that genetic variants influencing adult adiposity have an impact on disease development from early on in life.

Table 22. The five loci associating with early growth phenotypes from the discovery meta-analysis compared with the published GWAS studies (from GWAS Catalogue). The table shows the top five early growth SNPs (index SNPs) with their equivalent variant in LD with the index SNP (R2 > 0.5) from all published GWAS studies mapped to nearest gene. We provide distance of variant from the index SNP, r^2, risk allele, p values from early adiposity meta-analysis and GWAS catalogue, association to traits, direction of effect from early adiposity meta-analysis references. These SNPs give us good indication of the genetic susceptibility of top-five early growth phenotypes SNPs with various disease traits.

Early growth phenotypes	Position	In/ nearest Gene	Index SNP	GWAS Catalogue SNP	Distance from index SNP	R2 with index SNP	Risk Allele	*P* from early adiposity meta-analysis	*P* from GWAS catalogue	Disease or Trait	Direction of effect	References
BMI at AR	16q12.2	*FTO*	rs1421085	rs7202116	20661	0.935	G	NA	2.00E-10	Body mass index	NA	Yang J et al,2012
				rs8044769	38181	0.549	C	5.96E-05	4.00E-06	Osteoarthritis	-	arcOGEN Consortium, 2012
				rs9939609	19573	0.901	A	1.56E-07	1.00E-20	Type 2 diabetes	-	Perry JR et al, 2012
				rs9922619	30817	0.806	T	1.56E-06	6.00E-08	Subcutaneous adipose tissue	-	Fox CS et al, 2012
				rs9940128	200	0.905	A	9.87E-08	2.00E-09	Metabolic syndrome	-	Kristiansson K et al, 2012
				rs12149832	41954	0.868	A	1.13E-06	5.00E-22	Body mass index	-	Okada Y et al, 2012
				rs8050136	15321	0.901	C	1.01E-07	3.00E-26	Adiposity	+	Kilpelainen To et al,2011
				rs17817449	12413	0.901	NA	1.26E-07	2.00E-12	Obesity	NA	Wang K et al, 2011
				rs1558902	2620	1	T	NA	1.00E-07	Obesity	NA	Dorajoo R et al, 2011
				rs9939609	19573	0.901	A	1.56E-07	3.00E-08	Menarche (age at onset)	-	Elks CE et al, 2010
				rs8050136	15321	0.901	A	1.01E-07	4.00E-08	Body mass in chronic obstructive pulmonary disease	+	Wan ES et al, 2010
				rs1558902	2620	1	A	6.36E-08	5.00E-120	Body mass index	-	Speliotes EK et al, 2010
				rs11642841	44533	0.868	A	9.71E-07	3.00E-08	Type 2 diabetes	-	Voight BF et al, 2010
				rs1558902	2620	1	A	6.36E-08	7.00E-13	Obesity (extreme)	-	Scherag A et al,2010
				rs1558902	2620	1	NA	6.36E-08	5.00E-19	Waist	NA	Heard-Costa NL et al,

										circumference		2009
				rs9941349	24534	0.836	T	3.85E-07	6.00E-12	Obesity (extreme)	-	Cotsapas C et al, 2009
				rs9939609	19573	0.901	A	1.56E-07	2.00E-07	BMI	-	Cho YS et al, 2009
				rs8050136	15321	0.901	A	1.01E-07	1.00E-47	Body mass index	+	Thorleifsson G et al,2008
				rs6499640	31277	0.205	A	1.69E-01	4.00E-13	Body mass index	-	Thorleifsson G et al,2008
				rs8050136	15321	0.901	A	1.01E-07	5.00E-36	Weight	+	Thorleifsson G et al,2008
				rs9939609	19573	0.901	A	1.56E-07	4.00E-51	Body mass index	-	Willer CJ et al, 2008
				rs8050136	15321	0.901	NA	1.01E-07	2.00E-17	Type 2 diabetes	NA	Timpson NJ et al,2008
				rs1121980	8293	0.905	NA	1.05E-07	4.00E-08	Body mass index	NA	Loos RJ et al,2008
				rs8050136	15321	0.901	A	1.01E-07	7.00E-06	Type 2 diabetes	+	Zeggini E et al,2008
				rs1121980	8293	0.905	T	1.05E-07	1.00E-07	Obesity (early onset extreme)	-	Hinney A et al, 2007
				rs9930506	29511	0.806	A	1.64E-06	9.00E-07	Obesity-related traits	+	Scuteri A et al,2007
				rs9939609	19573	0.901	A	1.56E-07	2.00E-07	Type 2 diabetes	-	WTCCC, 2007
				rs8050136	15321	0.901	A	1.01E-07	1.00E-12	Type 2 diabetes	+	Scott LJ et al,2007
				rs8050136	15321	0.901	A	1.01E-07	7.00E-14	Type 2 diabetes	+	Zeggini E et al,2007
				rs9939609	19573	0.901	A	1.56E-07	2.00E-20	Body mass index	-	Frayling TM et al,2007
Age at AR	4p12	*GNPDA2*	rs10938397	rs10938397	0	1	G	5.41E-06	4.00E-31	Body mass index	+	Speliotes EK et al,2010
				rs10938397	0	1	G	5.41E-06	3.00E-16	Body mass index	+	Willer CJ et al, 2008

Table 23: Associations between 32 SNPs associating with adult BMI from GIANT consortium (N=249, 796) and early adiposity phenotypes, within the Stage 1 GWA cohorts (N=7,215). Results are from linear regression of z-score of early adiposity phenotypes against SNP (additive model), adjusted for gender and gestational age (GA). The gestational age was only adjusted in PWV, age and BMI at AP. The β are given in SD units. Results for PHV were removed as none of the results were significantly associated and due to lack of space.

			GIANT Consortium			Early adiposity phenotypes GWAS															
						N	PWV			Age at AP			BMI at AP			Age at AR			BMI at AR		
SNP	Nearest gene	Chr	Beta	SE	*P*		β	SE	*P*	β	SE	*P*	β	SE	*P*	β	SE	*P*	β	SE	*P*
rs1558902	FTO	16	0.39	0.02	4.80E-120	7215	-0.02	0.02	1.71E-01	-0.02	0.02	3.67E-01	-0.02	0.02	3.40E-01	-0.1	0.02	6.36E-08	0.09	0.02	5.29E-06
rs2867125	TMEM18	2	0.31	0.03	2.77E-49	7215	0.08	0.02	6.90E-05	0.05	0.02	4.14E-02	0.07	0.02	5.23E-03	-0.08	0.02	9.20E-04	0.1	0.03	1.36E-04
rs571312	MC4R	18	0.23	0.03	6.43E-42	7215	-0.01	0.02	6.49E-01	0.03	0.02	2.39E-01	-0.01	0.02	5.39E-01	-0.08	0.02	3.48E-04	0.05	0.02	2.90E-02
rs10938397	GNPDA2	4	0.18	0.02	3.78E-31	7215	0.03	0.02	1.20E-01	0.01	0.02	6.86E-01	0.03	0.02	8.47E-02	-0.08	0.02	1.70E-05	0.09	0.02	5.41E-06
rs10767664	BDNF	11	0.19	0.03	4.69E-26	7215	-0.01	0.02	4.83E-01	-0.03	0.02	1.67E-01	-0.02	0.02	4.81E-01	-0.05	0.02	2.31E-02	0.02	0.02	3.69E-01
rs543874	SEC16B	1	0.22	0.03	3.56E-23	7215	0.02	0.02	2.24E-01	0.04	0.02	8.33E-02	0.02	0.02	4.69E-01	-0.09	0.02	1.39E-04	0.1	0.02	1.96E-05
rs2815752	NEGR1	1	0.13	0.02	1.61E-22	7215	0.02	0.02	2.06E-01	0	0.02	8.22E-01	0.04	0.03	1.52E-01	-0.06	0.02	3.22E-03	0.04	0.02	5.50E-02
rs713586	RBJ	2	0.14	0.02	6.17E-22	7215	0.04	0.02	2.63E-02	0.02	0.02	1.85E-01	0.06	0.02	8.06E-04	-0.06	0.02	1.58E-03	0.09	0.02	6.23E-07
rs12444979	GPRC5B	16	0.17	0.03	2.91E-21	7215	0	0.02	8.70E-01	0.01	0.03	5.92E-01	0.02	0.03	3.61E-01	-0.02	0.03	4.22E-01	0.05	0.03	8.87E-02
rs7359397	SH2B1	16	0.15	0.02	1.88E-20	7215	0	0.02	8.85E-01	-0.01	0.02	7.31E-01	0.01	0.02	7.55E-01	-0.04	0.02	5.58E-02	0.04	0.02	1.90E-02
rs987237	TFAP2B	6	0.13	0.03	2.90E-20	7215	0.03	0.02	1.66E-01	0.04	0.02	9.91E-02	0.04	0.02	6.99E-02	-0.08	0.02	3.48E-04	0.07	0.02	2.09E-03
rs2241423	MAP2K5	15	0.13	0.02	1.19E-18	7215	-0.01	0.02	4.88E-01	0.01	0.02	7.64E-01	0	0.02	8.37E-01	-0.03	0.02	2.17E-01	0.02	0.02	4.86E-01
rs9816226	ETV5	3	0.14	0.03	1.69E-18	7215	0.06	0.02	7.28E-03	0.03	0.02	2.70E-01	-0.02	0.02	4.75E-01	-0.01	0.05	7.66E-01	0.02	0.03	4.08E-01
rs7138803	FAIM2	12	0.12	0.02	1.82E-17	7215	0.05	0.04	1.62E-01	0.02	0.02	2.54E-01	0.01	0.02	7.08E-01	-0.08	0.02	1.88E-05	0.07	0.02	4.24E-04
rs2287019	QPCTL	19	0.15	0.03	1.88E-16	7215	0	0.02	9.47E-01	-0.01	0.02	5.57E-01	-0.01	0.02	5.06E-01	-0.03	0.02	2.25E-01	0	0.02	9.16E-01
rs1514175	TNNI3K	1	0.07	0.02	8.16E-14	7215	0.05	0.02	2.40E-03	0.01	0.02	6.76E-01	0.05	0.02	4.54E-03	-0.07	0.02	5.02E-04	0.06	0.02	1.17E-03
rs13107325	SLC39A8	4	0.19	0.04	1.50E-13	7215	-0.05	0.05	3.23E-01	0.01	0.06	8.25E-01	0.03	0.05	5.95E-01	-0.02	0.06	7.54E-01	0.06	0.06	3.36E-01
rs2112347	FLJ35779	5	0.1	0.02	2.17E-13	7215	-0.01	0.02	7.42E-01	0.01	0.02	4.59E-01	0	0.02	8.41E-01	0	0.02	8.11E-01	0	0.02	9.96E-01
rs10968576	LRRN6C	9	0.11	0.02	2.65E-13	7215	0.02	0.02	3.39E-01	0.04	0.02	5.90E-02	-0.01	0.02	7.00E-01	-0.01	0.02	6.00E-01	0.02	0.02	2.08E-01
rs3817334	MTCH2	11	0.06	0.02	1.59E-12	7215	-0.01	0.02	6.32E-01	0	0.02	9.64E-01	-0.02	0.02	1.68E-01	-0.02	0.02	3.44E-01	0.01	0.02	6.61E-01
rs3810291	TMEM160	19	0.09	0.02	1.64E-12	7215	-0.01	0.02	5.68E-01	0.03	0.04	6.75E-01	-0.01	0.02	5.06E-01	-0.01	0.02	5.89E-01	0.03	0.02	1.67E-01
rs887912	FANCL	2	0.1	0.02	1.79E-12	7215	0.01	0.02	5.78E-01	-0.02	0.02	3.00E-01	0.01	0.02	6.69E-01	-0.01	0.02	6.46E-01	0	0.02	9.27E-01
rs10150332	NRXN3	14	0.13	0.03	2.75E-11	7215	0	0.02	9.82E-01	-0.02	0.02	2.98E-01	-0.02	0.02	4.77E-01	-0.03	0.02	1.96E-01	0.02	0.02	2.91E-01

rs13078807	CADM2	3	0.1	0.02	3.94E-11	7215	-0.02	0.02	4.52E-01	0	0.02	8.41E-01	0	0.02	9.85E-01	-0.03	0.05	7.75E-02	0.02	0.02	5.45E-01
rs11847697	PRKD1	14	0.17	0.05	5.76E-11	7215	0.08	0.05	9.81E-02	0	0.06	9.94E-01	0.08	0.06	1.60E-01	-0.16	0.07	2.22E-02	0.13	0.07	6.88E-02
rs2890652	LRP1B	2	0.09	0.03	1.35E-10	7215	0.01	0.02	6.42E-01	-0.01	0.02	6.29E-01	0.01	0.02	6.18E-01	-0.04	0.02	8.09E-02	0.06	0.02	9.58E-03
rs1555543	PTBP2	1	0.06	0.02	3.68E-10	7215	0	0.02	7.64E-01	0	0.02	9.79E-01	0	0.02	9.37E-01	-0.06	0.02	2.86E-03	0.02	0.02	2.62E-01
rs4771122	MTIF3	13	0.09	0.03	9.48E-10	7215	0.02	0.02	2.60E-01	0	0.02	8.26E-01	0.01	0.02	7.13E-01	-0.03	0.02	2.22E-01	0	0.02	9.22E-01
rs4836133	ZNF608	5	0.07	0.02	1.97E-09	7215	0.01	0.02	5.48E-01	0	0.02	9.75E-01	0.01	0.02	4.76E-01	-0.02	0.02	3.72E-01	0	0.02	8.43E-01
rs4929949	RPL27A	11	0.06	0.02	2.80E-09	7215	-0.02	0.02	1.35E-01	0	0.02	8.47E-01	-0.03	0.02	1.43E-01	-0.05	0.02	1.22E-02	0.01	0.02	7.59E-01
rs29941	KCTD15	19	0.06	0.02	3.01E-09	7215	0.02	0.02	1.79E-01	0.01	0.02	6.12E-01	0.05	0.02	6.15E-03	-0.01	0.02	6.63E-01	0.02	0.02	3.62E-01
rs206936	NUDT3	6	0.06	0.02	3.02E-08	7215	-0.02	0.02	4.08E-01	0.03	0.02	1.77E-01	0.01	0.02	6.98E-01	0.02	0.02	3.59E-01	-0.02	0.02	3.68E-01

Chapter 6: Conclusions

The first part of this chapter provides a summary of the methodologies used, conclusions reached, and discussion of the strengths and weaknesses of each study in the thesis. The second part of the chapter outlines work currently being undertaken that follows on directly from sections within the thesis and also work planned for the future. The thesis provides a background to the research topic by presenting published studies that have investigated several prenatal, postnatal, genetic and environment factors to understand the underlying causes of childhood and adulthood obesity. The literature review shows that there is still a gap in knowledge that needs to be filled. Each study included in this thesis provides evidence towards the 'programming effect' in the development of obesity. Section 6.1 discusses the general conclusion for the thesis, Section 6.2 discusses the methodology used, Section 6.3 and 6.4 examine the strengths and weaknesses of the study and finally section 6.5 discusses the further work from the thesis.

6.1 General conclusions

This is the first study that attempts to understand how prenatal predictors influence early growth, and how both the prenatal and postnatal environment can be responsible for the development of adverse metabolic health in later life. The study also confirms that leukocyte telomere length (LTL) is a biomarker of diseases associated with biological aging and that telomeres shorten in women who change substantially in body mass from childhood to adulthood. Furthermore, the study identifies several genetic variants associated with early growth measures, such as infant head circumference and early growth phenotypes: peak height and weight velocity, age and BMI at adiposity peak (AP) and age and BMI at adiposity rebound (AR). This study has shown that the effect of these genetic variants on adulthood life factors may be mediated via these early growth measures, indicating that these early growth measures could be important markers to explore in future in terms of individual health. This is the first kind of study that utilises dense longitudinal data to analyse extensive prenatal predictors with postnatal growth and adulthood health. It is also the first study to utilise longitudinal growth modeling in a GWAS context.

This study on prenatal predictors of infant height growth supports the hypothesis that smaller birth size results in faster postnatal growth. Moreover, it showed that several maternal and paternal factors, such as maternal age, weight, parity and smoking are directly associated with faster postnatal height growth. However, their association was found to be mediated by size-at-birth variables. The mediation by size-at-birth variables is indicative of an adverse intra-uterine environment during fetal development. The study provides evidence that

"damage" done during fetal development is compensated by increased early childhood growth, but can have an adverse effect later in life. These findings are important to public health knowledge and can be evaluated to form health policy aimed at improving the prenatal environment and also at monitoring an infant's growth during the first few years. However, further work will be required to disentangle more complex relationships between these infant height growth parameters and to test if they are markers of CVD in adults.

The findings from the study on the effect of an obesogenic environment *in utero,* and during childhood on adolescent metabolic health, highlights that there is still a gap in understanding the metabolic impact of maternal pre-pregnancy weight and postnatal obesogenic environment on offspring's health. This study attempts to understand how various metabolic phenotypes maybe 'programmed' *in utero* or whether they are affected by an obesogenic environment during postnatal development. Recent rodent studies show the 'programming' effect of the metabolic phenotypes in offspring and this study attempts to replicate the original findings from an experimental design using human observational data. The study indicates that the fetus of overweight and obese mothers will be exposed to higher concentrations of insulin and other adverse factors compared to the fetus of normal weight mothers, which could alter the development of key organs. There is a growing body of evidence indicating a 'programming' effect by maternal weight on the future health of the offspring however it is extremely challenging to segregate the complex interaction between the maternal factors and the postnatal obesogenic lifestyle factors, to uncover the true causal mechanisms for this programming effect. Despite the fact that such research is at an early stage, the indications of such a programming effect by maternal weight are substantial.

The study on the association of LTL at 31 years with adiposity-related anthropometric measurements and change in BMI distribution from childhood to adulthood suggests future investigation on the potential for using LTL as a biomarker for age related diseases such as CVD and obesity. Studies based on LTL measures can be effectively used to develop interventions to combat obesity, such as monitoring the maintenance of healthy weight in women and increased moderate physical activity[327]. The study shows that in women, a large increase in body mass after childhood (5years) to adulthood (31 years) may be detrimental and potentially result in earlier biological aging, which can be associated with CVD and obesity. Further longitudinal studies are required to replicate these long-term effects of weight gain after childhood on LTL, healthy aging and cardio-metabolic disease risk and to investigate the gender differences observed here.

The GWAS to find genetic variants affecting infant HC identified two genetic variants in/near genes *SBNO1*, *HMGA2* associates with infant HC that also associated with adult height. The study concluded that the associations between the genetic variants and adult height were mediated by HC. It is known that infant HC correlates with body size, and its role in regulation of body size is unknown. It has been proposed that the genes identified associating with HC also play a general role in skeletal growth and bone development. Moreover, one of the variants near *CRHR1* did show suggestive evidence of association with HC, and this gene has been found to be directly associated with brain development and specifically with skull growth. This was confirmed by animal studies, showing that absence of *CRHR1* gene can lead to mal-development of the central nervous system. Therefore, this study highlights that early brain growth may be associated with several neurological disease in later life and early markers such as infant HC needs to be further evaluated in larger samples and in samples of different ethnicity in order to confirm it as a marker for neurological diseases

The GWAS on early growth phenotypes, linking childhood growth with adulthood obesity, identified four genetic variants at genome-wide level and one at suggestive level. Three of these variants also associated with adult BMI, with the effects of these three genetic variants being partly mediated via early growth phenotypes, such as age and BMI at adiposity rebound. These findings suggest that early growth phenotypes can be used as a robust marker to understand the pathogenesis of obesity in adolescence and adulthood. Additionally, these indicators can be used in the clinical setting to monitor and implement interventions early in life. These new insights into the biology of growth in childhood and how they affect adulthood factors will be valuable as they can guide the development and targeting of future therapies and interventions for weight management.

6.2 Methodology considerations

The NFBCs have been an ideal source of population-based study for growth modeling process. They provide some of the densest data available with frequent height and weight measurements from birth to adulthood. This data is excellent for modeling and developing growth parameters for analysis. The selected population was also representative of the wider Northern Finland population as shown in the analysis (Chapter 5, Section 2) when selecting sample for peak height velocity analysis. The selection of a parametric model was done mainly on NFBC1966 data. Once selected, I applied the same model in NFBC1986. This was a collaborative effort from the team at Imperial. . The mean of early growth phenotypes was consistent with other participating cohorts, as seen in the GWAS of the early growth phenotype study, showing robust methodology. I also applied SuperImposition by Translational And Rotation (SITAR)[328] model to NFBC data but it did not fit well with the data. It was beyond the scope of this thesis to present the comparison between SITAR and growth models

selected in NFBCs. This model needs to be examined in detail in future work as well as other options for the analyses of longitudinal data, for example in terms of mediation[329].

6.3 Strengths of the study

The thesis provides one of the biggest studies conducted on fitting of growth curves on dense longitudinal data, as the NFBCs has the world's richest data in terms of data from early pregnancy until middle age with extensive genetic and other biological information. The NFBCs provide the ideal dataset for growth modeling at different age periods, such as infancy, childhood and adolescence. It also provided the most homogenous cohort in terms of ethnicity and a subsample available for the current study showed a high representativeness of the whole population. The GWAS analyses conducted were statistically powerful, and successful strategies were applied to identify the mediation of size at birth variables and early growth phenotypes.

The growth phenotypes modelled in this thesis, especially adiposity peak (AP) and adiposity rebound (AR), are considered occurring during the important developmental stage in a life course, and any physiological alternation during these times can increase the risk of later obesity. The age of AR has also been strongly associated with adiposity in adolescence and adverse metabolic health in adulthood. While there is still debate over the importance of these infant and childhood growth phenotypes; and the role of mediation by growth phenotypes between prenatal predictors and later obesity, the results from this thesis confirms a strong association. It indicates their importance in being early markers of disease risk. It is still early days to reach conclusions and further experimental study is required to assess these early risk markers in more detail.

6.4 Limitations of the study

The study has many strengths but it also has some limitations. One of the limitations could be due to a bias caused by missing values To address the issue of missing values, the current study applies complete case analysis for the growth modeling which restricts analysis to the subsample of subjects with complete cases, excluding all individuals who have missing values for any of the variables being considered, whether outcome or explanatory. However, attrition analyses suggested that samples included in the analyses here were an overall representation for the rest of the population.

Another limitation of this study is that maternal prenatal or postnatal nutritional information was not available within either NFBCs. Previous studies emphasize the role of maternal nutrition on birth outcomes [330]. It has been proposed that nutrition in fetal life acts as the central stimuli for programming of susceptibility to adult diseases[331]. For example, the Dutch hunger winter study shows that women exposed to limited nutrition during

pregnancy had smaller size babies and increased risk of obesity and T2D in adult life[332]. Several animal studies also provide evidence on the biological plausibility of regulation of mammalian fetal growth due to maternal nutrition[333]. Although reviews show that results of maternal nutrition on size-at-birth studies are inconclusive, as many of these studies linking maternal nutrition to body proportions lack a sound experimental basis. As a solution, the analyses in this thesis were adjusted for an extensive set of confounders such as maternal SES and education known to be correlated with maternal nutrition.

The thesis provides two robust GWAS identifying eight novel genetic variants associating with infant head circumference and early growth phenotypes. GWA studies have successfully identified multiple loci associated with common variants which potentially have clinical implications; however it has its own short comings. GWAS only detects the small structural variations at single nucleotide associating with disease phenotype, and the cumulative effects of multiple SNPs only explains a small fraction of an individual's risk for that trait. Additionally, current GWAS does not account for any environmental variations in the association, and therefore can confound the association. Despite this shortcoming, GWAS has been most successful in identifying robust associations with structural variants. Newer methodologies and approaches have been proposed to counter these shortcomings, for example, gene-environment interaction studies, as briefly discussed in future work below.

6.5 Future work

The research as part of this thesis has given me a good basis to identify further projects that I am now interested in working on. During the course of my PhD and then for a year on MRC Centenary Award, I have managed to acquire expertise in the area of GWAS which I have utilized to expand my research interest into several other areas. In this section, I have outlined some of the research projects I am currently involved in as well as those that I would like to undertake in the future.

6.5.1 Gene-environment (G x E) interaction studies on pre-pregnancy maternal smoking and offspring birth weight (BW).

The gene-environment (G x E) interaction studies are seen as the evolving methods of genetic epidemiology. Recently there has been a greater focus on these studies as they account for both genetic and environmental determinants to identify variants for common diseases[334], reducing confusion caused by environment or taking into account its contribution. In this G x E study, we examine the effect of pre-pregnancy maternal smoking on offspring birth weight (BW). The effect of smoke during pregnancy has been identified as one of the strongest risk factors for a number of adverse birth outcomes, including intra-uterine growth restriction (IUGR)[335]. Babies

whose mothers smoked during pregnancy are on average 150 to 200g lighter at birth than babies with non-smoking mothers[65,175]. Previous genetic studies have identified variations at *CHRNA5-CHRNA3-CHRNB4* locus on chromosome 15q25, robustly associated with the quantity of smoking in those who smoke [336-339]. Studies on women who smoked during pregnancy from the UK [340] and Netherlands [176] independently investigated association between the 15q25 variant and BW, showing reduction in BW with each additional copy of risk allele in women who smoked during pregnancy. However, both of these studies have been of limited power and potential association with fetal genotype has not been examined.

