Genetic Determinants of Obesity in Relation to Diet, Weight Gain and Mortality

Rukh, Gull

2016

Genetic Determinants of Obesity in Relation to Diet, weight Gain and Mortality

GULL RUKH | FACULTY OF MEDICINE | LUND UNIVERSITY

Genetic Determinants of Obesity in Relation to Diet, Weight Gain and Mortality

Genetic Determinants of Obesity in Relation to Diet, weight Gain and Mortality

Gull Rukh

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at Kvinnokliniken's Aula, 3rd floor, Jan Waldenströms gata 47,
Skåne University Hospital, Malmö, Sweden on April 15th, 2016 at 9:00 am.

Faculty opponent

Majken K Jensen, PhD
Assistant Professor of Genetic Epidemiology and Nutrition
Harvard T.H. Chan School of Public Health, Harvard University, USA

Organization DEPARTMENT OF CLINICAL SCIENCES LUND UNIVERSITY, SWEDEN	Document name Doctoral dissertation
	Date of issue 15th April, 2016
Author(s) GULL RUKH	Sponsoring organization

Title and subtitle: Genetic determinants of obesity in relation to diet, weight gain and mortality

Abstract

Obesity is one of the major health concerns that has reached epidemic proportions globally. It is generally believed to be a result of interactions between genetic and environmental factors. In this thesis we investigated the role of dietary factors in modifying the genetic susceptibility to obesity (papers I to III), studied the association between genetic susceptibility to obesity and weight gain at different time-points in life (paper IV) and tried to dissect the causality between cardiometabolic traits and mortality (paper V). The work in this thesis was conducted using data from the population based prospective Malmö Diet and Cancer Study (MDCS; N= ~30,000) and the Gene-Lifestyle interactions And Complex traits Involved in Elevated disease Risk (GLACIER; N= ~5000) cohorts. In paper I, we did not observe any evidence for macronutrient, fiber or total energy intake in modifying the genetic susceptibility to obesity when genetic susceptibility was represented as a Genetic Risk Score (GRS) based upon 13 BMI associated genetic variants. In individual SNP analyses, after correcting for multiple comparisons, some of the individual obesity loci such as *NEGR1* rs2815752 associated with fat, carbohydrate and fiber intakes ($P \leq 1x10^{-4}$ for all) and *BDNF* rs4923461 interacted with protein intake on BMI ($P_{interaction}=0.001$). In paper II, pooled analyses of MDCS and GLACIER suggested 0.16 (SE=0.04) kg/m^2 increase in BMI ($P=8x10^{-5}$) in the lowest quartile of GRS (comprised of 30 BMI-associated genetic variants) for each increment in category of sugar-sweetened beverages (SSB) intake vs. 0.24 (SE=0.04) kg/m^2 higher BMI in the highest GRS quartile ($P=1x10^{-7}$). We also observed evidence for the role of SSB intake in modifying the genetic susceptibility to obesity ($P_{interaction}=0.049$). In paper III, a copy number variant (CNV) in the salivary amylase gene (*AMY1*) did not associate with obesity traits neither in men nor in women ($P>0.05$ for all). However, upon stratification by dietary starch intake, BMI decreased with increasing *AMY1* CNV in low starch intake group ($P=0.035$) and increased with increasing *AMY1* CNV in the high starch intake group ($P=0.04$) among females. These results suggest a putative role of starch intake in modifying the association between *AMY1* CNV and obesity in women ($P_{interaction}=0.041$). In paper IV, a GRS based on 31 BMI-associated genetic variants was associated with increased annual weight change (β=0.003 kg; SE=0.01; $P=7x10^{-8}$) and increased odds for substantial weight gain (OR=1.01; 95% CI= 1.00-1.02; $P=0.013$) per risk allele from young to middle age in MDCS. However, the GRS was associated with decreased annual weight change (β=-0.005 kg; SE=0.002; $P=0.002$) and decreased risk for substantial weight gain (OR=0.97; 95% CI= 0.96-0.99; $P=0.001$) per risk allele during and after middle-age in the pooled analyses of MDCS and GLACIER. These results suggest a paradoxical inversed relationship between genetic susceptibility to obesity and weight gain during and after middle age compared to increased weight gain in younger age. In paper V, observations from multivariable Mendelian randomization analyses suggest a direct causal association of TG ($P=0.017$ and $P=0.028$) and an inverse association of HDLC ($P=0.049$ and $P=0.005$) with total- and cardiovascular mortality, respectively. In conclusion, the results from this thesis suggest a role of specific dietary factors in modifying the genetic susceptibility to obesity and that genetic variation affect weight gain differently at different time-points in life but the underlying mechanisms need to be further understood. Additionally,our findings points towards causal associations between TG and HDLC and mortality which can help to devise better treatment strategies in clinical practice.

Key words: genetic variants, copy number variants, genetic risk score, BMI, gene-diet interactions, annual weight change, substantial weight gain, sugar-sweetened beverages, mortality, Mendelian randomization.

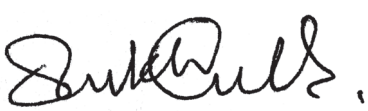
Classification system and/or index terms (if any)

Supplementary bibliographical information — Language: English

ISSN and key title 1652-8220 — ISBN 978-91-7619-269-6

Recipient's notes — Number of pages — Price

Security classification

Signature  Date 10th March, 2016

Genetic Determinants of Obesity in Relation to Diet, Weight Gain and Mortality

Gull Rukh

LUND
UNIVERSITY

Diabetes and Cardiovascular Disease - Genetic Epidemiology
Department of Clinical Sciences in Malmö
Faculty of Medicine, Lund University, Sweden

Cover picture: Abstract DNA

ISBN 978-91-7619-269-6
ISSN 1652-8220
Lund University, Faculty of Medicine Doctoral dissertation Series 2016:43

Printed in Sweden by Media-Tryck, Lund University
Lund 2016

To ammi, abbu (may their souls rest in peace) & my lovely family

"The real voyage of discovery consists of not in seeking new landscapes, but in having new eyes." Marcel Proust

Content

List of Papers

This doctoral thesis is based on the following five papers:

I. **Rukh G,** Sonestedt E, Melander O, Hedblad B, Wirfält E, Ericson U, Orho-Melander M. Genetic susceptibility to obesity and diet intakes: association and interaction analyses in the Malmö Diet and Cancer Study. ***Genes Nutr,*** 2013. 8(6): p.535-547.

II. Brunkwall L, Chen Y, Hindy G, **Rukh G,** Ericson U, Barroso I, Johansson I, Franks PW, Orho-Melander M, Renström F. Sugar-sweetened beverage consumption and genetic predisposition to obesity in two Swedish cohorts. ***Under revision in Am J Clin Nutr***, 2016.

III. **Rukh G,** Ericson U, Orho-Melander M, Sonestedt E. Role of dietary starch intake in modifying the relationship between *AMY1* copy number variant and obesity risk. ***Manuscript.***

IV. **Rukh G,** Ahmad S, Ericson U, Hindy G, Stocks T, Renström F, Almgren P, Nilsson PM, Melander O, Franks PW, Orho-Melander M. Inverse relationship between a genetic risk score of 31 BMI loci and weight change before and after reaching middle age. ***Int J Obes (Lond)***, 2015.

V. **Rukh G,** Hindy G, Almgren P, Schulz CA, Ericson U, Melander O, Orho-Melander M. Causal effects of cardiometabolic traits on cardiovascular and total mortality: A Mendelian randomization study. ***Manuscript.***

Papers not Included in the Thesis

I. Hindy G, Mollet IG, **Rukh G**, Ericson U, Orho-Melander M. Several type 2 diabetes associated variants in genes annotated to WNT signaling interact with dietary fiber in relation to incidence of type 2 diabetes. ***Genes Nutr***, 2016. In press.

II. Nettleton JA, Follis JL, Ngwa JS, Smith CE, Ahmad S, Tanaka T, Wojczynski MK, Voortman T, Lemaitre RN, Kristiansson K, Nuotio ML, Houston DK, Perälä MM, Qi Q, Sonestedt E, Manichaikul A, Kanoni S, Ganna A, Mikkilä V, North KE, Siscovick DS, Harald K, Mckeown NM, Johansson I, Rissanen H, Liu Y, Lahti J, Hu FB, Bandinelli S, **Rukh G**, Rich S, Booij L, Dmitriou M, Ax E, Raitakari O, Mukamal K, Männistö S, Hallmans G, Jula A, Ericson U, Jacobs DR Jr, Van Rooij FJ, Deloukas P, Sjögren P, Kähönen M, Djousse L, Perola M, Barroso I, Hofman A, Stirrups K, Viikari J, Uitterlinden AG, Kalafati IP, Franco OH, Mozaffarian D, Salomaa V, Borecki IB, Knekt P, Kritchevsky SB, Eriksson JG, Dedoussis GV, Qi L, Ferrucci L, Orho-Melander M, Zillikens MC, Ingelsson E, Lehtimäki T, Renström F, Cupples LA, Loos RJ, Franks PW. Gene × dietary pattern interactions in obesity: analysis of up to 68 317 adults of European ancestry. ***Hum Mol Genet***. 2015 May 20. pii: ddv186. [Epub ahead of print].

III. Shungin D, Cornelis MC, Divaris K, Holtfreter B, Shaffer JR, Yu YH, Barros SP, Beck JD, Biffar R, Boerwinkle EA, Crout RJ, Ganna A, Hallmans G, Hindy G, Hu FB, Kraft P, McNeil DW, Melander O, Moss KL, North KE, Orho-Melander M, Pedersen NL, Ridker PM, Rimm EB, Rose LM, **Rukh G**, Teumer A, Weyant RJ, Chasman DI, Joshipura K, Kocher T, Magnusson PK, Marazita ML, Nilsson P, Offenbacher S, Davey Smith G, Lundberg P, Palmer TM, Timpson NJ, Johansson I, Franks PW. Using genetics to test the causal relationship of total adiposity and periodontitis: Mendelian randomization analyses in the Gene-Lifestyle Interactions and Dental Endpoints (GLIDE) Consortium. ***Int J Epidemiol***. 2015 Apr;44(2):638-50. doi: 10.1093/ije/dyv075. Epub 2015 Jun 6.

IV. Smith JG, Luk K, Schulz CA, Engert JC, Do R, Hindy G, **Rukh G**, Dufresne L, Almgren P, Owens DS, Harris TB, Peloso GM, Kerr KF, Wong Q, Smith AV, Budoff MJ, Rotter JI, Cupples LA, Rich S, Kathiresan S, Orho-Melander M, Gudnason V, O'Donnell CJ, Post WS, Thanassoulis G; Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) Extracoronary Calcium Working Group. Association of low-density lipoprotein cholesterol-related genetic variants with aortic valve calcium and incident aortic stenosis. ***JAMA***. 2014 Nov 5;312(17):1764-71. doi: 10.1001/jama.2014.13959.

V. Stojkovic IA, Ericson U, **Rukh G**, Riddestråle M, Romeo S, Orho-Melander M. The PNPLA3 Ile148Met interacts with overweight and dietary intakes on fasting triglyceride levels. ***Genes Nutr***. 2014 Mar;9(2):388. doi: 10.1007/s12263-014-0388-4. Epub 2014 Feb 22.

VI. Ahmad S, **Rukh G**, Varga TV, Ali A, Kurbasic A, Shungin D, Ericson U, Koivula RW, Chu AY, Rose LM, Ganna A, Qi Q, Stančáková A, Sandholt CH, Elks CE, Curhan G, Jensen MK, Tamimi RM, Allin KH, Jørgensen T, Brage S, Langenberg C, Aadahl M, Grarup N, Linneberg A, Paré G; InterAct Consortium; DIRECT Consortium, Magnusson PK, Pedersen NL, Boehnke M, Hamsten A, Mohlke KL, Pasquale LT, Pedersen O, Scott RA, Ridker PM, Ingelsson E, Laakso M, Hansen T, Qi L, Wareham NJ, Chasman DI, Hallmans G, Hu FB, Renström F, Orho-Melander M, Franks PW. Gene × physical activity interactions in obesity: combined analysis of 111,421 individuals of European ancestry. ***PLoS Genet***. 2013 Jul;9(7):e1003607. doi: 10.1371/journal.pgen.1003607. Epub 2013 Jul 25.

VII. Ericson U, **Rukh G**, Stojkovic I, Sonestedt E, Gullberg B, Wirfält E, Wallström P, Orho-Melander M. Sex-specific interactions between the IRS1 polymorphism and intakes of carbohydrates and fat on incident type 2 diabetes. ***Am J Clin Nutr***. 2013 Jan;97(1):208-16. doi: 10.3945/ajcn.112.046474. Epub 2012 Dec 5.

VIII. Hertel JK, Johansson S, Sonestedt E, Jonsson A, Lie RT, Platou CG, Nilsson PM, **Rukh G**, Midthjell K, Hveem K, Melander O, Groop L, Lyssenko V, Molven A, Orho-Melander M, Njølstad PR. FTO, type 2 diabetes, and weight gain throughout adult life: a meta-analysis of 41,504 subjects from the Scandinavian HUNT, MDC, and MPP studies. ***Diabetes***. 2011 May;60(5):1637-44. doi: 10.2337/db10-1340. Epub 2011 Mar 11.

Abbreviations

ADP	Air displacement plethysmography
ADRB3	β-adrenergic receptor 3
AF	Atrial fibrillation
AMY1	Salivary amylase
ARID5B	AT rich interactive domain 5B
ASB	Artificial-sweetened beverages
BCDIN3D	BCDIN3 domain
BDNF	Brain derived neurotropic factor
BF%	Body fat percentage
BIA	Bioelectric impedance analyses
BMI	Body mass index
BMR	Basal metabolic rate
BP	Blood pressure
CETP	Cholesterylester transfer protein
CHD	Coronary heart disease
CHF	Congestive heart failure
CHO	Carbohydrates
CNV	Copy number variants
CT	Computed tomography
CVD	Cardiovascular disease
DBP	Diastolic blood pressure
ddPCR	Digital droplet polymerase chain reaction
DGKG	Diacylglycerol kinase
DOCK5	Dedicator of cytokinesis 5
DPP	Diabetes prevention program

DXA	Dual-energy X-ray absorptiometry
E%	Energy percent
EPIC	European prospective investigation of cancer
ETV5	Ets variant 5
FAIM2	Fas apoptotic inhibitory molecule-2
FBG	Fasting blood glucose
FFM	Fat free mass
FFQ	Food frequency questionnaire
FHS	Framingham heart study
FIL	Food intake level
FISH	Fiber-flourescence *in situ* hybridization
FM	Fat mass
FPG	Fasting plasma glucose
FTO	Fat mass and obesity associated gene
GBE1	Glucan branching enzyme-1
GEI	Gene-environment interaction
GIANT	Genomic investigation of anthropometric traits
GLACIER	Gene-lifestyle interactions and complex traits involved in elevated disease risk
GLM	Generalized linear model
GNPDA2	Glucosamine-6-phosphate deaminase-2
GPRC5B	G-protein coupled receptor, class C, group 5, member B
GRS	Genetic risk score
GWAS	Genome-wide association studies
HDLC	High density lipoprotein cholesterol
HFCS	High fructose corn syrup
HWE	Hardy Weinberg equilibrium
IARC	International agency for research on cancer
ICD	International classification of disease
IHD	Ischemic heart disease
INSIG2	Insulin induced gene 2
IRX3/5	Iroquois homeobox-3/5

IV	Instrumental variable
KCTD15	Potassium channel tetramerisation domain containing 15
LDLC	Low density lipoprotein cholesterol
LEP	Leptin
LEPR	Leptin receptor
LD	Linkage disequilibrium
Look AHEAD	Action for health in diabetes
LRP1B	Low density lipoprotein receptor-related protein 1B
LRRN6C	Leucine-rich repeat neuronal protein 6C
LTPA	Leisure time physical activity
MAF	Minor allele frequency
MAF	Musculoaponeurotic fibrosarcoma oncogene homolog
mCNV	Multiallelic copy number variant
MDC-CC	Malmo diet and cancer-cardiovascular cohort
MDCS	Malmö diet and cancer study
MI	Myocardial infarction
MMR	Multivariable Mendelian randomization
MR	Mendelian randomization
MRI	Magnetic resonance imaging
MTCH2	Mitochondrial carrier homologue-2
MTIF3	Mitochondrial translational initiation factor 3
MTNR1B	Melatonin receptor 1B
MVPA	Moderate to vigorous physical activity
NEGR1	Neuronal growth regulator-1
NHS	Nurses' health study
NPC1	Niemann-pick disease type C1
NRXN3	Neurexin-3
OR	Odds ratio
PC1	Prohormone convertase 1
PCSK1	Proprotein convertase subtilisin/kexin type 1
PLIN1	Perilipin 1

POMC	Proopiomelanocortin
PPYR1	Pancreatic polypeptide receptor 1
PRKD1	Protein kinase D1
PRT	Paralog ratio test
PTBP2	Polypyrimidine tract binding protein 2
PTER	Phosphotriesterase related
qPCR	Quantitative polymerase chain reaction
RCTs	Randomized controlled trial
RPGRIP1L	RPGR- interacting protein-1 like
RR	Relative risk
SBP	Systolic blood pressure
SEC16B	SEC-16 homologue-B
SH2B1	SH2B adaptor protein-1
SLC39A8	Solute carrier family 39, Member 8
SNP	Single nucleotide polymorphism
SSB	Sugar-sweetened beverages
T2D	Type 2 diabetes
TG	Triglycerides
TMEM18	Transmembrane protein-18
VHU	Västerbotten Halsö Undersökning
VLDLC	Very low density lipoprotein cholesterol
VNTR	Variable number tandem repeats
WC	Waist circumference
WHO	World health organization
WHR	Waist to hip ratio
ZNF608	Zinc finger protein 608

Introduction

Obesity is a serious health problem that has recently been increased dramatically in both children and adults, and no country to date has been able to reverse its obesity epidemic. Starting as the problem of high income countries, it is now recognized as equally affecting middle- and low income countries. According to World Health Organization (WHO), 65% of world´s population live in countries in which more people die of overweight and obesity compared to underweight. Obesity is a significant economic and health burden as it is associated with other comorbidities such as diabetes, cardiovascular disease (CVD), certain cancers, gynecological problems, osteoarthritis [1] and premature mortality [2].

In search for causes of the recent obesity epidemic, emphasis has been placed on the radical change in lifestyle during the last century promoting high intake of energy-dense foods and physical inactivity thus creating an 'obesogenic' environment which results in positive energy balance leading to weight gain. Yet, not everyone living in an obesogenic environment gets obese and genetic components play an important role in contributing to individual risk of obesity. The role of genes in the development of obesity has been established in several twin-, family- and adoption studies with heritability estimates ranging from 40-70% [3-5]. So far 97 body mass index (BMI) associated genetic variants have been identified which explain about 3% of the population variation in BMI [6] indicating that a large proportion of loci remain to be discovered. Thus an individual's risk of obesity is influenced by both genes and environment and possibly by interactions between the two. Despite an increased body of research in this area in the past few years, existing evidence is not strong enough to drive concrete policy recommendations for obesity prevention indicating a need for better phenotyped and well powered studies.

This thesis aims to investigate the role of gene-lifestyle interactions in the development of obesity as they may help to identify sub-groups of individuals in which effective strategies can be implemented for the prevention and treatment of obesity through individual genetic profiles. Further, this thesis investigates the effects of genetic variation on weight changes during different time points in life to understand how the genetic variants may vary in their association with weight changes at different ages. Finally, this thesis aims to explore causal links between cardiometabolic traits and mortality, as identifying the causal associations can decrease the disease burdens by improving prevention and treatment strategies.

Obesity

Definition

Generally, obesity can be defined as accumulation of excess body fat resulting from imbalance between energy consumption and expenditure [7]. However, no precise definition of 'excess' exists, and the degree of adiposity is a continuous trait without any clear division between normal and abnormal [8]. For practical purposes, obesity is defined as excess body weight rather than excess body fat, as direct measurement of body fat is more difficult than measuring weight. Thus, BMI, which is calculated as ratio of weight (in kilograms) to height squared (in meters), is commonly used to express height adjusted body weight both in children [9] and adults [10]. In most epidemiological studies, BMI is used as a surrogate marker of obesity. According to WHO, obesity is defined as BMI $\geq$30 kg/m^2 and overweight as BMI between 25 and 29.9 kg/m^2 [11].

Epidemiology

Prevalence of obesity has been increasing worldwide over the past three decades. Globally, the proportion of men having BMI >25 has increased from 28.8% in 1980 to 36.9% in 2013 and among women it has increased from 29.8% to 38.0% during the same period [12]. Overall, the combined prevalence of overweight and obesity worldwide has increased by 27.5% for adults and 47.1% for children between 1980 and 2013. The prevalence of overweight and obesity has increased both in developed and developing countries but the patterns are different in men and women. Men have higher rates of overweight and obesity in the developed countries whereas these rates are higher among women in the developing countries. However, the prevalence of obesity is higher among women both in the developed and developing countries. Surprisingly, according to 2013 estimates, more than half of the 693 million obese people in the world, live in just 10 countries including USA, China, India, Russia, Brazil, Mexico, Egypt, Pakistan, Indonesia, and Germany. Of the total, 13% of all obese people live in USA, 15% jointly in China and India, and 64% in developing countries. Largest increase in obesity rates in high income countries during 1980 -2013 have been observed in USA, followed by Australia and United Kingdom [12].

Obesity measurement methods

Accurate assessment of body composition is very important for both clinical and research settings. A number of established methods and techniques can be used to assess obesity such as anthropometry, densitometry, imaging and bioimpedance. Each method has its strengths and limitations, and scientific acceptability and appropriateness of each method depends upon the situation.

Anthropometry

Anthropometry is the measurement of body weight and dimensions including length, width, circumference and skin fold thickness [13]. Despite technological advances, anthropometric measurements are still widely used in large epidemiological studies as they are simple and inexpensive. Among various anthropometric measures, height and weight are measured with highest accuracy and precision and minimal technical error [14]. BMI is a simple and easy measure of overall obesity but cannot differentiate between fat mass (FM) and fat free mass (FFM). However, several studies have shown strong correlation between BMI and body fat percentage (BF%) [15, 16]. Validity of BMI as a marker of body fatness depends on age, sex and ethnicity [17]. Waist circumference (WC) and waist-to-hip ratio (WHR) are indirect measures of abdominal or central adiposity that are easier to obtain in large epidemiological studies, but the measurement procedure has greater between technician variability compared to weight and height [14]. Both WC and WHR have been validated against abdominal fat measured by magnetic resonance imaging and dual-energy X-ray absorptiometry [18, 19] and WHR has been shown not to be superior to WC alone in predicting abdominal obesity [20]. However, there is little evidence showing advantage of WC over WHR [18] and similar to BMI, same WC cutoff cannot be applied to all populations and ethnic groups [21]. Skinfold thickness is used as an indirect measure for body fat distribution because of two reasons: first, 40 to 60% of total body fat resides in the subcutaneous region and second, it can be directly measured with the help of a well-calibrated caliper. Skin fold thickness can be measured in more than 19 sites in the body [13]. Compared to other anthropometric measures they are less reproducible and more prone to interobserver errors [22].

Densitometry

Densitometry, also known as underwater or hydrostatic weighing, is a method to estimate body composition by means of total body density. It is based on the principle that fat is less dense than water, meaning that higher the amount of fat in the body, lower the density. For measuring body composition, densitometry has long been considered the “gold-standard” [23] because of its excellent precision and accuracy. However, this method is not suitable for children, older adults and

morbidly obese individuals as the procedure is complicated, time consuming and requires active cooperation of the participants [24]. The Air-displacement plethysmography (ADP) method is relatively quick and more comfortable and uses air instead of water for measuring body volume and density. It uses BodPod Body composition system (Life Measurement Instruments, Concord, California) and has become an excellent alternative to traditional densitometry, especially in younger children, pregnant women and morbidly obese individuals [25].