This current study hypothesises that maternal smoking is strongly associated with BW and there is evidence that genetic variation may decrease BW due to pre-pregnancy maternal smoking. The study selected ten GWAS cohorts (1958 T, 1958 W, ALSPAC, GENR, INMA, LISA, NFBC1966, RAINE, TEENAGE and PANIC) with maternal smoking information available as discovery cohorts (N=17,569). Genotypes were obtained using high-density SNP arrays and were then imputed for ~ 2.5 million HapMap SNPs [16]. The study sample excluded individuals who were born prematurely, those without any gestational age (GA) information, multiple deliveries and still born. Moreover, individuals with low BW (less than <500g) were also excluded, as these children may have other factors influencing the association. Individuals with Non-Caucasian ancestry or mixed ancestry were also removed from the analysis.

The outcome variable, BW (in grams), was quintile normalized adjusting for sex and gestational age (GA), major confounders, and then rescaled to the sample mean and standard deviation. Pre-pregnancy maternal smoking was divided into three categories: non-exposed group (those who never smoked and non-smokers) and exposed group (smokers). However, two different definitions of the non-exposed category were used in the analyses, to maximise the sample size. The never smokers were defined as mother's who did not smoke in their lifetime. Non-smokers were defined as mothers who did not smoke at the start of the pregnancy and during pregnancy. Smokers were defined as mothers who smoked one or more cigarettes per day at any time during pregnancy. Mothers who smoked tobacco in a pipe and/or cigars were removed from the analysis. The analyses were also adjusted for principal components; to remove inflation of the test statistics due to population stratification.

The linear additive models were run on QUICKTEST software[341]. Two sets of models were tested; main effects model stratified by maternal smoking groups and interaction model stratified by maternal smoking groups. In the main effects model, three different sets of models for never-smokers, non-smokers and smokers were fit for

transformed BW SD scores against genetic variants adjusting for gestational age, principal components and sex. The interaction model was fitted for transformed BW SD scores against and interaction term between genetic variant and smoking categories, with adjustments for GA, principal component and sex.

In the discovery stage, before meta-analysis, SNPs with MAF of < 1% and poorly imputed SNPs (proper_info of <=0.4 (SNTEST) or R2 of <=0.3 (MACH2QTL)) were filtered out. Fixed-effect meta-analyses were independently conducted on each study and genomic-control was applied twice, once for each discovery cohort and another for final discovery meta-analysis to adjust the statistics generated within each cohort. Meta-analysis was performed using the METAL software by me (lead analyst), as part of team at Imperial College London and cross-validated by another team at Kings College London, using R software. Meta-analysis was performed using the inverse-variance method: a fixed-effect model was assumed. SNPs available in less than 2 cohorts and N < 1000 were excluded. Meta-analysis results were compared for main effect models and interaction models to select top-SNPs for follow up.

From the final meta-analysis results, seven top-SNPs were selected for follow-up in another 12 independent studies. Table 24 presents the seven top-SNPs with information on the nearest gene, main effect *P* -value, interaction *P*-value, description of the gene selected and the functional role of the gene in published paper.

Table 24: SNP selected for maternal smoking and BW interaction G x E study for follow-up in 12 independent replication samples.

Rank	Chr	Nearest gene	Main effect , *P* value#	Interaction, *P* value*	Description and source	Function
1	2	KIF5C	7.56E-07	3.24E-05	Kinesin heavy chain isoform 5C; dbSNP	maintaining motor neuron; cancer (Cardoso et al, 2009)
2	7	PILRB	2.80E-06	1.56E-05	paired immunoglobin-like type 2 receptor beta; dbSNP	transmembrane receptor protein; cancer, autoimmune disorder (Wan et al, 2004; Thomas et al,2009)
3	4	GRID2	1.76E-05	9.07E-04	glutamate receptor, ionotropic, delta 2; dbSNP	neuronal apoptotic death; inherited diseases and cancer (Rozier et al, 2004)
4	2	LYPD6B	2.03E-05	2.77E-04	LY6/PLAUR domain containing 6B; dbSNP	expressed in CNS; ovarian cancer (Risinger et al, 2011)
5	7	CALCR	2.62E-05	8.59E-04	calcitonin receptor; receptor for peptide hormone calcitonin; dbSNP	maintaining calcium homeostasis; metabolic bone diseases (Wolfe et al, 2003)
6	10	CCDC3	6.84E-05	1.56E-04	Coiled-coiled domain containing 3; dbSNP	secreted by adipocytes and endothelial cells; obesity (Kobayashi et al, 2010)
7	22	RAC2	9.62E-05	5.26E-06	Ras-related C3 botulinum toxin substrate 2; dbSNP	control of cell growth, cytoskeleton reorganization, and the activation of protein kinases; human disease (Pai et al, 2010)

[#] and [*], The SNP were selected on a cut-off of $P < 1 \times 10^{-4}$ for main effects *P*-value and then their interaction *P*-value added (to double check association in interaction model as well), as presented in Table N. This step gave 222 SNPs. In the next step, LD for each SNP was calculated and SNPs were divided into blocks according to their genetic locations. Each top SNP from the block was selected for the Table above.

6.5.2. Association between birth and infant size measures and testosterone levels in adolescent with poly-cystic ovary syndrome (PCOS) symptoms in the NFBC1986.

Poly-cystic ovary syndrome (PCOS) can be diagnosed with the presence of hyper-androgenism(clinical and / or biochemical evidence of androgen excess) and chronic oligo-anovulation[342]. Obesity is strongly associated with PCOS[343], with reviews outlining that due to a lack of cross-sectional and longitudinal studies it is not possible to evaluate the prevalence rate of this association. Interestingly, some recent animal studies support the theory that *in utero* androgen excess may be an important factor programming subsequent PCOS development during puberty[344-346]. None of the previous studies assessed the relationship between birth or infant size and testosterone levels, which is a marker for PCOS.

The aim of this study is to determine the prevalence of hyper-androgenism in two large cohorts, NFBC1966 and ALSPAC, of post-menarchal adolescent girls followed prenatally and from birth in the UK and Finland. The study aims to determine the extent to which testosterone levels are related to early developmental changes such as prenatal androgen exposure, early postnatal weight gain, and other familial factors such as maternal history of PCOS. The study also investigated whether hyper-androgenism, independently of body mass index (BMI), is associated with cardiovascular risk factors during adolescence.

The main objectives of the study were to investigate if the associations between early developmental influences and combined primary outcomes remained after adjusting for confounders such as age at menarche and adolescence, both of which are strongly correlated with both testosterone and PCOS. The study defined cases and control using PCOS definition from Rotterdam workshop[347]. The study excluded individuals who were on the contraceptive pill, or were diagnosed with diabetes or were pregnant. Testosterone levels were assessed in girls at age 15-16 years and were selected as outcome measures (N=3000 in NFBC1986). The testosterone levels were log-transformed to reduce right skewness; hence the results were presented as % change. The predictor variables selected were BW, height and weight at infancy, delta change in infant height and weight, peak height velocity (PHV) and peak weight velocity (PWV). The standard deviation scores (SDS) were calculated for infant height and weight using the WHO anthro software (http://www.who.int/childgrowth/software/en/).

Three unadjusted and adjusted linear regression models were tested. In the unadjusted model direct association between testosterone levels and each of the predictors (BW, infant weight, infant height, delta weight, delta

height and PHV) was assessed. In the first adjusted model testosterone levels against predictors, adjusting for age at menarche, were assessed. In the second adjusted model, testosterone levels against predictors, after adjustments with age at menarche and adolescent BMI were assessed. Table 25 shows the association between these three models in detail.

Results show that, testosterone levels associated with weight, height and delta weight SD score at $P<0.05$ (Table 25). An increase in one unit of SD score of weight, height and delta weight increased the testosterone levels by 5% (95% CI: 2% to 8%), 3% (95% CI: 0.4% to 5%) and 3% (0.6% to 6%), respectively. None of the other predictors, such as BW, delta height, peak weight or peak height velocity, associated with testosterone levels at nominal significance values of $P<0.05$. In model 2, after adjustments with age at menarche, selected as a confounder, the association remained significant for weight, height and delta weight SD score at $P<0.05$. An increase in SD score of weight, height and delta weight increased the testosterone level by 5 %(2% to 8%), 3% (0.4% to 5%) and 3% (0.6% to 6%). However, in the last model, adjusting for BMI at 16 years the association between delta weight and testosterone level weakened and was non-significant ($P=0.08$). The association between weight and height SD score with testosterone levels remained unchanged. The testosterone level decreased to 4% (2% to 8%) by one unit change in weight SD score and no change in testosterone levels by one unit change in Height SD score(β=3% (0.4% to 5%)).

These are preliminary results showing that testosterone level associates with infant height and weight but not with BW. A further study using ALSPAC showed that height growth between 3-7 years and 7-10 years associate with testosterone levels (data not shown). Also, weight growth during 1-3 years is shown to associate with testosterone levels. These are very important findings as the age groups which shows significant association are within when the infancy and childhood growth spurts occurs, as shown in the thesis. A future study will investigate the role of adiposity peak (AP) in association with testosterone levels. It is clear from various studies in this thesis that AP may be having an indirect association with obesity in adolescence and in adulthood. Hormones, such as leptin, associating with AP may play an important role in PCOS as well, as shown in a previous study that demonstrates exercise increases leptin levels in PCOS women[348]. Therefore, these findings could give further understanding regarding the pathways of PCOS in adolescents.

Table 25: Association between birth and infant size measures and testosterone levels at 15/16 years in the NFBC1986.

	Unadjusted				Adjusted for age at menarche				Adjusted for age at menarche and BMI at 16 years			
	β	**95% CI**		***P*- value**	**β**	**95% CI**		***P*-value**	**β**	**95% CI**		***P*-value**
BW SDS	0.008	-0.015	0.031	0.479	0.012	-0.015	0.031	0.314	0.005	-0.015	0.031	0.681
Weight SDS	0.050	0.024	0.077	0.0002	0.049	0.024	0.077	0.001	0.043	0.024	0.077	0.004
Height SDS	0.031	0.005	0.058	0.021	0.033	0.005	0.058	0.018	0.031	0.005	0.058	0.031
Delta Weight	0.031	0.007	0.054	0.013	0.026	0.007	0.054	0.040	0.023	0.007	0.054	0.081
Delta Height	0.013	-0.010	0.035	0.271	0.013	-0.010	0.035	0.292	0.014	-0.010	0.035	0.239
Peak weight velocity (PWV)	0.005	-0.005	0.015	0.304	0.007	-0.005	0.015	0.176	0.006	-0.005	0.015	0.282
Peak height velocity (PHV)	0.003	0.000	0.007	0.073	0.003	0.000	0.007	0.138	0.003	0.000	0.007	0.116

SDS = standard deviation score; delta weight was calculated as a difference between BW and weight SDS; delta height was calculated as difference between BW and height SDS. β and 95% confidence interval (CI) were presented as % change in testosterone.

6.5.3. A genome-wide association studies (GWAS) on maternal gestational weight gain to identify novel genetic variants.

Maternal weight gain during pregnancy is an important risk factor for birth outcomes[349]. The study in this thesis confirms that maternal BMI programmes the offspring's later metabolic health, both due to *in utero* 'programming' as well as the childhood obesogenic environment. The study also showed that the obesity risk and associated cardio-metabolic risk is higher in children of overweight and obese mothers in comparison to children of lean mothers. Gestational weight gain (GWG) is a complex phenotype, it indirectly affects fetal growth via the *in-uterine* environment, which in turns influences the growth in childhood leading to adverse health effects in later life[90]. It is shown in studies that during the earlier period of pregnancy (up to ~18 weeks), the largest contribution to weight gain of the fetus is maternal fat deposition[350] however, a recent study using ALSPAC data did not associate with previous maternal or offspring adiposity variants to GWG[350].

The aim of this study was to conduct a GWAS of both maternal and fetal genome-wide data to identify genetic variants associated with GWG. The study selected five GWAS cohorts (ALSPAC, RAINE, Gen R, TEENS, INMA and NFBC1966) with GWG available; is led by the ALSPAC team and I am contributing on behalf of NFBCs as an analyst. Genotypes were obtained using high-density SNP arrays and were then imputed for ~ 2.5 million HapMap SNPs. The study included cohorts which had DNA available for both mothers and/or offspring. Several other inclusion criteria applied, for example maternal weight at three time points during pre-pregnancy, during gestation and before delivery were included. The study excluded twins or multiples, pre-term pregnancies (excluding deliveries < 37 weeks gestation), stillbirths / miscarriage, congenital anomalies such as Downs syndrome. Only samples of Caucasian origin were included in the study.

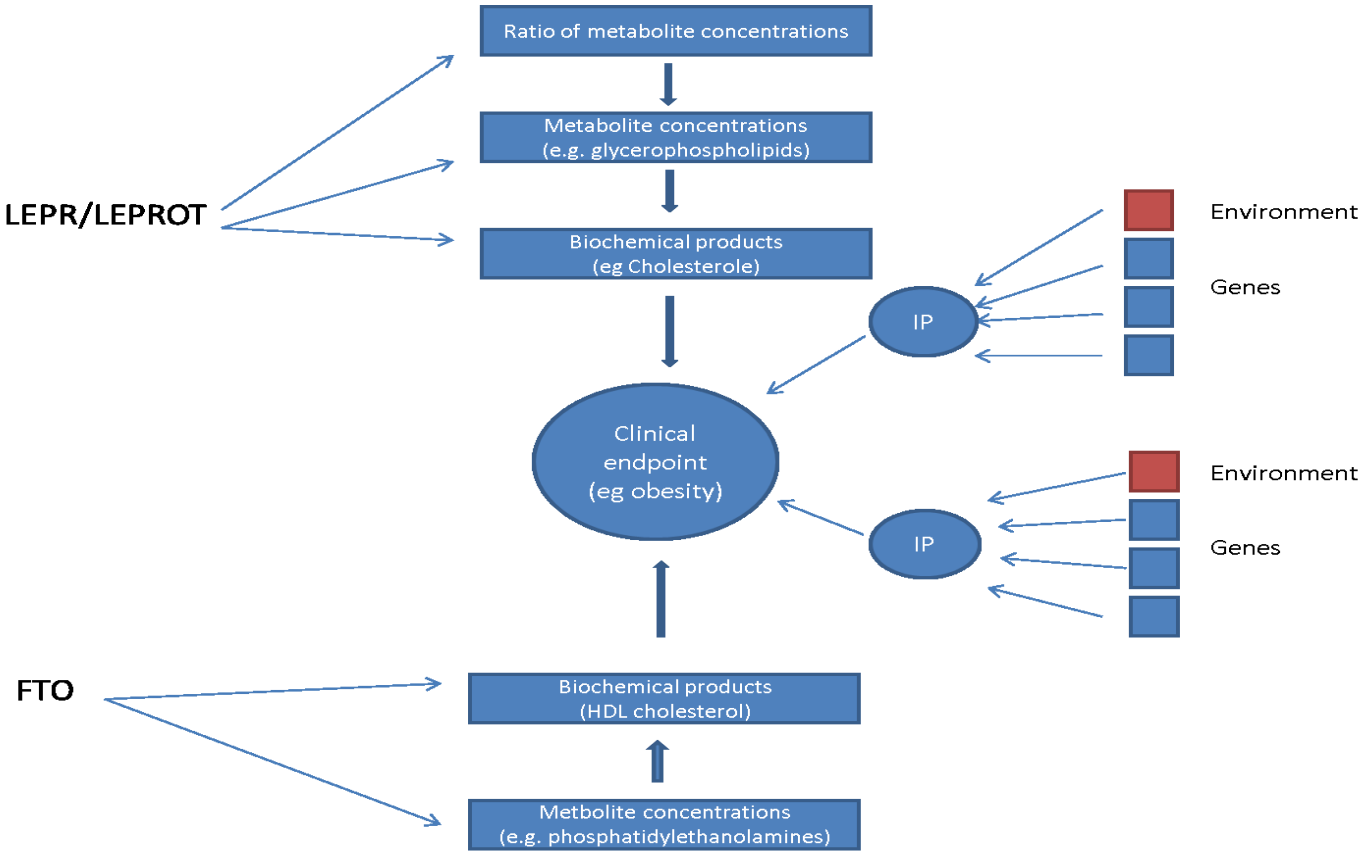
Three GWG variables were derived as outcome variables for the study – a total gestational weight gain (tGWG), early GWG per week gestation (eGWG) and late GWG per week gestation (lGWG). The three outcome variables were transformed to *z-scores* by subtracting the individual value subtracted by sample mean and divided by sample standard deviation. The association between each SNP and GWG variable was assessed in each study sample (including NFBC1966) using linear regression of GWG *z-score* against genotype, assuming an additive genetic model. This work is ongoing with follow-up SNPs for replication in NFBC1986 awaited.

6.5.4. The association between DNA methylation, early-life growth trajectories and adulthood metabolite levels.

This work is an extension of the GWAS conducted on early growth phenotypes. The aims of this study were to investigate the metabolic and epigenetic modification of the genetic variants associating with early growth phenotypes and adult BMI. Although GWA studies identify a number of genetic polymorphisms associating with complex diseases, little can be inferred of the underlying pathways for the disease-causing mechanisms. Moreover, large populations are needed to provide enough statistical power for identification of new disease-causing genetic variants. With the advances in gene expression and reduced costs, several new methodologies have been proposed to understand the causal mechanism of these disease-causing genetic variants. The studies on metabolomics measures can act as intermediary phenotypes (IP), which can associate a metabolite with disease end points and provide more details on potential pathways related to the etiology of the disease[351](**Figure 14**). Additionally, more recently epigenetic modifications have also been described to understand the molecular reprogramming of a DNA sequence, which describes the change in gene expression in response to environmental stimuli, such as change in diet and *in utero* environment[352]. Moreover these epigenetic markers are suggested to sustain long term change in gene expression and phenotypic variability[353], indicating more heritability of disease traits from parent to offspring.

The results from GWAS conducted on early growth phenotypes identified five genetic variants (in/near *LEPR/LEPROT*, *FTO*, *TFAP2B*, *GNPDA2* and *PCSK1*) regulating body weight, appetite and energy balance via the hypothalamic-pituitary-adrenal axis (HPA). Moreover, one SNP in *LEPR/LEPROT* regulates the growth hormone action in the liver and its over-expression reduces IGF-1 production, leading to growth retardation. Therefore the objective of this study is to understand the molecular mechanisms connecting these genetic variants with the metabolites and epigenetic markers, which may be associated with adverse health in adulthood. The main hypotheses which this study tests are: a) early growth parameters associate with metabolites and b) early growth phenotypes show epigenetic modification markers. The study aimed to use large-scale metabolomics data available in NFBC1966 (n=5500) for discovery stage analysis and then use NFBC1986 (n=6000) for validation. Within this ongoing work; the study will now extend to investigate the early growth phenotypes in DNA methylation data.

Figure 15: Schematic illustration of the role of intermediate phenotypes (IPs), such as metabolic traits or epigenetic markers, demonstrated by example of two genes that associated with early growth phenotypes and adult BMI, both markers for obesity. Adapted from Gieger et al, 2008[351].



Bibliography

1. Fernandez, J.R., Klimentidis, Y.C., Dulin-Keita, A. & Casazza, K. Genetic influences in childhood obesity: recent progress and recommendations for experimental designs. *Int J Obes (Lond)* **36**, 479-84 (2012).
2. Barker, D.J. The fetal and infant origins of adult disease. *BMJ* **301**, 1111 (1990).
3. Reilly, J.J. *et al.* Early life risk factors for obesity in childhood: cohort study. *BMJ* **330**, 1357 (2005).
4. Stettler, N. Nature and strength of epidemiological evidence for origins of childhood and adulthood obesity in the first year of life. *International Journal of Obesity* **31**, 1035-1043 (2007).
5. Wells, J.C., Chomtho, S. & Fewtrell, M.S. Programming of body composition by early growth and nutrition. *Proceedings of the Nutrition Society* **66**, 423-434 (2007).
6. Lucas, A. Programming by early nutrition: an experimental approach. *The Journal of nutrition* **128**, 401S-406S (1998).
7. Dietz, W.H. Critical periods in childhood for the development of obesity. *The American Journal of Clinical Nutrition* **59**, 955-9 (1994).
8. Monteiro, P.O. & Victora, C.G. Rapid growth in infancy and childhood and obesity in later life--a systematic review. *Obes Rev* **6**, 143-54 (2005).
9. Ong, K.K. & Loos, R.J. Rapid infancy weight gain and subsequent obesity: systematic reviews and hopeful suggestions. *Acta Paediatr* **95**, 904-8 (2006).
10. Baird, J. *et al.* Being big or growing fast: systematic review of size and growth in infancy and later obesity. *BMJ: British Medical Journal* **331**, 929 (2005).
11. Sovio, U. Imperial College London (2011).
12. Sovio, U. *et al.* Genetic determinants of height growth assessed longitudinally from infancy to adulthood in the northern Finland birth cohort 1966. *PLoS Genet* **5**, e1000409 (2009).
13. Warrington, N.M. *et al.* Modeling BMI Trajectories in Children for Genetic Association Studies. *PLoS ONE* **8**, e53897 (2013).
14. Tzoulaki, I. *et al.* Relation of immediate postnatal growth with obesity and related metabolic risk factors in adulthood: the northern Finland birth cohort 1966 study. *Am J Epidemiol* **171**, 989-98 (2010).
15. Sovio, U. *et al.* How do changes in body mass index in infancy and childhood associate with cardiometabolic profile in adulthood[quest] Findings from the Northern Finland Birth Cohort 1966 Study. *Int J Obes* (2013).
16. Gibbs, R.A. *et al.* The international HapMap project. *Nature* **426**, 789-796 (2003).
17. Consortium., T.E.P. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).
18. Schena, M. *et al.* Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proceedings of the National Academy of Sciences* **93**, 10614-10619 (1996).
19. Tanner, J.M. *Fetus into Man: Physical Growth from Conception to Maturity.*, (Castlemead Publications, Ware, 1989).
20. Alwan, A. *Global status report on noncommunicable diseases 2010*, (World Health Organization, 2011).
21. Ritchie, S.A. & Connell, J.M.C. The link between abdominal obesity, metabolic syndrome and cardiovascular disease. *Nutrition, Metabolism and Cardiovascular Diseases* **17**, 319-326 (2007).
22. Consultation, W. Obesity: preventing and managing the global epidemic. *World Health Organization technical report series* **894**(2000).
23. Fontaine, K. & Barofsky, I. Obesity and health-related quality of life. *Obesity reviews* **2**, 173-182 (2001).
24. Braet, C., Mervielde, I. & Vandereycken, W. Psychological aspects of childhood obesity: a controlled study in a clinical and nonclinical sample. *Journal of Pediatric psychology* **22**, 59-71 (1997).

25. Cole, T.J., Bellizzi, M.C., Flegal, K.M. & Dietz, W.H. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ* **320**, 1240 (2000).
26. WHO. Physical status: the use and interpretation of anthropometry. Geneva: WHO, 1995. (1995).
27. Deurenberg, P., Pieters, J. & Hautvast, J. The assessment of the body fat percentage by skinfold thickness measurements in childhood and young adolescence. *Br J Nutr* **63**, 293-303 (1990).
28. Han, T., Van Leer, E., Seidell, J. & Lean, M. Waist circumference action levels in the identification of cardiovascular risk factors: prevalence study in a random sample. *Bmj* **311**, 1401-1405 (1995).
29. Ahima, R.S. & Lazar, M.A. The Health Risk of Obesity—Better Metrics Imperative. *Science* **341**, 856-858 (2013).
30. Wang, Y. & Lobstein, T.I.M. Worldwide trends in childhood overweight and obesity. *International Journal of Pediatric Obesity* **1**, 11-25 (2006).
31. Kimm, S.Y. & Obarzanek, E. Childhood obesity: a new pandemic of the new millennium. *Pediatrics* **110**, 1003-1007 (2002).
32. Dietz, W.H. Health consequences of obesity in youth: childhood predictors of adult disease. *Pediatrics* **101**, 518-525 (1998).
33. Serdula, M.K. *et al.* Do obese children become obese adults? A review of the literature. *Preventive medicine* **22**, 167-177 (1993).
34. Center, T.H.a.S.C.I. Health Survey for England -2011, Health, social care and lifestyle [NS]. (2012).
35. Roseboom, T.J. *et al.* Effects of prenatal exposure to the Dutch famine on adult disease in later life: an overview. *Molecular and cellular endocrinology* **185**, 93-98 (2001).
36. de Rooij, S.R., Wouters, H., Yonker, J.E., Painter, R.C. & Roseboom, T.J. Prenatal undernutrition and cognitive function in late adulthood. *Proceedings of the National Academy of Sciences* **107**, 16881-16886 (2010).
37. Collins, F.S., Morgan, M. & Patrinos, A. The Human Genome Project: lessons from large-scale biology. *Science* **300**, 286-290 (2003).
38. Zhao, J. & Grant, S.F. Genetics of childhood obesity. *J Obes* **2011**, 845148 (2011).
39. Smith, G.D., Steer, C., Leary, S. & Ness, A. Is there an intrauterine influence on obesity? Evidence from parent–child associations in the Avon Longitudinal Study of Parents and Children (ALSPAC). *Archives of disease in childhood* **92**, 876-880 (2007).
40. Nelson, H.H., Kelsey, K.T., Mott, L.A. & Karagas, M.R. The XRCC1 Arg399Gln Polymorphism, Sunburn, and Non-melanoma Skin Cancer Evidence of Gene-Environment Interaction. *Cancer research* **62**, 152-155 (2002).
41. Müezzinler, A., Zaineddin, A. & Brenner, H. Body mass index and leukocyte telomere length in adults: a systematic review and meta-analysis. *Obesity Reviews* (2013).
42. Mohn, A. *et al.* Increased Oxidative Stress in Prepubertal Severely Obese Children: Effect of a Dietary Restriction-Weight Loss Program. *Journal of Clinical Endocrinology & Metabolism* **90**, 2653-2658 (2005).
43. Barker, D.J.P. *Fetal and infant origins of adult disease*, 343 (BMJ Publishing, London 1992).
44. Barker, D.J.P. The origins of the developmental origins theory. *Journal of Internal Medicine* **261**, 412-417 (2007).
45. Gillman, M.W. *et al.* Meeting Report on the 3rd International Congress on Developmental Origins of Health and Disease (DOHaD). *Pediatr Res* **61**, 625-629 (2007).
46. Kermack, W.O., McKendrick, A.G. & McKinlay, P.L. Death-rates in Great Britain and Sweden: Expression of Specific Mortality Rates as Products of Two Factors, and some Consequences thereof. *J Hyg (Lond)* **34**, 433-57 (1934).
47. Neel, J.V. The thrifty genotype revisited. *The genetics of diabetes mellitus* **3**(1982).
48. Neel, J.V. Diabetes mellitus: a 'thrifty' genotype rendered detrimental by 'progress'? *Am. J. Hum. Genet.* **14**, 353-362 (1962).
49. McCance, R.A. Food, growth, and time. *Lancet* **2**, 671-6 (1962).

50. Ravelli, G.P., Stein, Z.A. & Susser, M.W. Obesity in young men after famine exposure in utero and early infancy. *N Engl J Med* **295**, 349-53 (1976).
51. Forsdahl, A. Are poor living conditions in childhood and adolescence an important risk factor for arteriosclerotic heart disease? *Br J Prev Soc Med* **31**, 91-5 (1977).
52. Lucas, A. Programming by early nutrition in man. *The childhood environment and adult disease* **1991**, 38-55 (1991).
53. Barker, D.J. *et al.* Fetal nutrition and cardiovascular disease in adult life. *Lancet* **341**, 938-41 (1993).
54. Frankel, S. *et al.* Birthweight, body-mass index in middle age, and incident coronary heart disease. *The Lancet* **348**, 1478-1480 (1996).
55. Leon, D.A. *et al.* Failure to realise growth potential in utero and adult obesity in relation to blood pressure in 50 year old Swedish men. *BMJ* **312**, 401-406 (1996).
56. Whincup, P.H. *et al.* Size at Birth and Blood Pressure at 3 Years of Age: The Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC). *American Journal of Epidemiology* **149**, 730-739 (1999).
57. Huxley, R.R., Shiell, A.W. & Law, C.M. The role of size at birth and postnatal catch-up growth in determining systolic blood pressure: a systematic review of the literature. *Journal of hypertension* **18**, 815-831 (2000).
58. Barnett, A.H., Eff, C., Leslie, R.D.G. & Pyke, D.A. Diabetes in identical twins. *Diabetologia* **20**, 87-93 (1981).
59. Hales, C.N. & Barker, D.J.P. The thrifty phenotype hypothesis. *British Medical Bulletin* **60**, 5-20 (2001).
60. Jones, R.H. & Ozanne, S.E. Fetal programming of glucose–insulin metabolism. *Molecular and Cellular Endocrinology* **297**, 4-9 (2009).
61. Botton, J., Heude, B., Maccario, J., Ducimetiere, P. & Charles, M.A. 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* **87**, 1760-8 (2008).
62. Hyatt, M.A., Keisler, D.H., Budge, H. & Symonds, M.E. Maternal parity and its effect on adipose tissue deposition and endocrine sensitivity in the postnatal sheep. *J Endocrinol* **204**, 173-9 (2010).
63. Jarvelin, M.R., Hartikainen-Sorri, A.L. & Rantakallio, P. Labour induction policy in hospitals of different levels of specialisation. *Br J Obstet Gynaecol* **100**, 310-5 (1993).
64. Jaddoe, V.W. & Witteman, J.C. Hypotheses on the fetal origins of adult diseases: contributions of epidemiological studies. *Eur J Epidemiol* **21**, 91-102 (2006).
65. Jarvelin, M.R. *et al.* Ecological and individual predictors of birthweight in a northern Finland birth cohort 1986. *Paediatr Perinat Epidemiol* **11**, 298-312 (1997).
66. Jarvelin, M.R. *et al.* Early life factors and blood pressure at age 31 years in the 1966 northern Finland birth cohort. *Hypertension* **44**, 838-46 (2004).
67. Kinra, S., Baumer, J.H. & Davey Smith, G. Early growth and childhood obesity: a historical cohort study. *Arch Dis Child* **90**, 1122-7 (2005).
68. Kramer, M.S. *et al.* Why are babies getting bigger? Temporal trends in fetal growth and its determinants. *The Journal of Pediatrics* **141**, 538-542 (2002).
69. Hattersley, A.T. & Tooke, J.E. The fetal insulin hypothesis: an alternative explanation of the association of low birthweight with diabetes and vascular disease. *Lancet* **353**, 1789-92 (1999).
70. Hypponen, E., Smith, G.D. & Power, C. Parental diabetes and birth weight of offspring: intergenerational cohort study. *BMJ* **326**, 19-20 (2003).
71. Scholl, T.O., Sowers, M., Chen, X. & Lenders, C. Maternal glucose concentration influences fetal growth, gestation, and pregnancy complications. *Am J Epidemiol* **154**, 514-20 (2001).
72. Lawlor, D.A., Davey Smith, G. & Ebrahim, S. Birth weight of offspring and insulin resistance in late adulthood: cross sectional survey. *BMJ* **325**, 359 (2002).
73. Davey Smith, G., Sterne, J.A.C., Tynelius, P. & Rasmussen, F. Birth characteristics of offspring and parental diabetes: evidence for the fetal insulin hypothesis. *Journal of Epidemiology and Community Health* **58**, 126-128 (2004).

74. Barker, D.J., Osmond, C., Forsen, T.J., Kajantie, E. & Eriksson, J.G. Trajectories of growth among children who have coronary events as adults. *N Engl J Med* **353**, 1802-9 (2005).
75. Berkey, C.S. & Reed, R.B. A model for describing normal and abnormal growth in early childhood. *Hum Biol* **59**, 973-87 (1987).
76. Painter, R.C., Roseboom, T.J. & Bleker, O.P. Prenatal exposure to the Dutch famine and disease in later life: an overview. *Reprod Toxicol* **20**, 345-52 (2005).
77. Prader, A., Tanner, J.M. & von, H.G. Catch-up growth following illness or starvation. An example of developmental canalization in man. *J Pediatr* **62**, 646-59 (1963).
78. S., C., D., G. & F., B. Low birthweight and adult insulin resistance: the "catch-up growth" hypothesis. . *Arch Dis Child Fetal Neonatal Ed.* **81**, F71-3 (1999).
79. Fewtrell, M.S. *et al.* Catch-up growth in small-for-gestational-age term infants: a randomized trial. *The American journal of clinical nutrition* **74**, 516-523 (2001).
80. Monset-Couchard, M., De Bethmann, O. & Kastler, B. Mid-and long-term outcome of 166 premature infants weighing less than 1,000 g at birth, all small for gestational age. *Neonatology* **81**, 244-254 (2002).
81. Victora, C.G., Barros, F.C., Horta, B.L. & Martorell, R. Short-term benefits of catch-up growth for small-for-gestational-age infants. *International Journal of Epidemiology* **30**, 1325-1330 (2001).
82. Cameron, N. & Demerath, E.W. Critical periods in human growth and their relationship to diseases of aging. *Am J Phys Anthropol* **Suppl 35**, 159-84 (2002).
83. Law, C. *et al.* Fetal, infant, and childhood growth and adult blood pressure a longitudinal study from birth to 22 years of age. *Circulation* **105**, 1088-1092 (2002).
84. Eriksson, J.G., Forsen, T., Tuomilehto, J., Osmond, C. & Barker, D.J. Early growth and coronary heart disease in later life: longitudinal study. *BMJ: British Medical Journal* **322**, 949 (2001).
85. Crowther, N.J. *et al.* Influence of catch-up growth on glucose tolerance and β-cell function in 7-year-old children: results from the birth to twenty study. *Pediatrics* **121**, e1715-e1722 (2008).
86. Rolland-Cachera, M.F. *et al.* Adiposity rebound in children: a simple indicator for predicting obesity. *Am J Clin Nutr* **39**, 129-35 (1984).
87. Dorosty, A.R., Emmett, P.M., Cowin, S. & Reilly, J.J. Factors associated with early adiposity rebound. ALSPAC Study Team. *Pediatrics* **105**, 1115-8 (2000).
88. Whitaker, R.C., Wright, J.A., Pepe, M.S., Seidel, K.D. & Dietz, W.H. Predicting Obesity in Young Adulthood from Childhood and Parental Obesity. *New England Journal of Medicine* **337**, 869-873 (1997).
89. Williams, R.C., Long, J.C., Hanson, R.L., Sievers, M.L. & Knowler, W.C. Individual estimates of European genetic admixture associated with lower body-mass index, plasma glucose, and prevalence of type 2 diabetes in Pima Indians. *Am. J. Hum. Genet.* **66**, 527-538 (2000).
90. Whitaker, R.C. & Dietz, W.H. Role of the prenatal environment in the development of obesity. *Journal of Pediatrics* **132**, 768-776 (1998).
91. Buchan, I.E., Bundred, P.E., Kitchiner, D.J. & Cole, T.J. Body mass index has risen more steeply in tall than in short 3-year olds: serial cross-sectional surveys 1988-2003. *Int J Obes (Lond)* **31**, 23-9 (2007).
92. Freedman, D.S. *et al.* Inter-relationships among childhood BMI, childhood height, and adult obesity: the Bogalusa Heart Study. *Int J Obes Relat Metab Disord* **28**, 10-6 (2004).
93. Deheeger, M. & Rolland-Cachera, M.F. [Longitudinal study of anthropometric measurements in Parisian children aged ten months to 18 years]. *Arch Pediatr* **11**, 1139-44 (2004).
94. Mook-Kanamori, D.O. *et al.* Variants near CCNL1/LEKR1 and in ADCY5 and fetal growth characteristics in different trimesters. *J Clin Endocrinol Metab* **96**, E810-5 (2011).
95. Nobili, V., Alisi, A., Panera, N. & Agostoni, C. Low birth weight and catch-up-growth associated with metabolic syndrome: a ten year systematic review. *Pediatr Endocrinol Rev* **6**, 241-7 (2008).
96. Ong, K.K. Catch-up growth in small for gestational age babies: good or bad? *Curr Opin Endocrinol Diabetes Obes* **14**, 30-4 (2007).

97. Boney, C.M., Verma, A., Tucker, R. & Vohr, B.R. Metabolic syndrome in childhood: association with birth weight, maternal obesity, and gestational diabetes mellitus. *Pediatrics* **115**, e290-6 (2005).
98. Hendricks, K.A., Nuno, O.M., Suarez, L. & Larsen, R. Effects of hyperinsulinemia and obesity on risk of neural tube defects among Mexican Americans. *Epidemiology* **12**, 630-635 (2001).
99. Ehrenberg, H.M., Mercer, B.M. & Catalano, P.M. The influence of obesity and diabetes on the prevalence of macrosomia. *American journal of obstetrics and gynecology* **191**, 964-968 (2004).
100. Boney, C.M., Verma, A., Tucker, R. & Vohr, B.R. Metabolic syndrome in childhood: association with birth weight, maternal obesity, and gestational diabetes mellitus. *Pediatrics* **115**, e290-e296 (2005).
101. Wei, J.-N. *et al.* Low birth weight and high birth weight infants are both at an increased risk to have type 2 diabetes among schoolchildren in Taiwan. *Diabetes care* **26**, 343-348 (2003).
102. Feinleib, M. The NHLBI twin study of cardiovascular disease risk factors: methodology and summary of results. *Am. J. Epidemiol.* **106**, 284-285 (1977).
103. Stunkard, A.J. *et al.* An adoption study of human obesity. *N Engl J Med* **314**, 193-198 (1986).
104. Stunkard, A.J., Harris, J.R., Pedersen, N.L. & McClearn, G.E. The body-mass index of twins who have been reared apart. *N. Engl. J. Med.* **322**, 1483-1487 (1990).
105. Maes, H.H., Neale, M.C. & Eaves, L.J. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet* **27**, 325-351 (1997).
106. Segal, N.L., Feng, R., McGuire, S.A., Allison, D.B. & Miller, S. Genetic and environmental contributions to body mass index: comparative analysis of monozygotic twins, dizygotic twins and same-age unrelated siblings. *Int J Obes (Lond)* **33**, 37-41 (2009).
107. Wardle, J., Carnell, S., Haworth, C.M. & Plomin, R. Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment. *The American Journal of Clinical Nutrition* **87**, 398-404 (2008).
108. Rankinen, T. *et al.* The human obesity gene map: the 2005 update. *Obesity (Silver Spring)* **14**, 529-644 (2006).
109. Bell, C., Benzinou, M. & Siddiq, A. Genome-wide linkage analysis for severe obesity in French Caucasians finds significant susceptibility locus on chromosome 19q. *Diabetes* **53**, 1857-1865 (2004).
110. Walley, A.J., Asher, J.E. & Froguel, P. The genetic contribution to non-syndromic human obesity. *Nat Rev Genet* **10**, 431-442 (2009).
111. Clement, K. A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* **392**, 398-401 (1998).
112. Siddiq, A. Single nucleotide polymorphisms in the neuropeptide Y2 receptor (NPY2R) gene and association with severe obesity in French white subjects. *Diabetologia* **50**, 574-84 (2007).
113. Benzinou, M. Common nonsynonymous variants in PCSK1 confer risk of obesity. *Nature Genet.* **40**, 943-945 (2008).
114. Meyre, D. Variants of ENPP1 are associated with childhood and adult obesity and increase the risk of glucose intolerance and type 2 diabetes. *Nature Genet.* **37**, 863-867 (2005).
115. Blakemore, A.I. A rare variant in the visfatin gene (NAMPT/PBEF1) is associated with protection from obesity. *Obesity (Silver Spring)*.
116. Farooqi, I.S. Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N. Engl. J. Med.* **348**, 1085-1095 (2003).
117. Hinney, A. *et al.* Melanocortin-4 Receptor Gene: Case-Control Study and Transmission Disequilibrium Test Confirm that Functionally Relevant Mutations Are Compatible with a Major Gene Effect for Extreme Obesity. *Journal of Clinical Endocrinology & Metabolism* **88**, 4258-4267 (2003).
118. Cauchi, S. *et al.* Combined effects of MC4R and FTO common genetic variants on obesity in European general populations. *J Mol Med (Berl)* **87**, 537-46 (2009).
119. Meyre, D. Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nature Genet.* **41**, 157-159 (2009).

120. Haworth, C.M., Butcher, L.M., Docherty, S.J., Wardle, J. & Plomin, R. No evidence for association between BMI and 10 candidate genes at ages 4, 7 and 10 in a large UK sample of twins. *BMC Med Genet* **9**, 12 (2008).
121. Podolsky, R.H. *et al.* Candidate genes and growth curves for adiposity in African- and European-American youth. *Int J Obes (Lond)* **31**, 1491-1499 (2007).
122. Zhang, Y. Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425-432 (1994).
123. Herbert, A. A common genetic variant is associated with adult and childhood obesity. *Science* **312**, 279-283 (2006).
124. Scott, L.J. *et al.* A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants. *Science* **316**, 1341-1345 (2007).
125. Frayling, T.M. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889-894 (2007).
126. Scuteri, A. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet.* **3**, e115 (2007).
127. den Hoed, M. *et al.* Genetic Susceptibility to Obesity and Related Traits in Childhood and Adolescence. *Diabetes* **59**, 2980-2988 (2010).
128. Cecil, J.E., Tavendale, R., Watt, P., Hetherington, M.M. & Palmer, C.N. An obesity-associated FTO gene variant and increased energy intake in children. *N Engl J Med* **359**, 2558-66 (2008).
129. Hardy, R. *et al.* Life course variations in the associations between FTO and MC4R gene variants and body size. *Hum Mol Genet* **19**, 545-52 (2010).
130. Grant, S.F. *et al.* Association analysis of the FTO gene with obesity in children of Caucasian and African ancestry reveals a common tagging SNP. *PLoS ONE* **3**, e1746 (2008).
131. Ng, M.C. *et al.* Implication of genetic variants near NEGR1, SEC16B, TMEM18, ETV5/DGKG, GNPDA2, LIN7C/BDNF, MTCH2, BCDIN3D/FAIM2, SH2B1, FTO, MC4R, and KCTD15 with obesity and type 2 diabetes in 7705 Chinese. *J Clin Endocrinol Metab* **95**, 2418-25 (2010).
132. Yajnik, C.S. *et al.* FTO gene variants are strongly associated with type 2 diabetes in South Asian Indians. *Diabetologia* **52**, 247-52 (2009).
133. Loos, R.J. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nature Genet.* **40**, 768-775 (2008).
134. Chambers, J.C. *et al.* Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat Genet* **40**, 716-8 (2008).
135. Willer, C. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* **41**, 25 - 34 (2009).
136. Thorleifsson, G. *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* **41**, 18-24 (2009).
137. Lindgren, C.M. *et al.* Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS Genet* **5**, e1000508 (2009).
138. Heard-Costa, N.L. *et al.* <italic>NRXN3</italic> Is a Novel Locus for Waist Circumference: A Genome-Wide Association Study from the CHARGE Consortium. *PLoS Genet* **5**, e1000539 (2009).
139. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**, 937-948 (2010).
140. Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* **42**, 949-960 (2010).
141. Scherag, A. *et al.* Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and german study groups. *PLoS genetics* **6**, e1000916 (2010).
142. Jiao, H. *et al.* Genome wide association study identifies KCNMA1 contributing to human obesity. *BMC medical genomics* **4**, 51 (2011).

143. Sandholt, C., Hansen, T. & Pedersen, O. Beyond the fourth wave of genome-wide obesity association studies. *Nutrition & diabetes* **2**, e37 (2012).
144. Consortium, E.G.G. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nature genetics* **44**, 526-531 (2012).
145. Collerton, J. *et al.* Telomere length is associated with left ventricular function in the oldest old: the Newcastle 85+ study. *Eur Heart J* **28**, 172-6 (2007).
146. Bakaysa, S.L. *et al.* Telomere length predicts survival independent of genetic influences. *Aging Cell* **6**, 769-774 (2007).
147. Kimura, M. *et al.* Telomere Length and Mortality: A Study of Leukocytes in Elderly Danish Twins. *American Journal of Epidemiology* **167**, 799-806 (2008).
148. Samani, N.J., Boultby, R., Butler, R., Thompson, J.R. & Goodall, A.H. Telomere shortening in atherosclerosis. *Lancet* **358**, 472-3 (2001).
149. Brouilette, S., Singh, R.K., Thompson, J.R., Goodall, A.H. & Samani, N.J. White cell telomere length and risk of premature myocardial infarction. *Arterioscler Thromb Vasc Biol* **23**, 842-6 (2003).
150. Ogami, M. *et al.* Telomere shortening in human coronary artery diseases. *Arterioscler Thromb Vasc Biol* **24**, 546-50 (2004).
151. Wilson, W.R. *et al.* Blood leucocyte telomere DNA content predicts vascular telomere DNA content in humans with and without vascular disease. *Eur Heart J* **29**, 2689-94 (2008).
152. Valdes, A.M. *et al.* Obesity, cigarette smoking, and telomere length in women. *Lancet* **366**, 662-4 (2005).
153. Farzaneh-Far, R. *et al.* Telomere Length Trajectory and Its Determinants in Persons with Coronary Artery Disease: Longitudinal Findings from the Heart and Soul Study. *PLoS ONE* **5**, e8612 (2010).
154. Nordfjall, K. *et al.* The individual blood cell telomere attrition rate is telomere length dependent. *PLoS Genet* **5**, e1000375 (2009).
155. Epel, E.S. *et al.* Accelerated telomere shortening in response to life stress. *Proc Natl Acad Sci U S A* **101**, 17312-5 (2004).
156. Kim, S. *et al.* Obesity and weight gain in adulthood and telomere length. *Cancer Epidemiol Biomarkers Prev* **18**, 816-20 (2009).
157. O'Callaghan, N.J., Clifton, P.M., Noakes, M. & Fenech, M. Weight loss in obese men is associated with increased telomere length and decreased abasic sites in rectal mucosa. *Rejuvenation Res* **12**, 169-76 (2009).
158. Zannolli, R. *et al.* Telomere length and obesity. *Acta Paediatr* **97**, 952-4 (2008).
159. Al-Attas, O.S. *et al.* Telomere length in relation to insulin resistance, inflammation and obesity among Arab youth. *Acta Paediatr* **99**, 896-9 (2010).
160. Lee, M., Martin, H., Firpo, M.A. & Demerath, E.W. Inverse association between adiposity and telomere length: The Fels Longitudinal Study. *Am J Hum Biol* **23**, 100-6 (2011).
161. Sidorov, I., Kimura, M., Yashin, A. & Aviv, A. Leukocyte telomere dynamics and human hematopoietic stem cell kinetics during somatic growth. *Experimental hematology* **37**, 514-524 (2009).
162. Rantakallio, P. Groups of risk in low birth weight infants and perinatal mortality. *Acta Paediatr Scand* **193(suppl 193)**, 1-71 (1969).
163. Sabatti, C. *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* **41**, 35-46 (2009).
164. Kaakinen, M. *et al.* Life-course analysis of a fat mass and obesity-associated (FTO) gene variant and body mass index in the Northern Finland Birth Cohort 1966 using structural equation modeling. *American journal of epidemiology* **172**, 653-665 (2010).
165. Deurenberg, P., Weststrate, J.A. & Seidell, J.C. Body mass index as a measure of body fatness: age- and sex-specific prediction formulas. *Br J Nutr* **65**, 105-14 (1991).

166. Friedewald, W.T., Levy, R.I. & Fredrickson, D.S. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clinical chemistry* **18**, 499-502 (1972).
167. Karlberg, J. The human growth curve. in *Cambridge Encyclopedia of Human Growth and Development* (eds. Ulijaszek, S.J., Johnston, F.E. & Preece, M.A.) 108-115 (Cambridge University Press, Cambridge, 1998).
168. Loomba-Albrecht, L.A. & Styne, D.M. Effect of puberty on body composition. *Current Opinion in Endocrinology, Diabetes and Obesity* **16**, 10-15 10.1097/MED.0b013e328320d54c (2009).
169. Hauspie, R.C. & Molinari, L. Parametric models for postnatal growth. in *Methods in Human Growth Research* (eds. Hauspie, R.C., Cameron, N. & Molinari, L.) 205-233 (Cambridge University Press, Cambridge, 2004).
170. Muhlhausler, B.S., Hancock, S.N., Bloomfield, F.H. & Harding, R. Are Twins Growth Restricted? *Pediatr Res* **70**, 117-122 (2011).
171. Simondon, K.B., Simondon, F., Delpeuch, F. & Cornu, A. Comparative study of five growth models applied to weight data from congolese infants between birth and 13 months of age. *American journal of human biology* **4**, 327-335 (1992).
172. Sovio, U. *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* **7**, e1001307 (2011).
173. Pinheiro, J., Bates, D., DebRoy, S. & Sarkar, D. Linear and nonlinear mixed effects models. *R package version* **3**, 57 (2007).
174. HapMap. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861 (2007).
175. Jaddoe, V.W. *et al.* Active and passive maternal smoking during pregnancy and the risks of low birthweight and preterm birth: the Generation R Study. *Paediatr Perinat Epidemiol* **22**, 162-71 (2008).
176. Leermakers, E.T. *et al.* A common genetic variant at 15q25 modifies the associations of maternal smoking during pregnancy with fetal growth: the generation R study. *PLoS ONE* **7**, e34584 (2012).
177. Heinrich, J. *et al.* Die zwei deutschen Geburtskohorten GINIplus und LISAplus. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* **55**, 864-874 (2012).
178. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300 (1995).
179. Graham, J.W., Olchowski, A.E. & Gilreath, T.D. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev Sci* **8**, 206-13 (2007).
180. Sovio, U. Dissertation, Imperial College London, University of London (2006).
181. Baron, R.M. & Kenny, D.A. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* **51**, 1173-82 (1986).
182. Sobel, M.E. Some New Results on Indirect Effects and Their Standard Errors in Covariance Structure Models. *Sociological Methodology* **16**, 159-186 (1986).
183. Hayes, A.F. Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs* **76**, 408-420 (2009).
184. MacKinnon, D.P., Lockwood, C.M. & Williams, J. Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate behavioral research* **39**, 99-128 (2004).
185. Holbert, R.L. & Stephenson, M.T. The importance of indirect effects in media effects research: Testing for mediation in structural equation modeling. *Journal of Broadcasting & Electronic Media* **47**, 556-572 (2003).
186. Bush, W.S. & Moore, J.H. Chapter 11: Genome-Wide Association Studies. *PLoS Comput Biol* **8**, e1002822 (2012).

187. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559-575 (2007).
188. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-913 (2007).
189. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).
190. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
191. Gibson, G. Hints of hidden heritability in GWAS. *Nat Genet* **42**, 558-560 (2010).
192. Sanna, S. *et al.* Fine Mapping of Five Loci Associated with Low-Density Lipoprotein Cholesterol Detects Variants That Double the Explained Heritability. *PLoS Genet* **7**, e1002198 (2011).
193. Krestyaninova, M. *et al.* A System for Information Management in BioMedical Studies--SIMBioMS. *Bioinformatics* **25**, 2768-9 (2009).
194. Das, S. Imperial College London (2009).
195. Barker, D.J. Adult consequences of fetal growth restriction. *Clin Obstet Gynecol* **49**, 270-83 (2006).
196. Forsen, T., Osmond, C., Eriksson, J.G. & Barker, D.J. Growth of girls who later develop coronary heart disease. *Heart* **90**, 20-4 (2004).
197. Yajnik, C.S. Early life origins of insulin resistance and type 2 diabetes in India and other Asian countries. *J Nutr* **134**, 205-10 (2004).
198. Godfrey, K.M. & Barker, D.J. Fetal programming and adult health. *Public health nutrition* **4**, 611-624 (2001).
199. Valero de Bernabé, J. *et al.* Risk factors for low birth weight: a review. *European Journal of Obstetrics & Gynecology and Reproductive Biology* **116**, 3-15 (2004).
200. Newsome, C.A. *et al.* Is birth weight related to later glucose and insulin metabolism?—a systematic review. *Diabetic Medicine* **20**, 339-348 (2003).
201. Singhal, A., Cole, T.J., Fewtrell, M., Deanfield, J. & Lucas, A. Is Slower Early Growth Beneficial for Long-Term Cardiovascular Health? *Circulation* **109**, 1108-1113 (2004).
202. Rogers, I. The influence of birthweight and intrauterine environment on adiposity and fat distribution in later life. *Int J Obes Relat Metab Disord* **27**, 755-777 (2003).
203. WHO. Obesity: preventing and managing the global epidemic. in *World Health Organization technical report series* Vol. 894 (2000).
204. Rice, F. & Thapar, A. Estimating the relative contributions of maternal genetic, paternal genetic and intrauterine factors to offspring birth weight and head circumference. *Early Human Development* **86**, 425-432 (2010).
205. Stuebe, A.M., Forman, M.R. & Michels, K.B. Maternal-recalled gestational weight gain, pre-pregnancy body mass index, and obesity in the daughter. *Int J Obes* **33**, 743-752 (2009).
206. Voigt, M., Hermanussen, M., Wittwer-Backofen, U., Fusch, C. & Hesse, V. Sex-specific differences in birth weight due to maternal smoking during pregnancy. *European Journal of Pediatrics* **165**, 757-761 (2006).
207. Ong, K. & Dunger, D. Birth weight, infant growth and insulin resistance. *European Journal of Endocrinology* **151**, U131-U139 (2004).
208. Swamy, G.K., Edwards, S., Gelfand, A., James, S.A. & Miranda, M.L. Maternal age, birth order, and race: differential effects on birthweight. *Journal of Epidemiology and Community Health* **66**, 136-142 (2012).
209. Laitinen, S., Rasanen, L., Viikari, J. & Akerblom, H.K. Diet of Finnish children in relation to the family's socio-economic status. *Scand J Soc Med* **23**, 88-94 (1995).
210. Gasser, T., Muller, H.-G., Kohler, W., Molinari, L. & Prader, A. Nonparametric Regression Analysis of Growth Curves. *The Annals of Statistics* **12**, 210-229 (1984).
211. Dennison, B.A., Edmunds, L.S., Stratton, H.H. & Pruzek, R.M. Rapid infant weight gain predicts childhood overweight. *Obesity (Silver Spring)* **14**, 491-9 (2006).

212. Stettler, N., Zemel, B.S., Kumanyika, S. & Stallings, V.A. Infant weight gain and childhood overweight status in a multicenter, cohort study. *Pediatrics* **109**, 194-9 (2002).
213. Savage, T. *et al.* Increasing Maternal Age Is Associated with Taller Stature and Reduced Abdominal Fat in Their Children. *PLoS ONE* **8**, e58869 (2013).
214. Roberts, R.J., Leary, S.D., Smith, G.D. & Ness, A.R. Maternal age in pregnancy and offspring blood pressure in childhood in the Avon Longitudinal Study of Parents and Children (ALSPAC). *J Hum Hypertens* **19**, 893-900 (2005).
215. Shrout, P.E. & Bolger, N. Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychol Methods* **7**, 422-45 (2002).
216. Parsons, T.J., Power, C. & Manor, O. Fetal and early life growth and body mass index from birth to early adulthood in 1958 British cohort: longitudinal study. *BMJ* **323**, 1331-5 (2001).
217. Emanuel, I., Kimpo, C. & Moceri, V. The association of maternal growth and socio-economic measures with infant birthweight in four ethnic groups. *International Journal of Epidemiology* **33**, 1236-1242 (2004).
218. Ong, K.K., Preece, M.A., Emmett, P.M., Ahmed, M.L. & Dunger, D.B. 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* **52**, 863-7 (2002).
219. Eriksson, J., Forsen, T., Tuomilehto, J., Osmond, C. & Barker, D. Fetal and childhood growth and hypertension in adult life. *Hypertension* **36**, 790-4 (2000).
220. Srinivas, S.K. *et al.* Rethinking IUGR in preeclampsia: dependent or independent of maternal hypertension? *J Perinatol* **29**, 680-4 (2009).
221. Heslehurst, N., Rankin, J., Wilkinson, J.R. & Summerbell, C.D. A nationally representative study of maternal obesity in England, UK: trends in incidence and demographic inequalities in 619 323 births, 1989-2007. *Int J Obes (Lond)* **34**, 420-8 (2010).
222. O'Reilly, J.R. & Reynolds, R.M. The risk of maternal obesity to the long-term health of the offspring. *Clin Endocrinol (Oxf)* **78**, 9-16 (2013).
223. Catalano, P.M. & Ehrenberg, H.M. The short- and long-term implications of maternal obesity on the mother and her offspring. *BJOG* **113**, 1126-33 (2006).
224. Chopra, M., Galbraith, S. & Darnton-Hill, I. A global response to a global problem: the epidemic of overnutrition. *Bull World Health Organ* **80**, 952-8 (2002).
225. Sebire, N.J. *et al.* Maternal obesity and pregnancy outcome: a study of 287,213 pregnancies in London. *Int J Obes Relat Metab Disord* **25**, 1175-82 (2001).
226. Jolly, M.C., Sebire, N.J., Harris, J.P., Regan, L. & Robinson, S. Risk factors for macrosomia and its clinical consequences: a study of 350,311 pregnancies. *Eur J Obstet Gynecol Reprod Biol* **111**, 9-14 (2003).
227. Anzman, S.L., Rollins, B.Y. & Birch, L.L. Parental influence on children's early eating environments and obesity risk: implications for prevention. *Int J Obes* **34**, 1116-1124 (2010).
228. Park, M.H., Falconer, C., Viner, R.M. & Kinra, S. The impact of childhood obesity on morbidity and mortality in adulthood: a systematic review. *Obesity Reviews* **13**, 985-1000 (2012).
229. Gale, C.R. *et al.* Maternal size in pregnancy and body composition in children. *J Clin Endocrinol Metab* **92**, 3904-11 (2007).
230. Koupil, I. & Toivanen, P. Social and early-life determinants of overweight and obesity in 18-year-old Swedish men. *Int J Obes (Lond)* **32**, 73-81 (2008).
231. Mingrone, G. *et al.* Influence of Maternal Obesity on Insulin Sensitivity and Secretion in Offspring. *Diabetes Care* **31**, 1872-1876 (2008).
232. Samuelsson, A.M. *et al.* Diet-induced obesity in female mice leads to offspring hyperphagia, adiposity, hypertension, and insulin resistance: a novel murine model of developmental programming. *Hypertension* **51**, 383-92 (2008).
233. Drake, A.J. & Reynolds, R.M. Impact of maternal obesity on offspring obesity and cardiometabolic disease risk. *Reproduction* **140**, 387-98 (2010).

234. Alfaradhi, M.Z. & Ozanne, S.E. Developmental programming in response to maternal overnutrition. *Front Genet* **2**, 27 (2011).
235. Li, M., Sloboda, D.M. & Vickers, M.H. Maternal obesity and developmental programming of metabolic disorders in offspring: evidence from animal models. *Exp Diabetes Res* **2011**, 592408 (2011).
236. Poston, L. Maternal obesity, gestational weight gain and diet as determinants of offspring long term health. *Best Practice & Research Clinical Endocrinology & Metabolism* **26**, 627-639 (2012).
237. Taylor, P.D. & Poston, L. Developmental programming of obesity in mammals. *Exp Physiol* **92**, 287-98 (2007).
238. Chandler-Laney P.C & N.C., B. Maternal obesity, metabolic health, and prenatal programming of offspring obesity. *The open obesity journal* **3**, 42-50 (2011).
239. McDonald, S.D., Han, Z., Mulla, S. & Beyene, J. Overweight and obesity in mothers and risk of preterm birth and low birth weight infants: systematic review and meta-analyses. *BMJ* **341**(2010).
240. Ayonrinde, O.T. *et al.* Gender-specific differences in adipose distribution and adipocytokines influence adolescent nonalcoholic fatty liver disease. *Hepatology* **53**, 800-809 (2011).
241. Davies, M.J. & Norman, R.J. Programming and reproductive functioning. *Trends in Endocrinology & Metabolism* **13**, 386-392 (2002).
242. Rodriguez, A. Maternal pre-pregnancy obesity and risk for inattention and negative emotionality in children. *Journal of Child Psychology and Psychiatry* **51**, 134-143 (2010).
243. Pirkola, J. *et al.* Risks of overweight and abdominal obesity at age 16 years associated with prenatal exposures to maternal prepregnancy overweight and gestational diabetes mellitus. *Diabetes care* **33**, 1115-1121 (2010).
244. Kennedy, A. *et al.* The metabolic significance of leptin in humans: gender-based differences in relationship to adiposity, insulin sensitivity, and energy expenditure. *Journal of Clinical Endocrinology & Metabolism* **82**, 1293-1300 (1997).
245. Laws, A. *et al.* Differences in Insulin Suppression of Free Fatty Acid Levels by Gender and Glucose Tolerance Status Relation to Plasma Triglyceride and Apolipoprotein B Concentrations. *Arteriosclerosis, thrombosis, and vascular biology* **17**, 64-71 (1997).
246. Ford, S.P. & Tuersunjiang, N. Maternal obesity: how big an impact does it have on offspring prenatally and during postnatal life? *Expert Review of Endocrinology & Metabolism* **8**, 261-273 (2013).
247. Vaarasmaki, M. *et al.* Adolescent manifestations of metabolic syndrome among children born to women with gestational diabetes in a general-population birth cohort. *Am J Epidemiol* **169**, 1209-15 (2009).
248. F., F. & O., S. Telomere length and cardiovascular aging. *Annals of Medicine* **44**, S138-S142 (2012).
249. Saliques, S. *et al.* Telomere length and cardiovascular disease. *Archives of Cardiovascular Diseases* **103**, 454-459 (2010).
250. Bischoff, C. *et al.* No Association Between Telomere Length and Survival Among the Elderly and Oldest Old. *Epidemiology* **17**, 190-194 10.1097/01.ede.0000199436.55248.10 (2006).
251. Cui, Y. *et al.* Associations of leukocyte telomere length with body anthropometric indices and weight change in chinese women. *Obesity*, n/a-n/a (2013).
252. Aviv, A. *et al.* Leukocyte telomere dynamics: longitudinal findings among young adults in the Bogalusa Heart Study. *American journal of epidemiology* **169**, 323-329 (2009).
253. Buxton, J.L. *et al.* Childhood Obesity Is Associated with Shorter Leukocyte Telomere Length. *Journal of Clinical Endocrinology & Metabolism* **96**, 1500-1505 (2011).
254. Reilly, J.J. & Kelly, J. Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: systematic review. *Int J Obes* **35**, 891-898 (2011).
255. Rolland-Cachera, M.F. Rate of growth in early life: a predictor of later health? *Adv Exp Med Biol* **569**, 35-9 (2005).
256. Bergman, R.N. *et al.* A Better Index of Body Adiposity. *Obesity* **19**, 1083-1089 (2011).

257. Cawthon, R.M. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Res* **37**, e21 (2009).
258. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).
259. Gelber, R.P. *et al.* Measures of Obesity and Cardiovascular Risk Among Men and Women. *Journal of the American College of Cardiology* **52**, 605-615 (2008).
260. Sjöström, L., Smith, U., Krotkiewski, M. & Björntorp, P. Cellularity in different regions of adipose tissue in young men and women. *Metabolism* **21**, 1143-1153 (1972).
261. Barrett, E.L.B. & Richardson, D.S. Sex differences in telomeres and lifespan. *Aging Cell* **10**, 913-921 (2011).
262. Mather, K.A., Jorm, A.F., Parslow, R.A. & Christensen, H. Is Telomere Length a Biomarker of Aging? A Review. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences* **66**, 202-213 (2011).
263. Aviv, A. Commentary: Raising the bar on telomere epidemiology. *Int J Epidemiol* **38**, 1735-6 (2009).
264. Aviv, A. *et al.* Menopause modifies the association of leukocyte telomere length with insulin resistance and inflammation. *Journal of Clinical Endocrinology & Metabolism* **91**, 635-640 (2006).
265. Codd, V. Identification of seven loci affecting mean telomere length and their association with disease. *Nature Genetics* **in press**(2013).
266. Nordfjall, K. *et al.* Telomere length is associated with obesity parameters but with a gender difference. *Obesity (Silver Spring)* **16**, 2682-9 (2008).
267. Freedman, D.S. *et al.* The body adiposity index (hip circumference÷ height1. 5) is not a more accurate measure of adiposity than is BMI, waist circumference, or hip circumference. *Obesity* (2012).
268. Jousilahti, P., Vartiainen, E., Tuomilehto, J. & Puska, P. Sex, Age, Cardiovascular Risk Factors, and Coronary Heart Disease: A Prospective Follow-Up Study of 14 786 Middle-Aged Men and Women in Finland. *Circulation* **99**, 1165-1172 (1999).
269. Ahima, R.S. Connecting obesity, aging and diabetes. *Nat Med* **15**, 996-997 (2009).
270. von Zglinicki, T. Oxidative stress shortens telomeres. *Trends in Biochemical Sciences* **27**, 339-344 (2002).
271. Kurz, D.J. *et al.* Chronic oxidative stress compromises telomere integrity and accelerates the onset of senescence in human endothelial cells. *J Cell Sci* **117**, 2417-2426 (2004).
272. Kiesler, J. & Ricer, R. The abnormal fontanel. *American family physician* **67**, 2547 (2003).
273. Rosenberg, A. The IUGR Newborn. *Seminars in Perinatology* **32**, 219-224 (2008).
274. Samuelsen, S.O., Stene, L.C. & Bakketeig, L.S. Association of head circumference at birth among sibling pairs. *Paediatric and Perinatal Epidemiology* **18**, 26-32 (2004).
275. Hack, M. *et al.* Effect of Very Low Birth Weight and Subnormal Head Size on Cognitive Abilities at School Age. *New England Journal of Medicine* **325**, 231-237 (1991).
276. Ivanovic, D.M. *et al.* Head size and intelligence, learning, nutritional status and brain development: Head, IQ, learning, nutrition and brain. *Neuropsychologia* **42**, 1118-1131 (2004).
277. F., G. Hereditary Genius. *Macmillan, London* (1869).
278. Jensen, A.R. Psychometric< i> g</i> related to differences in head size. *Personality and Individual Differences* **17**, 597-606 (1994).
279. Bergvall, N., Iliadou, A., Tuvemo, T. & Cnattingius, S. Birth characteristics and risk of low intellectual performance in early adulthood: are the associations confounded by socioeconomic factors in adolescence or familial effects? *Pediatrics* **117**, 714-721 (2006).
280. Desch, L.W., Anderson, S.K. & Snow, J.H. Relationship of Head Circumference to Measures of School Performance. *Clinical Pediatrics* **29**, 389-392 (1990).

281. Schoenemann, P.T., Budinger, T.F., Sarich, V.M. & Wang, W.S.-Y. Brain size does not predict general cognitive ability within families. *Proceedings of the National Academy of Sciences* **97**, 4932-4937 (2000).
282. Bouthoorn, S.H. *et al.* Head Circumference of Infants Born to Mothers with Different Educational Levels; The Generation R Study. *PLoS ONE* **7**, e39798 (2012).
283. Ivanovic, D.M. *et al.* Nutritional status, brain development and scholastic achievement of Chilean high-school graduates from high and low intellectual quotient and socio-economic status. *British Journal of Nutrition* **87**, 81-92 (2002).
284. Smit, D. *et al.* Heritability of head size in Dutch and Australian twin families at ages 0–50 years. *Twin Research and Human Genetics* **13**, 370-380 (2010).
285. Bouchard Jr, T.J. Genetic and environmental influences on adult intelligence and special mental abilities. *Human Biology* **70**, 257-279 (1998).
286. Williams, C.A., Dagli, A. & Battaglia, A. Genetic disorders associated with macrocephaly. *American Journal of Medical Genetics Part A* **146A**, 2023-2037 (2008).
287. Kaindl, A.M. *et al.* Many roads lead to primary autosomal recessive microcephaly. *Progress in Neurobiology* **90**, 363-383 (2010).
288. Taal, H.R. *et al.* Common variants at 12q15 and 12q24 are associated with infant head circumference. *Nat Genet* **44**, 532-538 (2012).
289. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832-838 (2010).
290. Bruin, J. Introduction to STATA. UCLA: Statistical Consulting Group. (from http://www.ats.ucla.edu/stat/sas/notes2/ (accessed November 24, 2007). 2007).
291. Ikram, M.A. & Fornage M. Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nat Genet* **44**, 539-544 (2012).
292. Suzuki, C. *et al.* Identification of Myc-associated protein with JmjC domain as a novel therapeutic target oncogene for lung cancer. *Mol Cancer Ther* **6**, 542-51 (2007).
293. Coyle-Thompson, C.A. & Banerjee, U. The strawberry notch gene functions with Notch in common developmental pathways. *Development* **119**, 377-395 (1993).
294. Radtke, F. & Raj, K. The role of Notch in tumorigenesis: oncogene or tumour suppressor? *Nat Rev Cancer* **3**, 756-67 (2003).
295. Ehebauer, M., Hayward, P. & Martinez-Arias, A. Notch signaling pathway. *Science Signaling* **2006**, cm7 (2006).
296. Wikipedia. HMGA2. (http://en.wikipedia.org/wiki/HMGA2).
297. Fedele, M., Pierantoni, G.M., Visone, R. & Fusco, A. Critical Role of the HMGA2 Gene in Pituitary Adenomas. *Cell Cycle* **5**, 2045-2048 (2006).
298. Meyer, B. *et al.* HMGA2 overexpression in non-small cell lung cancer. *Molecular Carcinogenesis* **46**, 503-511 (2007).
299. Cattaruzzi, G. *et al.* The second AT-hook of the architectural transcription factor HMGA2 is determinant for nuclear localization and function. *Nucleic Acids Research* **35**, 1751-1760 (2007).
300. Saunders, C.L., Lejarraga, H. & Del Pino, M. Assessment of head size adjusted for height: an anthropometric tool for clinical use based on Argentinian data. *Annals of Human Biology* **33**, 415-423 (2006).
301. Sealfon, S.C., Weinstein, H. & Millar, R.P. Molecular mechanisms of ligand interaction with the gonadotropin-releasing hormone receptor. *Endocrine reviews* **18**, 180-205 (1997).
302. Barker, D.J.P., Osmond, C., Winter, P.D., Margetts, B. & Simmonds, S.J. WEIGHT IN INFANCY AND DEATH FROM ISCHAEMIC HEART DISEASE. *The Lancet* **334**, 577-580 (1989).
303. Sloboda, D.M., Hart, R., Doherty, D.A., Pennell, C.E. & Hickey, M. Age at Menarche: Influences of Prenatal and Postnatal Growth. *Journal of Clinical Endocrinology & Metabolism* **92**, 46-50 (2007).
304. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research* **42**, D1001-D1006 (2014).

305. Higgins, J.P.T., Thompson, S.G., Deeks, J.J. & Altman, D.G. Measuring inconsistency in meta-analyses. *BMJ* **327**, 557-560 (2003).
306. Friedlander, Y. *et al.* Candidate Molecular Pathway Genes Related to Appetite Regulatory Neural Network, Adipocyte Homeostasis and Obesity: Results from the CARDIA Study. *Annals of Human Genetics* **74**, 387-398 (2010).
307. Gallicchio, L. *et al.* Single nucleotide polymorphisms in obesity-related genes and all-cause and cause-specific mortality: a prospective cohort study. *BMC Medical Genetics* **10**, 103 (2009).
308. Mammès, O. *et al.* LEPR gene polymorphisms: associations with overweight, fat mass and response to diet in women. *European Journal of Clinical Investigation* **31**, 398-404 (2001).
309. Wen, W. *et al.* Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat Genet* **44**, 307-311 (2012).
310. Heni, M. *et al.* Association of obesity risk SNPs in PCSK1 with insulin sensitivity and proinsulin conversion. *BMC Medical Genetics* **11**(2010).
311. Bouchard, C. Childhood obesity: are genetic differences involved? *The American Journal of Clinical Nutrition* **89**, 1494S-1501S (2009).
312. Zhang, Y.-Y. *et al.* Genetic variability at the leptin receptor (LEPR) locus is a determinant of plasma fibrinogen and C-reactive protein levels. *Atherosclerosis* **191**, 121-127 (2007).
313. Kassem, A., Soliman, A. & Yasin, M. *Leptin in pediatrics: A hormone from adipocyte that wheels several functions in children*, 577-587 (2012).
314. Shalitin, S. & Phillip, M. Role of obesity and leptin in the pubertal process and pubertal growth[mdash]a review. *Int J Obes Relat Metab Disord* **27**, 869-874 (2003).
315. Tabassum, R. *et al.* Common Variants of IL6, LEPR, and PBEF1 Are Associated With Obesity in Indian Children. *Diabetes* **61**, 626-631 (2012).
316. Enns, J.E., Taylor, C.G. & Zahradka, P. Variations in Adipokine Genes AdipoQ, Lep, and LepR Are Associated with Risk for Obesity-Related Metabolic Disease: The Modulatory Role of Gene-Nutrient Interactions. *Journal of Obesity* **2011**(2011).
317. Wheeler, E. *et al.* Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. *Nat Genet* **45**, 513-517 (2013).
318. Morton, G.J., Cummings, D.E., Baskin, D.G., Barsh, G.S. & Schwartz, M.W. Central nervous system control of food intake and body weight. *Nature* **443**, 289-295 (2006).
319. Shalitin, S. & Phillip, M. Role of obesity and leptin in the pubertal process and pubertal growth[mdash]a review. *Int J Obes Relat Metab Disord* **27**, 869-874 (0000).
320. Williamson, J.A. *et al.* Chromosomal Mapping of the Human and Mouse Homologues of Two New Members of the AP-2 Family of Transcription Factors. *Genomics* **35**, 262-264 (1996).
321. Arreola, R., Valderrama, B., Morante, M.L. & Horjales, E. Two mammalian glucosamine-6-phosphate deaminases: a structural and genetic study. *FEBS Letters* **551**, 63-70 (2003).
322. Maeda, S. *et al.* Genetic variations in the gene encoding TFAP2B are associated with type 2 diabetes mellitus. *J Hum Genet* **50**, 283-92 (2005).
323. Mumby, H.S. *et al.* Mendelian Randomisation Study of Childhood BMI and Early Menarche. *J Obes* **2011**, 180729 (2011).
324. Lasky-Su, J. On the replication of genetic associations: timing can be everything! *Am. J. Hum. Genet.* **82**, 849-858 (2008).
325. Sovio, U. *et al.* Genetic Determinants of Height Growth Assessed Longitudinally from Infancy to Adulthood in the Northern Finland Birth Cohort 1966. *PLoS Genet* **5**, e1000409 (2009).
326. Eriksson, J.G. Early growth and coronary heart disease and type 2 diabetes: findings from the Helsinki Birth Cohort Study (HBCS). *Am J Clin Nutr* **94**, 1799S-1802S (2011).
327. Ludlow, A.T. & Roth, S.M. Physical activity and telomere biology: exploring the link with aging-related disease prevention. *J Aging Res* **2011**, 790378 (2011).
328. Cole, T.J., Donaldson, M.D. & Ben-Shlomo, Y. SITAR—a useful instrument for growth curve analysis. *International journal of epidemiology* **39**, 1558-1566 (2010).