Radiation and imaging techniques

Dual-energy X-ray absorptiometry (DXA) is frequently used in clinical studies for estimating body composition and can provide estimates of the three components of the whole body including FM, FFM and bone mineral density. It is based on the principle that two x-ray beams, of different but very low energy, attenuate differently by different body tissues when passing through the body. The method has been extensively validated and is highly accurate and reproducible and because of very low radiation exposure can be used in children. However, it cannot be used in pregnant women, and the equipment is expensive and immobile, which further limits its use in large epidemiological settings. Further limitation is that DXA cannot accurately discriminate visceral fat from subcutaneous fat [24].

Computed tomography (CT) and magnetic resonance imaging (MRI) have excellent accuracy and reproducibility in measuring body composition at tissue and organ levels. Measurements obtained from CT and MRI can be classified into visceral adipose tissue, subcutaneous adipose tissue, interstitial adipose tissue and total adipose tissue [26]. MRI has an advantage over CT as it does not involve radiation exposure and can be used in pregnant women. Both methods are highly expensive and not readily accessible, which limits their use in large studies, but are methods of choice in calibration and validation of simple and inexpensive measures of body fat distribution.

Bioelectric impedance analysis

Bioelectric Impedance analysis (BIA) is used to estimate body composition by measuring resistance to a small electrical current passed across body tissues. The principle of BIA is based on the electrical conductive properties of the human body i.e., the higher the fatty tissue content, the greater will the resistance to the applied alternating current be; or the greater the lean body mass or water content of a person, the faster will the current pass through [26]. BIA can be used in large studies as the equipment is simple, inexpensive and portable. BIA works well in healthy subjects and in patients without significant fluid and electrolyte abnormalities, when using a validated BIA equation appropriate for age, sex and ethnicity. However, clinical use in subjects at extremes of BMI ranges or with abnormal hydration is not recommended. Technological advances in the past few decades have developed

multifrequency and segmental BIA techniques which provide more accurate measurement of body composition than single-frequency BIA [27]. By comparing BF% estimated by multifrequency BIA and DXA in healthy subjects, it has been shown that BIA is a good alternative for estimation of BF% in subjects within normal body fat range. However, BIA tends to underestimate BF% in obese subjects and overestimate it in lean subjects [28].

Obesity, cardiometabolic traits and mortality

The global increase in the incidence and prevalence of obesity is occurring in parallel with the increasing burden of CVD and metabolic diseases such as type 2 diabetes (T2D), hypertension and dyslipidemia.

Obesity and cardiovascular disease

Obesity is a major risk factor for CVD which includes coronary heart disease (CHD), myocardial infarction (MI), angina pectoris, congestive heart failure (CHF), stroke, hypertension and atrial fibrillation (AF) [29, 30]. Adverse effects of obesity on CVD have been confirmed in large prospective and observational studies. Association of obesity with CVD risk factors, and subclinical vascular disease (coronary artery calcium, carotid artery intimal medial thickness and left ventricular mass), was assessed among 6814 participants free of CVD at baseline in the Multi-Ethnic Study of Atherosclerosis. The study showed that hypertension, diabetes and subclinical vascular disease were more prevalent among obese compared to non-obese subjects [31]. In the Framingham Heart Study (FHS), the associated effects of obesity on the risk of CVD (stroke, MI, CHD and angina pectoris), hypertension, diabetes and hypercholesterolemia were prospectively evaluated. After 44 years of follow-up, the age adjusted relative risk (RR) for CVD was 1.46 among obese men and 1.64 among obese women as compared to non-obese [32]. Data from FHS has also shown that the lifetime risk of CVD is higher among individuals with diabetes and the risk is further accentuated by obesity status [33].

Despite being recognized as a risk factor for CHD, obesity is not included in the global risk assessment tool, such as the Framingham risk score [34] because the consequences of obesity have been thought to be entirely mediated through the established risk factors like diabetes, dyslipidemia and hypertension. Recently, a large study by Lu *et al.* evaluated the magnitude of the effect of BMI on CHD and stroke that is mediated through blood pressure (BP), glucose and cholesterol, by using data from 97 prospective cohorts with 1.8 million participants [35]. The results

of this study suggested that 46% (95% CI: 42 to 50) of the excess risk of BMI for CHD and 76% (65 to 91) for stroke is mediated by these three factors, with BP being the most important mediator explaining 31% (28 to 35) of the excess risk of CHD and 65% (56 to 75) of the stroke. In addition, compared to normal weight individuals, these three mediators were observed to mediate 50% (44 to 58) of the excess risk of overweight and 44% (41 to 48) of the excess risk of obesity for CHD. The corresponding values for stroke were 98% (69 to 155) for overweight and 69% (64 to 77) for obesity [35]. Moreover Lu *et al.*, by analyzing data from 9 prospective cohorts, concluded that the metabolic mediators (BP, glucose and cholesterol) explain about half of the adverse effects of BMI on CHD whereas the contribution via the inflammatory (C-reactive protein) and prothrombotic (fibrinogen) mediators is much smaller [36]. These results suggest that the metabolic mediators are important in a pathway between obesity and CHD, but that the excess risk cannot entirely be explained by them. It is thus very important from both clinical and public health point of view to add obesity into a CHD risk assessment tool.

Obesity paradox: Obesity is an independent risk factor for CHF [37], but in patients with established CHF and other chronic diseases, obesity has been associated with lower mortality. This phenomenon is known as "obesity paradox" [38]. A recent meta-analysis has shown that the risk of hospitalization, CVD- and total mortality was highest among underweight and lowest among overweight CHF patients [39].

The relative importance of BMI and body fat distribution varies with age, sex and ethnicity. Independent of BMI and other CVD risk factors, body fat distribution measured by WC and WHR has also shown association with CHD and stroke [40-43]. Moderate weight gain independent of BMI during young adulthood (since age 21 for men and 18 years for women) is associated with increased risk of CHD and stroke at young age [44, 45]. All obesity measures (BMI, WC and weight gain since young adulthood) are thus very important in assessing the relationship between adiposity and CVD.

Obesity and hypertension

Obesity stands out as the major preventable contributor to hypertension. The prevalence of hypertension substantially increases with increasing BMI and hypertension greatly increases the risk for CVD. The obesity attributable burden of hypertension is very high and estimated to be approximately 80% and 60% for men and women, respectively [46]. The National Health and Nutrition Examination Survey (NHANES) observed that when compared to normal weight individuals, the odds ratio (OR) for hypertension was 1.7 for overweight, 2.6 for individuals with BMI 30.0-34.9 kg/m^2, 3.7 for those with BMI 35.0-39.9 kg/m^2 and 4.8 for those with BMI ≥40 kg/m^2 [47]. Furthermore, data from the prospective Nord–Trondelag

Health Study (HUNT) reported that changes in BMI were significantly associated with changes in BP and that the risk of hypertension increased among those who increased their BMI [48].

Obesity and type 2 diabetes

Obesity is a major risk factor for T2D [49]. In a recent study from FHS cohort, where they looked at the trend in incidence of diabetes over the four decades (1970-2000), diabetes incidence remained highest among obese individuals [50]. In line with this, results from the Nurses' Health Study (NHS) [51] and the Health Professional Follow-up Study [52] have shown more than a tenfold increased risk of developing T2D among both men and women with BMI >35kg/m^2 and >29kg/m^2, respectively, compared to individuals in lower BMI categories. Apart from BMI, WC and WHR are independent risk predictors of T2D. In NHS, after adjusting for BMI, RR for T2D was 5.1 (95% CI: 2.9 to 8.9) for WC and 3.1 (2.3 to 4.1) for WHR for the participants in the 90th percentile of these traits compared to those in the 10th percentile. In addition, moderate weight gain during adulthood has been associated with increased risk of diabetes. In data from NHS, Colditz *et al*. compared the women who kept their weight stable (±5kg), between the age of 18 years and the baseline examinations in 1976, with the women who gained 5.0-7.9 kg, 8.0-10.9 kg and ≥20.0 kg, and the corresponding RRs for diabetes were 1.9 (95% CI: 1.5 to 2.3), 2.7 (2.1 to 3.3) and 12.3 (10.9 to 13.8), respectively. Additionally, the risk for diabetes was reduced to 50% or more among women who lost more than 5.0 kg of their weight [49]. Reduced risk of T2D in relation to weight loss has also consistently been observed in randomized controlled trials (RCTs) like the Diabetes Prevention Program (DPP) [53] and the Finnish Diabetes Prevention Study [54]. Mendelian randomization studies have provided further support for a direct causal association between BMI and T2D [55, 56].

Obesity and dyslipidemia

Dyslipidemia related to obesity is characterized by high levels of triglycerides (TG), low levels of high density lipoprotein cholesterol (HDLC) and abnormal composition of the low density lipoprotein (LDLC) particles (small dense LDL particles with normal or slightly elevated levels of LDL cholesterol) [57]. Dyslipidemia is an important component of the metabolic syndrome [58] and plays an important role in the development of CVD [57]. However, the link between obesity and dyslipidemia is complex and although evidence has been presented suggesting insulin resistance as the underlying mechanism [58], more studies are needed to distinguish between the role of insulin resistance and body fatness for the

lipid and lipoprotein profile. All the components of dyslipidemia have been associated with atherogenicity, and weight loss and exercise have been shown to reduce the risk of CVD by improving the atherogenic lipid/lipoprotein profile [57].

Obesity and mortality

Association between obesity and mortality has been well established [59], majority of the studies using BMI as the measure of obesity. Until quite recently, the nature of this relationship has however, remained unclear due to inconsistent results between different studies: everything from U-shaped and J-shaped to linear relationship have been described [60, 61]. In 2009, the Prospective Studies Collaboration published a large study including 57 studies with almost 900,000 participants, originating mainly from Western Europe and North America with a median follow-up of 8 years. This study observed lowest mortality among individuals within a BMI range between 22.5 to 25 kg/m^2. BMI both above and below this normal range associated with higher overall mortality thus strongly supporting the U-shape association between BMI and mortality [62]. Further, a recent meta-analysis of 97 studies with 2.88 million individuals and 270,000 deaths reported that overall obesity (BMI ≥30) as well as grade 2 and 3 obesity (BMI ≥35) associated with higher mortality, while overweight (BMI 25 to <30) associated with lower total mortality, compared to normal weight (BMI 18.5 to <25). Surprisingly, grade 1 obesity (BMI 30 to <35) was not found associated with higher mortality [63] suggesting contribution of mainly higher levels of BMI to excess mortality. However, there are several discrepancies in the findings of the studies that have investigated association between BMI and mortality such as contrasting associations of overweight with mortality and wide variation in the estimated numbers of obesity associated deaths in different studies [64-66]. This lead to methodological challenges in analyzing the relationship between BMI and mortality including effect modification by age, confounding by smoking, reverse causation, over-adjustment for intermediate variables (T2D, hypertension and dyslipidemia) and imperfect measures (over and under reporting in self-reported data) of adiposity [61, 67-69]. Apart from this, intentional weight loss has shown to be associated with decreased mortality while unintentional weight loss associated with increased mortality among overweight and obese adults [70]. However, healthy dietary and lifestyle behaviors may confound the relationship between intentional weight loss and mortality, and morbidity related weight loss (for example in cancer) can confound the relationship between unintentional weight loss and mortality.

Apart from BMI, several studies have demonstrated an important role of abdominal or central obesity (measured as WC and/or WHR) in predicting mortality. In NHS, both WC and WHR showed strong association with total-, CVD- and cancer mortality independent of BMI. Moreover, WC associated with increased CVD

mortality even among normal weight women [71]. Thus all three measures (BMI, WC and WHR) are important to be able to comprehensively evaluate the impact of obesity on mortality.

Lifestyle factors that associate with obesity

The prevalence of overweight and obesity is increasing and it is now well established that rapid globalization of westernized lifestyle is fueling this growing problem. However, lifestyle factors are not the only culprit as both genetic and environmental factors play major roles in weight gain and obesity. The lifestyle factors that have major impact on obesity are diet and physical activity and both of these factors are in turn influenced by genetic traits [72].

Diet

Diet plays a major role in weight control but it is still unclear how and which specific dietary factors, apart from excess calories, are important for weight gain. One major contributing change over time worldwide has been the general increase in portion sizes that lead to increased energy intake and subsequent weight gain [73]. Even small positive daily energy balance, whatever dietary factors are behind the excess energy intake, results in weight gain and contribute to increased risk of obesity overtime.

Dietary fat intake: Because of high density and high palatability of high-fat foods, it is generally believed that high fat intake lead to weight gain and obesity. However, the evidence on the relationship between fat intake and obesity, based on both epidemiological studies and clinical trials, has remained controversial. Cross-sectional studies have suggested a positive association between dietary fat concentration and relative weight, results from prospective studies of diet in relation to subsequent weight change have been inconsistent, and intervention studies have provided evidence for a consistent but short lived effect of low fat diets on weight loss [74]. A review evaluating RCTs of low fat diets on weight loss suggested that such diets are not better than other calorie restricted diets in achieving long term weight loss in overweight or obese subjects [75].

Dietary carbohydrate intake: Despite decreasing fat content in diet, prevalence of obesity has increased in both USA and Europe which has drawn attention to the alternative hypothesis that the corresponding increase in carbohydrate (CHO) content may be the reason behind the obesity epidemic [76]. Cross-sectional studies have shown negative association between CHO intake and BMI but this relationship

may reflect confounding by health-conscious behaviors used to control weight [77]. Recently, CHO restriction has been promoted as an alternative strategy for weight loss. A meta-analysis of five RCTs, comparing the effects of ad libitum low-CHO with low-fat calorie restricted diets on weight loss and CVD risk factors, reported that low-CHO resulted in higher weight loss after 6 months but not after 12 months, compared to low-fat diets [78]. A recent RCT has shown that low-CHO diet, resulted in higher weight loss, and similar or greater improvement in inflammation, adipocyte dysfunction and endothelial dysfunction compared to standard low-fat diet among obese subjects [79].

Dietary protein intake: There is some evidence that high-protein diets may facilitate weight loss due to their association with greater satiety and lower energy intake compared to low-protein diets [80]. Protein intake is inversely associated with abdominal obesity in cross-sectional studies [81] but prospective studies in this regard are lacking. A randomized fat reduced weight loss trial comparing protein vs CHO in ad libitum diet found that replacement of some dietary CHO by proteins in such diet improved weight loss as well as adherence to low-fat diet [82].

Sugar-sweetened beverages: Consumption of sugar-sweetened beverages (SSB) has increased in parallel to the increase in overweight and obesity. A great body of epidemiological and experimental evidence have shown that a greater consumption of SSB is associated with weight gain and obesity [83]. In addition to SSB, consumption of high-fructose corn syrup (HFCS) has greatly increased in US representing >40% of the caloric sweeteners in the soft drinks [84]. The digestion, absorption and metabolism of fructose differ largely from glucose as it does not stimulate insulin secretion or enhance leptin production which are the key signals in regulation of food intake and body weight [85]. Furthermore, as compared to glucose, it can be easily incorporated into triglycerides backbone (glycerol) during fat synthesis, thus facilitating synthesis of fatty acid [85].

Physical activity

The role of physical activity in the regulation of body weight has long been recognized. There is an inverse relationship between physical activity and adiposity [86]. In USA, trend for leisure-time physical activity (LTPA) has been stable or slightly increased over time compared to substantial decline in physical activities related to work, household and transportation. In addition, sedentary behaviors like television watching and computer use have substantially increased resulting in an overall decline in total physical activity [87]. A cross-sectional study in Australian workers has reported that both leisure time sitting and occupational sitting independently associate with obesity risk but the higher risk for obesity was observed for leisure time sitting [88]. Recently, International Physical activity and

the Environment Network study observed almost linear negative association between BMI and 0-50 min/day of moderate to vigorous physical activity (MVPA) measured by accelerometer, but this relationship was weakened at higher levels of MVPA and supports the current recommendation of Institute of Medicine that 60 minutes of MVPA everyday prevents weight gain in normal weight adults. Additionally, country- and gender dependent relationship between physical activity and BMI has been reported [89]. Hemmingsson *et al*. studied the impact of obesity on the relationship between physical activity and BMI and observed a weak association among non-obese subjects but a highly significant association among obese subjects [90]. A recent meta-analysis of 45 RCTs on long term weight loss with non-surgical interventions in obese adults has shown that interventions that deal with both diet and physical activity have small but significant benefits on weight loss maintenance [91].

Smoking

Smoking is an important contributing factor for obesity mainly because of the association of smoking with lower weight and BMI, and of smoking cessation with weight gain [92, 93]. Smoking reduces appetite and increases basal metabolic rate by its thermogenic effects resulting in lower weight [94]. Although the underlying mechanisms relating smoking cessation to weight gain is not clear, nicotine has been considered the thermogenic agent that effects the peptides involved in the feeding behavior [95]. Data from large epidemiological studies have shown that smoking cessation is a plausible contributor to increasing rates of overweight and obesity [92]. However, the interaction between smoking and overweight in terms of CVD is frightening as the risk of coronary disease (MI and CHD) among non-smokers and smokers with BMI >29 increases to 2-fold and 12-fold, respectively, compared to their normal weight counterparts [96].

Genetic contribution to obesity

Obesity is a complex multifactorial condition with an important genetic component. Evidence for genetic contribution to obesity comes from both descriptive epidemiological and heritability studies.

Descriptive epidemiological studies

Epidemiological studies of families and migrants provided the first evidence of a genetic contribution to susceptibility of obesity. Data from twin and family studies have suggested that an individual with a family history of obesity has 1.5-5 times higher risk of obesity compared to the risk in population at large [97-100]. The familial risk increases with the increase in degree of relatedness and doubles if the related individual is extremely obese (BMI≥45 kg/m^2) [98]. However, the increased familial risk cannot be entirely attributed to a shared genetic background as the shared non-genetic factors also contribute, but to a lesser extent [97].
Studies investigating migrants are based on comparing the risk of disease in migrants to the risk of disease in individuals in the country of their origin, and to the native born population of the country where they have migrated. A classic example of this are American Indians (Pima Indians) living in central and southern Arizona (U.S) and in Sonora (Mexico). In Arizona, despite of living in the same 'obesogenic environment', Pima Indians have double prevalence of obesity (69%) compared to the white American of European descent (33%), suggesting a higher genetic susceptibility to obesity in Pima Indians [101]. Despite sharing the same genetic background with those living in Arizona, Pima Indians living in the 'restrictive' original environment of Mexico have much lower prevalence of obesity (13%) [102], suggesting interaction between genetic susceptibility to obesity and lifestyle factors.

Heritability studies

Heritability can be defined as the proportion of phenotypic variation among individuals in the population due to genetic contribution. Heritability of obesity has now been widely accepted as between 40%-70% based on the estimates from twin-, family- and adoption studies [3-5]. This wide range of heritability estimates is partly explained by the study design as the estimates from twin studies (47%-90%) are higher than those from family- (24%-81%) or adoption studies (20%-60%) [4, 103], although the range is wide also within studies of similar design. Furthermore, heritability estimates are population specific (e.g. the heritability estimated from a population with little variation in environmental factors will be higher than that of a population with diverse lifestyles) and may vary with age, which could also explain the large variation in reported estimates. Thus, the reported estimates should be interpreted after taking into account both study design and population. A review of twin- and adopted children studies suggested that genetic factors have much stronger effect than environmental factors on BMI in children up to the age of 18 years [104].

Approaches to identify human obesity genes

Monogenic obesity

First insights into the genetics of obesity come from studies of single gene disorders. Monogenic obesity is caused by a single mutation that occurs *de novo* or segregates in the family and disrupts the expression of the gene in which it is located, or the function of the gene product. Several different monogenic forms of obesity have been described, of which the first, caused by mutations in the leptin gene (*LEP*) leading to severe obesity, was identified in 1997 [105]. Later on reversal of severe obesity upon administration of recombinant leptin in leptin deficient patients proved that a therapy can be highly effective if a clear molecular basis for an individual's obesity is identified [106]. Apart from leptin, several obesity-causing genes have been identified mostly within the leptin-proopiomelanocortin (*POMC*)-melanocortin pathway, such as leptin receptor (*LEPR*), *POMC*, prohormone convertase 1 (*PC1*) and melanocortin receptor 4 (*MC4R*) [107] as shown in Figure 1. However, despite providing valuable insights into biological pathways and mechanisms leading to excessive weight gain, monogenic forms of obesity are rare and affect only a small fraction of population. The most common are *MC4R* mutations with a prevalence of at least 0.05% in normal weight population, 0.5%-1% among obese adults and 1%-6% among obese children [108].

Identification of genetic variation involved in common forms of obesity applicable to general population proved to be an arduous task. In order to identify genetic variants associated with common obesity in the general population, two main approaches have been used: hypothesis-driven approaches by using candidate gene studies, and the hypothesis-generating approaches by using genome-wide screening studies.

Candidate gene studies

As the name indicates, candidate gene approach is hypothesis-driven and depends upon the currently available information related to the biology and pathophysiology underlying obesity disease or a trait.

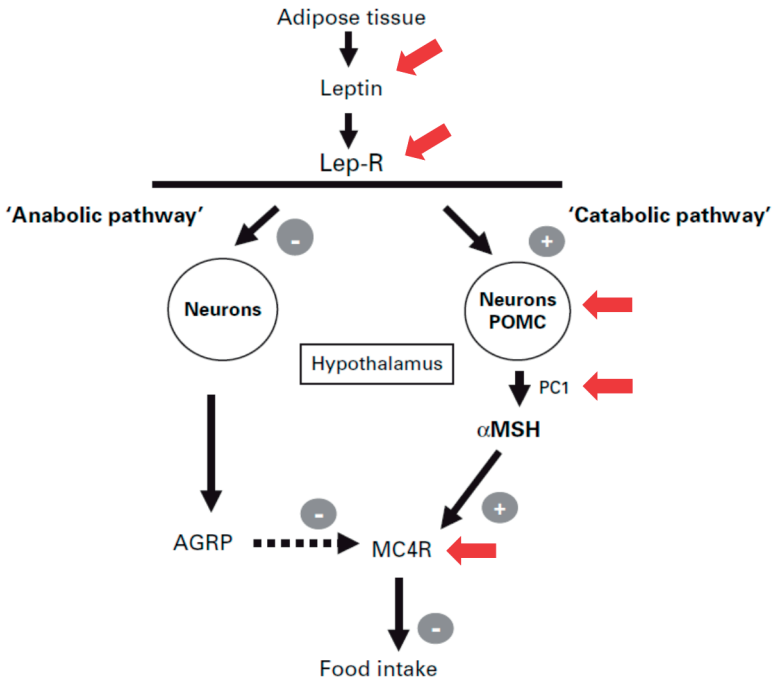


Figure 1. Leptin and melanocortin pathways
Lep-R, leptin receptor; POMC, proopiomelanocortin; α-MSH, α-melanocyte-stimulating hormone; AGRP, agouti-related protein; MC4R, melanocortin-4 receptor; PC1, proconvertase 1; red arrow, location of mutations responsible for monogenic obesity in humans; dashed arrow, AGRP is a natural antagonist of MC4R; +, pathway activated; –, pathway inhibited. Reproduced with permission from the publisher [109].

A candidate gene can be either a functional or a positional candidate. Functional candidates can for example be genes known or thought to be involved in regulation of energy balance based on animal models or human physiology, or genes that have been implicated in monogenic/extreme obesity. Positional candidate genes are based on locations within the genomic regions implicated by genome-wide linkage analyses (see below) to be connected to obesity phenotypes [110]. Once a candidate gene is selected, genetic variations in the gene need to be identified and tested for association with obesity related traits at the population level. There are various forms of genetic variations including single nucleotide polymorphisms (SNPs), small insertions/deletions (indels), copy number variants (CNV) and microsatellites. Of all these, SNPs, that are most common and account for >90% of the genetic variation in the human genome, have most commonly been used in candidate gene studies. Most robust associations have been observed for non-synonymous variants (the variants that bring about a change in the amino acid sequence) in the *MC4R*, β-adrenergic receptor 3 (*ADRB3*), proprotein convertase subtilisin/kexin type 1 (*PCSK1*), brain-derived neurotrophic factor (*BDNF*), melatonin receptor 1B

(*MTNR1B*) genes and for a functional variant near the lactase gene (*LCT*) [111]. The main reasons for the very limited success of candidate gene studies for obesity performed before the genome-wide association studies (GWAS) era initiated in 2007, include the small sample sizes that made them insufficiently powered to detect modest effect sizes expected for common obesity, incomprehensive coverage of the genetic variants in the genes of interest due to high sequencing and genotyping costs, focus in protein-coding regions in identification and analysis of polymorphisms, and the selection of candidate genes based on the very limited knowledge about biological mechanisms.

Genome-wide linkage studies

Genome-wide linkage approach is a hypothesis-generating method that relies on studies in families or affected siblings and originally tests whether certain chromosomal regions co-segregate with a trait or disease across generations [111]. However, as the genetics of multifactorial diseases is complex, with among other things, unknown or low penetrance and lack of Mendelian inheritance structures, it is not possible to define unaffected family members as these may get affected later on. Therefore, the most commonly used approach was affected-only analysis, based on identifying chromosomal regions shared by the affected family members more often than expected by chance. For this purpose, around 300-500 microsatellite markers, that are highly polymorphic, were genotyped through all human autosomes. Because of low power mainly due to low resolution and limited number of affected families and family members, genome-wide linkage approach identified broad intervals that covered usually very many genes and usually further fine mapping of the region could not successfully define the regions of interest. In addition, the statistical significance of the identified regions, was commonly not very strong, and difficult to replicate. First genome-wide linkage study on body fat percentage in Pima Indians was published in mid-1990s [112] and after that a number of chromosomal loci were linked to obesity-related traits. Human obesity gene map reported identification of 253 loci of interest from 61 genome-wide linkage scans, of which only 15 had some evidence of replication in at least three studies [113]. Furthermore, a meta-analysis of 37 genome-wide linkage scans comprising of 10,000 families of European origin with 31,000 participants, despite having sufficient power to identify loci with small effects, could not pinpoint a single BMI or obesity locus with convincing evidence [114]. Thus, genome-wide linkage approach was not found to be an effective approach for identifying genetic variants associated with multifactorial diseases as common obesity.

Genome-wide association studies

GWAS are based on hypothesis-generating approach that test the association between millions of SNPs and a particular disease in a case-control set-up, or a quantitative trait in population, by screening the entire genome at a much higher resolution compared to genome-wide linkage studies. GWAS has replaced the genome-wide linkage approach and as it does not rely on families, very large sample sizes can be achieved and the high resolution screening helps to narrow down and identify the associated locus and variants [111]. GWAS became possible thanks to sequencing of the human genome through the Human Genome Project [115], the identification of common genetic variants and linkage disequilibrium (LD) mapping by the International HapMap Consortium [116] and more recently the 1000-Genomes project [117], together with advances in technological development of comprehensive, affordable and high-throughput genotyping technologies [118].

GWAS is based upon the principle of LD (nonrandom association between alleles at different loci) at the population level. Loci that are physically closer (physical distance is measured in base pairs) together on a chromosome exhibit stronger LD than those that are farther apart. LD structure also varies with the size of population, number of generations a population lived and ancestry such as older African population has smaller LD regions because of more recombination events compared to younger European or Asian populations. The genomic distance at which LD decays represent the number of genetic markers required to 'tag' a haplotype and the number of these tagging markers is usually much lower than the total number of segregating variants within the population. For example, nearly 500,000 common SNPs [minor allele frequency (MAF) > 1%] are sufficient to tag common variation in a non-African population, despite the fact that total number of common SNPs exceeds 10 million [119].

GWAS typically comprises of two stages; a discovery stage and a replication stage. In the discovery stage, millions of common genetic variants (e.g., SNPs with MAF $\geq$1) are tested for association with the trait or disease of interest. The SNPs that show significant association are further tested for association in independent samples in the replication stage to confirm or refute the findings from the discovery stage. To account for multiple testing, a stringent *P*-value of $<5x10^{-8}$ (corresponds to a 5% genome-wide type I error rate) is considered as the minimum threshold to be reached after validation of association in the replication stage [120].

GWAS discoveries

Since the introduction of GWAS approach in 2005 [121], a number of GWAS on obesity and related traits have been conducted predominantly in European

populations. However, the number of GWAS in African and Asian populations is also growing. The very first locus reported in a GWAS of obesity was Insulin induced gene 2 (*INSIG2*) that was found to be associated with obesity in FHS [122] but was not identified in subsequent larger meta-analyses. Year 2007 mark the first revolutionary breakthrough in the genetics of common obesity when three independent studies simultaneously reported a strong association between genetic variants within the fat mass and obesity associated gene (*FTO*) and BMI and obesity [123-125]. The association of *FTO* with obesity and related traits has been replicated in several populations both in children and adults. To date >500 studies have examined the association of *FTO* with obesity and >60 SNPs in this gene have shown significant associations with obesity [126]. All BMI-associated *FTO* SNPs lie within a 47 kilobase (kb) LD block encompassing parts of the first two introns and exon 2 of *FTO* and are highly correlated (LD of r^2 >0.8) [127].

Soon scientist realized the need of collaborative efforts to increase sample size and thus power of the studies to identify more common variants with effect sizes smaller than that of *FTO* variants. The GIANT (Genomic Investigation of Anthropometric Traits) consortium was formed by collaboration between the research groups from Europe and USA, and data from seven GWAS for BMI comprising of ~17,000 individuals was meta-analyzed. This first joint effort resulted in identification of variants in the melanocortin-4 receptor gene (*MC4R*) associated with BMI, in addition to confirmation of *FTO* as a BMI locus [128]. Represented by rs17782313 SNP, this newly identified locus was associated with obesity among both children and adults [128]. Simultaneously, another GWAS in Indian Asians found significant association of common variation near *MC4R* with obesity [129]. Thus, in addition to the well-established role of rare mutations within *MC4R* in the development of extreme monogenic obesity, GWAS provided convincing evidence that also common variation in *MC4R* contributes to the susceptibility of obesity in population.

In their second study in 2009, GIANT consortium meta-analyzed data from 15 GWAS of BMI in Caucasians comprising of 32,387 individuals in the discovery stage and 35 significant SNPs were taken forward for replication in ~59,000 individuals [130]. This effort resulted in identification of six new loci associated with BMI where the associated SNPs were located near or in the genes encoding the neuronal growth regulator-1(*NEGR1*), the transmembrane protein-18 (*TMEM18*), SH2B adaptor protein-1 (*SH2B1*), the glucosamine-6-phosphate deaminase-2 (*GNPDA2*), the potassium channel tetramerisation domain containing-15 (*KCTD15*) and the mitochondrial carrier homologue-2 (*MTCH2*), in addition to *FTO* and *MC4R*. At the same time, another meta-analysis was published including four GWAS for BMI comprising of ~30,000 individuals of European and 1160 of African American origin by an Icelandic company, the deCODE Genetics [131]. A total of 43 most significant signals were taken forward for replication in ~5,500

Danish samples and for further confirmation in the discovery stage data of the GIANT consortium. In addition to *FTO* and *MC4R*, eight loci reached genome-wide significance in this study. Four of these loci were novel and had BMI associated SNPs in genes encoding the SEC-16 homologue-B (*SEC16B*) and *BDNF*, and between genes encoding the ets variant-5 (*ETV5*) and the diacylglycerol kinase (*DGKG*) genes and between the BCDIN3 domain (*BCDIN3D*) and the Fas apoptotic inhibitory molecule-2 (*FAIM2*), while the remaining four loci (*NEGR1*, *TMEM18*, *SH2B1* and *KCTD15*) were also identified by the GIANT consortium [131].

In 2010, the GIANT consortium further expanded the discovery stage sample to ~124,000 individuals from 46 studies, and 42 most significant SNPs were taken forward for replication in another set of ~126,000 individuals of white European descent [132]. They confirmed all of the 12 previously identified BMI-associated loci as well as two previously identified WC-associated loci [near the transcription factor AP-2 beta gene (*TFAP2B*) [133] and the neurexin-3 gene (*NRXN3*) [134]], and additionally revealed 18 new BMI-associated loci [132]. Thus, all the GWAS efforts identified 32 loci unequivocally associated with BMI by the end of 2010.

In their most recent meta-analysis in 2015, the GIANT consortium included GWAS and Metabochip data from 125 studies of European (~322,000 individuals) and non-European ancestries (~17,000 individuals) [6]. Analyses in European individuals identified 77 BMI-associated loci while inclusion of non-European individuals identified additional 10 loci. Secondary analyses (such as European sex-specific and population-based analyses) revealed 10 more loci. Thus, overall this effort yielded 97 BMI-associated loci of which 56 loci were novel and 41 had previously been associated with one or more obesity traits [132, 135-139]. Comparison of the effect estimates of these 97 BMI associated SNPs across ancestries revealed two loci [*SEC16B* and zinc finger protein 64 (*ZFP64*)] that showed evidence for heterogeneity between sexes and three loci [*NEGR1*, protein kinase D1 (*PRKD1*) and glucan branching enzyme 1 (*GBE1*)] that showed evidence for heterogeneity between ancestries. Apart from these, the effect estimates of 79% of the BMI-associated SNPs in Africans and 91% in East Asians showed directional consistency with Europeans suggesting that common BMI-associated SNPs have comparable effects across ancestries [6].

The role of GWAS in unraveling common variants in complex diseases like obesity has been remarkable, but the so far identified loci still explain only 2.7% of the population variation in BMI and majority of the genetic variation remain unexplained. For most of the identified loci, there is uncertainty over which exact SNP is causal given the large number of variants in LD and further functional studies and fine mapping is required to identify causal variants. Clinical relevance and functional mechanisms of even the best-documented “hits” still remain question marks. Moreover, the vast majority of potentially causal candidate SNPs are located

in non-coding regulatory regions of the genome and provide very limited information about which cell types may be most relevant for a SNP, further limiting the identity and function of the causal SNP.

For understanding this complexity, the *FTO* locus, which has been consistently identified by GWAS and contains intronic SNPs that strongly associate with obesity in diverse populations provides an excellent example. The evidence for that *FTO* is highly expressed in hypothalamus and controls appetite and energy expenditure suggested it to be the relevant target gene. However, the link between the intronic SNPs and *FTO* activity has remained unclear [140]. In 2014, two studies [141, 142] provided convincing evidence that obesity associated SNPs in the *FTO* region appear to be functionally connected with two neighboring genes; iroquois homeobox 3 (*IRX3*) and RPGR-interacting protein-1 like (*RPGRIP1L*) rather than *FTO* itself as shown in Figure 2.

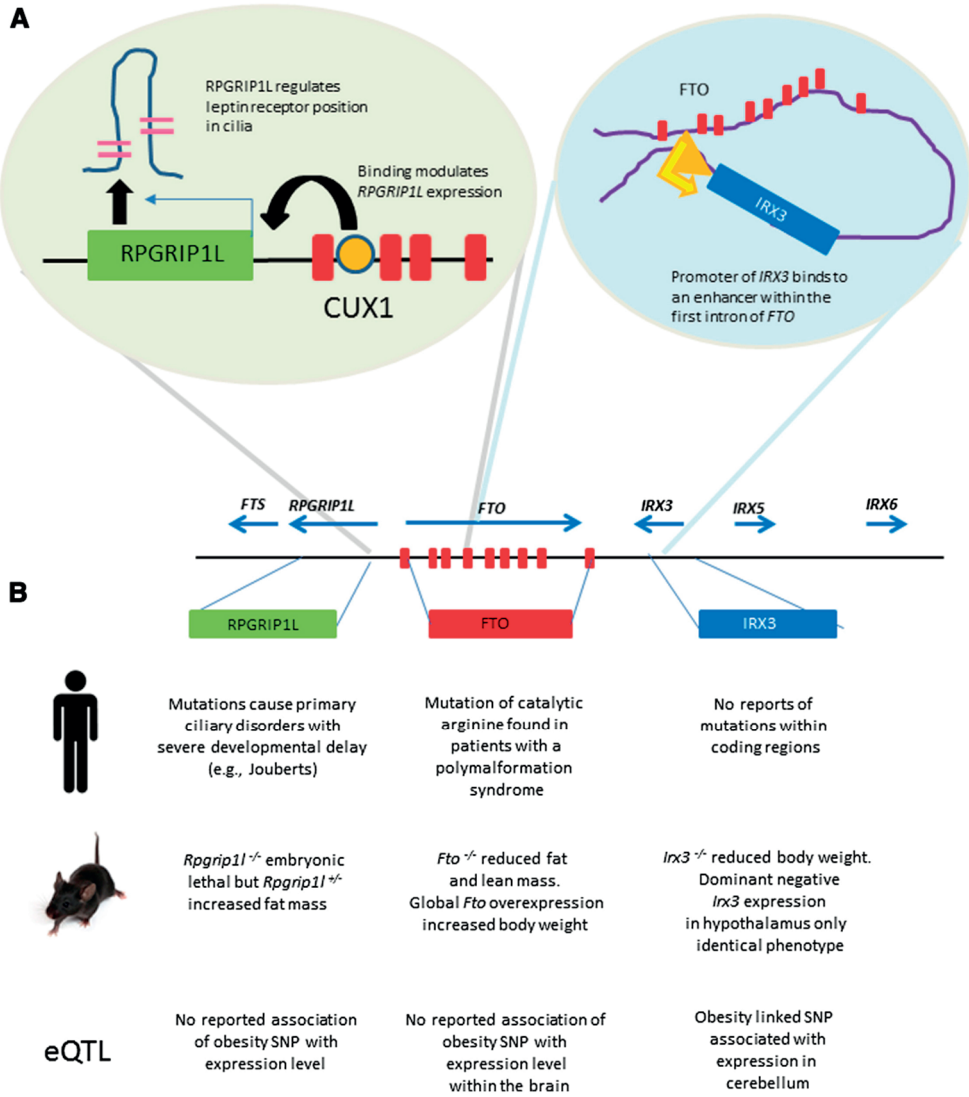


Figure 2. Increasing functional complexity around the *FTO* locus
(A) Emerging data indicate that the links between intronic variance within *FTO* and body composition are mediated through funstional interactions with neighbouring genes. The first intron of *FTO* contains a binding site for the transcription factor *CUX1* (cut-like homeobox 1) which, through regulation of *RPGRIP1L* expression, modulates leptin receptor localization within neurons. This intron also contains an enhancer sequence which directly binds to the promotor of *IRX3*. (B) Summary of data on *FTO*, *RPGRIP1L* and *IRX3* from human genetic and mouse model studies. Data on *IRX3* are notable in that eQTL (quantitative trait loci) mapping demonstrates an association of obesity-linked SNPs with *IRX3* expression. Adapted from Tung *et al*. [143] with permission from the publisher.

In 2015, Claussnitzer *et al*. [144] made further progress in the field by identifying the causal SNP and proposing a mechanism by which this SNP could affect body weight as shown in Figure 3. Similar to Smemo *et al*. [142], Claussnitzer *et al.*

provided compelling evidence that SNPs in this region associate with the expression of *IRX3* and /or iroquois homeobox 5 (*IRX5*) in human cells but not with *FTO* [144]. Taken together, the cell type and mechanism by which the *FTO* variant affects body weight is still unsettled but these studies demonstrated that only by combining bioinformatics and experimental approaches, biology associated with GWAS loci can be elucidated.

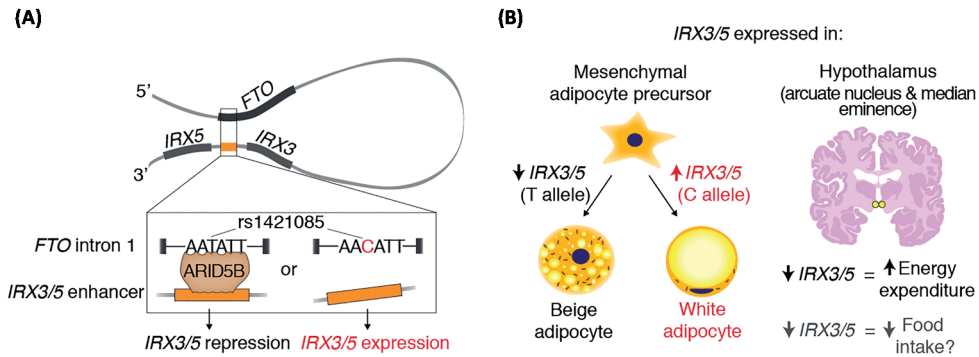


Figure 3. Variants in the *FTO* locus regulate *IRX3/5* expression to exert effects on body weight
(A) Variants in the *FTO* locus regulate the expression of *IRX3* and *IRX5* via binding the transcription factor *ARID5B* (AT rich interactive domain 5B). (B) The consequences of altered *IRX3/5* expression may be manifested in the adipose lineage, with altered beige fat development, and/or in the hypothalamus, via changes in food intake and energy expenditure. Adapted from Herman and Rosen [145] with permission from the publisher.

Thus, the puzzle of common obesity genetics cannot be solved through a single approach and technological advances enabling the sequencing of entire genomes to identify rare variants with potentially large effects at affordable prices are needed to identify elusive obesity associated genes, causal mutations, pathways and biological mechanisms.

Copy number variants in obesity

Copy number variants (CNV) are defined as genomic structural variations in which one kilobase to several megabases long segments of DNA are either deleted, duplicated or multiplicated. Thus, CNVs can be simple bi-allelic, deletions and duplications or more complex, multiallelic variants (mCNV) [146] as shown in Figure 4.

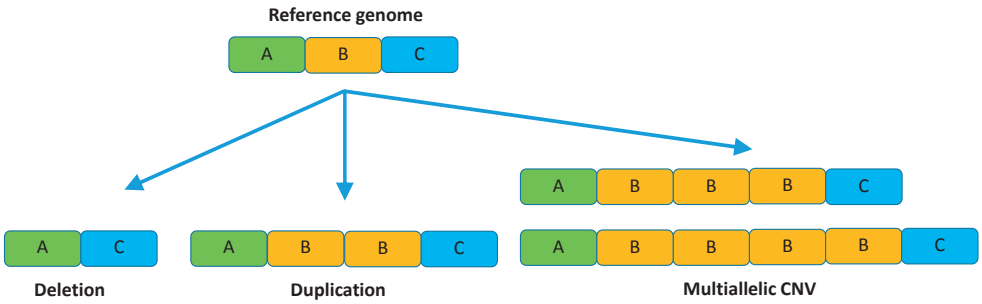


Figure 4. Copy number variant types
Copy number variants may consist of simple deletions and duplications of DNA segments or they may vary in the number of copies of the duplicated region such as multiallelic CNV.

CNVs are widespread throughout the genome and are common variations in human population [146]. The highest resolution genomewide CNV discovery study carried out todate [147] using oligonucleotide microarrays comprising of 42 million probes, generated a comprehensive map of 11,700 CNVs greater than 443 bp of which 8,599 have been independently validated. The validated CNVs cover 3.7% of the genome with a median CNV size of 2.7 kb, and a median number of 1117 CNVs were found in Europeans and 1488 in Yoruban Africans. Further investigation of ~5,000 validated CNVs revealed that 77% were deletions, 16% were duplications and 7% were mCNVs [147]. Despite being minority, the mCNVs account for seven times more variation in gene dosage compared to combined contribution of deletions and bi-allelic duplications [148] resulting in abundant variation in gene expression. Difference in copy numbers for any two individuals account for 0.78% of the difference in genome and affect structures of nearly 2.7% of the gene transcripts [147]. Common bi-allelic CNVs are well tagged by surrounding SNPs but significantly less LD has been detected between mCNVs and their surrounding SNPs. CNVs may influence gene expression levels in a variety of ways by either deletion,duplication or multiplication of entire genes or by disruption or insertion of regulatory elements such as enhancers or repressors [146].

Recently, several associations have been reported between common CNVs and complex diseases like obesity. As some CNVs are in strong LD with common SNPs, some studies have identified common CNVs potentially contributing to disease susceptibility through LD with SNPs. The association of a common deletion upstream *NEGR1* gene and body weight was detected in a SNP GWAS for BMI by Willer *et al*. [130]. Another common CNV near G-protein coupled receptor, class C, group 5, member B gene (*GPRC5B*) was identified through association of tag SNP with BMI in GWAS meta-analyses by Speliotes *et al*. [132]. However, the effect sizes for BMI observed at each of these loci were small.

Similar to SNP GWAS, genome-wide CNV association studies have focused on common CNVs (CNVs with population frequency >5%). Several common CNVs have been identified in studies that directly measure copy numbers. A common CNV on chr10q11.2 encompassing pancreatic polypeptide receptor 1 gene (*PPYR1*) has been identified where low copy numbers were marginally associated (P=0.011) with increased BMI among chinese contributing 1.6% of variation in BMI [149]. In 2011, a genome-wide study searching for common CNVs in early-onset extreme obesity identified common CNV encompassing olfactory receptor genes (*OR4P4*, *OR4S2* and *OR4C6*) [150]. This study also confirmed the CNVs at *NEGR1* and *PPYR1*. In addition to these common bi-allelic CNVs, a complex copy number variable region on chr8p21.2 encompassing dedicator of cytokinesis 5 gene (*DOCK5*) has been shown to be associated with severe obesity [151]. This region includes two variable number tandem repeats (VNTR) flanked by a 3975 bp deletion. The VNTRs explains 0.8% of the phenotypic variance while the 3975 bp deletion explains 0.46% [151]. Despite all this evidence, the role of common CNVs in disease susceptibility remains an issue because of low replication of initial findings. In general, the main challenges are confounding due to various sources of systemic errors such as source and quality of DNA, batch effects and difficulties in genotyping complex mCNVs [146].

Rare CNVs (CNVs with frequencies <1% in general population) are not well-tagged by surrounding common SNPs genotyped on GWAS panels. Due to the difficulties in genotyping CNVs, studies analyzing rare CNVs have primarily focussed on variants of large sizes (200-500kb) [146]. Several large rare CNVs have been identified within the chr16p11.2 in relation to body weight and risk of obesity. In 2010, Walters *et al*. reported association of a 593kb deletion at chr16p11.2 with 30-fold increased risk for obesity and 43-fold increased risk for morbid obesity and additionally with macrocephaly [152]. Simultaneously, another study reported the association of 593kb deletion at this locus with severe obesity [153]. In 2011, Jaquemont *et al*. observed that the reciprocal duplication of 16p11.2 was associated with 8.3 fold increased risk of being underweight in adults as well as with reduced postnatal weight and BMI compared to non-duplication carriers [154]. The 16p11.2 duplication was also associated with microcephaly mirroring the macrocephaly in deletion carriers. Additionally on chr16p11.2, a 220 kb deletion encompassing nine genes including *SH2B1* (which is involved in leptin signalling and insulin resistance) has been associated with hyperphagia, severe early onset obesity and developmental delay [153, 155].