329. Imai, K., Keele, L. & Yamamoto, T. Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science* **25**, 51-71 (2010).
330. Harding, J. The nutritional basis of the fetal origins of adult disease. *International Journal of Epidemiology* **30**, 15-23 (2001).
331. Barker, D.J. Maternal nutrition, fetal nutrition, and disease in later life. *Nutrition* **13**, 807-813 (1997).
332. Hart, N. Famine, maternal nutrition and infant mortality: a re-examination of the Dutch Hunger Winter. *Population Studies* **47**, 27-46 (1993).
333. Gluckman, P.D. & Hanson, M.A. Maternal constraint of fetal growth and its consequences. in *Seminars in fetal and neonatal medicine* Vol. 9 419-425 (Elsevier, 2004).
334. Thomas, D.C., Lewinger, J.P., Murcray, C.E. & Gauderman, W.J. Invited Commentary: GE-Whiz! Ratcheting Gene-Environment Studies up to the Whole Genome and the Whole Exposome. *American Journal of Epidemiology* **175**, 203-207 (2012).
335. Hackshaw, A., Rodeck, C. & Boniface, S. Maternal smoking in pregnancy and birth defects: a systematic review based on 173 687 malformed cases and 11.7 million controls. *Hum Reprod Update* **17**, 589-604 (2011).
336. Berrettini, W. *et al.* Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol Psychiatry* **13**, 368-73 (2008).
337. Thorgeirsson, T.E. *et al.* Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet* **42**, 448-53 (2010).
338. Thorgeirsson, T.E. *et al.* A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452**, 638-42 (2008).
339. Ducci, F. *et al.* TTC12-ANKK1-DRD2 and CHRNA5-CHRNA3-CHRNB4 Influence Different Pathways Leading to Smoking Behavior from Adolescence to Mid-Adulthood. *Biological Psychiatry* **69**, 650-660 (2011).
340. Freathy, R.M. *et al.* A common genetic variant in the 15q24 nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNB4) is associated with a reduced ability of women to quit smoking in pregnancy. *Hum Mol Genet* **18**, 2922-7 (2009).
341. Manning, A.K. *et al.* Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP× environment regression coefficients. *Genetic epidemiology* **35**, 11-18 (2011).
342. Azziz, R. Diagnosis of polycystic ovarian syndrome: The Rotterdam criteria are premature. *Journal of Clinical Endocrinology & Metabolism* **91**, 781-785 (2006).
343. Pasquali, R., Gambineri, A. & Pagotto, U. Review article: The impact of obesity on reproduction in women with polycystic ovary syndrome. *BJOG: An International Journal of Obstetrics & Gynaecology* **113**, 1148-1159 (2006).
344. Abbott, D., Dumesic, D. & Franks, S. Developmental origin of polycystic ovary syndrome - a hypothesis. *Journal of Endocrinology* **174**, 1-5 (2002).
345. Eisner, J.R., Dumesic, D.A., Kemnitz, J.W. & Abbott, D.H. Timing of Prenatal Androgen Excess Determines Differential Impairment in Insulin Secretion and Action in Adult Female Rhesus Monkeys. *Journal of Clinical Endocrinology & Metabolism* **85**, 1206-1210 (2000).
346. Robinson, J.E., Forsdike, R.A. & Taylor, J.A. In Utero Exposure of Female Lambs to Testosterone Reduces the Sensitivity of the GnRH Neuronal Network to Inhibition by Progesterone. *Endocrinology* **140**, 5797-5805 (1999).
347. group, T.R.E.A.s.P.c.w. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Human Reproduction* **19**, 41-47 (2004).
348. Ashawesh, K., Aghilla, M. & Randeva, H. Exercise increases plasma adiponectin and bound leptin, and decreases Plasma free leptin in Overweight Young Women with Polycystic Ovary Syndrome. (2012).

349. Oken, E., Taveras, E.M., Kleinman, K.P., Rich-Edwards, J.W. & Gillman, M.W. Gestational weight gain and child adiposity at age 3 years. *American Journal of Obstetrics and Gynecology* **196**, 322.e1-322.e8 (2007).
350. Lawlor, D.A. *et al.* Maternal and offspring adiposity-related genetic variants and gestational weight gain. *The American Journal of Clinical Nutrition* **94**, 149-155 (2011).
351. Gieger, C. *et al.* Genetics Meets Metabolomics: A Genome-Wide Association Study of Metabolite Profiles in Human Serum. *PLoS Genet* **4**, e1000282 (2008).
352. Jaenishch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics* **33**, 245-254 (2003).
353. Kangaspeska, S. *et al.* Transient cyclical methylation of promoter DNA. *Nature* **452**, 112-115 (2008).
354. Bouatia-Naji, N. *et al.* Smallness for gestational age interacts with high mobility group A2 gene genetic variation to modulate height. *European Journal of Endocrinology* **160**, 557-560 (2009).
355. Pillas, D. *et al.* Genome-Wide Association Study Reveals Multiple Loci Associated with Primary Tooth Development during Infancy. *PLoS Genet* **6**, e1000856 (2010).
356. Consortium, T.I.H. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-58 (2010).

Appendix

1 Appendix note 1. Reclassification of maternal socio-economic status (SES) in NFBC1986

Maternal SES was re-classified into eight categories according to the occupation of the mother and father. It is well known that the socio-economic status (SES) according to occupation is a good reference to detect the general trends of difference in health. Since only few studies focus on related variables such as size of dwelling or number of persons per room, this study will exclude the related variables for current analysis. In the original status, maternal SES is classified based on maternal and paternal occupation as:

1. Class I – Professionals (leading position)
2. Class II – Professionals (Small entrepreneurs)
3) Class III – Skilled worker
4) Class IV (Unskilled worker)
5) Farmer >=8
6) Farmer <8
7) Farmer's wife, size of the farm >=8 ha
8) Farmer's wife, size of the farm <8 ha

The classification 5 and 6 are derived from paternal occupation and was empty for this dataset. In future, this study would like to see the consequences for missing dataset. The distribution of frequencies for the maternal SES is shown in the table 3a and for analysis purposes was further grouped as shown in Table 3b.

1. Class I and II – Professionals
2. Class III and IV – Skilled and unskilled workers
3. Class V, VI, VII and VIII – Farmer

2 Appendix Note 2. MSc Dissertation

Abstract

Background and aims: The aim of the study is to develop height growth models for prenatal/postnatal data to explore association between demographic, biological and socio-economic determinants of fetal and postnatal growth velocity (peak height velocity, PHV). This study explores applicability of existing growth models so that they could be later tested to study relationship between infant growth and health phenotypes.

Methods: This study explored data on 9203 singleton individuals born in Northern Finland from 1 July 1985 until 30 June 1986. A total of 5644 infants with full growth data were included for growth modeling and to analyse the relationship between prenatal biological and socio-demographic factors and PVH from birth until 24 months of age. Three different models were tested, one for prenatal growth (Rossavik model) and two for postnatal growth (Count and Reed models).

Results: Reed 1 model was selected to develop postnatal height growth curves of boys and girls separately and showed good fit into the data (Goodness of fit, $P<0.0001$). PVH was derived from the model as its first derivatives and was used as an outcome. Rossavik model did not fit well into the prenatal data; will need further development.

Conclusion: It was observed that Reed 1 model was well fitted to the data. The association analyses showed that many potential determinants of early growth associated independently of other factors with PHV. It was slightly unexpected that many factors, particularly those related to size at birth, had such a strong negative association with PHV. Although this provides a first of its kind of study and a better understanding of association between determinants of fetal growth and growth velocity in infants, further work is needed to understand more complex relationship between early life course factors, potential markers of adult diseases, and growth velocity.

Introduction

Growth

Growth has been defined as increase in size or mass of tissue, primarily resulting at the cellular level (Thomson et al, 1971). This is a complex process which takes almost 20 years to complete and believed to be regulated through multiple endocrinological pathways. Growth descends from the combination of the genetic factors, the action of environment factors, and the interaction between the two. Silventoinen *et al*, 2006, show that 90% of variation of height in childhood and adulthood is explained by genetic factors. An individual obtains a genetic blueprint at conception that includes potential to achieve a particular size and shape, however, environment can alters this potential and reduce our ability to achieve them.

Growth is majorly controlled by the hormones secreted by the endocrine system. The key hormones which play important function are: sex hormones (estradiol and testosterone), thyroid hormone, growth hormone, insulin-like growth factor-1(IGF1) and insulin. Sex hormones promote growth in height in childhood and early puberty and accelerate skeletal maturation and epiphyseal closure in late puberty resulting in lengthening of the long bones. Thyroid hormones are necessary for normal growth in early infancy and they stimulate osteoblast maturation and acceleration ossification by increasing the levels of growth hormone and IGF1. Growth hormone is the most important determinant of linear growth and it mainly acts by increasing the proliferation of chondrocytes directly and in concert with IGF1. Growth hormone increases the production of insulin, IGF1 and sex hormones which in turn reduce growth hormone release by a negative feedback system (Noel *et al*, 2004).

The process of human growth is not uniform throughout life. Each region and part of body can be shown to have its own growth rate and this can vary according to age. Human growth occurs in spurts and can be divided into infancy, childhood and pubertal growth. Lampl et al,1992, demonstrates from his work done in human growth curve, that human growth may not be a continuous process but actually occur in short bursts of activity, called “saltation”, that punctuate periods of no growth, called “stasis”. Tanner *et al*, 1990*,* described the term *“*Growth velocity” or “height velocity” as the pattern created by changing rate of growth from the continuous distance curve of varying magnitudes. In recent times, velocity curve has become the preferred tool of studying individual growth pattern as it shows the growth in a clear and intuitive way.

Dimensions of growth

Growth can be measured in terms of changes in one of the many anthropometric variables. However, Dietz *et al*, 1998, have shown that only height and weight are generally routinely measured in the clinical setting and included in medical records. This study will focus on height as the dimension of growth to advance the future studies on the role of growth velocity in diseases later in life. Height has also been shown as an excellent indicator of general child health and nutritional status which has been found to inversely associate with initially with total mortality, coronary heart disease, type 2 diabetes and cancer (Rolland-Cachera, 1995).

Growth pattern in fetus and infancy

According to "fetal hypothesis", it is important to understand the fetal development and find patterns that can influence adverse effects in later life. Tanner, 1978, was the first to illustrated fetal growth in crown-heel length peaks at about 20 week gestation. Fetal growth is generally considered in terms of birthweight; therefore, there is considerably little information available on growth patterns in terms of size. Fetal length relies largely on two sources of information: extra uterine anthropometric measurements and intrauterine ultrasound measurement of fetus, therefore, in the study, it uses the details on the extra uterine anthropometric measurements of fetus as a marker for fetal length.

The growth pattern in fetus follows a slope of decreasing velocity between 20^{th} and 30^{th} weeks of gestation, however, the rate at which the length of the fetus increases reaches a peak at approximately 120 cm/yr-1. Nevertheless, the only element that can be observed postnatal is the slope of decreasing velocity lasting until about 4 years of age. The velocity of in utero linear growth is maximal at about 18 wk gestational age in the human. At this time, the fetus grows four times more rapidly than at any time postnatal but the rate declines sharply during the last weeks of gestation.

Infancy is a high velocity, rapidly decelerating, nutrition-dependent phase of growth. Cole, 2000, defined that adult height trends matches that at age 2 years, so that the increment in adult height has already been achieved by age 2 years. Hence Infancy in this study is defined as the period between

birth and 24 months of age. Infants grow very fast during the first year of life, at approximately 25cm/yr; and surging the first half of this year, velocity may be even faster, around 30 cm/yr. However, growth does not occur at a regular rate and is episodic during infancy, where increase in length occurs in a few days which can be followed by quiet long periods of no growth. At rather step and continuous deceleration can be observed from birth up to the third year. Thereafter, there is a much milder decay in velocity during school years before the adolescent growth spurt which is very weak compared to fetal and infancy growth. Growth during infancy is thought to be especially important because of high growth velocity and sensitivity to external factors in this age range (Waterlow, 1988; Forman et al, 1990).

Why model growth?

Hauspie *et al*, 2004, stated that for describing growth patterns, it is important to use mathematical modeling of the growth curve as they describe, summarize and quantify features that are usually based on longitudinal data for individuals. There are multiple reasons for studying human growth, but two main areas of interest can be distinguished. First, one is interested in obtaining age-dependent reference curves for various somatic variables, usually based on cross-sectional data. The reference curve will be helpful in clinical setting to obtain comparable values at different ages. The second motivation is get a scientific understanding of the growth process over certain phases, such as, birth to adulthood. By fitting the growth model, one can also gain insight into functional relationship between the physical aspect of growth and the parameters provide e.g. biochemical concentrations, maturity indices etc.

Although there are many mathematical ways to modeling the human growth curve, one particular model by Karlberg *et al*, 1989 is particularly relevant to the purpose of explaining biological factors. Karlberg, 1989, designed the Infancy-Childhood-Puberty (ICP) model, which reflects, the main physiological features of the growth process and is compatible with the underlying endocrine and biological influences (Figure1). This model defines infancy as a high growth velocity immediately after birth and rapid decelerating until about 3 years of age. It uses polynomial functions to describe the height velocity and the following gradual decelerations. This is followed by childhood, a period of lower, slowly decelerating velocity which lasts until the onset of puberty,

although a slight increase in velocity, referred to as the mid-growth spurt, occurs between age 6 and 8 years in many children. This is defined by a second degree polynomial function that describes height velocity as following a gradual deceleration similar to infancy. The puberty component was modelled with a logistic expression describing additional growth induced by pubertal hormones which produce acceleration up to peak height velocity (PHV) and then a deceleration until end of growth.

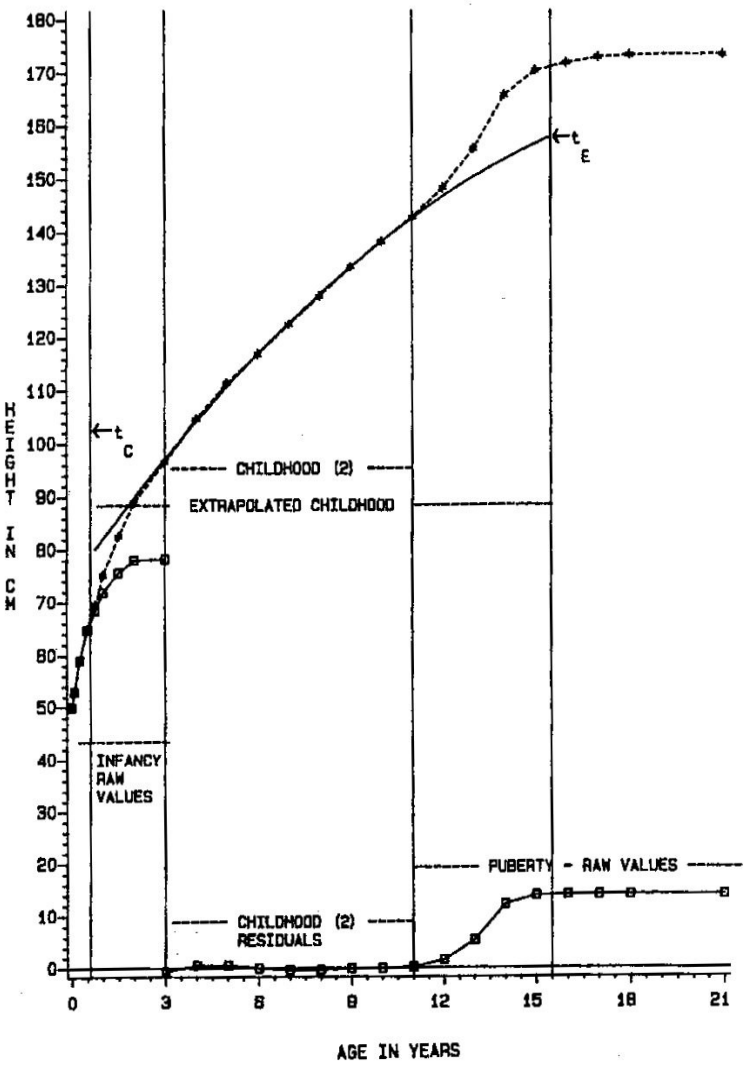


Figure 1: Showing the three phases of postnatal growth, according to Karlberg et al, 1989. The growth curve is distinguished into an infancy cycle (beginning from birth and ending before 1.5 to 2 years), the childhood cycle and pubertal phase. The first of these "growth spurts" is called the juvenile or mid-growth spurt and second is called the adolescent growth spurt (Karlberg et al, 1989). Each phase is separately modelled and the figure presents the raw values for each growth curve.

Methods

Study population and design

The NFBC 1986 study was originally set up to enable the study of the epidemiology of disease, risk factors for pregnancy outcomes and school achievements later in life. All expectant mothers of the Northern provinces of Oulu and Lapland, whose expected date of delivery was between the 1 July 1985 and 30 June 1986, were eligible for the study. In total there were 9362 women, which made up 99% of all births in the study area, with 9479 live and 47 still births, of these, 9203 were singletons births.

In this study recruitment and data collection began in the 12th week of pregnancy when the mothers visited their municipal maternity health centre, continuing after birth with main data collection at 7, 8 and 16 years. Information on health and socio-economic situation was collected through specially designed self-completed questionnaires filled out by the expectant mothers. The questionnaires being handed out at the first antenatal visit were returned by the 24th gestational week. All expectant women had maternity cards containing part of their medical records during pregnancy, filled in and updated by midwives on each visit during pregnancy, these and hospital records were used to extract obstetrical and birth data. Birth data were collected from the records after each delivery. Permission was taken from the children and their parents during the follow-up at 16 years.

In the present study we excluded all multiple births (n=232), still birth (n=47) and those participants who had withheld consent of the study group. In total there were 9203 individuals for the final dataset of which 4726 were boys and 4477 were girls. This study restricted the analysis to birth to 24 months of age, hence final analysis, selects 5466 infants, including 2799 boys and 2865 girls. Appendix A describes the nature of the dataset.

Models

Growth modeling has been seen to a useful method for summarizing childhood growth data and derives growth features of interest for further analysis. Table 2, represents previous models used to describe early growth (0-24 months) data. Parametric growth curve were fitted separately for boys and girls using nonlinear random effects models (Lindstrom & Bates, 1990). The best-fitting curves for each individual were estimated by maximum likelihood method. Height growth velocity curves were obtained as first derivatives of the height growth curves. Velocity curves were then used to

draw biologically meaningful parameters, such as, PHV. The NLME function in R program (version 2.7.0) was used to fit the model.

Assumptions for non-linear regression were check for each model. The Gaussian distribution of the residuals was checked by scatter plot of height in infants, homoscedasticity was verified graphically by the inspection of residuals versus the predicted values and independence of residuals.

The present study fitted three growth models - Rossavik model (Rossavik et al, 1984) for fetal growth data, Count model and Reed 1 model for birth to infant data(age 0-24 months). The models were selected for better fit to the data in terms of Akaike Information Criterion (AIC) and Bayesian Information Criterion or Schwarz Criterion (BIC) (Akaike, 1974; Schwarz, 1978). Both AIC and BIC are closely related to model selection by penalising for additional parameters. AIC gives an asymptotic unbiased estimate of the expected log likelihood and BIC uses Bayesian arguments to measures the efficiency of the parameterized model in terms of predicting the data. For the purpose of better fit age were expressed in months for both Count and Reed 1 model.

Table 2: Summary of model developed specifically for growth in infants.

Model	Year	Dimension(s) modelled	Parameters	Growth model
Jenss-Bayley	1937	Height, Weight	Four parameter model	Birth-6 years
Count A-curve	1943	Height, Weight	Three parameter model	Birth-6 years
Berkley-Reed	1973	Weight, Head circumference	Four parameter model	Birth-6 years
Preece-Baines	1978	Height	Five parameter model	2 year-final adult height
Bock-Thissen	1980	Height	Six parameter triple logistic model	1 year-final adult height
Karlberg	1989	Height	Three parameter model	Birth-final adult height
JPA-2	1992	Height	Eight parameter model	Birth-final adult height

The Rossavik model was chosen as it shows a good fit to fetal data as previously described by Simon et al, 1989. Berkley, 1982, compared the reliability, efficiency, precision and goodness-of-fit of the Count and Jenss-Bayley models and concluded that Jenss-Bayley model fitted the growth data better than Count model, especially prior to 1 year of age. However, for the purpose of this study, the aim was to compare a model on growth data which provides good fit for birth to 24 month of data and Count model has been applied several times by previous study. Another model was selected on the basis of previous work by Simondon et al, 1992, who compared five growth models to fit weight data between birth and 13 months of age and concluded that Reed 1 model shows a better fit to the early height growth data than Kouchi, Karlberg or Jenss-Bayley model.

Rossavik Model for Prenatal data

Rossavik et al, 1984, developed a growth model capable of describing the individual growth curve standards for various fetal parameters using the data of ultrasound evaluations obtained between 15 weeks to 30 weeks of menstrual age. The basis of Rossavik model uses polynomials, which are frequently used to describe fetal growth. The general form of polynomials used to describe fetal growth is:

$$P = a_0 + a_1 (t) + a2 (t)^2 + \ldots + an (t)^n$$

Rossavik model uses this polynomial equation to fit a linear parameter growth curves with three-dimensional parameter. This three dimensional parameter has been found to follow sigmoid-shaped curves similar to that for birthweight (Brenner et al, 1976). The general form of the Rossavik equation is:

$$P = c (t)^{k+st}$$

Where P stands for the ultrasound parameter, *k* a fixed coefficient determined by the anatomical characteristics of the parameter, c is related to genetic regulators of growth and *s* represents an unknown regulatory system that modifies genetically determined growth and t is the duration of growth of the parameters. Rossavik further defined a logarithmic form for his equation which can be easily estimated using least square fitting procedure and is linear function:

$$Log\ P = log\ (c) + k\ log\ (t) + s\ t\ log\ (t)$$

The current study uses the log form to fit fetal height growth derived from the Symphysis-fundus (SF) measurements of fetus during pregnancy for the NFBC 1986 (NFBC did not have any more accurate measurement for fetal height). The SF distance is generally regarded as an acceptable screening instrument for antenatal detection of intrauterine growth retardation (IUGR), however, its validity as a screening instrument of IUGR still need to be investigated further (Steingrimsodottir et al, 1995).

Count Model for Infancy

The Count model (Count et al, 1943) is a linear model with three parameters which has been described to fit early childhood data fairly well:

$$Y = A + Bt + C\ ln\ (t)$$

Where Y is size, t is age, and A, B and C are the three functional parameters. The model is fitted by standard least square regression methods and is linear in its parameters. It was proposed for modeling anthropometric variables such as weight, height, and head circumference (Count et al, 1943).

Livshits et al, 2000, proposed a slightly modified version of Count Model in order to allow the inclusion of birth data:

$$Y = A + Bt + C\ ln\ (t+1)$$

Where Y is size, t is age, and A, B and C are the functional parameters. This model has been previously been applied to infancy weight growth in developing countries (Kim and Pollitt, 1987; Wohlleb et al, 1983).

Reed 1 model for Infancy

The Reed 1 model (Berkley & Reed, 1989) is a four parameter extension of the three parameter Count model and its functional form is:

$$Y = A + Bt + Cln\ (t) + D/t$$

Since this model is not defined at birth (t = 0), it was modified for this study to include birth (t = 0 + 1).

$$Y = A + Bt + Cln\ (t + 1) + D/\ (t + 1)$$

where *Y* represents the height reached at age *t* and *A, B, C* and *D* are the function parameters and *t* is the postnatal age. Of the functional parameters, A is related to the baseline height at birth, B to the linear component of the growth velocity, C to the decrease in the growth velocity over time, and D to the inflection point that allows growth velocity to peak after birth rather than exactly at birth. The Reed 1 model is linear in its constants and can accommodate one or more inflection points(depending on the number of reciprocal terms), allowing the description of one or more periods of growth acceleration and thus fitting a wider variety of both normal and abnormal growth patterns in early childhood.

Correlation between all variables (independent and dependent) selected for the study

The spearman correlation coefficient was used to check the collinearity between the variables as it does not require the population to have a normal distribution, nor that the data is distributed with equal intervals - assumptions that need to be fulfilled when using Pearson correlation coefficients. It does however require that the variables are continuous. The correlation coefficient ranges from +1 to -1, where a score of ±1 shows a prefect positive or negative linear relationship and a score of zero shows no association between the variables in question. Scatter plots were also used to look for possible linear or non linear relationships between dependent and independent variables (As explained in Appendix C). Variables whose correlation was not shown to be statistically significant were still included in our analysis if there was a biological plausibility or graphically a non-linear relationship was seen.