In addition to investigating contribution of individual CNVs to obesity susceptibility, global burden of rare CNVs has been assessed by comparing the number of rare CNVs in obese with normal weight subjects. Analyses have shown the enrichment of large, rare deletions in obese cases and larger effect was associated with CNVs that disrupt genes [153, 156]. Despite the mounting evidence

for the contribution of both common and rare CNVs to obesity susceptibility, our understanding of the contribution of CNVs to obesity remain incomplete. Novel technical and statistical methodologies are needed to investigate the role of structural variants particularly complex mCNV followed by well-powered and well-designed replication studies to confirm the initial signals. Lastly, functional studies are needed to uncover the underlying mechanisms.

Gene-lifestyle interactions in obesity

Rapid rises in the prevalence of overweight and obesity clearly exemplify the importance of lifestyle factors in the etiology of obesity. As both genetic factors and lifestyle factors affect the risk, the challenge is to understand the interplay between genes and lifestyle factors, and identify interactions of importance. Gene-environment interaction (GEI) can be defined as "*a different effect of a genotype on disease risk in persons with different environmental exposures*" or "*a different effect of environmental exposures on disease risk in persons with different genotypes*" [157]. Thus, not everyone living in the present obesogenic environment (which in general promotes excess calorie intake and discourages physical activity) become obese and the response to such environment can be expected to be dependent on genetic-susceptibility to obesity, and additionally other genetic and environmental factors.

Two main types of study designs for identifying or testing the effects of GEIs are observational association studies and RCTs. There is limited evidence for testing GEI on the genome-wide scale due to computational and statistical difficulty. Winkler *et al*. recently conducted a genome-wide interaction study to examine the influence of age and sex on genetic association with adult body size and shape [158]. The most studied environmental factors in context of GEI in obesity include age, sex, physical activity, total energy intake, dietary fats, dietary CHO and SSB.

Evidence from observational studies

In observational studies, GEI has been tested using various study designs such as cohort studies, case-control studies, case-only studies [159], each having advantages and limitations and may be more or less suitable for different scientific questions. Cohort studies can be either cross-sectional or prospective (that follows subjects over time) and are less prone to selection and recall bias but are expensive, require large sample sizes and long enough follow-up time specially in case of chronic conditions with low incidence [160]. Case-control studies in which subjects with a certain disease (cases) are compared to unaffected individuals (controls) are

relatively easy to conduct but they are prone to several potential sources of bias such as selection bias (differences in source populations of cases and controls), survival bias (when subjects die from disease of interest before enrollment in the study), recall bias (differential reporting of past behaviors e.g. diet, exercise among cases and controls, or lifestyle changes due to disease among cases) and population stratification bias (when cases and controls differ in ethnicity/ancestry). These biases reduce the power to detect GEIs [161]. Case-only study design is simple and efficient when interest is limited to study GEI and there is no need for control subjects. Its validity is highly dependent on the assumption that environmental exposures and genotypes are independent of each other so that the exposure should not differ among different genotypes and such design performs poorly if this assumption is violated. The main limitation of this method is that it cannot assess the main effects of neither genotype nor the exposure [162].

GEI studies for candidate genes: GEI studies of biological candidate genes have not been very successful, mainly due to that candidate gene approach was not successful in identifying obesity genes. Thus, only a few GEI findings of candidate obesity genes have been replicated in independent studies due to the small effect sizes and modest levels of significance for most of the obesity associated candidate genes [163]. As the interaction effect sizes are likely to be of even smaller magnitude, many of the small scale GEI studies were underpowered and hence false positive [164].

GEI studies for GWAS genes: Since GWAS identified significant and replicated loci are known to reliably associate with the disease traits that have been studied, the possibilities to detect GEI may be better as causal inference for an interaction may be increased [165]. Among GWAS identified loci, *FTO* is the most studied locus and accumulating evidence supports its involvement in eating behavior, satiety and dietary intake [166]. In terms of GEI, most of the studies have examined variation in *FTO* with dietary components and physical activity in context of obesity. Andreasen *et al*. reported that low physical activity accentuate the effect of *FTO* rs9939609 on body fat accumulation [167]. In a Swedish study, Sonestedt *et al*. reported that high intakes of fat, low intakes of CHO and low physical activity accentuate the association between *FTO* rs9939609 SNP and obesity traits (i.e. BMI and FM) [168, 169]. A large meta-analysis of 45 adult and 9 pediatric cohorts found that physical activity decreases the odds for obesity by 27% among *FTO* risk allele carriers [170]. Recent large scale meta-analyses evaluating the interaction between *FTO* rs9939609 and dietary intake on obesity found that lower dietary protein intake attenuates the association between *FTO* genotype and adiposity in children and adolescents [171] but no such significant interaction was observed in adults [172]. Vimaleswaran *et al*. investigated the association of *FTO* rs9939609 with changes in weight and WC during 6.8 years of follow-up in data from European Prospective Investigation of Cancer (EPIC) and examined whether these associations were

modified by dietary energy percentage from fat, CHO, protein or glycemic index (GI), but found no influence of dietary factors on associations of *FTO* with obesity traits [173]. Similarly, a recent study has reported no influence of Mediterranean or Nordic diet scores on associations between *FTO* and changes in weight or WC [174].

Corella *et al*. investigated whether the independent and joint associations of *FTO* and *MC4R* are modulated by diet and physical activity. They observed a significant interaction between *FTO* genotype and LTPA on BMI, WC and risk of obesity, but no interactions with the studied environmental factors and the *MC4R* genotype were observed. This was the first study to investigate the joint effect of obesity associated SNPs and a significant interaction between an additive genetic score of the two SNPs and physical activity on BMI and obesity. They also reported that adherence to Mediterranean diet could modify the genetic susceptibility to obesity, however this interaction with diet was not statistically significant [175]. Li *et al*. by using a genetic risk score (GRS) of 12 BMI-associated SNPs showed that physically active lifestyle attenuates the genetic predisposition to obesity in ~20,000 participants from EPIC-Norfolk cohort [176]. Ahmad *et al*. tried to replicate this finding in ~111,000 participants of European ancestry and although their meta-analyses provided further support for interaction between genetic susceptibility to obesity and physical activity, the findings were observed to hinge on the inclusion of North American cohorts indicating the results to be population-specific or not causal [177]. A Danish prospective population based Inter99 study examined whether the effect of lifestyle changes on body weight fluctuations could be modulated by a GRS based on 30 GWAS identified BMI-associated loci. They observed no significant interaction between lifestyle changes (diet, physical activity and smoking) and GRS on body weight changes [178].

Qi *et al*. by using data from multiple US cohorts, created a weighted GRS of 32 BMI associated SNPs and demonstrated that genetic association with adiposity was strengthened among participants with higher consumption of SSB [179], fried food [180] and increased hours of television watching, while it was weakened with increased levels of LTPA [181]. Recently, Nettleton *et al*. investigated whether a diet score calculated from self-reported intakes of healthy (fruits, vegetables, fish, whole grain and nuts) and unhealthy (fried potatoes, sweets, SSB and red/processed meat) food items modifies the association of 32 BMI-associated SNPs (independently as well as when combined to a GRS) with obesity traits and found no significant interaction between diet score and GRS on obesity. However, two of the SNPs [leucine-rich repeat neuronal protein 6C (*LRRN6C*) rs10968576 and mitochondrial translational initiation factor 3 (*MTIF3*) rs4771122] indicated nominally significant interactions with the diet score on BMI [182].

Evidence from intervention studies

Most reliable evidence for interactions between genetic components and lifestyle factors can be derived from RCTs due to random allocation of exposures that eliminates confounding. However, these studies are expensive and difficult to perform, and therefore usually small short-term studies. Another drawback is that study dropouts are common and compliance with dietary interventions are poor which complicates the interpretation of the results [160].

GEI studies for candidate genes: In one of the very first GEI studies in obesity, carriers of the Trp64Arg variant in the *ADRB3* were found to have less weight loss upon three month low-calorie diet and exercise intervention compared to non-carrier obese women [183]. Several intervention studies have studied GEI for variants in the *LEP* and perilipin 1 (*PLIN1*) genes and found them to be associated with differences in weight loss in response to calorie restricted diets [184, 185]. One RCT reported no major effect of variants in the melanocortin-3 receptor gene (*MC3R*) on weight loss after a 10-week intervention with low-calorie diets in obese Europeans [186].

GEI studies for GWAS genes: A couple of intervention studies have investigated interaction between variants in *FTO* and lifestyle factors on obesity traits. One nutritional intervention study that evaluated the associated effect of *FTO* rs9939609 on body weight changes after three years of intervention with a Mediterranean style diet in high CVD risk subjects, did not find any significant interaction. However, risk allele carriers gained less weight compared to non-risk homozygotes regardless of the nutritional intervention [187]. Finish Diabetes Prevention Study did not find the *FTO* rs9939609 to modify the success of lifestyle modification on weight loss [188]. Huang *et al*. examined interaction between *FTO* genotype and protein intake on long term changes in 737 overweight adults in the two year Preventing Overweight by Using Novel Dietary Strategies (POUNDS) trial and suggested that individuals having the risk A-allele of *FTO* rs9939609 might reduce their appetite and food cravings by choosing low-calorie high-protein weight loss diet [189]. In another study where 742 obese subjects were randomly assigned to one of four diets differing in the proportions of fats, proteins and CHO, a high-protein diet was found to facilitate weight loss and improvement of body composition only in individuals with risk-allele of rs1558902 variant in *FTO* and not among other genotype carriers [190]. In a one year lifestyle intervention study, SNPs in *FTO* and *INSIG2* collectively associated with lowest degree of overweight reduction or even aggravated their effect on overweight reduction [191]. Jääskeläinen *et al*. examined whether 26 BMI-associated SNPs individually and as a weighted GRS associate with obesity and weight change after one and three years, and further tested whether these associations were modified by dietary factors or physical activity among 459 participants of Finish Diabetes Prevention Study. The study results suggested that

the association between BMI GRS and obesity might be attenuated by high fibre diet, however this interaction was not statistically significant ($P_{interaction}$= 0.065) [192].

Recently, Papandonatos *et al*. examined whether 91 BMI-associated SNPs influence weight change or modify the response to weight loss interventions in two large RCTs of lifestyle modification i.e., in people with pre-diabetes in DPP and in T2D patients (in the Action for Health in Diabetes (Look AHEAD). For most of the SNPs they did not find any significant interaction with lifestyle on weight loss or weight regain, but a variant in *MTIF3* rs1885988 was observed to modify the effects of lifestyle intervention on weight loss. None of the SNPs modified treatment response on weight regain [193].

Mendelian randomization studies

Many behavioral, physiological and pharmacological measures that show significant associations in observational studies fail to do so in RCTs [194]. Residual confounding, reverse causation and incorrect causal inference may explain the associations found in observational studies [195]. Mendelian randomization (MR) is a technique that make use of genetic data to examine causal relationship between risk factors (usually circulating biomarkers) and disease outcomes [196]. Since genes are randomly allocated at conception, they can be used as natural experiments to prove causation. In MR analysis, a genetic variant associated with the biomarker of interest is used as a proxy for the biomarker and randomly divides the study population into groups based on genotypes similar to the arms of RCT [196, 197].

Assumptions of MR studies

In MR studies genetic variants such as SNPs are used as instrumental variables (IV) to infer causal association between a trait and an outcome because of several reasons. First, genetic variant is associated with the trait of interest in only one direction and eliminates the possibility for reverse causation. Secondly, measurement of genetic variants usually involve very low levels of errors. Thirdly, genetic variants in high LD with causal variant can be used as instruments [196]. However, in order to be used as IV, genetic variants must fulfill three IV assumptions: (i) the genetic variant is reliably associated with the exposure; (ii) the genetic variant is associated with the outcome through the studied exposure only; and (iii) the genetic variant is independent of other factors which affect the outcome (confounders) [198]. Furthermore, using MR for accurate estimation of effect sizes

in mediation analyses using a measured exposure requires that the measured exposure accurately captures true causal exposure, and that the genetic variants should be robustly associated with the exposure of interest [198].

MR studies vs RCTs

MR study design is considered as a non-experimental alternative to RCTs as both share many commons features. In a RCT participants are randomly assigned to treatment or control groups to balance confounders and establish causal inference whereas in MR, alleles are randomly allocated at conception [197]. Thus, similar to RCT, groups defined by genotype in MR will not differ with respect to confounding factors but will experience an on-average difference in exposure to the studied trait [197]. These similarities between the two study designs suggest that MR can predict the outcome of an RCT as long as it strictly meets the criteria of MR assumptions. However, in contrast to RCT, MR studies have much lower cost and once a population is genotyped on genome-wide level, basically every biomarker that is measured in the population, that is at least partially modulated by genetic factors, can be studied *in silico* by simple exploration of the data set [199].

Limitations of MR studies

There are several potential threats to IV assumptions such as pleiotropy, LD, population stratification, canalization and weak instruments and breaking against any of these assumptions interferes testing for causal relationship.

Pleiotropy

The most important limitation of MR studies is pleiotropy, which in this context means that a genetic variant not only associates with the exposure trait of interest but with one or more other traits. Pleiotropic associations can affect the results of MR analysis in multiple ways. Pleiotropic associations can counteract any effect of the variant acting through the biomarker of interest on the disease resulting in a null finding when a true causal relationship exists between biomarker and disease, but also result in positive association between the genetic variant and a disease that could mistakenly be interpreted as causal association, although the causal connection would be mediated by association with biomarker(s) other than the biomarker of interest [199].

Linkage disequilibrium

LD means correlation, allelic association between physically close alleles that are inherited together. If there is LD between a genetic variant being used as IV and a

genetic variant that is causally associated with the exposure, then none of the MR assumptions will be violated and the association between IV and exposure can be estimated. This form of LD is exploited in many MR studies when a genetic variant in LD with the often unknown functional variant is used as IV. However, when a genetic variant being studied is correlated with another polymorphic locus and this locus influences the outcome of interest in the genetic association study, then this will violate the assumption (iii) of MR and IV analyses will be confounded [196]. Thus, when conducting MR studies, it would be ideal to use only those SNPs that lie in genomic regions without any further proximity to loci that might circumvent SNP-disease association [199].

Population stratification

Population stratification occurs when the population under investigation is not homogenous, but rather based on several subgroups that experience both different disease rates as well as have different frequencies of alleles of interest. Population stratification in MR studies could result in confounded associations. This potential limitation requires special attention when genetic variant-biomarker-disease relationships are not studied in one population but in several cohorts of different origin. Combining different data sets to increase statistical power, such as combining findings from a GWAS meta-analyses on a biomarker with another GWAS meta-analyses on an outcome of interest, brings its own challenges [199] as some studies may introduce heterogeneity in the meta-analysis.

Canalization

Canalization or developmental compensation refers to alteration in the expected disease-genotype association by adaptation to a genetically determined phenotype. A genetic variant can affect the biomarker already during the childhood or even earlier while a clinical biomarker can usually only be of relevance later in life. Thus the association of the genetic variant with the disease may be blurred by the counter-regulatory mechanisms that compensate for the effects related to SNPs *in utero* or during childhood [199]. Canalization could invalidate findings from MR studies by altering the effect of a genotype on the outcome of interest in adulthood without any effect on the association between genotype and modifiable exposure of interest. Hence the estimate of association between genetic variant and exposure would be valid whereas that between genetic variant and outcome would not be valid, and consequently the IV estimate of causal effects would be biased [196].

Weak instruments

A weak instrument is defined as one that explains only small proportion of variance in the exposure and may lead to very imprecise estimates of the causal effects. F-statistics from first stage regression of the 2-stage least square regression can be

used to check instrument strength. A value of F-statistics >10 is taken as indicative of sufficient strength to ensure the validity of IV method. Thus, it is now suggested that MR studies always report F-statistics to examine the strength of the instrument and seek expert advice when this value is close or less than one [196].

Use of GRS in MR studies

Multifactorial polygenic traits (such as obesity, BP, T2D) are affected by many common variants with small effects. GRS is a convenient way of combining a large number of genetic variants that together may explain a considerable proportion of variation in the risk factor. Both un-weighted GRS (constructed by summing up the total number of risk alleles present in the genotype of an individual) and weighted GRS (where each risk allele is weighted on the basis of effect estimate of the corresponding SNP on the trait) are used as IVs in MR studies for reasons of simplicity, increased power and avoidance of weak instrument bias. For the variants with different effect sizes on the risk factor, use of estimated weights rather than an un-weighted GRS gave some improvement in power [200]. However, care should be taken while constructing the GRS as each of the variants should satisfy the assumptions of the IV, and use of invalid genetic variants (with pleiotropic effects) can severely bias estimates of causal effects even when 90% of the variants in the score are valid instruments [200].

MR studies of BMI and cardiometabolic traits

In the recent years, several MR studies have attempted to evaluate the associations between obesity measures (especially BMI) and several diseases including cardiometabolic traits and outcomes. Freathy *et al.* used *FTO* rs9939609 as IV to study causal association between BMI and 10 metabolic traits in ~17000 European participants and found that each copy of *FTO* A-allele was associated with higher fasting insulin, glucose and TG and with lower HDLC. No associations were found for fasting alanine-transferase, gamma-glutamyl transferase, LDLC, A1C and systolic and diastolic blood pressure (SBP and DBP) [201]. Timpson *et al*. by using a combination of *FTO* rs9939609 and *MC4R* rs17782313 as IV, studied if the earlier observed association between BMI and BP could be causal in the Copenhagen General Population Study. The study confirmed the observational associations between BMI and BP by finding 3.85 mmHg (95%CI: 1.88 to 5.83 mmHg; *P*=0.0002) increase in SBP and 1.79 mmHg (95% CI: 0.68 to 2.90; *P*=0.002) increase in DBP for each 10% increase in BMI [202]. Fall *et al*. in a large MR effort from the European Network for Genetic and Genomic Epidemiology (ENGAGE) consortium investigated if the association between BMI and 24 cardiometabolic

traits is causal, using *FTO* rs9939609 as IV in ~199,000 European individuals. Apart from replicating earlier findings from Freathy *et al*. [201] and Timpson *et al*. [202], they reported novel causal associations of BMI with incident heart failure (HR: 1.19 per unit increase in BMI; 95% CI: 1.03 to 1.39) and increased liver enzymes (alanine-aminotransferase and gamma-glutamyl transferase) [55]. Nordestgaard *et al*. using a GRS comprised of *FTO* rs9939609, *MC4R* rs17782313 and *TMEM18* rs6548238 SNPs as IV found a causal link between BMI and ischemic heart disease (IHD). They reported a 26% and 52% increase in odds for IHD in observational analyses and MR analyses, respectively, for every 4 units increase in BMI [203].

Aims

General aims

The general aim of this doctoral thesis was to investigate how common BMI-associated genetic polymorphisms interact with dietary intakes to modify the genetic susceptibility to obesity, and how they affect the weight gain in different ages. Further, genetic variants that associate with cardiometabolic traits were used to dissect causality between cardiometabolic traits and mortality.

Specific aims

Paper I: To test whether genetic susceptibility to obesity associates with dietary intake levels of fats, CHO, protein and fibre and total energy intake, and whether diet intakes modify the genetic susceptibility by using a GRS comprised of 13 replicated BMI-associated SNPs.

Paper II: To replicate the recent report that genetic susceptibility to obesity modifies the association between SSB intake and obesity risk in two large Swedish cohorts using a GRS of 30 BMI associated SNPs.

Paper III: To investigate the association of salivary amylase (*AMY1)* CNV with obesity traits (BMI, WHR and BF%) and to test whether starch intake modifies these associations.

Paper IV: To investigate how genetic susceptibility to obesity measured as GRS comprised of 31 BMI-associated loci affects weight gain from young adulthood to middle age and later life.

Paper V: To understand the causal nature of the associations of common cardiometabolic traits (BMI, SBP, LDLC, HDLC, TG and FPG) with CVD- and total mortality using trait specific GRS as IV in a Mendelian Randomization approach.

Participants and methods

The different hypotheses of this thesis were tested analyzing two Swedish cohorts (Figure 5). Three of the papers, i.e. the study evaluating the association and interaction between the genetic susceptibility to obesity and diet intakes (Paper I), the study investigating the role of dietary starch in modifying the association between *AMY1* CNV and obesity (Paper III) and the study dissecting the causal associations between cardiometabolic traits and mortality (Paper V), were all based on the cohort of 30,447 subjects from the population based Malmö Diet and Cancer Study (MDCS). The two remaining studies i.e., the replication study of interaction between BMI-GRS and SSB on obesity (Paper II) and the association study of BMI-GRS and weight gain at different time points (Paper IV) also included participants from the Gene-Lifestyle interactions And Complex traits Involved in Elevated disease Risk (GLACIER) cohort in addition to participants from MDCS.

Malmö Diet and Cancer Study

MDCS is a population based prospective cohort study from Malmö, a city in southern Sweden with about 250,000 inhabitants. The study was planned and initiated in collaboration with the International Agency for Research on Cancer (IARC), Lyon, France, the Swedish Cancer society and the Faculty of Medicine, Lund University, Sweden. The main aim of MDCS was to investigate the relationship between diet, life-style factors and various forms of cancers [204].

MDCS baseline examinations

Baseline examinations for MDCS were conducted between March 1991 and October 1996. In 1991, by using both letter of invitation and information campaign (including advertisements in newspapers and television and posters in public places), all men and women born between 1926 and 1945 (n=53,325) were invited to participate in the study. In May 1994, the invited population was extended to include all men born between 1923 and 1945 and all women born between 1923 and 1950 (n=74,138) with intention to increase the number of younger women to study

breast cancer. The exclusion criteria for recruitment were limited Swedish language skills and mental incapacity. Of those invited, 41% attended the baseline examinations, yielding a cohort of 30,447 subjects [205]. Ethical permission of MDCS protocols was obtained from ethical committee at Lund University (LU 51-90). All participants provided written informed consent.

Each participant visited the study center twice. During the first visit, anthropometric measurements (weight, height, waist and hip circumference, and body lean mass and fat mass) were taken, BP was measured and non-fasting blood samples were collected by trained nurses. Participants were provided with the extensive dietary, lifestyle and socioeconomic questionnaire, menu book and detailed instructions about the dietary data collection procedure. The questionnaire included items such as place of birth, education, employment, social network and support, LTPA, smoking, alcohol use, previous weight change, dietary change in the past, previous and current disease, use of medications etc. During the second visit, after nearly two weeks, trained dietary interviewers conducted individual interviews to complete the diet history and to check the correctness of the completed questionnaires [205].

In 1993, MDCS became an associated member of EPIC, which is one of the largest cohort studies in the world comprising of >521,000 participants from 23 centers across 10 Western European countries. EPIC is organized by the IARC, WHO, Lyon, France [206, 207].

Reproducibility of the MDCS questionnaire

In 1994, three weeks after the first invitation, 232 randomly selected participants were invited to complete the questionnaire a second time. Of the 211 subjects who responded to the questionnaire twice (participation rate 90.9%), 209 were complete participants. The kappa values for concordance between two questionnaires showed high reproducibility of the questionnaire for most of the factors. For example, kappa coefficients among men/women were as follows: 0.96/0.84 for education, 0.73/0.77 for alcohol, 0.70/0.84 for weight change, 0.76/0.68 for dietary change, 0.97/0.96 for hypertension, 0.90/0.85 for diabetes [208].

The biobank

Blood samples from each participant were fractionated i.e., 10 ml blood without anticoagulant was used to store serum sample (at -80^{o}C) and 28 ml of heparinized blood was used to purify and store mononuclear leucocytes (at -140^{o}C), granulocytes, erythrocytes and plasma (all at -80^{o}C). An alteration in this method was made in August 1995 after which buffy coats were stored (at -140^{o}C) instead of mononuclear leucocytes and granulocytes. The yield, purity and storage of blood samples were all assessed by a quality control program and no differences in terms of purity and yield were found [209, 210].

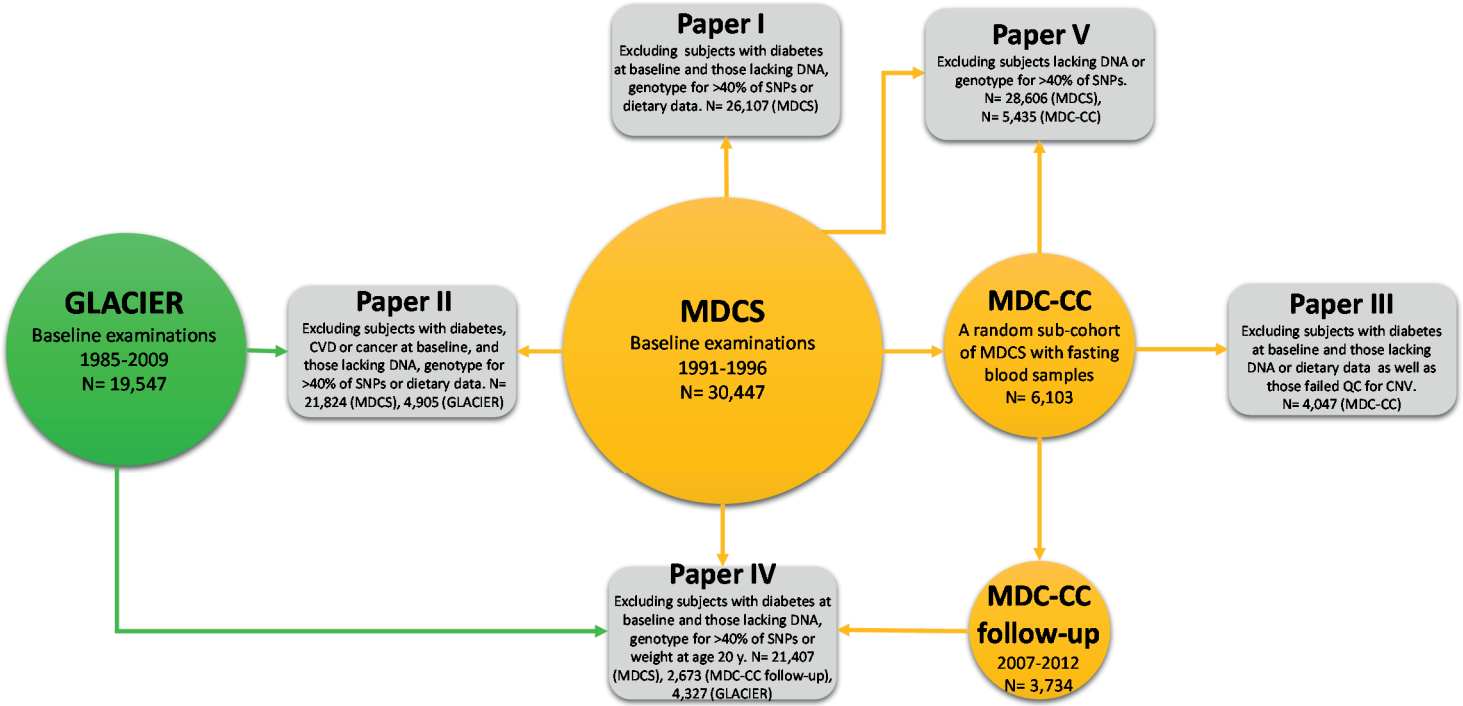


Figure 5. Study populations in papers I-V

Malmö Diet and Cancer Study-Cardiovascular Cohort

Of the MDCS participants that were enrolled between October 1991 and February 1994 (n=12,445), approximately 50% (n=6,103) were randomly selected and invited to participate in a study for the epidemiology of carotid artery disease forming a sub-cohort known as the Malmö Diet and Cancer Study- Cardiovascular Cohort (MDC-CC). These participants additionally underwent a review of their medical history, a physical examination and provided fasting blood samples for the assessment of cardiovascular risk factors such as fasting glucose and fasting serum lipid and lipoprotein concentrations [211].

MDC-CC follow-up

During 2007 to 2012, a re-examination of MDC-CC participants was conducted after a mean follow-up of 16 years (range 13 to 20 years). A total of 4,924 participants who were still alive and had not emigrated from Sweden were invited to participate in a follow-up examination using the same methods as during the baseline. Of those invited, 3,734 participants attended and underwent re-examination including anthropometric measurements, a questionnaire on lifestyle factors and fasting blood samples.

Diet assessment in MDCS

A modified diet history method (comprising of a 7-day menu book, an extensive 168-item dietary questionnaire and a 1-hour diet history interview) was specially designed for MDCS. The menu book covered information about meals that vary from day to day such as cooked lunches and dinner meals as well as cold and alcoholic beverages, medications, natural remedies and dietary supplements. The dietary questionnaire covered information about the food items regularly consumed during the past one year, not covered by the menu book. Participants were asked to report frequency and estimated intake of each food by using a booklet containing 48 photographs, each showing four different potion sizes of the respective food item. Both menu book and questionnaire were filled at home by the participants. During their second study visit which was about two weeks after the first visit, participants were interviewed by trained interviewers to collect information about their food preparation methods, portion sizes of the foods collected in the menu book (using a more extensive booklet of photos) and detailed food choices (e.g., type of bread and fat) as well as to check for the correctness of information provided in the menu book

and questionnaire. By combining the information from menu book, questionnaire and interview, the average daily food intakes (grams per day) were calculated and were converted into nutrient and energy intakes by using MDC Food and Nutrient Database specifically developed for MDCS from the PC KOST-93 of the Swedish National Food administration [212].

Validity and reproducibility

The relative validity of the modified diet history methods used in MDCS has been evaluated among 206 residents of Malmö using an 18-day weighed food record (obtained by collecting 3-days weighed records every second month over one year to equally represent week days and weekend as well as seasonal variations) as a reference method. Energy–adjusted Pearson correlation coefficients in men/women were 0.64/0.69 for fat, 0.66/0.70 for CHO, 0.54/0.53 for protein and 0.74/0.69 for fiber [213]. Reproducibility of the method was examined approximately one year after the first diet assessment among 126 men and 115 women residing in Malmö. Energy-adjusted Pearson correlations were between 0.50 and 0.80 for most of the nutrients [214].

Energy misreporting and food habit change

To identify the participants misreporting their energy intake, the physical activity level (PAL) was calculated from self-reported information on physical activity at work, household and leisure time, sleeping hours, self-care and passive time. The individuals having ratio of energy intake to basal metabolic rate (BMR) outside the 95% CI limits of calculated PAL, were classified as misreporters [215]. Information about the change in food habit was obtained through a question stating ‘Have you substantially changed your eating habits because of illness or for other reasons?’ in the questionnaire.

Seasonal variation and diet assessment method version

Since the dietary intake may change with the season, the season of the dietary interview was recorded as spring (Mar-May), summer (Jun-Aug), autumn (Sep-Nov) and winter (Dec-Feb). A slight alteration (replacement of individualized with standardized recipes and portion sizes for some foods) in diet assessment method was made in September 1994 to shorten the interview time. The diet assessment method version variable indicate if the data was collected before or after this change but it appears to have no major influence on the ranking of participants [213].

The GLACIER Study

The GLACIER study comprising of 19,547 participants is a population based prospective cohort nested within the ongoing Västerbotten health survey (VHU, Västerbotten Hälso Undersökning) of >100,000 adults from northern Swedish county of Västerbotten. GLACIER participants underwent the baseline health examination between 1985 and 2004 when all residents of the Västerbotten County were invited to visit their primary health care center in connection with their 40th, 50th and 60th birthdays. A sub-group of GLACIER participants (n=5010) attended a 10-year follow-up examination between 1995 and 2007. During baseline examinations basic anthropometric measurements (weight, height and WC) and BP were taken, fasting blood samples were drawn and information about physical activity, lifestyle factors and diet was obtained using validated self-administered questionnaires [216, 217]. Initially an 84-item food frequency questionnaire (FFQ) was used to capture habitual diet intakes but in 1996, some of the food items were combined and the questionnaire was reduced to 66-items. Participants were asked to record food intake on a 9-point frequency scale ranging from never to >4 times per day as well as portion sizes from three food groups including potatoes/rice/pasta, meat/fish and vegetables. Energy and nutrient intake was calculated by the nutritional values available through the National Food Administration's database (www.slv.se). Food intake level (FIL) was calculated by dividing total energy intake with the BMR and was used to identify and exclude participants with unreliable data (top 1% and bottom 5% of the FIL distribution within entire VHU population were excluded). Information about lifestyle such as family history of diseases (CVD, diabetes), quality of life, social support, tobacco and alcohol use etc. was obtained through the Short Form-36 and information about physical activity was obtained through a modified version of International Physical Activity questionnaire. All participants provided written informed consent as part of VHU and GLACIER study protocols were approved by the regional ethical review board in Umeå [217].

Study Specific Materials

Dietary variables

Macronutrients, starch and fiber (Paper I and III)

The variables for total energy intake, macronutrients (fat, CHO and protein) and fiber were used in paper I. Total energy intake (kcal/day) included energy from fat, CHO, protein, fiber and alcohol. Macronutrient intakes were converted into

percentage of non-alcohol and non-fibre total energy (E%) by using a conversion factor of 4kcal/g for CHO and proteins and 9kcal/g for fats while fiber intake was studied as fiber density (g/1000kcal). Fat intake included saturated-, mono-unsaturated- and poly-unsaturated fats and cholesterol. CHO intake included monosaccharides, disaccharides and starch but not fiber. Fibre intake included all types of fiber as data on fiber subtypes was not available in MDCS. In paper III, starch intake was used which was calculated by subtracting monosaccharides and disaccharides from total CHO intake. Starch intake was also studied as E%.

Sugar-sweetened and artificial-sweetened beverages (Paper II)

In paper II, dietary data on SSB and artificial-sweetened beverages (ASB) was studied. In MDCS, SSB include all carbonated and non-carbonated beverages sweetened with energy containing sweeteners except juice. ASB include all beverages with non-energy artificial sweeteners, for example fruit drinks, sodas and pops. Reported intakes of SSB and ASB in g/day were converted into serving/day by using a conversion factor of 250g/serving. Further, SSB (or ASB) intake was stratified into tertiles after putting zero/seldom consumers in a separate category. In GLACIER, information on SSB was collected via two questions (one on carbonated and other on non-carbonated beverages) in the 88-item FFQ but later on in 66-item FFQ, both questions were combined and juice was added. The 9-point scale of FFQ was combined to form 4 categories similar to MDCS. No information on ASB was available.

Lifestyle variables

The information regarding socioeconomic and lifestyle factors was obtained through a self-administered questionnaire.

Smoking (Paper II and III)

In both MDCS and GLACIER, the smoking variable was categorized into never, former and current smokers.

Alcohol (Paper II and III)

Alcohol consumption was quantified as gram/day. In MDCS, participants reporting no alcohol consumption in the previous year in the questionnaire as well as in the menu book, were considered as zero consumers while the rest of the individuals were further categorized into gender specific low, moderate and high consumers based on information from the menu book. The cut-off levels for males were <20g/day, 20-40g/day and >40g/day and for females were <15g/day, 15-30g/day and >30g/day.

Education (Paper III)

Based on the type of education attained, the variable was categorized as elementary (≥8years), primary or secondary (9-10 years), upper secondary (11-13 years) and university degree.

Leisure time physical activity (Paper II and III)

In MDCS, information on LTPA was obtained from 17 different pre-specified activities and one open activity in the questionnaire. The questionnaire was adapted from Minnesota leisure time physical activity questionnaire [218]. Thus the time spent on each activity was multiplied with an intensity factor to calculate LTPA score which was further stratified into gender-specific tertiles. In GLACIER, information on LTPA regarding past three months was gathered through a modified International Physical Activity questionnaire. The five point scale representing never, occasionally, 1-2 times/week, 2-3times/week and >3 times/week was combined to form a two level scale (low level (<1 time/week) and a medium/high level (≥1-2 times/week)) for analyses in paper II.

Clinical measurements

Anthropometric measurements (Paper I to V)

In both cohorts, weight (kg) and height (cm) were measured by trained staff using a calibrated balance-beam scale (with participant wearing light indoor clothing) and a wall-mounted stadiometer respectively. In MDCS, participants also provided self-reported data during baseline examinations regarding their weight at the age of 20 years. They also provided the information whether their weight has been stable, increased or decreased since then. BMI was calculated by dividing weight (kg) with height (meters^2) and obesity was defined according to WHO criteria [11]. WC (cm) was measured midway between the lowest rib margin and iliac crest and hip circumference (cm) was measured at the level of greatest lateral extension of the hip by trained staff. WHR was calculated by dividing WC with hip circumference. Body composition (FM and FFM) was estimated by using BIA (BIA 103; JRL Systems, Mt. Clemens, MI, USA) and BF% was calculated using manufacturer provided algorithm.

Laboratory measurements (Paper V)

BP was measured in supine position after resting for ten minutes by using a mercury-column sphygmomanometer. Fasting blood samples were available only in MDC-CC for the measurement of fasting blood glucose (FBG) and fasting lipids and lipoproteins (TG, LDLC and HDLC) which were analyzed by routine standard methods at the Department of Clinical Chemistry, Malmö University Hospital. FBG

was measured by hexokinase method and was multiplied by 1.13 to be converted into fasting plasma glucose (FPG). TG and total cholesterol (TC) were measured using reagents and calibrators from the supplier on a DAX 48 automatic analyzer (Bayer AB, Göteborg, Sweden). HDLC was first treated with dextran sulphate to precipitate LDLC and very low-density lipoprotein cholesterol (VLDLC) and then measured in the same manner as TC. LDLC was calculated using Friedewald formula: LDLC= TC – HDLC - (TG/2.2) [211], excluding individuals with TG>4.0 mmol/l.

Genetic variants and genotyping

In MDCS, blood samples (non-fasting) were collected and stored at baseline and DNA was extracted from frozen granulocyte or buffy coat samples using QIAamp96 spin blood kits (QIAGEN, VWR, Gaithersburg, MD, USA). Genetic variants and genotyping methods used in papers I to V are summarized in Table 1. In all the studies genotypes were recoded as 0, 1 and 2 according to the number of risk increasing alleles for the corresponding trait.

Table 1. Summary of genotyping methods and genetic variants in paper I-V

Paper	Genetic variant type	Main trait	Total no. of loci used	Genotyping method	Reference
Paper I	SNP	BMI, obesity	16	Sequenom iPLEX, TaqMan, KASPar	Meyre *et al*., 2009 [219]; Thorleifsson *et al*., 2009 [131]; Willer *et a*l., 2009 [130]
Paper II	SNP	BMI	30	Sequenom iPLEX, TaqMan, KASPar, *Metabochip array	Speliotes *et al*., 2010 [132]
Paper III	CNV	BMI	1	TaqMan	Falchi *et al*., [220]
Paper IV	SNP	BMI	31	Sequenom iPLEX, TaqMan, KASPar, *Open array, *Metabochip array	Speliotes *et al*., 2010 [132]
Paper V	SNP	BMI, SBP, TG, LDLC, HDLC, FPG	153	Sequenom iPLEX, TaqMan, KASPar	Speliotes *et al*., 2010 [132]; Newton-Cheh *et al*., 2009 [221]; Ehret *et al*., 2011 [222]; Wain *et al*., 2011 [223]; Teslovich *et al*., 2010 [224]; Dupuis *et al*., 2010 [225]

SNPs: single nucleotide plymorphisms; CNV: Copy numbervariant; BMI: Body mass index; SBP: Systolic blood pressure; TG: Triglycerides; LDLC: Low-density lipoprotein cholesterol; HDLC: High-density lipoprotein cholesterol; FPG: Fasting plasma glucose. *Open array and MetaboChip array genotyping methods were used only in GLACIER.

BMI-associated SNPs (Paper I, II, IV and V)

In paper I, 16 GWAS identified BMI and/or obesity associated SNPs were studied. These included *FTO* rs9939609, *MC4R* rs17782313, *SH2B1* rs7498665, *GNPDA2* rs10938397, *BDNF* rs4923461, *AIF1* rs2844479, *MTCH2* rs10838738, *FAIM2* rs7138803, *SEC16B* rs10913469, *TEM18* rs6548238, *NEGR1* rs2815752, *SFRS10* rs7647305, *KCTD15* rs29941 [130, 131], *MAF* rs1424233, *NPC1* rs1805081 and *PTER* rs10508503[219]. The SNPs were genotyped either by TaqMan or KASPar allelic discrimination assay-by-design method using an ABI 7900 PCR system (Applied Bio-systems, foster City, CA, USA) or by Sequenom iPLEX method using a MALDI-TOF mass spectrometer (Sequenom, San Diego, CA, USA) depending on the availability of assays and reagents. Average successful genotyping call rate was 98.4% and all SNPs were in Bonferroni corrected Hardy-Weinberg equilibrium (HWE) ($P>0.0031$ for 16 independent tests at $\alpha=0.05$). In paper II, IV and V, 32 BMI associated SNPs identified in GWAS by Speliotes *et al*. in 2010 were studied [132]. Of these, one of the SNPs representing *ZNF*608 rs4836133 was not successfully genotyped in MDCS and hence was excluded in all the three studies. In MDCS, all SNPs were genotyped using the same methods as stated in Table 1 with average genotyping success rate of 97.2% and all SNPs were in Bonferroni corrected HWE ($P>0.001$ for 31 independent tests at $\alpha=0.05$). In GLACIER, genotyping was performed by MetaboChip array (Illumina Inc., San Diego, CA, USA) with a genotyping success rate of 96.0% and all SNPs were in HWE ($P>0.001$). Additionally one more SNP representing *LRP1B* rs206936 was not in Bonferroni corrected HWE and as no proxy was available for the GLACIER sample used in paper II, in total 30-BMI associated SNPs were included in paper II.

Lipids-, BP- and FPG-associated SNPs (Paper V)

In paper V, SNPs associated with cardiometabolic traits other than BMI were also analyzed. All SNPs were genotyped by Sequenom iPLEX, TaqMan or KASPar methods as described above. As a quality control, all SNPs with genotype success rate of <90% and deviating HWE for each trait-specific set of SNPs were excluded. Thus, in total 153 GWAS identified SNPs were studied including 31 SNPs for BMI [132], 29 for SBP [221-223], 26 for TG [224], 32 for LDLC [224], 41 for HDLC [224] and 15 for FPG [225].

AMY1 CNV (Paper III)

Copy numbers of *AMY1* were determined by TaqMan assay using AB 7900HT Real-time PCR system (Applied Biosystems, Foster City, CA, USA). All samples were run twice in triplicate and copy numbers were estimated using $\Delta\Delta$Ct method using a reference DNA sample with 14 copies of *AMY1*. As a quality control, all samples for which the difference in estimated copy numbers is >1 between the two runs and

those that were outside 99 percentile for ΔCt.SD and/or predicted copy numbers were excluded.

Since the reliability of qPCR method for genotyping CNVs has been questioned in the recent literature [226-228], we additionally genotyped *AMY1* CNV by another method reported to be more reliable known as droplet digital PCR (ddPCR) [226, 227]. In ddPCR, reaction mixture is partitioned into thousands of nanoliter-sized droplets each with 0 or 1 copy of locus of interest and after thermocycling number of florescent droplets are counted. Each of the 96-well plate was run with three reference DNA samples containing 4, 8 and 17 copies of *AMY1*. All the samples with unreliable number of droplets (too few or too many) were excluded. However we did not find any significant difference in genotypes by the two methods as they were highly correlated (Pearson correlation coefficient =0.98). However, upon stratification into low (≤10) and high (>10) copy numbers, the correlation coefficients were 0.98 and 0.78 respectively.

Clinical end points

Diabetes (Paper I to V)

In MDCS, diabetes cases at baseline examination were identified by self-reported diagnosis in the questionnaire, by use of anti-diabetic medication and additionally by a FBG ≥6.1 mmol/L (equivalent to FPG ≥7.0 mmol/L) in MDC-CC. In GLACIER, diabetes cases at baseline were identified through self-reported data provided in the questionnaire.

CVD- and total Mortality (Paper IV and V)

In papers IV and V the information regarding total mortality and CVD mortality were retrieved by linking the individual's 10-digit civil registration number with Swedish National Cause of Death Register. Mortality was attributed to CVD causes when the main International Classification of Disease (ICD) code was ICD:9(390-459) or ICD:10I(00-99) on the cause of death certificate.

Study specific methods and analyses

Paper I

Study participants

For paper I, the whole MDCS cohort (n=30,447) was included. After excluding participants lacking DNA, genotype information for more than 40% of the studied 16 SNPS or crucial phenotype information, 29,480 participants remained that constituted the study sample for genetic analyses (association analyses of SNPs/GRs with obesity and related traits). For association and interaction analyses with dietary variables, participants with incomplete dietary data (n=2,258) and with diabetes at baseline (n=1,115) were excluded. Thus, 26,107 participants constituted the study sample for analyses with dietary variables.

GRS calculation and statistical analyses

A non-weighted GRS comprising of 13 of the 16 SNPs (three SNPs were not replicated in Paper I) was calculated using PLINK (version 1.05). All other analyses were performed using SPSS version 20 (IBM Corp. Armonk, NY, USA). All variables were logarithmically transformed to normalize the distributions. Assuming additive model and adjusting for age and sex, logistic regression was used to analyze association between SNP/GRS and dichotomous variables (overweight and obesity) and linear regression was used to analyze association between was SNP/GRS and continuous variables (height, weight, WC, hip circumference, FM, FFM, BF%). Association analyses with dietary variables [total energy intake (kcal/day), fat (E%), CHO (E%), protein (E%) and fiber (g/1000 kcal)] using linear regression were additionally adjusted for season, diet assessment method version and total energy intake when applicable. Association between GRS and BMI, FM and FFM was also evaluated in population specific quintiles of macronutrients (E%) and fiber (g/1000 kcal). Generalized linear model (GLM) was used to study interactions between SNP/GRS and diet quintiles on BMI, FM and FFM. Since diet reporting and body composition differ between male and females, we run all SNP/GRS x diet interaction analyses separately in males and females using sex-specific diet quintiles. Power calculations were performed using Quanto (Quanto version 1.2.4: http://hydra.usc.edu/gxe) and the present study was found to have 80% power to detect a gene x diet interaction of at least 0.022 on BMI at α level of 0.05. A P value of <0.05 was considered significant in the association analyses of SNP/GRS with BMI and related traits and in the GRS x diet interaction analyses on BMI and associated traits. However, in the association analyses of 16 SNPs with diet intakes and SNP x diet interaction analyses on BMI and related traits, a

Bonferroni corrected P value of ≤0.003 (0.05/16) was considered significant. Sensitivity analyses were performed after excluding energy misreporters (~19%).

Paper II

Study participants

In paper II, participants from MCDS and GLACIER were included. In MDCS, the study population comprised of 21,824 healthy participants (free of diabetes at baseline, CVD and cancer) with complete dietary data and genotype information. In GLACIER, 4,905 participants with available genotype and phenotype data for variables used in this study were included.

GRS calculation and statistical analyses

All statistical analyses were performed by using SPSS version 20 (IBM Corp, Armonk, NY, USA) (in MDCS), SAS version 9.4 (SAS Institute Inc., Cary, NC, USA) (in GLACIER) and STATA (Stata Corp, College station, TX, USA) (in meta-analyses). GWAS identified 30 BMI-associated SNPs were used to calculate both weighted (where each BMI-increasing risk allele was weighted by their previously reported effect sizes [132]) and un-weighted (by summing up the BMI-increasing risk alleles) GRSs. To facilitate interpretation, weighted GRS was rescaled to reflect the number of risk alleles by using a previously described method [229]. However, the results for both GRSs were overall similar, so results for weighted GRSs were presented throughout the paper. Assuming an additive effect of alleles, GLM was used to study interaction between GRS and SSB/ASB on BMI by including a multiplicative term (GRS x SSB or GRS x ASB) in addition to marginal effect terms in the model. All models were adjusted for age, sex and study specific covariates (MDCS: season, diet assessment method; GLACIER: FFQ version). To account for potential confounding factors, analyses were additionally adjusted for physical activity, smoking, alcohol intake and total energy intake in a second model. For illustration of interaction, the association between GRS and BMI were analyzed stratified by SSB intake categories (four categories) and association of SSB with BMI was stratified by cohort-specific GRS quartiles. To combine the cohort-specific effect estimates from main and interaction analyses and their respective variance estimates, inverse-variance weighted fixed effects meta-analysis was performed using *metan* command in STATA.