The chi-square test ($\mathbf{X^2}$) was performed to examine association between those covariates chosen and outcome variables, that were both categorical in nature. Collinearity occurs when two variables are a near linear combination of one another. When this occurs in the process of performing linear regression, the predicted outcome will not be affected, the coefficients of the individual variables will, and therefore the contribution of an independent variable on the model will not be correctly

measured. The variable that causes Collinearity can be left outside for the model or some interaction between the variables can be set. The independent variables, birthweight, birth length, infant's BMI, gestational age, maternal age and maternal BMI were looked at.

Univariate and Multivariate Analysis

Building the model

The statistical analysis begins with an exploration of univariate association between outcome and each exposure variable independently. Using the linear regression command LM in R univariate linear regression analyses was first performed with PHV as the only outcome variable and birthweight, birth length, infant's BMI, maternal age, marital status, education, SES, height, weight and BMI (details in Appendix C). The SUMMARY command was used to examine the distribution of result between the outcome and exposure groups and scatter plot are given to check the association graphically. This study tested the association between each birth measurement (birthweight, birth length and infant's BMI) separately; however, birthweight was excluded from the final analysis as it has high correlation with BMI and birth length.

The next step was to understand factors which influence the association most between PHV and all birth length, while including all other variables as confounders (Appendix D). When categorical variables were used, the most common or the lowest measurement was marked as baseline for comparison. The final step was to compute a stepwise multivariate regression analysis, by analysing effect of each independent variable on the outcome variable (PHV) by adding one by one in the model. The objective of the multiple regression analysis was to predict the change in the outcome variable in response to change in the independent variable when added in succession.

Testing the model

Once the exposure variable to be included were finalised, the final model was tested using standard procedure. The results were interpreted by looking at the point estimates using their P-values and confidence intervals to check the level of significance. Significance was considered if the P-value was <0.05 and the 95% confidence interval did not cross the zero. The F-test was also looked at to examine the significance of the model itself, here also a value of <0.05 was considered significant. Model selection was based on significant improvement in the adjusted R-square for our models. All

models were then checked for independence, linearity, homoscedasticity and normality, which are assumptions for regression analysis. Normality was checked using QQ-plots and to check whether there is any notable homoscendesticy, we plotted residuals of the model against each predictor variable.

Result

The current study consists of 9302 singleton births with approximately 51% (n=4726) boys and 49% (n=4477) girls and utilises the data collected for 99% (n=9431) live births in the original sample. For the purpose of the current final analysis, the postnatal data was restricted on the period from birth to 24 months of age (n=5644), of which, boys (n=2779) and girls (n=2865).

Modeling of prenatal and infant growth

Both Count and Reed model were fitted for the data. For Rossavik model concerning prenatal period each individual with at least four SF measurements four weeks apart were chosen for analysis. In spite of the original aim, Rossavik model could not be fitted by the non-linear procedure, as the extremely high parameter set for this procedure issued error correlations and aborted the fitting procedure.

Table 3 presents the mean, median and standard deviation of residual variances for each growth model and parameter values obtained from the final fitted model (height units are in cm and age units are in months). Only paired comparisons for boys and girls were carried out because the aim was to rank the models according to their goodness of fit. The closest fit for infancy growth was provided by Reed 1 model in comparison to Count model for both boys and girls with statistically significant results for the goodness of fit ($P < 0.0001$). The Residual Standard Deviation (RSD) for Reed 1 model was 0.90 in boys and 0.83 in girls. In Count model the RSD was 0.93 for both boys and 0.86 in girls. The RSD shows the goodness of fit measure, lower values of RSD in Reed 1 model show that the fit to the data is closer. Both models are fairly robust towards the choice of starting values. The initial parameter values for the models were altered to test for absolute true minimum, the parameter for Count model were changed to a=50, b=0.4 and c=9.5 and the parameter in Reed 1 model were changed to a=48, b=0.5, c = 3.4 and d = -0.09, nevertheless, the fitted parameter values, AIC and BIC remained unchanged for both the models.

Count model shows a better RSD, but on examining the AIC and BIC for models, Reed 1 model gives a better fit for both boys and girls. Figure 2a, 2b presents predicted response trajectories and observed height curve with standard error for Count model and Reed 1 model separately for boys and girls. Peak height velocities estimated for all infants from the models separately for boys and girls are presented in Table 4. The three-parameter Count model had decreasing velocities throughout infancy, because no inflexion point is allowed by these models. The mean age at inflexion of the Reed curves was 0.7 months of age.

Table 3: Functional parameter values, Residual Standard deviation (RSD), Akaike Information Criterion(AIC) and Bayesian Information Criterion (BIC) for the Count model and Reed 1 model on infant height data using numeric minimization algorithm technique. Count model: Y=A + Bt + C ln(t); Reed 1 model: Y=A + Bt + C ln(t) + D/t.

	Males (n=2779)						Females (n = 2865)					
	Mean	SD	Standardized Residual		Within-Group	AIC	BIC	Mean	SD	Standardized Residual		Within-Group	AIC	BIC
			Min	Median	Max					Min	Median	Max		
Count						86057.2	86114.31						86994.94	87052.41
A*	48.61	2.16						48.07	2.06					
B*	0.46	0.09						0.58	0.10					
C*	8.70	0.75						7.68	0.70					
Residual		0.93	-27.36	-0.003	19.23				0.86	-11.14	0.001	17.16		
Reed						84781.0	84854.44						85677.6	85751.5
A**	46.37	1.73						44.19	1.75					
B**	0.39	0.09						0.45	0.10					
C**	9.86	0.51						9.65	0.51					
D**	3.14	3.58						5.56	2.99					
Residual		0.90	-28.36	0.003	19.71				0.83	-11.23	0.004	17.81		

Count model : A*- baseline height; B*- linear component of growth velocity; C*- decrease in growth velocity

Reed 1 model: A**-baseline height;B**- linear component of growth velocity;C**- decrease in growth velocity;D**- inflection point

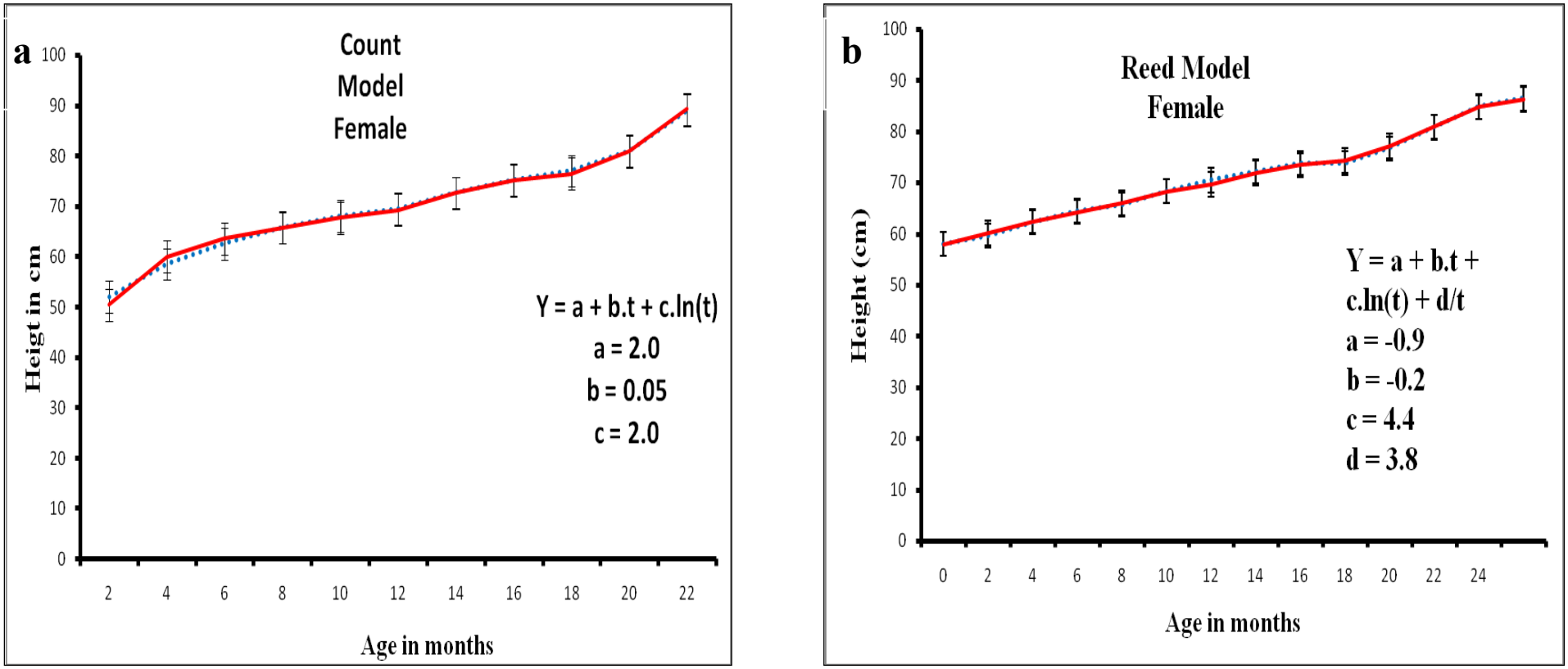


Figure 2a: Predicted (solid red curve) and observed (blue circles) response trajectories for 2865 girl infants with standard error bar for (a) Count model and (b) Reed 1 model. Note that the data shown is from birth to 24 months of age.

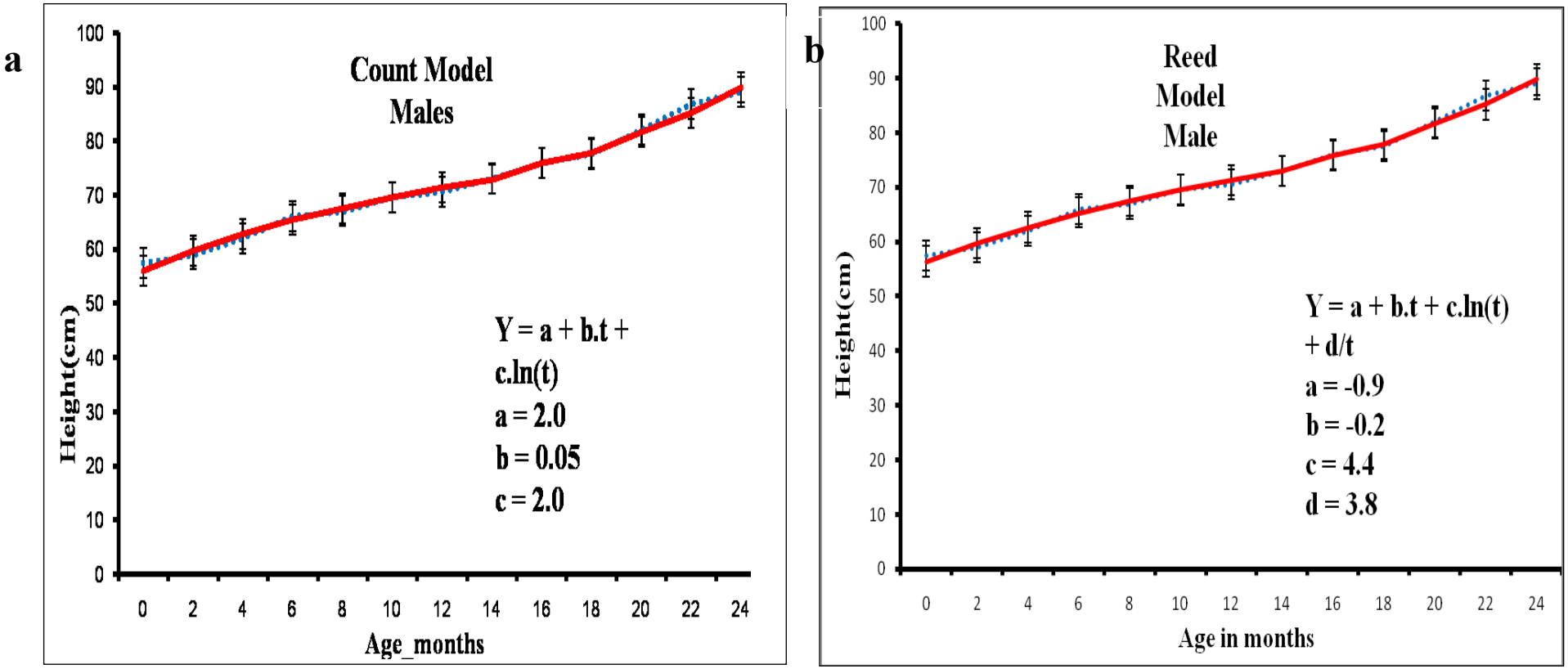


Figure 2b: Predicted(solid red curve) and observed(blue circles) response trajectories for 2799 boys with standard error bars for (a) Count model and (b) Reed 1 model. Note that the data shown is from birth to 24 months of age.

Phenotype (Outcome) Variable estimation and definition

The outcome of interest was PHV for boys and girls derived from the growth models. Table 4 shows the distribution and mean of outcome variable, PHV, in the study population from Count and Reed models. This study uses the PHV from Reed 1 model as it has a better fit to the data. Between boys and girls there was a significant difference in mean of 2.26 cm/age in months in the PHV. In boys, the PHV ranges from a minimum of 1.1 cm/month to a maximum of 25.31 cm/month and from 1.1 cm/month to a maximum of 17.0 cm/month for girls(from Reed 1 model). Figure 2 shows the distribution of PHV in boys and girls. PHV data was analysed separately for boys and girls to assess whether untransformed or log-transformed data gave a better fit to a Gaussian distribution (Figure 3). Means, medians, SD values, and indices of skewness and kurtosis were compared with transformed and untransformed data (Table 4). After log-transformation, data from boys and girls in infants were compared with unpaired t-tests. Compared to the PHV values from NFBC 1966(Sovio et al, 2009) these values were higher than expected and will need further review later because this data was newly used for these analyses. However, slight differences should not have major effect in the association analyses between prenatal/perinatal factor and infant growth.

Discussion

Growth models

This study utilises data from NFBC 1986 to study the growth patterns in prenatal and postnatal growth in infants. It uses longitudinal height measurements with average 14 measurements per person obtained from health clinic records from the original cohort and chooses standard parametric approach to model growth data. Fitting parametric models to growth data gives the advantage of natural biological interpretability from the obtained parameters and it appears to fit well with this data as well. For fetal data, the study uses Rossavik model to fit height growth curves using SF-measurements with average of 4 measurements per person obtained from ultrasound examination of pregnant women from original dataset. The program sets initial values for its parameter and least square techniques to fit values. However, in the current analyses it was not possible to achieve a complete fit for Rossavik model on the data due to restraints in selection of technique for analysis, nature or prenatal growth data and time constraint for the project. In future, this study would like to

focus on developing technique for fitting Rossavik model for fetal growth and consider using factors which can give precise measurements for fetal size.

The initial analysis was done using Reed 1 model, and the growth curves were compared with Count model for goodness of fit. Table 3 shows the results of the comparisons between models for infant height, showing significant RSD, AIC and BIC for both boys and girls in REED 1 model. The results were consistent with the model comparison on PHV for growth in another study by Sovio *et al*, 2009, where the Reed1 model showed the best fit for infancy data. The goodness of fit of the liner Count model was not satisfactory and showed age related bias, similar to what was shown by another comparison by Berkley, 1982. However, unexpectedly the mean PHV was higher than expected and needs further exploration as this data was used for the first time for these purposes. Moreover, this should not have any major effect on the results from the association analyses.

Table 4: Mean, median, SD values and unpaired t-tests for Peak Height velocity (cm/month)

Outcome variable	Boys(n=2779)				Girls(n=2865)				P-value*
Peak height velocity	Min	Median	Mean (SD)	Max	Min	Median	Mean (SD)	Max	
Count Model	4.6	9.2	9.2 (0.6)	12.1	4.8	8.3	8.3 (0.5)	11.1	
Reed Model	1.7	6.7	7.2 (2.5)	25.3	1.1	4.6	5.0 (1.7)	17.0	
Log transformed	0.1	1.9	1.9(0.3)	3.2	0.1	1.5	1.5(0.3)	2.8	<0.0001

*P-value - derived from the un-paired t-test for the difference between boys and girls.

Methodological discussion

There were approximately equal distribution of for boys (50%) and girls (51%) in the analysed dataset. Girls were 140g lighter than boys with significant results (P<0.0001). The regression analysis shows that the birthweight were inversely related to the outcome PHV with a reduction of 0.03 % in every 1g increase in birthweight in both boys and girls. This finding could combine with findings from previous studies, such as, Eriksson et al, 1999 and Forsen et al, 2000, showing that lower birthweight children and subsequently become overweight or obese either during childhood or adult life due to rapid increase in their growth pattern to catch up their genetically programmed potential. Gestation age, maternal age, maternal marital status and maternal education did not form a strong confounder between the association of PHV and birth length (Appendix D), however, it is important to include in the analyses in order to demonstrate any effect they would have had on the regression coefficient if acting as a confounder.

It was observed that the children of mothers who either divorced or widowed had a positive association with PHV with an increase by 9 % in boys (P=0.007,95% CI 0.02,0.14) and 6.5 % increase in girls (p=0.01, 95%CI 0.014,0.113). PHV showed a relationship with maternal education so that the children of those mothers had a university degree had PHV of 5.3 % (P=0.03, 95% CI -0.103, -0.006) but the results were not significant in girls (P=0.24,95%CI -0.10,-0.0006). Baxter-Jones et al, 1999, examined in an 1921 Aberdeen birth cohort that growth in 1st year of life is likely to reflect a number of environmental influences, some of which may continue to have effects throughout early life and beyond. However, in this analysis, adjustments for maternal SES did not show any significant results for boys or girls, which could be explained as poor SES stagnates growth in utero and may triggers catch up growth in later childhood which could be a factor for diseases in later life. Maternal alcohol consumption after 16 weeks showed positive association with in both boys and girls, although, this variable is of interest for the infant's growth it was removed from the further analyses due to high number of missing values and time constraints related to this project. A new variable has been recently classified in NFBC 1986 dataset, which could be used for future analyses.

Ong et al, 2001, examined the relationship the relationship between maternal smoking and size at birth and childhood growth between 0 and 5 y in a large retrospective birth cohort and concluded that infant's of maternal smokers were symmetrically small at birth compared with infants of non-smokers, however, showed complete catch-up growth over the first 12 months. In this study when adjusted for smoking at 2nd months of pregnancy,

shows a positive association with increase in PHV by 5 % in boys (P=0.002, 95% CI 0.01, 0.07) and 6 % in girls (P=<0.0001, 95% CI 0.029, 0.082). This is first evidence on how potential intrauterine growth retardation in utero due to environmental factors, such as, smoking, leads into change in speed of height growth. The smaller children grow faster to reach their genetic growth potential. Adding maternal height to the relationship with outcome PHV and main exposure birth length, it was observed that the association becomes even more significant when adjusted for all other factors in the regression model with an increase of 0.34% in both boys and girls (P =0.003).

Limitations of this study

The growth models were checked for true minimum by changing the initial parameters values but the analysis returned same results for all the parameters. It is important to check the true minimum as good starting values will often allow an iterative technique to converge to a solution much faster, also, if multiple minima exist or if there are several local minima in addition to an absolute minimum, poor starting values may results in convergence to an unwanted stationary point of the sum of squares surface. Hence for future growth modeling different strategies could be compared to see goodness of fit for these techniques, for example, model to see goodness of fit between parametric and non-parametric models.

Addition of several confounding variables into the regression analyses decreases the power to detect the differences especially once stratified by sex. The relatively wide confidence intervals reflect this. In other words, it may not be possible to see the difference especially after adjusting for several confounders in spite of that there still might be a difference. Though this study population, with its large size and homogenous distribution is an ideal sample for longitudinal, prospective, observational studies, in this study we however had a few important limiting factors. The nutritional factor is an important factor for infants and no information was available on that. The final task involves the validation process of the regression model which could not be performed due to time constraints. The primary concern of this process is to ensure that the results are genralizable to the population and not specific to the sample used in estimation. In this analysis, the whole cohort data was selected and hence reduces any bias due to sampling.

Bias

As seen in previous work in the area, it impossible to fit growth models without dense measurement points. Longitudinal height measurements from birth to adulthood are rarely available in large population based studies and any replication of the results can be challenging. In many growth studies uses only a few measurements per child to calculate growth velocity (REF?), hence, do not provide measurements that are independent; thus variability estimators derived from these data sets are biased. The current analysis shows high quality of measurements derived for growth in this cohort. This study avoids the problems of biased estimates by applying the statistical analysis to the entire longitudinal growth curve, rather than individual measurements or even pairs of measurements.

Missing Data

Another important factor for bias is this study is due to missing data. In this study, missing data is not possible to state or mark any "missed appointments" as missing, since there are no pre-determined times when those appointments should have happened. Since the study uses unbalanced longitudinal growth data, which occur when there is no intention to observe the anthropometric variable at a common set of ages for each subject and "missing data" cannot be defined in the same way one defines missing data for a cross-sectional variable such as social class at birth (it either has a value or is missing). As a solution, the current study applies complete case analysis for the growth modeling which restricts analysis to the subsample of subjects with complete cases, excluding all individuals who have missing values for any of the variables being considered, whether outcome or explanatory.

Rounding Errors

In this dataset, some of the selected variables had been over represented compared to those from specific places. Thus, to overcome this problem and to allow the results to be more representative at the population level, categorization was used. This inflated the under-sampled cases and deflated the over-sampled ones. However, the R program introduces some minor systematic rounding errors which were therefore, unavoidable. For instance, the regression coefficient values of -0.0003 was rounded to 0.01 with two decimal places; however, it is unlikely to have a significant impact on the results.

Measurement issues

The analysis examined for errors made in measuring outcome or exposure variables, and the implications of such errors for the results of statistical analyses. The current analysis does not report for instrumental errors, underlying variability, respondent errors, and data processing errors currently and would like to explore in future.

Future work

This is the first analyses to try and understand the association between wide varieties of prenatally and infant growth factors. It uses longitudinal data collected over 20 years and is a novel study of its kind using extensive prenatal and perinatal growth data. In future, this study will focus on check on growth models and find biological or measurement reasons behind obtaining higher values for PHV when compared to other epidemiological studies in this area. There are numerous possibilities one could perform in further analysis in this area. One possibility could be to look for good fitting with other mathematical model in fetal data and perform a meta-analysis of epidemiological studies on growth.

Analyses of the effects of the different covariates chosen here and possible covariate from other studies on the outcomes of growth velocity need to be examined further. Some factors, such as hypertension and gestational diabetes, could not be followed here due to time constraint can be possibly important in understanding the relationships for later health. This study would like to explore the replacement of missingness by multiple imputations in future work. As stated previously, many diseases in adulthood are being linked to prenatal development and childhood; further research will need to focus on utilising the growth velocity as a marker of growth patterns in early life in understanding specific character of growth abnormalities and diseases in later life.

Conclusions

In summary, after adjustment for all variables birth length as our main factor of interest was associated statistically significantly with PHV. The interest was, however, to study a broader spectrum of early life factors in relation to other birth measures and potential prenatal determinants of growth. The analyses showed that birth weight, gestational age, maternal age, maternal smoking and maternal height independently of other factors associate with PHV. This project has been highly ambitious in achieving the aims and objective and showed

that modeling of growth itself was even more challenging than originally expected and needs further work. Although this provides a basic understanding between the determinants of fetal growth and growth velocity in infants, further work needs to be done to understand more complex relationship between factors which are markers of adult diseases such as, hypertension and diabetes.

References

Akaike, H. (1979) A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. Biometrika, 66 (2), 237-242.

Baird, J., Fisher, D., Lucas, P., Kleijnen, J., Roberts, H. & Law, C. (2005) Being big or growing fast: systematic review of size and growth in infancy and later obesity. BMJ, 331 (929), .

Barker, D. J. P. (1995) Fetal under nutrition and adult disease. Highlights, 3, 12-15.

Barker, D. J. P., Gluckman, P. D., Godfrey, K. M., Harding, J. E., Owens, J. A. & Robinson, J. S. (1993) Fetal nutrition and cardiovascular disease in adult life. Lancet, 341, 938-941.

Barker, DJP. (ed.) (1998) Mothers, babies and health in later life. 2nd edition. Edinburgh, Churchill Livingstone.

Berkley, C. S. & Reed, R. B. (1987) A Model for Describing normal and abnormal growth in early childhood. Human Biology, 59 (6), 973.

Berkley, C. S. (1982) Comparison of two longitudinal growth models for preschool children. Biometrics, 38 (1), 221-234.

Cameron N (2002) Human growth and development. London: Academic Press. 432 p.

Cooper. (2001) Maternal height, childhood growth and risk of hip fracture in later life: a longitudinal study. Osteoporosis International, 12 (8), 623.

Count W. E. (1943) Growth Patterns of the human physique: An approach to kinetic anthropometry. Human Biology, 15 (1), 1.

Eriksson, J. G., Forsen, T., Tuomilehto, J., Winter, P. D., Osmond, C. & Barker, D. J. P. (1999) Catch-up growth in childhood and death from coronary heart disease: longitudinal study. BMJ, 318, 427-431.