Paper III

Study participants

In paper III, 4,047 participants from the sub-cohort MDC-CC were included after excluding participants lacking DNA and information about dietary variables, failing quality control of *AMY1* CNV genotyping and having diabetes at baseline.

Statistical analyses

In paper III, all statistical analyses were performed separately in men and women using SPSS version 22 (IBM Corp, Armonk, NY, USA). *AMY1* CNV and dietary starch intake (E%) were used both as continuous variables and as stratified into sex-specific tertiles when indicated. Across tertiles of starch intake, ANOVA was used to compare mean (±SD) values for continuous variables (age, BMI, WHR, BF% and total energy intake; they were all normally distributed) and Chi-square test was used to compare distribution [n (%)] for categorical variables (LTPA, alcohol, smoking and education). Association of *AMY1* CNV with obesity traits (BMI, WHR and BF%) was tested by using linear regression adjusted for age. GLM was used to calculate age-adjusted mean values and 95% CI for obesity traits in tertiles of starch intake and tertiles of *AMY1* CNV. The association analyses of *AMY1* CNV with obesity traits in strata of starch intake were adjusted for age. The association analyses of starch intake with obesity traits were adjusted for age season and total energy intake in basic model and additionally adjusted for LTPA, smoking, alcohol intake and education in the fully adjusted model. Similarly, using both basic and fully adjusted models, GLM was used to test interaction between *AMY1* CNV and starch intake on obesity traits by including a multiplicative term in the model. A two sided P value <0.05 was considered significant. In sensitivity analyses, the participants identified to have changed their food habits in the past and/or misreport their energy intakes were excluded (~32%).

Paper IV

Study participants

I Paper IV, participants from both MDCS and GLACIER were included. In MDCS, from the total 30,447 participants, 21,407 participants were included after excluding participants lacking DNA, crucial basic phenotypic information and reliable self-reported weight information at age of 20 years, with poor genotyping (lacking genotype information for >40% of the SNPs) and with diabetes at baseline. Of these, 2673 participants also had follow-up information taken during MDC-CC re-examination. In GLACIER, after excluding participants with poor genotyping and

with diabetes at baseline, 4,327 participants with 10 year follow-up data were included in the study.

GRS calculation

GRSs, both weighted (where each risk allele was weighted by their previously published effect sizes [132]) and un-weighted (where BMI-increasing risk alleles were summed up) were calculated from 31 BMI-associated SNPs. The weighted GRS was rescaled so that each point of the GRS corresponded to one risk allele. Since both GRSs produced similar results, results from weighted GRS were presented. GRS was used as both continuous variable and stratified into quintiles.

Annual weight change

Information about weight was available at three time points in MDCS [at age 20 years (young age), at baseline (late middle age) and at follow-up (old age)] and at two time points in GLACIER [at baseline (early middle age) and at follow-up (late middle age)]. The weight change was calculated from the difference in weight between (i) age 20 years and baseline and (ii) baseline and follow-up in MDCS and (iii) baseline and follow-up in GLACIER. The resulting weight change was divided by the follow-up time corresponding to each time period to calculate annual weight change. Annual weight change from baseline to follow-up was further stratified into weight gain (when weight at follow-up was higher than weight at baseline) and weight loss (when weight at follow-up was lower than weight at baseline) groups.

Substantial weight gain

Substantial weight gain was defined as at baseline, having gained ≥10% of their weight at age 20 years in MDCS and at follow-up, having gained ≥10% of their weight at baseline in both MDCS and GLACIER. Participants who gained or lost <10% were considered as having stable weight and those who lost ≥10% were few (<3%) and were included in the stable weight group in all analyses (since excluding them did not influence the results).

Statistical analyses

Data was analyzed using SPSS version 20 (IBM Corp, Armonk, NY, USA), PLINK version 1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/), and STATA version 13 (Stata Corp, College station, TX, USA) in MDCS and replication analyses in GLACIER were performed using SAS version 9.3 (SAS Institute Inc., Cary, NC, USA), STATA version 12.1 and PLINK version 1.07. In MDCS, linear regression was used to analyze association between GRS and BMI at age 20 years, at baseline and at follow-up. Association between annual weight change and substantial weight gain (from age 20 years to baseline and from baseline to follow-up) and SNP/GRS (per unit increase in GRS, per GRS quintile and comparing extreme GRS quintiles)

was analyzed by linear regression and logistic regression respectively. In order to investigate changes in body composition, we additionally performed similar analyses for WC and WHR and annual changes in these traits in relation to BMI GRS (these analyses were not performed in GLACIER as longitudinal data was unavailable). Analyses were adjusted for age, sex and follow-up time when applicable. Model coefficients [effect size (β) or odds ratio (OR)] from these analyses (from age 20 years to baseline and from baseline to follow-up) were compared by implementing seemingly unrelated estimation. GLM was used to test for interaction between GRS and weight gain and weight loss groups on annual weight change. Using time to follow-up as time axis and adjusting for age at baseline and sex, Cox proportional hazard model was used to calculate the hazard ratios (HR) and 95% CI for GRS in relation to total- and CVD mortality. In analyses with GRS, a two sided P value of <0.05 was considered significant and in analyses with single SNPs, Bonferroni corrected P value of <0.0016 (0.05/31) was considered significant. Replication analyses in GLACIER were performed using the same statistical models. Cohort-specific estimates were combined by performing both fixed-effect and random-effect meta-analyses using *metan* command in STATA. Since results were similar using both models, only fixed-effect results were presented in paper IV.

Paper V

Study participants

In paper V, 28,606 MDCS participants with genotype data were included. All of these participants had information about BMI and SBP and of these, 5435 randomly selected participants had information about fasting lipids, lipoproteins and glucose.

GRS calculation

Trait-specific weighted GRSs were calculated using PLINK such that BMI, SBP, TG, LDLC, HDLC and FPG were comprised of 31, 29, 26, 32, 41 and 15 SNPs respectively.

Statistical analyses

All statistical analyses were performed using STATA version 13 (Stata Corp, College station, TX, USA), PLINK version 1.07 and R version 3.1.0.

For comparability, all study traits (BMI, SBP, TG, LDLC, HDLC and FPG) were log transformed and changed to z-scores prior to analyses. Linear regression was used to test association between GRSs and their respective traits. Analyses were adjusted for age and sex and additionally for lipid lowering therapy or antihypertensive therapy when outcome variable was lipid (TG, LDLC or HDLC)

or SBP respectively. Participants with diabetes at baseline were excluded when the outcome variable was FPG.

Observational analyses: Cox proportional hazard regression using age as time-scale and stratifying the analyses by sex and medication use (when applicable), was used to assess associations between trait specific z-scores and outcome traits (total- and CVD mortality). Data were treated as left-truncated and right censored.

Instrumental variable analyses: Causal effect of cardiometabolic traits on total- and CVD mortality was evaluated by using two-stage least square regression. In the first stage, for each cardiometabolic trait, linear regression with trait-specific z-score as the dependent variable and respective GRS (instrument) as the independent variable was performed to obtain predicted fitted values (exposure variables) based on the instrument. In the second stage, Cox regression was performed with mortality as the dependent variable and genetically predicted exposure level from the first stage as the independent variable using same conditions as described above.

Multivariable Mendelian Randomization analyses: To control for the pleiotropic effects, we used multivariable Mendelian randomization (MMR) method [230] and further modified it by using inverse-variant of the outcome as weights [231]. First, each of the 153 SNPs was regressed on each of the cardiometabolic trait using linear regression and each of the outcome trait (total- and CVD mortality) using Cox regression as described above. The resulting β coefficients of each SNP for each cardiometabolic trait were inverse-weighted. Finally, using β coefficients of the outcome traits as the outcome variables, and weighted β coefficients of the cardiometabolic traits as the predictor variables in a single multivariable regression model and using inverse variance of the outcome as weights, weighted linear regression analyses were performed.

Results

Paper I: Genetic susceptibility to obesity and diet intakes

Associations of obesity susceptibility SNPs and GRS with obesity and related traits

In paper I, 16 BMI and/or obesity associated SNPs were first tested for the association with obesity and related traits in 29,480 MDCS participants. More than 50% of the participants were either overweight or obese and 4.2% had diabetes at baseline. Of the tested 16 SNPs, 14 showed directionally consistent associations with BMI as reported in the original GWAS. Apart from *PTER* and *MAF* that did not show any significant associations, and *AIF1* that associated with lower BF% and show a negative trend for association with BMI, rest of the SNPs associated with at least one or more of the anthropometric (BMI, waist and hip circumference) and body composition traits (BF%, FM and FFM) ($P<0.05$ for all) to expected direction. *FTO* rs9939609 explained the highest per-risk allele increase for each of these traits.

For the association with overweight and obesity, six of the loci in/near *FTO, GNPDA2, SEC16B, BDNF, TMEM18* and *NPC1* associated with increased risk for both overweight and obesity. SNPs in/near *SH2B1, SFRS10* and *KCTD15* associated with increased risk for overweight and *MC4R* and *NEGR1* associated with increased risk for obesity. Three loci (*AIF1*, *PTER* and *MAF*) did not associate with risk for overweight or obesity.

To study the cumulative associated effect of all the obesity susceptibility loci, a GRS comprising of 13 SNPs was calculated. Since three loci (*AIF1, PTER* and *MAF*) were not replicated in the present study, they were not included in the GRS. Of these three loci, *PTER* and *MAF* being originally identified as loci for morbid obesity, did not associate with BMI and/or obesity while *AIF1* which was originally identified for association with weight associated with lower BMI in this study. The GRS strongly associated with all the studied anthropometric and body composition traits but not with hip circumference. Each additional BMI-increasing allele of the GRS associated with 0.12 kg/m^2 increase in BMI which corresponds to 347g increase in body weight per allele in a 170 cm tall person. A difference of 2.2 kg in total weight,

1.2 kg in FM and 1.0 kg in FFM was observed when individuals having a low number of risk alleles (GRS≤11) were compared with those having a high number of risk alleles (GRS≥16).

Each unit increase in GRS was associated with 5% increased odds of being overweight and 8% odds of being obese (Table 6). Using individuals with GRS≤11 as the reference group, individuals having GRS≥16 had 1.38- and 1.64 fold higher risk for being overweight or obese, respectively.

Associations of obesity susceptibility SNPs and GRS with dietary intakes

For the associations of individual SNPs with total energy intake, macronutrients and fiber intake, *FTO, GNPDA2, MTCH2, NEGR1, MAF* and *NPC1* showed nominal significant associations with at least one of these dietary traits. However, after correction for multiple comparisons, only the association of *FTO* with lower total energy intake (P=0.001) and *NEGR1* with lower fat intake (P=3.2x10^{-5}), higher CHO intake (P=3.3x10^{-5}) and higher fiber intake (P=1.1x10^{-4}) remained significant. In sensitivity analyses, after excluding potential misreporters of energy, the association of *FTO* with total energy intake became non-significant (P=0.083) but the association of *NEGR1* with fat (P=5.5x10^{-5}), CHO (P=1.8x10^{-4}) and fiber (P=4.0x10^{-6}) intakes remained unchanged.

Mean total energy intake, protein and fiber intakes differed significantly across GRS groups but no significant differences in fat and CHO intakes were observed. Individuals in the high GRS group (GRS≥16) had on average lower total energy intake (P=0.001) and higher intakes of protein (P=0.011) and fiber (P=2.3x10^{-4}). Upon additional adjustment with BMI, the association of GRS with protein intake was no more significant (P=0.25), while the associations with total energy intake and fiber intake remain unaffected. Similarly, in the sensitivity analyses, association of GRS with protein intake became non-significant (P=0.11) but remained unchanged for total energy intake (P=0.019) and fiber (P=2.1x10^{-4}).

Interaction between obesity susceptibility SNPs or GRS and dietary intakes on obesity-related traits

Nominal significant interactions were observed for 11 loci with at least one of the dietary intakes on BMI, FM and FFM. However, after correction for multiple testing, only one locus (*BDNF* rs4923461) showed significant interaction with protein intake on BMI (P_{int}=0.001). High protein intake was significantly associated with higher BMI in all *BDNF* genotype groups but was stronger among BMI-

increasing A-allele homozygotes (P=1.1x10^{-60}) compared to non-risk G-allele homozygotes (P=2.7x10^{-4}). These associations remained similar in sensitivity analyses.

GRS was significantly associated with BMI in each dietary intake and total energy intake quintiles but the associated effect sizes did not significantly differ across the quintiles of fat (P_{int}=0.83), CHO (P_{int}=0.49), protein (P_{int}=0.27), fiber (P_{int}=0.67) or total energy intake (P_{int}=0.63). Similarly no significant interactions between GRS and dietary intakes on odds of being overweight or obese were observed. Additional adjustments with fiber intake or with physical activity levels did not change the results. The results remained similar in the sensitivity analyses.

Paper II: Genetic predisposition to obesity and beverages consumption

In 2012, Qi *et al*. using three large prospective U.S. cohorts of European ancestry have shown that common genetic susceptibility to obesity modifies the association between SSB intake (and not ASB intake) and obesity risk [179]. We sought to replicate this finding in two large Swedish cohorts: MDCS and GLACIER.

Association of beverage intakes and GRS with BMI

Mean daily intake of SSB was similar overall as well as in each of the four categories of SSB intake between MDCS (n=21,824) and GLACIER (n=4905) cohorts. After adjusting for age, sex, cohort specific covariates and putative confounders (see methods for more details), each increment in category of SSB intake was associated with 0.19 (SE=0.02; P=1.2x10^{-16}), 0.04 (SE=0.06; P=0.55) and 0.18 (SE=0.02; P=3.0x10^{-20}) kg/m^2 higher BMI in MDCS, GLACIER and pooled analyses, respectively. After same adjustments, each increment in category of ASB intake was associated with 0.64 (SE=0.04; P=3.9x10^{-58}) kg/m^2 higher BMI in MDCS and this association remained unchanged upon additional adjustment for SSB intake.

Unweighted GRS comprising of 30 BMI associated SNPs was significantly associated with BMI in MDCS [β= 0.09 (SE=0.01) per risk allele; P=5.5x10^{-29}] and GLACIER [β= 0.16 (SE=0.02) per risk allele; P=4.3x10^{-22}]. GRS was not associated with SSB (or ASB) intake in both cohorts. Analyses with weighted GRS produced similar results.

Interaction between GRS and beverage intake on BMI

In the interaction analyses, beverage intake was defined as either four categories (seldom, low, medium and high intake) or two categories (seldom to low and medium to high intake). In the pooled analyses, for each 10 unit increase in GRS, the magnitude of association with BMI increased with each increment in category of SSB intake (P_{int}=0.030). Additional adjustments with putative confounding lifestyle factors reduced the statistical significance of the observed interaction (P_{int}=0.049). In the lifestyle adjusted pooled analyses with two categories of SSB intakes, individuals reporting medium to high intake had on average 1.29 (SE=0.11; P=8.3x10^{-33}) kg/m^2 higher BMI for each 10 unit increase in GRS. Participants reporting seldom to low intake had 0.83 (SE=0.09; P=9.6x10^{-21}) kg/m^2 higher BMI for each 10 unit increase in GRS. Similar results were obtained with weighted GRS.

Paper III: *AMY1* CNV and dietary starch intake

In 2014, Falchi *et al*. indicated a very strong association between CNV in *AMY1* and obesity, and reported this locus as having the largest genomic influence on obesity [220]. However, GWAS studies for obesity have not detected this locus and larger, well powered studies failed to replicate these findings [227]. We set up to investigate this conflicting relationship of *AMY1* CNV with BMI, WHR and BF% in 4047 participants of MDC-CC, challenging the question of if starch intake may modify association between the CNV and obesity traits.

No significant differences were observed in age, BF%, LTPA, and education across starch intake tertiles in neither men nor women. However, in the high starch intake tertile, men had significantly lower WHR (P=0.040), women had significantly lower BMI (P=0.009) and both men and women had a lower total energy intake (P=3x10^{-6} for men; P=5x10^{-15} for women) and lower proportion of smokers (P=0.002 for men; P=2x10^{-7} for women) compared to those in the low starch intake tertile.

Association of *AMY1* CNV with obesity traits

No significant associations were observed between *AMY1* CNV and BMI, WHR or BF% in either men or women (P>0.05 for all). After stratifying by starch intake, we observed opposite direction for association between *AMY1* CNV and BMI in the low and high intake tertiles. BMI decreased with increasing *AMY1* CNV in the low starch intake tertile (P=0.035) while BMI increased with increasing *AMY1* CNV in the high starch intake tertile (P=0.040). Moreover, we observed that participants

with low *AMY1* CNV and high dietary starch intake had the lowest BMI (P=0.024 for men; P=0.005 for women).

No association between *AMY1* CNV and WHR was observed across tertiles of starch intake. However, men with low *AMY1* CNV and high starch intake had significantly lower WHR compared to those with high starch intake (P=0.011). No significant associations between *AMY1* CNV and BF% in strata of starch intake were observed among men, while *AMY1* CNV significantly associated with BF% (P=0.031) among women with high starch intake. Moreover, women with low *AMY1* CNV and high starch intake had significantly lower BF% compared to those with high starch intake (P=0.004).

Interaction between *AMY1* CNV and starch intake on obesity traits

A significant interaction between *AMY1* CNV and starch intake on BMI was observed among women (P_{int}=0.041) in the basic model. However, interaction did not remain significant in the sensitivity analyses when participants with energy misreporting and food habits change were excluded. No significant interactions between *AMY1* CNV and starch intake on BMI, WHR or BF% was observed in men or women when using fully adjusted model (P_{int} >0.05 for all). Interaction results remained unchanged in the sensitivity analyses.

Paper IV: BMI GRS and weight change

In paper IV, we investigated the association between a weighted GRS comprised of 31 BMI associated SNPs and changes in weight at different time points in life using data from MDCS (n=21,407) and GLACIER (n=4,327) cohorts. Mean weight, BMI and proportion of normal weight and obese subjects at baseline and at follow-up were similar in both cohorts. In MDCS, the GRS was significantly associated with higher BMI at age 20 years (P=7x10^{-35}), at baseline (P=2x10^{-34}) and at follow-up (P=0.005) with comparable effect sizes of 0.15, 0.23 and 0.19 kg/m^2 per GRS quintile. The GRS was also significantly associated with higher BMI in GLACIER, both at baseline (P=1x10^{-17}) and at follow up (P=5x10^{-13}) with somewhat higher effect sizes of 0.33 and 0.31 kg/m^2 per GRS quintile as compared to MDCS.

MDCS

Weight change from age 20 years to baseline in MDCS

Self-reported weight information at the age of 20 years was only available in MDCS. The GRS was significantly associated with an increased annual weight per risk allele of GRS [β= 0.003 (SE=0.001); P=7x10^{-8}], per GRS quintile [β= 0.007 (SE=0.001); P=5x10^{-7}] and comparing the extreme quintiles [β= 0.032 (SE=0.006); P=2x10^{-7}]. Among the individual SNPs, only *FTO* rs1558902 showed significant association with increased annual weight gain [β= 0.013 (SE=0.003) kg per risk allele; P=5x10^{-6}] after correction for multiple comparisons.

Of the 21,407 MDCS participants, 29% (n=6182) maintained a stable weight between age 20 years and baseline and 71% (n=15225) gained ≥10% of their weight until baseline. The GRS was associated with 1%, 3% and 12% increased odds for substantial weight gain from age 20 years to baseline per risk allele (P=0.013), per GRS quintile (P=0.021) and comparing the highest and lowest quintiles (P=0.020). *FTO* rs1558902 (P=0.047), *BDNF* rs10767664 (P=0.034) and *TMEM160* rs3810291 (P=0.008) showed nominal significant associations with substantial weight gain, however, these associations were not significant after correction for multiple comparisons.

Weight change from baseline to follow-up in MDCS

The GRS was associated with reduction in weight from baseline to follow-up by -0.006 kg (SE=0.002; P=0.009) per risk allele, -0.018 kg (SE=0.006; P=0.004) per GRS quintile and -0.062 kg (SE=0.03; P=0.026) for comparing highest and lowest quintiles. Nominally significant associations were observed for *MC4R* rs571312 (P=0.042) with weight reduction and for *QPCTL* rs2287019 (P=0.021) and *FLJ35779* rs2112347 (P=0.037) with annual weight gain.

Among MDCS participants (n=2673) with follow-up data, 78% (n=2083) maintained a stable weight between baseline and follow-up and 22% (n=587) gained ≥10% of their baseline weight. The GRS was associated with 4%, 10% and 32% decreased odds for substantial weight gain per risk allele (P=0.001), per GRS quintile (P=0.002) and for comparing highest and lowest quintiles (P=0.011). Additional adjustments with baseline weight did not change the associations of GRS with annual weight change or substantial weight gain. When stratified into weight loss and weight gain groups, the association of GRS with substantial weight gain remained unchanged but the association with the annual weight change was significant only in the weight loss group. However, the associations did not significantly differ between the two groups (P_{int}=0.26). Among the individual SNP associations with substantial weight gain, *FTO* rs1558902 (P=0.008) and *GNPDA2* rs10938397 (P=0.037) showed nominal while *SLC39A8* rs13107325 (P=0.001) showed significant associations. In order to further understand these results, we

performed association analyses between GRS and annual changes in WC and WHR. Similar to annual decrease in weight, GRS was significantly associated with annual reduction in WC and showed similar though non-significant trend for WHR.

Comparison of estimates

We further investigated whether the effect size (β) for association of GRS with annual weight change or odds (OR) for substantial weight gain differ between the two time periods. The effect size for association of GRS with annual weight change from age 20 years to baseline was significantly different from the effect size from baseline to follow-up (P=0.0002 per risk allele and P=0.0001 per GRS quintile). Among the individual SNPs, only nominally significant differences in effect estimates were observed for *FTO* rs1558902 and *MC4R* rs571312 (P>0.01 for both). Similarly, the OR for association of GRS with substantial weight gain from age 20 years to baseline was significantly different from the OR from baseline to follow-up (P=0.0001 per risk allele and P=0.0002 per GRS quintile). Among the individual SNPs, the OR for *FTO* rs1558902, *GNPDA2* rs10938397, *SH2B1* rs7359397 and *TMEM18* rs3810291 were nominally significantly different (P<0.05 for all) and OR for *SLC39A8* rs13107325 was significantly different (P=0.001) when the results from the two time periods were compared.

Association of GRS with mortality

We further tested the association between GRS and total- and CVD mortality in order to clarify if the observed inverse associations between GRS and weight gain during and after middle age could be due to association of GRS with mortality. During a mean follow-up of 15 years, 3879 participants died and of these 1217 died of CVD. No significant association were observed between GRS and total mortality, but higher number of BMI-increasing alleles associated with increased risk of CVD mortality [HR (95% CI)=1.02(1.00-1.03); P=0.029]. Upon further stratifying participants into weight stable and substantial weight gain groups from age 20 years to baseline, GRS associated with CVD mortality particularly among those in the highest GRS quintile compared to those in the lowest GRS quintile [1.33(1.10-1.56); P=0.003] irrespective of the weight gain or weight stable group, while no association with total mortality was observed.