Eriksson, J., Forsen, T., Osmond, C. & Barker, D. (2003) Obesity from cradle to grave. International Journal of Obesity, 27, 722-727.

Forsen, T., Eriksson, J. G., Tuomilehto, J., Osmond, C. & Barker, D. J. P. (1999) Growth in utero and during childhood among women who develop coronary heart disease: longitudinal study. BMJ, 319, 1403-1407.

Frankel, S., Elwood, P., Sweetnam, P., Yarnell, J. & Davey, S. G. (1996) Birth weight, body-mass index in middle age, and incident coronary heart disease. Lancet, 348, 1478-1480.

Gideon Schwarz. (1978) Estimating the Dimension of a Model. The Annals of Statistics, 6 (2), 461-464.

Hakansson, A., Aberg, A., Nyberg, P. & Schersten, B. (1995) A new syphysis-fundus height growth chart based on a well-defined female population with ultrasound-date singleton pregnancies. Acta Obstericia Et Gynecologica Scandinavica, 74 (9), 682-686.

Hauspie RC, Molinari L (2004) Parametric models for postnatal growth. In: Hauspie RC, Cameron N, Molinari L, eds. Methods in Human Growth Research. Cambridge: Cambridge University Press. pp 205–233.

Jarvelin, M-R, Elliott, P., Kleinschmidt, I., Martuzzi, M., Grundy, C., Harkitkainen, A-L & Rantakallio, P. (1997) Ecological and individual predictors of birth weight in a northern Finland birth cohort 1986. Paediatric and Perinatal Epidemiology, 11, 298-312.

Jenss R, Bayley N. A mathematical method for studying the growth of a child. Hum Biol l937;9:556-61

Karlberg, J. (1989) On the construction of the Infancy-childhood-puberty growth standard. Acta Paediatrica, 78, 0803.

Kouchi M, Mukherjee D, and Roche Al? (1983a) Curve fitting for growth in weight during infancy with relationships to adult status and familial associations of the estimated parameters. Hum. Biol. 57:245-263.

Kowarski, . (1971) Determination of integrated plasma concentrations and true secretion rates of human growth hormone. The Journal of Clinical Endocrinology and Metabolism, 32 (3), 356.

Kramer, M. S. (1987) Determinants of low birth weight: methodological assessment and meta-analysis. Bulletin of the World Health Organisation, 65 (5), 6634-737.

Lampl, M., Veldhuis, J. D. & Johnson, M. L. (1992) Saltation and Statis: A model of human growth. American Association for The Advancement of Science, 258 (5083), 801-803.

Launer, . (1992) Maternal recall of infant feeding events is accurate. Quality Safety in Health Care, 46 (3), 203.

Lindstrom, M. J. & Bates, D. M. (1990) Nonlinear Mixed Effects Models for Repeated Measures Data. Biometrics, 46 (3), 673-687.

Liu, Y. X., Jalil, F. & Karlberg, J. (1998) Risk factors for impaired length growth in early life viewed in terms of infancy-childhood-puberty(ICP) growth model'. Acta Paediatrica, 87 (3), 237-243.

Livshits, . (2000) Genetic analysis of growth curve parameters of body weight, height and head circumference. Annals of Human Biology, 27 (3), 299.

Lucas, A., Fewtrell, M. S. & Cole, T. J. (1999) Fetal origins of adult disease - the hypothesis revisited. BMJ, 319, 245.

Michaelsen, A.. (1991) Short-term measurement of linear growth in preterm infants: validation of a hand-held knemometer. Paediatric Research, 30 (5), 464.

Ong,B. (2002) Size at Birth and Early Childhood Growth in Relation to Maternal Smoking, Parity and Infant Breast-Feeding: Longitudinal Birth Cohort Study and Analysis. Paediatric Research, 52 (6), 863.

Osmond, C., Barker, D. J. P., Winter, P. D., Fall, C. H. D. & Simmonds, S. J. (1993) Early growth and death from cardiovascular disease in women. BMJ, 307, 1519-1524.

Parsons, T. J., Power, C. & Manor, O. (2001) Fetal and early life growth and body mass index from birth to early adulthood in 1958 British cohort: longitudinal study. BMJ, 323, 1331-1335.

Parsons, T. J., Power, C., Logan, S. & Summerbell, C. D. (1999) Childhood Predictors of adult obesity: a systematic review. International Journal of Obesity., 23 (8), S1-S107.

R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rich-Edward, J. W., Stampfer, M. J., Manson, J. E., Rosner, B., Hankinson, S. E., Colditz, G. A., Willett, W. C. & Hennekens, C. H. (1997) Birthweight and risk of cardiovascular disease in a cohort of women followed up since 1976. BMJ, 315, 396-400.

Rolland-Cachera, . (1995) Influence of macronutrients on adiposity development: a follow up study of nutrition and growth from 10 months to 8 years of age. International Journal of Obesity, 19 (8), 573.

Rossavik , I. K. & Deter, R. L. (1984) The effect of abdominal profile shape changes on the estimation of fetal weight. J Clin Ultrasound, 12 (57), .

Silventoinen, K., Pietilaninen, K. H., tynelius, P., Sorensen, T. I. A., Kaprio, J. & Rasmussen, F. (2008) Genetic regulation of growth from birth to 18 years of age: The Swedish young male twins study. American Journal of Human Biology, 20, 292-298.

Simon, N. V., Deter, R. L., Shearer, D. M. & Levisky, J. S. (1989) Prediction of Normal Fetal Growth by the Rossavik Growth Model Using Two Scans Before 27 Weeks, Menstrual Age. Journal of Clinical Ultrasound, 17 (4), 237-243.

Simondon, K. B., Simondon, F., Delpeuch, F. & Cornu, A. (1992) Comparative Study of Five Growth models applied to weight data from congolese infants between birth and 13 months of age. American Journal of Human Biology, 4, 327-335.

Sovio, U., Bennett, A. J., Milwood, I. Y., Molitor, J., O'Reilly, P. F., Timpson, N. J., Kaakinen, M., Laitinen, J., Haukka, J., Pillas, D., Tzoulaki, I., Molitor, J., Hoggart, C., Coin, L. J. M., Wittaker, J., Pouta, A., Hartikainen, A-L, Freimer, N. B., Widen, E., Peltonen, L., Elliott, P., McCarthy, M. I. & Jarvelin, M-R. (5) Genetic determinants of height growth assessed longitudinally from infancy and adulthood in the northern Finland birth cohort 1966. PLOS Genetics, 3 (e1000409), .

Stein, C., Fall, C. H. D., kumaran, K., Osmond, C., Cox, V. & Barker, D. J. P. (1996) Fetal growth and coronary heart disease in South India. Lancet, 348, 1269-1273.

Steingrimsdottir, T., Canttinguis, S. & Lindmark, G. (1995) Symphysis-fundus height: construction of a new Swedish reference curve, based on ultrasonically dated pregnancies. Acta Obstertricia Et Gynecologia Scandinavica, 74 (5), 346-351.

Stettler, N., Zernel, B. S., Kumanyika, S. & Stallings, V. A. (2002) Infant weight gain and childhood overweight status in a multicenter, cohort study. Pediatrics, 109, 194--9.

Tanner, . (1970) Standards for children's height at ages 2-9 years allowing for height of parents. Archives of Disease in Childhood, 45 (244), 755.

Tanner, . (1990) Fetus into man: physical growth from conception to maturity.

Waterlow, J. 1988. Linear growth retardation in less developed countries (Nestle Nutrition workshop series Vol 14, New York: Raven Press).

Whitaker, R. C. & Dietz, W. H. (1998) Role of the prenatal environment in the development of obesity. The Journal of Pediatrics, 132 (5), .

Wohlleb, , Pollitt, , Mueller, & Bigelow, . (1983) The Bacon Chow study: maternal supplementation and infant growth. Early Human Development, 9 (1), 79.

3 Appendix Note 3. Classification of hypertension disorder during pregnancy

1. **GH = gestational hypertension**
- BP 140/90 mmHg or more after 20th gestational week. Normal BP in the early pregnancy i.e. (<20th gestational week). No proteinuria.
2. **PE = pre-eclampsia**
- BP 140/90 mmHg or more after 20th gestational week. Normal BP in the early pregnancy i.e. (<20th gestational week). Proteinuria (no information on g/L) at least in one sample during pregnancy.
3. **Chronic hypertension**
    a. BP elevated already in early pregnancy (<20th gestational week), 140/90 mmHg or more, and BP elevated also during whole pregnancy or after pregnancy (6weeks after pregnancy)
    b. Chronic hypertension (hypertension essentialis) diagnoses registered in questionnaire or taking antihypertensive medication.
4. **Superimposed pre-eclampsia**
- Chronic hypertension + proteinuria
5. **Proteinuria**
- Includes also women with Proteinuria (F8V023 ==2)!
6. **Normotensive**
- BP under 140/90 and is not included in any of the above mentioned groups.

4 Appendix Note 4. Example analysis plans

I. EGG-GWAS within cohort additional analysis plan for infant HC

In this document:

1. Traits of interest

2. Participating studies

3. Number of subjects

4. Genotyping + Imputation

5. Model of association.

6. Data exchange

7. Analysis plan

8. Proposed deadline

Background:

The main idea of this additional analysis is that the curve of head circumference growth follows a single percentile line during infancy. With this assumption we can widen the age window and thus increase the sample size. We chose to not include head circumference measurements below the age of 6 months, since the shape/size of the head might still be influenced by the delivery. Based on the data inventory, it seems that we will include the maximum number of subjects by using an age range of 6-30 months (median 18 months). Cohorts with more than one measurement per individual in this time window should use the measurement closest to the age 18 months.

In infancy, the relationship between age and head circumference is non-linear and the standard deviation score changes over time. To standardize our outcome, it is best to create sex- and age-adjusted standard deviation scores for each head circumference measurement. For this purpose, we will need to use a reference curve. For this analysis we will use the reference curve of the Netherlands, 1997. These growth curves have been integrated in an easy to use software package called Growth Analyser. Since we are looking at the SNP

difference within each cohort, it does matter that your subjects are not from the Netherlands (1997). We are only using this software to standardize our outcome.

1. Traits of interest

Head circumference in infancy (HC18) (6 to 30 months).

2. Participating studies (as of 26-01-2010)

- ALSPAC (2,500)
- Generation R (2,240)
- NFBC66 (4,300)
- Raine (759)
- LISA (350)
- CHOP (1733)
- COPSAC (411)

3. Projected number of subjects (as of 26-01-2010)

- HC18 ~ 12,300 individuals

4. Genotyping + Imputation

- Genotyped SNPs (Affymetrix, Illumina, Perlegen)
- Imputation HapMap Phase II CEU SNPs. Preferred release 22 of HapMap, build 36. Predefined marker filters to apply before imputation (HWEP>10-6, MAF>0.01, SNP-call>95%). Suggestion left open to cohorts to apply specific filters but should be reported.
- Analyse all SNPs, no filtering on call rate/HWE/MAF/imputation quality (QC metrics to be reported, and filtering will be done at meta-analysis stage)

5. Model of association

- Additive model (SNP coded as allele dose from 0 to 2), which accounts for genotype imputation uncertainty by use of linear regression onto estimated dose (as included in MACHQTL, ProbABEL, SNPTEST), adjusting for population structure and covariates.

6. Data exchange

- See separate RESULTS_FORMAT file for details of results file formatting and file naming.
- Summary statistics to be uploaded to the AIMS system (link will be provided)
- Only summary statistics will be transferred, not individual level genotype or phenotype data.

7. Analysis outline

- HC18
- Transformation of outcome variable into sex- and age-adjusted standard deviation scores (SDS): use Growth Analyser (see text below).
- If there are multiple measurements of an individual, please take the measurement nearest to 18 months.

- Model:
  - HC18 SDS* = SNP
  - NOTE: YOU DO NOT HAVE TO ADJUST FOR SEX OR AGE ANYMORE!

* Determined by Growth Analyser (see below)

- Inclusions:
  - One of each twin (or any multiple births). This can be random, or based on best genotyping quality.
  - Premature subjects: children born before 37 weeks of gestation. Do not adjust for gestational age!

- Exclusions:
  - Non-Caucasian

8. Deadline

- The proposed deadline for uploading the within cohort results will be 15-02-2010

GROWTH ANALYSER

Creating SD scores using Growth Analyser (max. 15 minutes):

Installing the program:

1. Go to www.growthanalyser.org

2. Click on 'register', and register your name. You will receive an e-mail confirmation and you will be asked to confirm you registration in the e-mail.

3. Log in using your e-mail address and password.

4. Go to 'download' and download the Growth Analyser. It's available for most operating systems.

5. Install Growth Analyser.

6. Open the program. It will ask you some basic info and to enter another password. Once you've entered these, just cancel through any other options and you'll be ready to analyse.

Analysis:

1. Open your data (.txt, .csv. and many others). Data needs to contain sex (coded 1 or 2), gestational age (weeks) and weight (kg).

2. Go to the 'analyse' menu and then click on 'standard deviation scores'.

3. Click on 'reference', and scroll down to 'Netherlands'. Then choose the 'head circumference for age'. Make sure that you indicate gender, age (years) and head circumference (cm) variables.

4. Click 'OK'. Now you should have an additional column in your table with SD scores, ranging from about -3 to +3. If not, make sure that your variables were set in the correct units.

5. Save you new data set.

6. If you have any questions, please e-mail me at d.mook@erasmusmc.nl

II. EGG Consortium Replication of GWAS Meta-Analysis Hits for Growth Parameters

[A] ONE SNP REPRESENTING EACH SIGNAL GENOTYPED (=fetal genotype)

NB: If the proxy SNP is already typed, use best proxy instead of LEAD SNP. Example: FTO rs9939609 (best proxy) has previously been typed. Use this instead of the untyped lead SNP rs1421085. If none of the SNPs have been previously genotyped, type the lead SNP. If the assay fails, use a proxy instead (primarily the best proxy).

For adiposity rebound (AR) analyses: please analyse both Age at AR and BMI at AR.

http://www.evachan.org/calc_snp_stats.R

Table 1: SNP-phenotype pairs required for replication.

PHENOTYPE	SIGNAL	LOCUS	LEAD SNP	BEST PROXY SNP	2ND-BEST PROXY SNP
PWV1	1	*TMEM18*	**rs2860323**	rs6548238	rs7567570
BMI at AP	2	*LEPR*	**rs9436303**	rs9436301	rs9436302
BMI at AP	3	*PCSK1*	**rs10515235**	rs13153770	rs13158163
Age at AR	4	*FTO*	**rs1421085**	rs9939609	rs1558902
Age at AR	5	*RANBP3L/SLC1A3*	**rs2956578**	rs2956577	rs2967066
Age at AR	6	*TFAP2B*	**rs2817419**	rs2635727	rs4715208
BMI at AR	7	*DLG2*	**rs2055816**	rs7116340	rs1354397
BMI at AR	8	*GNPDA2*	**rs10938397**	rs12641981	rs13130484

NB for GNPDA2, published SNP (rs10938397) was chosen as a lead SNP instead of the lowest p-value SNP (rs12641981).

[B] ANALYSIS

To enable meta-analysis using the inverse-variance fixed effects method (based on beta and SE), please perform the analysis outlined below for each SNP related to corresponding phenotype(s).

For these analyses, dense growth measurements at specific age windows are needed.

1. **Peak weight velocity in infancy (PWV1)**

- *The median number of measurements per person in the cohort should be at least 4 within the first 2 years of life including birth (some individuals may have less than 4 but all those with less than 3 will be excluded from the analyses). The focus is on very early growth (PWV occurs at around 1 month of age), and therefore this cannot be reliably estimated in the absence of measurements at birth and the first few months of life (e.g. if all measurements are taken at 1-2 years of age, the cohort cannot take part in this analysis).*

- Derive PWV1 for each individual using Reed1 model (see specification in R below) for boys and girls separately. Fit the Reed1 model by sex on all weight measurements taken at 0-2 years of age, including birth weight. Assume both fixed and random component for all four parameters in the model. For each person, take the first derivative of the fitted distance curve to get the weight velocity curve, and take the maximum of this curve to get peak weight velocity (PWV) in infancy.

- Transformation of outcome variable: First log-transform the outcome using natural logarithm: LOGPWV1 = LN(PWV1)

  PWV1 Z-score = ZLOGPWV1= [LOGPWV1 - mean]/standard deviation

  (based on the mean and SD of LOGPWV1 in the sample)

- Model:

ZLOGPWV1 = SNP + sex + gestational age

- SNP = genotype
- Sex: coded 1 for male, 2 for female
- Gestational age in weeks
- Additionally adjust the model for population structure, e.g. PCs 1-3

- Exclusions:
  - Twins (or any multiple births)
  - Children born before 37 weeks of gestation (exclude <37 weeks, include >= 37 weeks)
  - Those with less than 3 weight measurements at 0-2 years of age

NB Please see example R code for PWV modeling in File 2.

2. BMI at adiposity peak (AP) in infancy

- *The median number of measurements per person in the cohort should be at least 4 at age 14 days – 1.5 years (some individuals may have less than 4 but all those with less than 3 will be excluded from the analyses). The focus is on growth at 6-12 months where BMI peaks, and therefore cohorts that do not have any measurements at 6-12 months cannot take part in this analysis (see the note below).*

- Calculate BMI from height and weight measurements taken at the same time: BMI = weight [kg] / (height [m])^2. Fit a cubic mixed effects model (see specification in R below) on LN(BMI) at 14 days – 1.5 years in the whole sample, using sex as a covariate (this allows each sex a separate BMI baseline). When fitting the model, centralise age to 0.75 years. In addition to fixed effects, include random effects for the constant and the slope in the model. Assume autoregressive AR(1) within-person correlation structure between the measurements. For each individual, derive age and BMI at the point where the curve reaches its maximum, i.e. at infant adiposity peak (AP). It is enough to use the age window 0.25-1.25 years for this (the population average should be near the mid-point, around 0.75 years, i.e. 9 months). BMI at AP is likely to be skewed to the right and therefore a natural logarithmic transformation will be used: LOGBMIAP = LN(BMIAP).

- LOGBMIAP Z-score ZLOGBMIAP = [LOGBMIAP – mean]/standard deviation
  (based on the mean and SD of LOGBMIAP in the sample)

- Model:

  ZLOGBMIAP = SNP + sex + gestational age

- Exclusions:
  - Twins (or any multiple births)
  - Children born before 37 weeks of gestation (exclude <37 weeks, include >= 37 weeks)
  - Those with less than 3 BMI measurements at 14 days - 1.5 years of age

- ***Note**: If you have on average less than 4 measurements at age 14 days to 1.5 years in your study but instead you have a single BMI measurement at 6–12 months of age, please fit the following model using the measurement closest to age 9 months (here called BMIAPproxy) for each individual: ZLOGBMIAPproxy = SNP + sex + age + age^2 + gestational age, where ZLOGBMIAPproxy = Z-score of LOGBMIAPproxy = [LOGBMIAPproxy – mean]/SD and age = measurement age (centered, i.e. observed age – population mean age).*

NB Please see example R code for BMI modeling in File 2.

3. Adiposity rebound (AR) in childhood

- *The median number of measurements per person in the cohort should be at least 4 at age 1.5 - 13 years (some individuals may have less than 4 but all those with less than 3 will be excluded from the analyses). The focus is on growth at 4-7 years around adiposity rebound, and therefore cohorts that do not have any measurements at 4-7 years cannot take part in this analysis.*

- Calculate BMI from height and weight measurements taken at the same time: BMI = weight [kg] / (height [m])^2. Fit a cubic mixed effects model (see specification in R below) on LN(BMI) at >1.5 – 13 years in the whole sample, using sex as a covariate. When fitting the model, centralise age to 7.25 years. Include also sex-age and sex-age^2 interactions. In addition to fixed effects, include random effects for the constant and the slope in the model. Assume autoregressive AR(1) within-person correlation structure between the measurements. For each individual, derive age and BMI at the point where the

curve reaches its minimum, i.e. at adiposity rebound (AR). It is enough to use the age window 2.5-8.5 years for this (the population average should be near the mid-point, around 5.5 years). Age at AR (AGEAR) should be fairly normally distributed but BMI at AR is likely to be skewed to the right and therefore a natural logarithmic transformation will be used: LOGBMIAR = LN(BMIAR).

- AGEAR Z-score = ZAGEAR = [AGEAR – mean]/standard deviation
  (based on the mean and SD of AGEAR in the sample)
- LOGBMIAR Z-score ZLOGBMIAR = [LOGBMIAR – mean]/standard deviation
  (based on the mean and SD of LOGBMIAR in the sample)
- Model 1:
  ZAGEAR = SNP + sex
- Model 2:
  ZLOGBMIAR = SNP + sex
- Exclusions:
  - Twins (or any multiple births)
  - Those with less than 3 BMI measurements at >1.5-13 years of age

NB Please see example R code for BMI modeling in File 2.

[C] RESULTS FILE FORMAT

Please provide:

1. EXCEL or TEXT file containing data in the following columns

****Column headings and their descriptions****

SNP	rsNUMBER of SNP
STRAND	The strand on which the alleles are reported: we request SNPs to be aligned to the forward (+) strand of the chromosome on HapMap. State the strand as a single character: - or +
N	Number of individuals in the analysis for that SNP
EFFECT_ALLELE	Allele for which the effect (BETA) is reported (A/T/G/C)
NON_EFFECT_ALLELE	Second allele at the SNP
EFFECT_ALLELE_FREQ	Allele frequency of the allele for which the effect (BETA) is reported

BETA	Effect size
SE	Standard error on the estimate of the effect size
P_VAL	The 2-sided p-value of the association test
CALL_RATE	Genotyping call rate for the SNP (N_successful_genotypes / N_attempted_genotypes)
HWE_P	Hardy-Weinberg P value in those genotyped

2. README file containing
    a. Ancestry (e.g. European) and location (e.g. UK) of cohort
    b. How growth measurements were ascertained (e.g. self-report / clinical measure / medical records)
    c. Any genotyping QC in addition to call rate and HWE (e.g. % included duplicates and duplicate concordance)

    In addition, if a single BMI measurement at 6-12 months was used instead of estimated BMI at AP, give also details of the BMI distribution in the included sample (mean, SD, min, max).

Please name files as follows:

STUDY-PHENOTYPE-EGG-CONTENTS-DATE.txt

STUDY is a short identifier for the population studied.

PHENOTYPE is one of:

"LOGPWV1", "LOGBMIAP" (or "LOGBMIAPproxy"), "AGEAR", "LOGBMIAR"

DATE is the date on which the file was prepared, in the format "DDMMYYYY".

CONTENTS indicates whether the file is a README or RESULTS file.

For example:

NFBC1986-AGEAR-EGG-RESULTS-01082010.txt

If you have any questions on growth modeling and derivation of the required parameters, please email: shikta.das08@imperial.ac.uk, nicole.warrington@uwa.edu.au AND h.taal@erasmusmc.nl

5 Appendix Table 1. The SNP selection criteria for candidate loci, utilised in GWAS of infant head circumference in Stage 1 (discovery) meta-analysis data, for validation in Stage 2 (replication). SNPs with P-value P < 1 x 10-7 or P < 1 x 10-5 in/near candidate genes were selected. One exception for peak weight velocity (PWV) was included with did not fulfill the selection criteria due to strong biological candidacy.

CHR	Phenotype	Index SNP	Proxy1	Proxy2	Nearest gene	Top SNP m/a p-value	Justification
12	HC	rs7980687	rs12322888	rs12316131	near SBNO1	3.3 x 10-7	Low P-value
12	HC	rs1042725	rs7970350	rs1351394	near HMGA2	6.6 x 10-7	Low P-value and biology. Known SNPs has been associated with height [354] and primary tooth development during infancy [355]
15	HC	rs11655470	rs12938031	-	near CRHR1/MAPT	1.4 x 10-6	Low P-value

6 Appendix Table 2. Description of the top-hit loci associated with infant head circumference.

CHR	Location	Coded Allele	Phenotype	Index SNP	Nearest Gene	Other names	Description	Function of gene	Type
12	12q24	A/G	HC	rs7980687	near SBNO1	Sno; MOP3	Protein strawberry notch homolog 1		protein coding
12	12q15	T/C	HC	rs1042725	near HMGA2	BABL; LIPO; HMGIC; HMGI-C; STQTL9	high mobility group AT-hook 2	A non-histone chromosomal high mobility group (HMG) protein family. HMG proteins function as architectural factors and are essential components of the enhancesome. This protein contains structural DNA-binding domains and may act as a transcriptional regulating factor. Identification of the deletion, amplification, and rearrangement of this gene that are associated with myxoid liposarcoma suggests a role in adipogenesis and mesenchymal differentiation. A gene knock out study of the mouse counterpart demonstrated that this gene is involved in diet-induced obesity. Alternate transcriptional splice variants, encoding different isoforms, have been characterized	protein coding
17	17q21	T/C	HC	rs11655470	near CRHR1	CRF1; CRHR; CRF-R; CRFR1; CRF-R1; CRFR-1; CRH-R1; CRHR1L; CRHR1f; CRF-R-1; CRH-R-1; CRH-R1h	corticotropin releasing hormone receptor 1	A G-protein coupled receptor that binds neuropeptides of the corticotropin releasing hormone family that are major regulators of the hypothalamic-pituitary-adrenal pathway. The encoded protein is essential for the activation of signal transduction pathways that regulate diverse physiological processes including stress, reproduction, immune response and obesity. Alternative splicing results in multiple transcript variants, one of which represents a read-through transcript with the neighboring gene MGC57346.	protein coding

7 Appendix Table 3. Loci associated with infant head circumference (P<1 x 10-5), which were not taken forward for replication.