GLACIER

Weight change from baseline to follow-up in MDCS

In GLACIER, no significant association between GRS and annual weight change was observed, and among the individual SNPs only nominally significant

association was observed between *PTBP2* rs1555543 and reduction in weight from baseline to follow-up (*P*=0.030).

Of the 4327 study participants, 79% (3442) maintained stable weight and 21% (885) gained ≥10% weight during follow-up. The GRS was associated with 2% decreased odds for substantial weight gain per risk allele (*P*=0.034) and showed a trend for 5% and 18% decreased odds for substantial weight gain per GRS quintile (*P*=0.055) and for comparing highest and lowest quintiles (*P*=0.087) from baseline to follow-up. None of the SNPs showed association with substantial weight gain. Additional adjustment with baseline weight did not change the results. Further stratification into weight gain and weight loss groups did not change the association of GRS with substantial weight gain but a significant inverse association between GRS and annual weight change was observed in the weight loss group (*P*=0.080 per risk allele; *P*=0.047 per GRS quintile; *P*=0.016 comparing highest and lowest quintiles).

Meta-analyses

In the pooled analyses of MDCS and GLACIER, in total comprising of 7000 participants, the GRS was significantly associated with annual weight reduction per risk allele of GRS [β= -0.005 kg (SE=0.002); *P*=0.002], per GRS quintile [β= -0.012 kg (SE=0.005); *P*=0.007] and comparing the extreme quintiles [β= -0.052 kg (SE=0.020); *P*=0.011] from baseline to follow-up. Among individual SNPs, *FLJ35779* rs2112347 (*P*=0.009) and *PTBP2* rs155543 (*P*=0.023) showed nominally significant association with annual weight change from baseline to follow-up.

Similar to association with annual weight reduction, the GRS was associated with 3%, 7% and 24% decreased odds for substantial weight gain per risk allele (*P*=0.001), per GRS quintile (*P*=0.001) and for comparing highest and lowest quintiles (*P*=0.004). In meta-analysis, *FTO* rs1558902 was the only variant that showed nominally significant associations with decreased odds for substantial weight gain from baseline to follow-up (*P*=0.019).

Paper V: Mendelian randomization analyses for the role of cardiometabolic traits in mortality

In paper V, including 28606 MDCS participants and employing Mendelian randomization approach, we tried to investigate causal associations between cardiometabolic traits and total- and CVD mortality. At baseline, the mean age of the participants was 58 years (SD=7.7 years) and 60.3% of the study participants were women. Of all the participants at baseline, 3.0% (n=868) were on lipid

lowering therapy, 16.9% (n=4841) on antihypertensive therapy and 4.4% (n=1261) had diabetes. During a mean follow-up period of 17.6 years (until 31st December 2013), 27.1% (n=7760) of the participants died of which 9.0% (n=2578) died of CVD. Trait specific GRSs were significantly associated with their respective traits ($P \leq 3x10^{-16}$ for all) and explained 0.8, 0.5, 7.2, 5.7, 4.3 and 2.2% of variation in BMI, SBP, LDLC, HDLC, TG and FPG, respectively.

Observational analyses

Higher levels of BMI, SBP, TG and FPG and lower levels of HDLC were significantly associated with increased risk for total- and CVD mortality ($P<0.05$ for all) in the observational analyses. No significant association were observed between LDLC and risk for total- or CVD mortality ($P>0.05$ for both).

Instrumental variable analyses

Genetically elevated BMI [HR (95% CI)=1.37 (1.06-1.77); $P=0.015$], LDLC [1.10 (1.01-1.20); $P=0.027$] and TG [1.15 (1.04-1.28); $P=0.007$] were associated with higher risk whereas genetically elevated HDLC [0.90 (0.82-0.98); $P=0.020$] was associated with lower risk for total mortality in the IV analyses. Similarly, genetically elevated BMI [1.83 (1.17-2.86); $P=0.008$], LDLC [1.18 (1.02-1.36); $P=0.028$] and TG [1.30 (1.09-1.55); $P=0.004$] were associated with higher risk whereas genetically elevated HDLC [0.78 (0.66-0.91); $P=0.002$] was associated with lower risk for CVD mortality.

Multivariable Mendelian randomization analyses

In line with the associations observed in instrumental variable analyses, MMR analyses suggested a direct causal association between TG and total- ($P=0.017$) and CVD mortality ($P=0.028$), and an inverse causal association between HDLC and total- ($P=0.049$) and CVD mortality ($P=0.005$). However, the significant associations observed in instrumental variable analyses between BMI and LDLC and total- and CVD mortality did not remain significant in MMR analyses. Similar results were observed in the sensitivity analyses, when participants using lipid lowering medications and antihypertensive medications at baseline were excluded when estimating β coefficients for lipid/lipoprotein and SBP SNPs, respectively.

Discussion

The overall purpose of this thesis was to contribute to understanding of reasons behind the rapid rises in the prevalence of obesity, and possibly to assist in devising strategies for the prevention and treatment of this major global health problem. This thesis deals with two broad questions. First, by using lifestyle factors, specially diet, and available genetic information from recent GWAS discoveries, we tried to investigate if and how genetic susceptibility to obesity can be modified by dietary factors (paper I to III) and how genetic susceptibility to obesity relates to weight gain at different time points in life (paper IV). Understanding gene-diet interactions may help to identify sub-groups of people that may benefit from specific lifestyle modifications. Further, understanding the associated effects of genetic variants on weight changes at different ages may help to appropriately implement the lifestyle modifications in people with different age groups. Secondly, we investigated if cardiometabolic traits causally link with mortality (paper V) as identifying causal associations can decrease the disease burden by improving strategies for prevention and treatment.

Paper I

In paper I, 13 of the 16 obesity susceptibility variants were replicated for their cross-sectional associations with BMI, overweight, obesity or body fat distribution. The SNPs in/near *AIF1*, *PTER* and *MAF* were not replicated in the middle-aged Swedish population in paper I and similarly have not been replicated in other studies [232, 233]. GRS comprising of 13 replicated variants was associated with an increase of 347 g in body weight per BMI-increasing allele. GRS was found to associate with higher fiber intake and lower total energy intake but the dietary macronutrients, fiber or total energy intake did not seem to modify the association between GRS and obesity related traits (BMI, FM or FFM). Several of the SNPs showed nominal significant associations with dietary intakes and nominal significant interactions with dietary intakes on obesity related traits. However, after Bonferroni correction, only *NEGR1* locus showed significant associations with fat, CHO and fiber intakes and *BDNF* locus showed significant interaction with protein intake on BMI.

All the studied 16 variants were identified through GWAS for BMI/weight and early-onset morbid adult obesity [130, 131, 219]. Replication studies following GWAS are very important in providing convincing statistical evidence for association, to rule out associations due to biases and to improve effect estimations. Additionally, extension of original findings to more detailed and complex phenotypes is necessary to ascertain the clinical relevance of genetic variation [234]. In paper I, for the replicated loci, the per-allele effect sizes observed for BMI were modest but quite similar to those observed in discovery studies [130, 131]. *FTO* rs9939609 explained the highest per-allele change in BMI in paper I (0.32 kg/m^2) which is comparable to effect sizes observed in GWAS (0.33 kg/m^2) [130]. The second largest per-allele effect size for BMI was observed for *TMEM18* rs6548238 both in our study (0.19 kg/m^2) and in the discovery studies (0.26 kg/m^2) [130, 131]. In line with our study, associations between quantitative measures of obesity such as BMI, weight and WC and/or risk for obesity have also been confirmed in other studies for variants in/near *FTO, MC4R, TMEM18, SEC16B, NEGR1, SH2B1, MTCH2, GNPDA2, FAIM2, BDNF* and *KCTD15* [235-242]. In total these loci explain only 1% of the total genetic variation in BMI in the population [243]. Later on, more recent GWAS efforts have extended the list to 97 BMI-associated genetic variants, that explain 2.7% of the genetic variation in BMI but still a large part of the genetic variation controlling obesity awaits discovery. It is expected that rare variants with potentially large effects and copy number variants may cover part of the missing heritability.

Many of the studied variants are in/near the genes that are expressed particularly in the hypothalamus which is a crucial center for energy balance and regulation of food intake [130]. It is well established that genetic variants do not solely result in obesity without the exposure of an obesogenic environment i.e., increased energy intake and reduced energy expenditure that are considered to be the main culprits in global obesity epidemic in the past decades [244]. Apart from total energy intake, which is the most important aspect of food intake in weight control and obesity development, the macronutrient composition may be an important factor, and may be dependent on genetic factors. In our study, we hypothesized that the relative intakes of fat, CHO, protein and fiber could influence the genetic predisposition to obesity. Among the individual SNPs after accounting for multiple comparisons, *FTO* rs9939609 was significantly associated with lower total energy intake and *NEGR1* rs2815752 was significantly associated with lower fat and higher CHO and fiber intakes. *FTO* variant has earlier shown to be associated with increased total energy intake in children [245-247] and adults from multiple ethnicities [248]. We could not confirm these findings which could be due to the differences in the studied variants, type of participants (children, multiple ethnicities), smaller study populations in the earlier studies as compared to our, and differences in the dietary assessment methods. In line with our results, despite the mentioned differences, other studies both in

European adults [240, 249] and European and African-American children [250, 251] did not find association between *FTO* and increased total energy intake. Moreover, BMI-increasing allele of *FTO* variant was found to be significantly associated with decreased total energy intake [172, 252] in adults and with increased total energy intake in children and adolescents [171] in large scale meta-analyses.

Misreporting is a common inevitable measurement error found in any dietary assessment that relies on self-reports such as FFQ and dietary records. After exclusion of misreporters, the association of *FTO* rs9939609 became non-significant but *NEGR1* rs2815752 results remained unchanged. In MDCS cohort misreporting of energy has been closely examined in relation to other characteristics. Under-reporting of energy was significantly associated with higher BMI, larger WC, lower education and a-blue collar profession while over-reporting of energy was associated with lower BMI, living alone and current smoking in both genders compared to adequate energy reporters [215]. There is no gold standard approach for measuring dietary intake and all diet assessment methods are prone to measurement errors [253]. Measurement errors can be because of random errors in dietary measurement (which occur when there is large day-to-day variation in dietary intake making it difficult to estimate mean daily intake and when dietary intakes are inaccurately measured) or systemic errors (which occur when diet assessment method fail to cover frequently consumed foods in a specific group resulting in underestimation of dietary intakes) [253]. Measurement error in the assessment of dietary data may significantly attenuate the diet–disease association [254]. However, the diet assessment method used in MDCS is a combination of FFQ, 7-days dietary record and extensive interview and proved to have high validity [213] and reproducibility [214]. Energy-adjusted Pearson correlation coefficients in men/women were 0.64/0.69 for fat, 0.66/0.70 for CHO, 0.54/0.53 for protein and 0.74/0.69 for fiber [213].

The association between genetic variation in *FTO* and obesity has been clearly established but the underlying mechanism by which *FTO* variants influence adiposity is unknown. Animal studies have suggested a role of Fto in regulating energy homeostasis but whether it does so by influencing energy intake [255, 256] or energy expenditure [257, 258] is unclear. Moreover, a debate is ongoing concerning the role of the different genes in the FTO locus and it is currently not clear if it is *FTO* itself or one of the nearby genes such as *IRX3* [142] or *RPGRIP1L* [141], or if genetic variation in this locus has functional effects on several genes in the region. Most recently, Claussnitzer *et al*. proposed the switch like behavior of *FTO* rs1421085 variant in browning of white adipocytes and thermogenesis through an evolutionary conserved motif for the *ARID5B* repressor that regulates the downstream expression of *IRX3* and *IRX5* [144]. The variant can disrupt the *ARID5B* repressor binding during early adipocyte differentiation resulting in lipid and fat storage and weight gain by decreasing adipocyte browning and

thermogenesis [144]. Animal studies have proposed a role of *NEGR1* in body weight control, food intake and regulation of energy balance [259, 260].

A higher GRS was associated with lower total energy intake and higher intakes of protein and fiber. These associations could be secondary to the associations between GRS and BMI or could be due to under-reporting. We further tested the associations by additionally adjusting for BMI and physical activity and after excluding energy misreporters but the association of GRS with lower total energy intake and higher fiber remained significant in all analyses. One possibility for the association of GRS with low total energy intake could be that some of the obesity susceptibility variants are involved in regulation of BMR, but this warrants further investigation. The unanticipated association of higher GRS with higher fiber intake remained unanswered. However, the associations between GRS and BMI and related traits were not significantly modified by dietary macronutrients or fiber intakes. Prior to this study, only a few studies have reported interactions between obesity GRS and lifestyle factors. By using a GRS comprised of 12 BMI-associated SNPs, Li *et al*. showed that physically active lifestyle attenuates the genetic predisposition to obesity in ~20,000 participants from EPIC-Norfolk cohort [176]. Qi *et al*. by using a weighted GRS of 32 BMI associated SNPs, demonstrated that genetic association with adiposity is accentuated among participants with higher consumption of SSB [179] and increased hours of television watching while is attenuated among those with increased levels of LTPA [181]. Our study is among the first ones to investigate the role of dietary intakes in modifying the association between GRS and obesity. Simultaneously, an intervention study led by Jääskeläinen *et al*. assessed the interaction between a GRS comprised of 26 BMI-associated SNPs and dietary macronutrient composition on BMI during a 3-year follow-up among 459 participants of the Finish diabetes prevention study. They observed higher BMI by GRS among those who reported diet low in fiber and suggested that the genetic predisposition to obesity could be attenuated by high fiber diet, however the interaction was not significant (P_{int}=0.065) [192].

We further conducted single SNP analyses to examine the contribution of each locus in interaction with dietary exposures on anthropometric traits. We observed a significant interaction between *BDNF* rs4923461 and protein intake on BMI, where risk allele carriers had stronger association between higher protein intake and higher BMI. *BDNF* is a neurotropin that plays a key role in the development of central nervous system and participates in energy metabolism and food intake regulation [261]. An interaction between *BDNF* rs6265 and PUFA intake on WC has been reported among Boston Puerto Rican men [262]. However, the interaction between *BDNF* and protein intake on obesity should be confirmed in other studies to exclude any chance findings. None of the other SNPs showed significant interactions with dietary intakes on BMI and related traits.

There are several challenges in studying gene-lifestyle interactions. Findings are often inconsistent across studies and replication studies are usually lacking. One major challenge in replication studies is the publication bias. Thus, reporting negative results is equally as important as reporting positive results [263]. Another limiting factor is inadequate statistical power due to modest sample sizes and measurement errors in environmental exposures [263]. GEI studies may benefit more from better measurement of environmental exposures rather than increasing sample size. Simulation studies have demonstrated that smaller studies with more precise and repeated measures of exposures and outcomes may be equally powerful as 20 times bigger studies with imprecisely measured exposures [264]. In many studies, like ours, SNP-diet interactions are nominally significant but do not survive a correction from multiple testing and thus it is not clear whether statistical power is an issue. Most likely, the challenges faced by main-effect association studies are also faced by gene-diet interaction studies. As the for almost 95% of the loci associated with chronic diseases including obesity, the underlying mechanism is unknown and it is challenging to select which putative loci are involved in the disease pathway and should be analyzed in context of gene-diet interactions. Simulation studies have shown that variants that show most significant associations in GWAS are not likely to be those which interact with the environment to influence phenotypes of interest. GWAS detects variants of relatively large effect but presence of an interaction with diet is likely to weaken the main effect, thus it is likely that genetic loci that are not the top GWAS hits are those which hold promise for gene-diet interactions [265]. However, there is no empirical or theoretical threshold for selection of these variants.

Since MDCS is a large well phenotyped cohort with highly validated dietary data, it is unlikely that the lack of significant interaction was due to low statistical power. According to our statistical power calculations we had adequate sample size to detect interactions of small magnitude (>80% power to detect gene-diet interaction effect size of 0.022 kg/m^2). We observed some nominal interactions between individual SNPs and dietary intakes on obesity traits, and these could possibly reflect different underlying physiological mechanisms. However, upon combining into GRS, the potential individual SNP interactions may get diluted or neutralized by interactions with specific dietary factors in specific directions. Moreover, we only investigated the interaction with total energy, macronutrients and fiber intake but investigating dietary patterns or examining how overall diet interacts with genetic variation in context of chronic diseases is an important direction for public health and should be further studied. Nevertheless, there is a possibility that interaction effect is an underestimation of the true effect because of measurement error in dietary intakes. Additionally, the present analyses are based on cross-sectional data which limits the interpretation of our findings. Thus, using

longitudinal data with repeated measures might be helpful in reducing measurement errors and improving power to detect interactions.

Paper II

In 2012, Qi *et al*. published a paper showing that SSB intake accentuates genetic predisposition (based upon 30 BMI-associated genetic variants) to obesity [179]. In paper II, we replicated this finding in a pooled sample of 26,729 adults from two large Swedish cohorts, where the mean reported SSB intakes were comparable to those in the American study. In line with that study, we observed that the magnitude of association between SSB intake and BMI was stronger in people with higher number of BMI associated risk alleles. Additional adjustments for potential confounding lifestyle factors had no material impact on the results. Increment in each category of SSB intake was associated with higher BMI in both cohorts. ASB intake was available only in MDCS and was significantly associated with BMI but no significant interaction with genetic predisposition was observed.

A strong body of evidence both from observational studies and randomized intervention trials suggest a positive association between SSB consumption and BMI [266-268]. However, a recent systematic review reported inconsistent evidence from observational studies as in many studies total energy intake was not taken into account making it difficult to evaluate energy-independent role of SSB intake in obesity risk, and none of the intervention studies passed the inclusion criteria [269]. However, another recent review evaluating the results from several systematic reviews and meta-analyses reported discrepant results regarding the association between SSB and obesity. They highlighted several factors such as use of inappropriate study design, lack of energy adjustment and limitations in the dietary assessment methods, that may have affected the observed relationship between SSB and obesity [270]. As reported by Qi *et al*. [179] and further supported by our findings from Swedish cohorts in paper II, these inconsistent findings can partly be explained by taking into account individual genetic susceptibility to obesity.

Since common obesity is a polygenic disease, information from multiple genetic variants is required to characterize genetic susceptibility to obesity. A GRS created by combining several small-effect SNPs may potentially increase power to determine genetic risk for complex diseases [271]. Compared to weighted GRS used by Qi *et al*. we calculated both weighted (see methods for details) and unweighted GRS and found no marked differences in the results. Similarly, previous studies that have compared weighted and unweighted GRS have reported similar effects for both models [236]. A possible explanation could be that in majority of the populations, effect of each allele tends to be normally distributed so the alleles with larger effects

are counter balanced by the alleles with smaller effects. Thus, upon summing these effects, the weighted means approximate that of the unweighted mean [272].

Our results have some public health implications. Since genetic predisposition to obesity is modifiable, principle objective of studying gene lifestyle interaction is to identify high-risk individuals for more efficient and targeted diet and/or lifestyle interventions. In paper II, we observed that higher consumption of SSB increases the risk of obesity particularly among those with higher genetic predisposition to obesity. As SSB contain a large amount of rapidly absorbable sugars despite containing a lot of calories their intake is associated with less satiety and incomplete compensatory reduction in energy intake at subsequent meals [273]. This may lead to over consumption of total daily calories and higher SSB intake has been observed to increase the risk of insulin resistance, beta-cell dysfunction, visceral adiposity, inflammation and other metabolic disorders. However, since biological functions of the studied genetic loci are mostly unknown [132], the mechanisms underlying the observed interaction between SSB intake and genetic predisposition to obesity require further studies.

Our study has some limitations that need to be acknowledged. The main limitation is the cross-sectional study design and the self-reported intakes of SSB and ASB. Additionally, due to the differences in diet assessment methods, SSB intake could not be uniformly defined across the two cohorts. Moreover, seldom consumers in MDCS were those who reported zero consumption of SSB (or ASB) and might harbor certain degree of misclassification. However, similar results in both cohorts reduce the possibility of measurement errors confounding the results. The modified diet history method used in MDCS captured both current (by using a 7-day food record) and habitual dietary intake (by means of FFQ) while in GLACIER only FFQ was used that more likely capture habitual diet intake. Thus differences in total energy intake between the two cohorts, owing to different diet assessment methods, might partly explain why we observed a dose-dependent effect modification in GLACIER similar to the US cohort, but a threshold-effect in MDCS.

The main caloric sweetener used in US is HFCS (which is a mixture of free glucose and fructose) while in Europe it is sucrose (a disaccharide made up of glucose and fructose) [266]. It is not known whether HFCS and sucrose affect obesity risk through the same mechanisms or follow different biological pathways. As it has been shown that added caloric sweeteners irrespective of the type (sucrose, HFCS or fruit-juice concentrates) all result in similar metabolic effects [266] and the similar interaction between GRS and SSB intake on obesity was observed in US and Swedish cohorts may imply that the underlying mechanism may be independent of the sweetener type. Despite differences in culture and various aspects of lifestyle between Europe and US, agreement in the results from both studies warrant the generalizability of these results to other populations of European ancestry.

Paper III

The *AMY1* in humans exhibit extensive copy number variation and recent studies have implicated this variation in adaptation to high starch diets [274] and in association with BMI [220]. In paper III, we investigated whether CNV in *AMY1* associates with obesity traits (BMI, WHR or BF%) and if this association is modified by dietary starch intake. Overall, *AMY1* CNV was not significantly associated with obesity traits in our study population. However, stratification by *AMY1* CNV and starch intake revealed some significant associations and interactions. For participants with low *AMY1* CNV, high starch intake was associated with lower WHR in men, lower BF% in women and lower BMI in both genders. Significant interaction between *AMY1* CNV and starch intake on BMI was observed among women, but this result did not remain significant after adjustment for potential confounding lifestyle factors.

Human amylase locus has been shown to have a very complicated structure. Salivary amylase is encoded by three closely related genes, namely *AMY1A*, *AMY1B* and *AMY1C*, which are 99.9% identical in DNA sequence and are collectively referred as *AMY1*. Pancreatic amylase genes (*AMY2A* and *AMY2B*) are located on the telomeric end of the amylase gene cluster. *AMY2A* and *AMY2B* genes share 94% similar sequence identity with each other, and are 93.2% and 93.6% similar to *AMY1* [275]. Salivary and pancreatic amylase genes encode enzymes that digest starch into sugars and these genes vary widely in their copy numbers (*AMY1*: 2-17 copies; *AMY2A*: 0-8 copies; *AMY2B*: 2-6 copies) [274, 275]. Thus, it has been hypothesized that *AMY1* shapes the metabolic response to diet due to its role in starch metabolism, and a greater average copy number of *AMY1* has been observed in populations with starch-rich diets [274].

AMY1 CNV received special attention as its copy number varies widely and a recent study demonstrated 1.2-fold decrease in obesity for each copy increase in *AMY1* [220] reflecting a profound effect. Another study by the same group reported association between *AMY1* CNV and obesity risk in a smaller sample of Mexican children [276] where higher copy numbers were found to associate with lower risk of obesity (OR per estimated copy = 0.84). However, they did not observed significant associated effect of low *AMY1* copy numbers on increased risk of obesity in children. A Finnish study did not observe any difference in mean *AMY1* copy numbers between 61 childhood-onset obesity cases and 71 controls [277]. Moreover, Falchi *et al*. reported that *AMY1* CNV explains around 11% of the genetic contribution to obesity [220]. However, so far GWAS studies have identified 97 BMI-associated loci and together they explain about 2.7% of the variation in BMI and suggest that common variation account for 21% of BMI variation [6]. Surprisingly, *AMY1* despite explaining such a large proportion of BMI variation in

the Falchi *et al.* study, has not been detected in a GWAS of ~340,000 people [6]. GWAS relies on tag SNPs to capture the extent of variation in human genome and identifying association signals. Variants poorly tagged by GWAS SNPs can be of high importance for association studies such as putative rare variants with high effect sizes. Many CNVs are also poorly tagged by GWAS design. The effectiveness of GWAS SNPs to tag CNV is highly dependent on the type of CNV. For example, tagging properties of bi-allelic CNVs usually closely match those of SNPs as they generally arise from a single ancestral mutation, while multiallelic, highly polymorphic CNVs (like *AMY1* CNV) mutate more frequently, and are therefore often in low LD with GWAS tag SNPs [278].