Index SNP	Locus	Effect on Head Circumference	Se	P value	Nearest Gene	Gene function
rs3094072_C	6p21	0.103	0.022	$1.5x10^{-6}$	*HLA-L*	Part of the major histocompatibility complex genes. SNPs in linkage disequilibrium have been associated with Systemic Lupus Erythematosus
rs12438760_T	15q14	0.109	0.023	$2.1x10^{-6}$	*C15orf41*	Unknown
rs17134374_T	15q11	-23.293	5.033	$3.7x10^{-6}$	*POTEB*	Unknown
rs238150_T	20q13	-0.078	0.0171	$4.5x10^{-6}$	*DDX27*	DEAD box proteins, characterized by the conserved motif Asp-Glu-Ala-Asp (DEAD), are putative RNA helicases. They are implicated in a number of cellular processes involving alteration of RNA secondary structure such as translation initiation, nuclear and mitochondrial splicing, and ribosome and spliceosome assembly. Based on their distribution patterns, some members of this family are believed to be involved in embryogenesis, spermatogenesis, and cellular growth and division. This gene encodes a DEAD box protein, the function of which has not been determined. [provided by RefSeq, Jul 2008]
rs9940645_A	16q12	0.066	0.0145	$5.8x10^{-6}$	*ZNF423*	The protein encoded by this gene is a nuclear protein that belongs to the family of Kruppel-like C2H2 zinc finger proteins. It functions as a DNA-binding transcription factor by using distinct zinc fingers in different signaling pathways. It is thought that this gene may have multiple roles in signal transduction during development. [provided by RefSeq, Jul 2008]
rs11683142_A	2q23	0.293	0.065	$6.1x10^{-6}$	*RBM43*	Unknown
rs1385504_A	8q24	0.094	0.021	$7.1x10^{-6}$	*TNFRSF11B*	The protein encoded by this gene is a member of the TNF-receptor superfamily. This protein is an osteoblastsecreted decoy receptor that functions as a negative regulator of bone resorption. This protein specifically binds to its ligand, osteoprotegerin ligand, both of which are key extracellular regulators of osteoclast development. Studies of the mouse counterpart also suggest that this protein and its ligand play a role in lymph-node organogenesis and vascular calcification. Alternatively spliced transcript variants of this gene have been reported, but their full length nature has not been determined. [provided by RefSeq, Jul 2008] SNPs in linkage disequilibrium have been associated with osteoporosis in genome wide association study meta-analysis.
rs2056666_T	8p23	0.067	0.015	$7.3x10^{-6}$	*CSMD1*	Unknown
rs9675157_C	17q23-24	0.185	0.042	$9.6x10^{-6}$	*AXIN2*	The Axin-related protein, Axin2, presumably plays an important role in the regulation of the stability of betacatenin in the Wnt signaling pathway, like its rodent homologs, mouse conductin/rat axil. In mouse, conductin organizes a multiprotein complex of APC (adenomatous polyposis of the colon), beta-catenin, glycogen synthase kinase 3-beta, and conductin, which leads to the degradation of beta-catenin. Apparently, the deregulation of beta-catenin is an important event in the genesis of a number of malignancies. The AXIN2 gene has been mapped to 17q23-q24, a region that shows frequent loss of heterozygosity in breast cancer, neuroblastoma, and other tumors. Mutations in this gene have been associated with colorectal cancer with defective mismatch repair. [provided by RefSeq, Jul 2008]

8 Appendix Table 4. The SNP selection criteria for candidate loci, utilizing Stage 1 meta-analysis data, for Stage 2 (replication). SNPs with P-value P < 1 x 10-7 or P < 1 x 10-5 in/near candidate genes were selected. One exception for peak weight velocity (PWV) was included with did not fulfill the selection criteria due to strong biological candidacy.

Chr	Phenotype	Index SNP	Proxy1	Proxy2	In/ nearest gene	Top SNP m/a p-value	Justification for selection	Proxy linkage disequilibrium (LD) with index SNP and proxy *P*-value fromStage1 meta-analysis
1	BMI at AP	rs9436303	rs9436301	rs9436302	*LEPR/LEPROT*	4.15E-09	Low P-value	Proxy1 LD R^2=1 and P=6.4x10-8, proxy2 LD R^2=0.786 and p=1.41x10-7
16	Age at AR	rs1421085	rs9939609	rs1558902	*FTO*	6.14E-08	Low P-value	If already typed, use either of the published SNPs (index SNP or Proxy1). Proxy1 LD R^2=0.931 and P=1.56x10-7, proxy2 LD R^2 =1 and P=6.36x10-8
5	Age at AR	rs2956578	rs2956577	rs2967066	*Intergenic between RANBP3L and SLC1A3*	6.73E-08	Low P-value	Proxy1 LD R^2=1 and P=6.73x10-8, proxy2 LD R^2=1 and P=6.75x10-8
4	BMI at AR	rs10938397	rs12641981	rs13130484	*GNPDA2*	1.41E-07	Biology: known BMI locus (Willer et al, 2009), LD with rs10938397 is very strong (R2=1) with top SNP. Rs10938397 BMI increasing allele (G) is associated with higher BMI at AR ($p=5.4x10^{-6}$)	Published SNP chosen as index SNP instead of top SNP. Proxy1 (top SNP) LD R2=1 and P=3.11x10-6, proxy2 LD R2=1 and P=3.37x10-6
5	BMI at AP	rs10515235	rs13153770	rs13158163	*PCSK1*	1.23E-06	Biology : mutations in this gene are associated with extreme obesity	Proxy1 LD R2=0.961 and P=1.26x10-6, proxy2 LD r2=1 and P=3.39x10-6
6	Age at AR	rs2817419	rs2635727	rs4715208	*TFAP2B*	2.96E-06	Biology: known adiposity locus (Lindgren et al, 2009). rs987237 G = WC increasing allele is associated with lower age at AR (p=3.5x10-4), but LD is weak (R2=0.044) between the SNPs	Proxy1 LD R2=0.865 and P=3.16x10-6, proxy2 LD R2=0.861 and P=3.17x10-6
11	BMI at AR	rs2055816	rs7116340	rs1354397	*DLG2*	5.41E-06	Biology: DLG2 may be involved in maintenance of glucose homeostasis (Palmer et al 2010). However, our SNP is not in the region of DLG2 they identified for association	Proxy1 LD R2=1 and P=1.70x10-7, proxy2 LD R2=1 and P=1.72x10-7
2	PWV*	rs2860323	rs6548238	rs7567570	*TMEM18*	5.91E-05	Biology: known BMI locus (Willer et al, 2009)[135]. rs6548238 is in high LD (R^2=0.85) with rs2860323. rs6548238 BMI increasing allele (C) is associated with higher PWV1 ($P=1.6x10^{-4}$)	If already typed, use the published SNP (proxy1 rs6548238). Proxy2 is in perfect LD (r2=1) with index SNP and its P=5.92x10-5

*rs2860323 was selected as exception to the selection criteria due to strong candidacy as loci does not fulfill *P*-value criterion for candidates: $1.0 \times 10^{-5} < P < 1.0 \times 10^{-5}$

9 Appendix Table 5. Description of the five top-hit loci associated with early growth phenotypes.

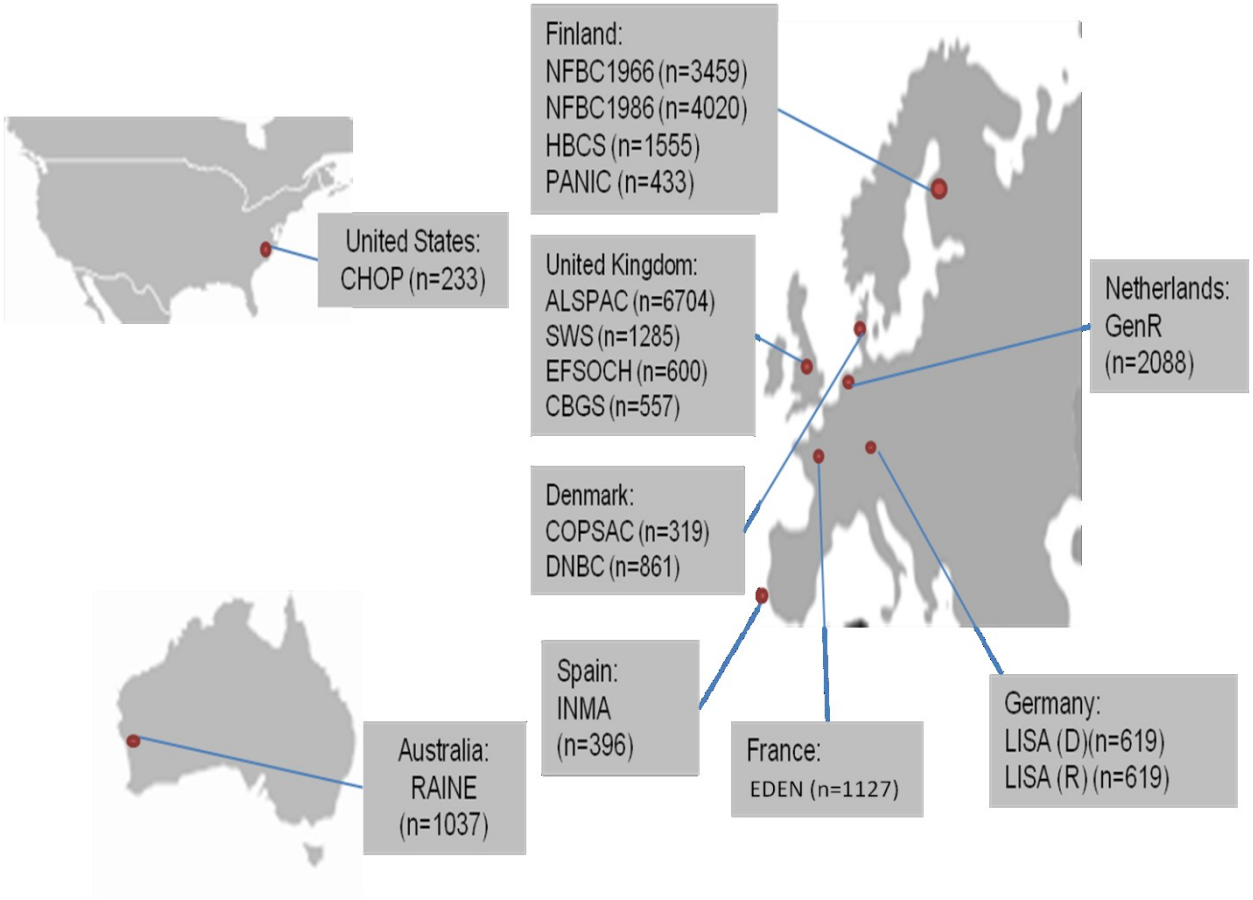
CHR	Location	In/nearest gene	Coded allele	Measure	Other Names	Description	Cellular component	Biological processes	Molecular processes	Type
1	1p31	*LEPR/ LEPROT*	G	BMIAP z-score	*CD295; OBR*	Leptin receptor and leptin receptor overlapping transcript	extracellular region, integral to plasma membrane, plasma membrane	cell surface receptor linked signal transduction, energy reserve metabolic process, multicellular organismal development	hematopoietin/interferon-class cytokine receptor activity , protein binding	protein-coding
5	5q15	*PCSK1*	G	BMIAP z-score	*PCSK1; NEC1; PC1; PC3; SPC3*	Proprotein convertase subtilisin/kexin type 1	cytoplasmic vesicle	cell-cell signaling, metabolic process, proteolysis	calcium ion binding, peptidase activity, proprotein convertase 1 activity, subtilase activity	protein-coding
16	16q12.2	*FTO*	C	AGEAR z-score	*FTO; KIAA1752; MGC5149*	Fat mass and obesity associated	NA	determination of left/right symmetry, embryonic limb morphogenesis	NA	protein-coding
6	6p12	*TFAP2B*	G	AGEAR z-score	*TFAP2B; AP-2B; AP2-B; MGC21381*	Transcription factor AP-2 beta (activating enhancer binding protein 2 beta)	nucleus	kidney development, nervous system development, positive regulation of global transcription from RNA polymerase II promoter, regulation of transcription from RNA polymerase II promoter, transcription	protein dimerization activity, sequence-specific DNA binding , transcription coactivator activity, transcription factor activity	protein-coding
4	4p12	*GNPDA2*	G	BMIAR z-score	*GNPDA2; SB52*	Glucosamine-6-phosphate deaminase 2	NA	N-acetylglucosamine metabolic process , carbohydrate metabolic process	glucosamine-6-phosphate deaminase activity , isomerase activity	protein-coding

10 **Table 6. Heterogeneity analysis of the five top-SNPs with the early growth phenotypes. For rs9436303, heterogeneity was observed between studies; the Table shows the three actions taken to address this.**

SNP	Nearest gene	Allele[a]	β(F)[b]	95% CI[b]	β(R)[c]	95% CI[c]	I^{2d}	P_{HET}[e]	Steps taken[f]
rs10515235	*PCSK1*	A/G	-0.06	(-0.08, -0.03)	-0.06	(-0.08, -0.03)	15.60%	0.2785	
rs1421085	*FTO*	C/T	-0.12	(-0.14, -0.10)	-0.12	(-0.14, -0.09)	17.80%	0.2892	
rs2817419	*TFAP2B*	G/A	-0.08	(-0.05, -0.10)	-0.08	(-0.05, -0.11)	30.20%	0.1866	
rs10938397	*GNPDA2*	G/A	0.08	(0.05, 0.10)	0.08	(0.05, 0.11)	30.20%	0.1866	
rs9436303	*LEPR and LEPROT*	G/A	0.07	(0.06, 0.08)	0.08	(0.05, 0.11)	66.90%	0.0002	Action 1
rs9436303	*LEPR and LEPROT*	G/A	0.07	(0.04, 0.09)	0.07	(0.03, 0.11)	61.80%	0.0017	Action 2
rs9436303	*LEPR and LEPROT*	G/A	0.05	(0.02, 0.07)	0.05	(0.02, 0.07)	0%	0.8419	Action 3

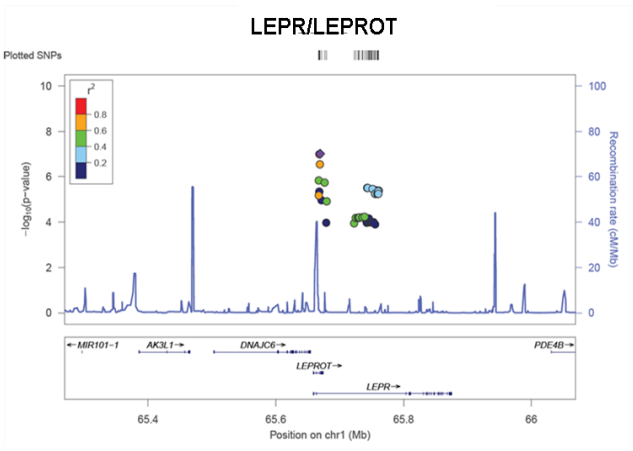
[a]Minor allele/Major alllele.
[b]The value of beta (B) and 95% confidence interval (CI) under fixed model.
[c]The value of beta (B) and 95% confidence interval (CI) under random effect model.
[d]I^2 describes the proportions of the total variability caused by heterogeneity by various studies in the meta-analysis. I^2 <25% is considered as no heterogeneity, I^2 =25-50% is considered as low heterogeneity, I^2=50-70% is considered as moderate heterogeneity and I^2 >75% is considered as strong heterogeneity.
[e]The P value of heterogeneity.
[f]Action 1. Both fixed effect and random effect model was used.
Action 2. Removing EDEN from the study as it was showing most inflated results.
Action 3. Removing EDEN and NFBC1966 (lambda=1.04) as both studies showing high variability inflation factor (VIF).

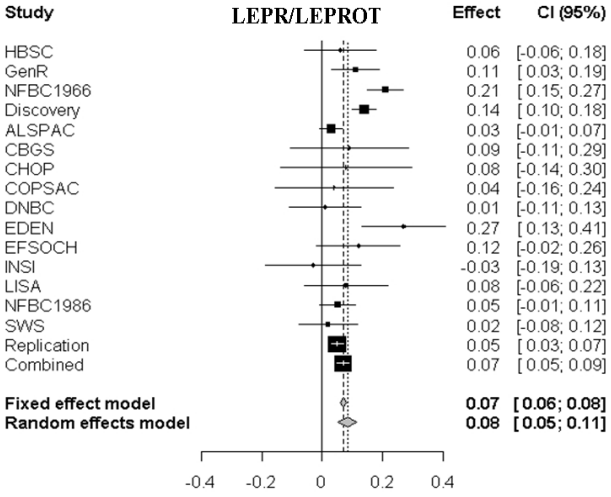
11 **Appendix Figure1. World map showing countries which participated in the study.**

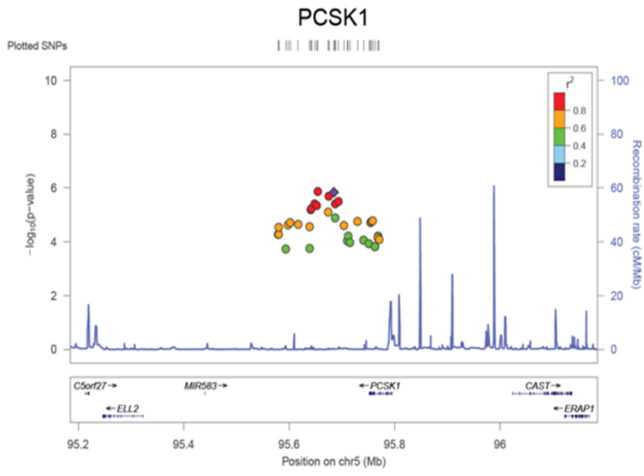


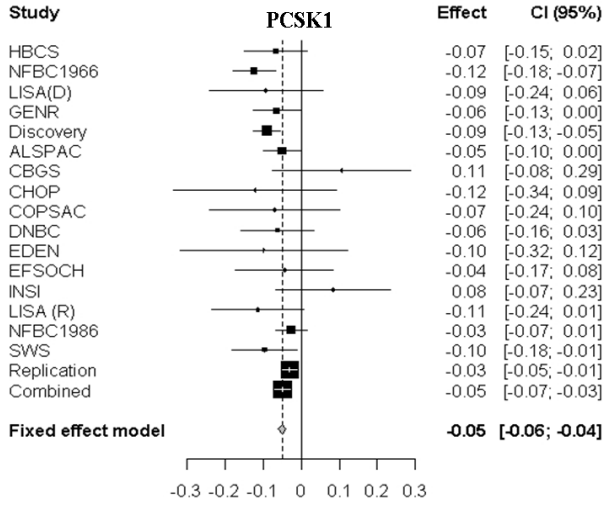
12 **Appendix Figure 2. Figure 3: Regional association and Forest plots of the five loci associated with early growth phenotypes (A, B, C). Blue diamond indicates the most statistically significantly associated SNP, and other SNPs in the region are presented by circles with coloring from blue to red corresponding to r^2 values from 0 to 1. The SNP position refers to the NCBI build 36. Estimated recombination rates are from HapMap build 36 [356]. Forest plots from the meta-analysis for each of the identified loci are plotted abreast. Results by study, discovery, replication and combined meta-analyses presented (for heterogeneity, see note[#])**

A) BMI at adiposity peak (AP)

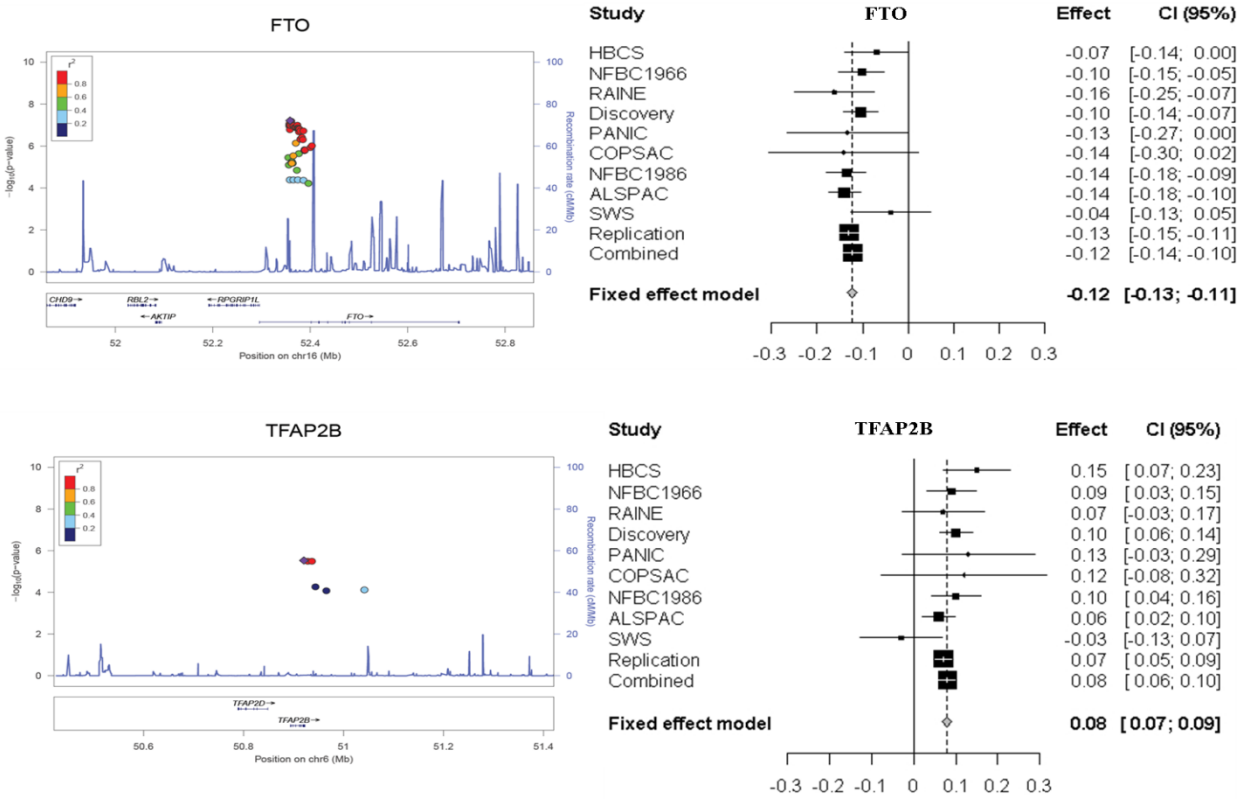




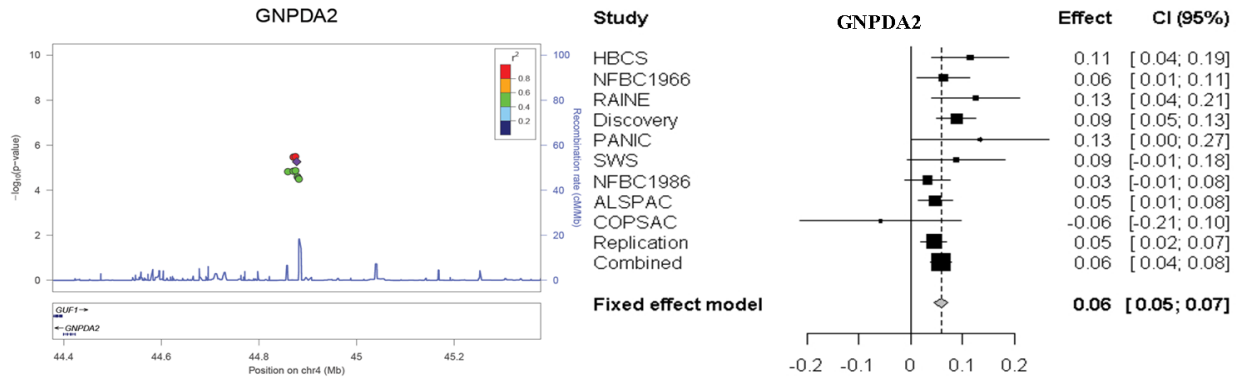




B) **Age at adiposity rebound (AR)**



C) BMI at adiposity rebound (AR)



#There was no evidence of heterogeneity between the studies, except at *LEPR/LEPROT* (I^2=66.9%; P_{HET}=0.0002), by two large and well-defined studies, NFBC1966 (n=3459) and EDEN (n=1127) (P_{HET}=0.84 without these two cohorts). However, associations are overall directionally consistent between all the studies and further the GWAS analyses without NFBC1966 and EDEN show that results are not driven only by them. For *LEPR/LEPROT* both fixed and random effect models give very similar results (shown above).