In Paper III, *AMY1* CNV was genotyped by quantitative polymerase chain reaction (qPCR), similar to the method used by Falchi *et al*. but we observed no significant association between *AMY1* CNV and BMI. Similar to our findings, a very large study led by Usher *et al*. with high statistical power, could not replicate this finding [227]. Complex mCNVs are notoriously difficult to measure (as counting copies is much more difficult than registering presence or absence of an allele) and association studies often involve rough copy number estimates which can be confounded by technical issues with genotyping and can create the false impression of strong association [279]. Carpenter *et al*. reported that qPCR used by Falchi *et al*. for estimation of *AMY1* CNV is less accurate than high resolution molecular analyses [such as paralog ratio tests (PRT) and fiber-fluorescence in situ hybridization (FISH)] [228]. PRT is a precise molecular method which uses paralogous (copy number invariant sequences elsewhere in the genome as embedded controls) to calibrate copy number measurements. Another precise method is the ddPCR in which PCR reaction mixture is partitioned into thousands of nanoliter-sized droplets, each with 0 or 1 copy of the locus of interest, and after thermocycling the number of florescent droplets are counted. Analyses using whole genome sequencing, ddPCR and PRT revealed that odd copy numbers of *AMY1* CNV are four times more common than even copy numbers, and this was not observed with qPCR [226]. Analyses with more precise methods have shown that some SNPs modestly correlate with *AMY1* CNV and if the *AMY1* CNV associates with BMI then the GIANT consortium comprising of >300,000 participants would had 99.9% power to detect this association, but Usher *et al*. found no such association [227]. Similarly, in a sub-sample of ~3000 MDC-CC participants, we did not observe any significant association between these SNPs and BMI.

Apart from genotyping by qPCR method, another limitation of our study is the estimation of dietary intakes from self-reported data, which can be prone to measurement errors. However, as the relative validity of our diet assessment method is quite high, we do not expect the measurement error to have a major impact on our observed results [280]. We observed significant differences in BMI across *AMY1* CNV tertiles in women such that among women with low starch intake, high *AMY1*

CNV associated with lower BMI, while among women with high starch intake *AMY1* CNV associated with higher BMI. These results suggest that starch intake modifies the association between *AMY1* CNV and BMI. The possible explanation for the association of high copy number of *AMY1* with high BMI among those with high starch intake could be the facilitated digestion of starch among high copy number carriers resulting in higher glucose absorption and subsequently more available energy to the cells. This indicates that the amount of starch that is digested is dependent on *AMY1* CNV, if diet is rich in starch.

Moreover, all the significant associations between starch intake and obesity traits were limited to the participants with low *AMY1* CNV. Salivary amylase is the most abundant protein in the human saliva and responsible for the pre-digestion of starch in the oral cavity [281], and *AMY1* CNV is directly proportional to the salivary amylase content [274]. Thus, among participants with low *AMY1* CNV, starch may not be fully hydrolyzed into glucose due to limited amount of salivary amylase. Previous reports demonstrating that direct delivery of starch in the small intestine results in decreased starch digestion and glucose absorption [282], which further highlights the important role of salivary amylase in starch digestion and metabolism. Starch that has not been digested and absorbed by the small intestine, also called resistant starch, is thus transported through the gastro-intestinal tract. Importantly, it has been reported that resistant starch may increase satiety [283], which further supports our results as we observed low BMI among individuals reporting high starch intake but having a low number of copies of *AMY1* CNV.

Recently, *AMY1* CNV was reported to be significantly correlated with BF% and BMI only among obese women with childhood onset obesity [277]. In line with these gender-specific findings, our observations of significant interaction between *AMY1* CNV and starch intake on BMI were restricted to female participants of MDCS. However, the interaction did not remain significant after excluding ~32% of the participants that were categorized as potential misreporters of their dietary intake or who reported having changed their dietary habits, as well as after adjusting for putative lifestyle confounding factors.

Finally, as an attempt to examine if the observed gender specific associations in paper III could be related to differences in food sources of starch, we compared the different starch sources in males and females. A subgroup of MDCS participants (n=3,132) underwent a face-to-face 24-hour dietary recall in 1996-1997 [207]. The average daily starch intake was 142g for males and 105g for females. The major sources of starch were cereal and cereal products (55% in males and 52% in females), bread (35% in males and 30% in females), potatoes (15% in males and 13% in females) and rice (7% in both males and females) [284]. Thus, although men had a higher starch intake as compared to women, the percentage of starch obtained from various food sources was similar in men and women, and it is unlikely that the

observed gender-specific associations in paper III might be due to differential intake of starch among men and women. Our findings suggest a putative role of starch intake in modifying the link between *AMY1* CNV and obesity, particularly in females.

Paper IV

In paper IV, we investigated how established BMI-associated genetic variants affect changes in weight across different adult ages. To clarify whether the genetic variants that positively associate cross-sectionally with BMI, also influence weight gain over time, we studied their association with annual weight change and substantial weight gain (≥10%) between different adult age periods in two prospective Swedish cohorts. The GRS strongly associated with BMI at all adult ages and with increased weight gain both annually and substantially from young age to late middle age. In contrast, higher GRS was associated with significantly decreased annual weight gain and less substantial weight gain after middle age.

Identifying the age associated effects of genetic variants on body weight may help in understanding the mechanisms underlying age related weight changes and devise intervention strategies. Earlier evidence suggest that genetic effects of BMI associated variants vary by age and most of these studies have focused on childhood and adolescence or compared these periods with adulthood [285-287]. There is a limited evidence for genetic influence of BMI-associated variants on changes in weight during adult life course. Nevertheless, although the inverse and significantly different association between GRS and weight change at different adult age periods observed in paper IV have not been reported previously, some evidence for differential genetic effects at different time–points in life have been recognized. Murphy *et al*. investigated the cross-sectional associations between 10 BMI-associated SNPs and anthropometric measurements among Europeans and African Americans aged >65 years. Their findings that genetic loci related to BMI in middle age do not associate with weight and adiposity in older age support our results that genetic variants may have different influences at different ages [288]. Similarly, a Swedish study investigating multifactorial causes of change in BMI over the adult life course demonstrated that a GRS comprising of 32 BMI-associated variants predicted an accelerated increase in BMI until the middle age but not later in life [289]. However, it needs to be recognized that the design of the Swedish study was markedly different from our study and that they did not analyze the associated effect of GRS on weight gain. Among the individual SNPs in our study, the *FTO* variant was significantly associated with decreased odds for substantial weight gain during and after middle age and with significantly different effect estimates on weight

change before and after middle age. In line with this, Hardy *et al*. tested association between *FTO* and *MC4R* variants with BMI and weight at 11 time-points from 2 to 53 years of age and showed that associations reached peak strength at age 20 years and then weakened during adulthood [290].

Further support to our findings comes from a recent work by Winkler *et al*. where they examined the influence of both age and sex, simultaneously, on associated effects of genetic variants affecting BMI and WHR_{adjBMI} in a large scale genome-wide interaction study. They discovered that the association of 15 BMI-associated loci with BMI were influenced by age, and that the association of 44 WHR_{adjBMI}-associated loci with WHR_{adjBMI} were influenced by sex. Further, none of the BMI loci were influenced by sex and none of the WHR_{adjBMI} loci were influenced by age. Of the 15 BMI-associated loci identified to be modified by age in that study, 11 had 1.5 to 3.5 times smaller effect on BMI in older adults (>50 years) compared to younger adults (≤50 years) [158]. The smaller or inverse effects of BMI associated variants on weight gain or BMI in older adults can be expected to reflect a greater influence of environmental and lifestyle factors on adiposity in older adults, thus overwhelming the genetic effects. Sandholt *et al.* investigated the effect of GRS (comprised of 30 BMI-associated loci) on weight change among 3982 individuals from the Danish Inter99 cohort (mean age 46.7 years) during a five year period, and did not find any significant associations [178]. Although these results can be interpreted to be in contrast to our results, they may also be explained by some important differences between the studies. Firstly, Inter99 cohort was on average 11 years younger at baseline and had a 10 years shorter follow-up than MDCS. Secondly, despite the comparable ages of our replication cohort GLACIER (mean age 45.2 years) and Inter99 (mean age 46.3 years) at baseline, the follow-up time of GLACIER was almost twice that of Inter99. Thus, the age of the Inter99 participants may have potentially coincided with a time-point when the increased risk of weight gain associated with GRS may had begun to shift towards decreased risk, resulting in null association. However, in line with findings from Inter99, GRS did not associate with annual weight gain but only with substantial weight gain in GLACIER, putatively indicating a higher statistical power in the latter model.

The mechanisms explaining the significantly inversed consequence of carrying more BMI-increasing alleles on weight gain until and after reaching middle age could not be addressed in our study. The participants gained 0.34kg/year from age 20 years to baseline in MDCS while weight gain from baseline to follow-up was 0.32 and 0.15 kg/year in MDCS and GLACIER, respectively. Thus one might expect smaller effect estimates owing to lesser weight gain during the later time period compared to weight gain in early adulthood, but not the significantly inversed associations we observed. A possible explanation for the differential associated effects of genetic variants on weight gain at different ages could be due to differences in the genetic background for body weight as compared to body weight

fluctuations. It is important to emphasize that all the variants used to calculate GRS were identified in GWAS for cross-sectional associations with BMI, and our results thus indicate that these variants have different effects on weight gain at different points in life.

In order to differentiate genetic effects from non-genetic effects, we stratified the analyses by weight gain and weight loss status during the follow-up. We observed comparatively stronger associations among the group of individuals who lost weight, as compared to those who gained weight from baseline to follow-up, which could suggest a greater influence of non-genetic effects (e.g. dieting and unidentified diseases) on the observed findings. As additional adjustments for baseline weight in the sensitivity analyses did not affect the association of GRS with weight loss after the middle-age argues against the possibility that this finding could be a result of more weight loss in obese people in old age. Both physiological and genetic factors related to aging might explain this relationship. Loss of bone and muscle tissues and increase in body fat has been shown to associate with ageing leading to clinical hazards of obesity at lower BMI in elderly [291-293]. Furthermore, rapid weight gain during midlife has been suggested to have negative health consequences which may lead to weight loss in later life leading to increased mortality [294].

We also tested the association of GRS with mortality and found evidence for association between GRS and CVD mortality independently of whether the study participants gained weight or kept their weight stable until baseline. This association was more pronounced among those with higher number of BMI-increasing alleles. Thus, an implication of the findings may be that the influence of common genetic variants on weight gain is not uniform throughout the life and lifestyle interventions may be particularly important during the young to middle-age period to avoid or dilute the negative effects of increased weight gain during this period on older age, especially among those with higher genetic burden for obesity.

Strengths of our study include a large sample size, longitudinal data, long follow-up period and the availability of a replication cohort from Sweden. However, certain limitations of our study also need to be recognized. First, weight information at 20 years of age in MDCS was self-reported and recalled several decades later by the participants. However, the strong association of GRS with self-reported BMI suggests that recall bias is unlikely to have significantly affected the results. Second, the GLACIER participants were younger at baseline and had a shorter follow-up time, which probably affected the statistical power of the replication analyses, which was reflected in the comparatively weaker associations. Finally, analyzing association between GRS and age-related changes in body composition traits (e.g. FM, FFM and BF%) could have provided insights into the underlying mechanisms behind the observed associations and thus had facilitated the

interpretation of our results, but both cohorts lack data for such analyses. Nevertheless, longitudinal data on WC and WHR was available in MDCS and showed similar association patterns with GRS as the annual weight change, indicating that part of the weight loss was because of decrease in abdominal fat. Lastly, genetic loci could affect weight changes differently in men and women, but we did not perform sex-stratified analyses because this had markedly decreased the statistical power to detect significant associations.

Paper V

In paper V, we attempted to understand the causal role of cardio-metabolic traits in total- and CVD mortality by using the MR approach. We used data from a large cohort of middle-aged Swedish adults that were followed up for more than 17 years. Our findings suggest an inverse causal role between HDLC and both total- and CVD mortality, and a direct causal role between TG and both total- and CVD mortality.

Observational studies provide important insights into disease etiology but their validity has been questioned lately. This is due to the fact that many findings from observational studies have not been confirmed in subsequent large RCTs [295, 296]. However, research in humans relies on observational studies for identification of risk factors as many risk factors cannot be randomized using controlled trials in humans because of technical and ethical reasons. Furthermore, observational studies are prone to spurious results because of confounding, reverse causation and selection bias making it difficult to firmly establish causal relationship between risk factor and disease. This is especially true for common diseases such as obesity, CVD etc. which are complex and influenced by multiple risk factors that may be correlated with each other and with modest influence on disease. Thus, it is hard to find a risk factor that is independently (of all other risk factors) and causally associated with a disease in observational studies. MR is a promising approach to deal with this difficult task of establishing causal inference [297].

Although free from most forms of confounding, the validity of MR findings suffers from two key threats; population stratification and pleiotropy [196]. Population stratification is unlikely to be an issue in MDCS cohort as majority of the participants were either from Sweden or from countries in geographic proximity to Sweden [204]. However, we have used trait-specific GRSs (as IVs) comprising of genetic variants that have, or may have pleiotropic effects. We therefore used two different methods to dissect causality; a traditional MR method employing IVs and a modified MMR method developed to control for potential pleiotropic effects.

BMI significantly associated with increased risk for both total- and CVD mortality in IV analyses. However, no significant association between BMI and total- or CVD mortality was observed after controlling for pleiotropic effects in MMR analyses. Similarly, Fall *et al*. examined causal associations between adiposity and cardio-metabolic traits using *FTO* rs9939609 genotype as the instrumental variable and did not observe causal association between BMI and total mortality [55]. In line with our results, intervention studies have shown no evidence for association between weight loss and total- or CVD mortality [298, 299]. Moreover, evidence for causality regarding adiposity and CVD from earlier MR studies has been inconsistent as some studies support a direct causal relationship [203, 300] while others do not [55, 56]. Thus, larger and well-designed studies are needed to provide conclusive evidence.

SBP strongly associated with increased risk for both total-and CVD mortality supporting the earlier evidence from observational studies [301-303]. However, we did not observe any significant association between SBP and mortality outcomes in either instrumental variable- or MMR analyses. In contrast to our results, RCTs have clearly demonstrated that BP lowering treatment reduces total- and CVD mortality [304]. In the present study, since the SBP GRS only explains about 0.5% of the total genetic variation in SBP, the lack of association between genetically elevated SBP and mortality could be attributed to low power to detect causal effects due to weak instrument.

We did not find any evidence for causal association between LDLC and mortality outcomes in the present study. However, findings from clinical trials have established that lipid lowering therapy with statins reduces the risk of MI and stroke in addition to lowering the risk of total- and CVD mortality in both primary and secondary prevention [305]. In addition to the effects of statins on LDLC lowering, it is well established that they also exert TG lowering effects, and in our study we observed a causal role of TG in total- and CVD mortality. However, statins have not been shown to reduce total mortality in low risk primary prevention setting, but instead have been shown to increase the risk for side effects [306]. Moreover, findings from RCTs [307, 308] and large cohort studies [309] suggest a link between statin use and elevated risk for T2D.In line with this, we [310] and others [311] have shown that genetically lower LDLC associates with increased risk for T2D. This could partially explain the lack of association between genetically elevated LDLC and mortality over a long follow-up period. However, these findings should be interpreted with caution as our results could be biased by unknown confounding. Nonetheless, “negative” results from MR analyses are suggested to be less prone to biases related to violation of assumptions and provide robust evidence when effects are entirely absent or very small [170].

In contrast to previous MR studies [312] and RCT [313] where no causal link between HDLC and coronary events could be established, we observed a putative role for genetically elevated levels of HDLC in significantly lowering the risk for total- and CVD mortality. However, a trial with Torcetrapib, which is a cholesterylester transfer protein (CETP) inhibitor that lowers LDLC and raises HDLC, was prematurely terminated due to increased mortality possibly as a result of off-target drug effects [313]. Thus, our result support a role of HDLC in CVD mortality, including deaths not only due to coronary events but also from stroke and other vascular and heart related complications, and total mortality. In addition to HDLC, we observed a direct causal relationship between TG levels and total- and CVD mortality. In agreement with our findings, MR studies have shown that higher concentration of remnant cholesterol, marked by increased levels of TG, is an additional causal risk factor for total- and CVD mortality [314-316]. However, a recent meta-analysis of RCTs of HDLC increasing drugs such as niacin, fibrates, CETP inhibitors (that also lower TG levels) on stroke, MI and total- and CHD mortality did not observed any additional benefits of these drugs on the top of statins , although when used without statins, niacin and fibrates have shown reduced risk for non-fatal MI [317]. A subgroup analysis of large RCTs has shown that in patients with high TG and low HDLC levels, fibrates reduce the risk of cardiovascular events [318]. Thus, for patients with low HDLC and high TG who are resistant to statin treatment, further studies are needed to clarify the role of fibrates. Furthermore, two large trials with CETP inhibitors; evacetrapib [ACCELERATE (NTC01687998)] and anacetrapib [REVEAL (NTC01252953)] are underway and will report their findings in 2016 and 2017, respectively [319].

The major limitation of our study is the pleiotropic effects of the genetic variants that were used to create the GRS. Although it is difficult to completely exclude confounding by pleiotropy, we exerted efforts to correct for it, but still residual pleiotropic effects could bias our observations. Another limitation of our study can be the relatively low statistical power to detect causal associations as the used GRSs explain only a limited proportion of variance of their respective cardiometabolic traits, especially for BMI and SBP. Thus, in order to reliably confirm the true negative findings, larger studies with better statistical models to control for pleiotropy and stronger non-pleiotropic genetic instruments are needed.

Conclusions

The work presented in this thesis was aimed at identifying the role of dietary factors in modifying the genetic susceptibility to common obesity and to investigate whether genetic effects of the identified loci vary in relation to weight gain at different age periods during a life course. We further aimed to understand if obesity and other cardiometabolic traits are causally linked to mortality. From the findings in papers I to V, we can conclude that:

I. Dietary macronutrients and fiber, or total energy intake levels, do not play a major role in modifying the overall genetic susceptibility to obesity. However, some of the individual obesity loci like *NEGR1* may have a role in the regulation of food intake and *BDNF* may interact with protein intake on BMI.

II. SSB consumption may play a role in modifying the genetic susceptibility to obesity. Thus, reduction in SSB intake may reduce the risk for obesity, especially among people with higher number of BMI-increasing risk alleles.

III. Dietary starch intake may play a putative role in modifying the association between *AMY1* CNV and obesity, at least in females where low copy number of *AMY1* was associated with higher BMI in the lowest tertile of starch intake, but with lower BMI in highest tertile of starch intake.

IV. Genetic susceptibility to obesity associates with higher BMI at all adult ages cross-sectionally. However, we present convincing evidence for a paradoxical inversed relationship between higher number of BMI-increasing risk alleles and reduced weight gain during and after middle age in contrast to increased weight gain in younger age.

V. Our study provides evidence for causal association between TG and HDLC and both total- and CVD mortality. Further evidence is required to understand the causal impact of other cardiometabolic traits on mortality outcomes.

Future perspectives

We are living in exciting era of genetic research when GWAS are making significant development in dissecting the genetic basis of human obesity, and has been successful in identifying almost 100 BMI-associated loci in adults. Despite all the discoveries, identified genetic variants explain only a small percentage of obesity susceptibility and the mechanisms by which they lead to the development of obesity are largely unknown. Of all the identified loci, *FTO* has the largest effect on obesity susceptibility explaining 0.31% of the inter-individual variation in BMI. Thus, we can speculate that rare low frequency variants may contribute more to the obesity-susceptibility. In order to find rare variants, exome genotyping arrays in sample cohorts that have been previously studied by GWAS should be applied. Moreover, in addition to SNP associations, focus should also be given to CNVs, which remain largely unexplored for their association with common obesity. Exome genotyping is cost-effective when applied to large samples. On the other hand, despite being extremely expensive, whole-genome sequencing remains the most powerful approach to study the different types of genetic variants including SNPs, small insertions/deletions, CNVs and other structural variants in a single dataset and cover both common and rare variants. Epigenetics is another promising area that might explain individual differences in obesity risk. Most recently, possible influence of human gut microbiota in obesity has opened new window in understanding obesity etiology.

Similar to other complex traits, dietary and lifestyle factors play a major role in the etiology of obesity. The identification of interactions between genetic markers and environmental exposures may be informative for clinical and/or public heath interventions. However, there are several methodological challenges that need to be overcome before we can identify putative variants that interact with dietary patterns. These include very large sample size (a sample size four times larger than that required to detect the marginal effect), study design, frequency of interacting allele and accurately measured exposures. These issues may be solved by:

- Increasing sample size through collaborations of international consortia similar to GWAS.
- Identifying genetic variants through longitudinal GWAS and conducting genome-wide interaction analyses.
- New statistical methods for studying interactions.

- Studying interactions in different ethnic groups to identify causal genetic and environmental factors.
- Using better measured exposures.
- Testing results of interaction in intervention settings whenever possible and
- Complementing with other 'omics' approaches such as transcriptomics, proteomics and metabolomics to understand how diet and lifestyle alter the expression of our genomes.

Another major challenge of all the observational studies is that they are prone to confounding and reverse causation and randomized control trials still remain the golden standard. However, owing to the extremely high costs of conducting clinical trials and several ethical and technical issues, it is not always feasible or possible to conduct one. An alternating approach is MR that uses genetic variants as instrumental variables to identify causal associations by taking care of confounding and reverse causation. However, to effectively use MR approach in dissecting causal inference, we need new stronger genetic instruments and more reliable and better statistical methods to control for pleiotropic effects and unmeasured confounding.

Populärvetenskaplig sammanfattning

Fetma är ett världsomfattande hälsoproblem som drabbar rika och fattiga länder i samma utsträckning och leder till minskad förväntad livslängd. Detta är inte bara ett problem i sig utan för med sig ytterligare följdsjukdomar så som diabetes, hjärt-kärlsjukdomar, vissa former av cancer och artros. Det är en allmän uppfattning att den oroväckande ökningen av fetma under de två senaste decennierna beror på ohälsosamma matvanor och fysisk inaktivitet, men trots samma levnadssätt drabbas inte alla av fetma. Studier i familjer och tvillingar visar att 40-70% av fetman förklaras av genetiska faktorer. De hittills identifierade genetiska varianterna förklarar dock enbart en liten del av den sammanlagda genetiska risken. Den individuella risken att utveckla fetma beror således på både gener och miljön och möjliga interaktioner mellan dessa. I nuläget finns inte tillräckligt starka bevis för att kunna ge tydliga rekommendationer syftande till att förebygga fetma och visar att stora studier med tydliga sjukdomsdefinitioner behövs.

Att studera interaktioner mellan gener och miljö kan hjälpa till att identifiera grupper av människor där man kan skräddarsy förebyggande och behandling av fetma baserat på individens genetiska profil. I avhandlingen fokuserade vi på genetiska varianter med koppling till BMI och studerade dessa både individuellt och tillsammans i så kallad genetisk riskberäkning (GRS). GRS byggs upp genom att summera antal riskvarianter hos varje individ till ett individuellt score. Vi undersökte om matvanor kan förändra den sammanlagda genetiska risken för fetma. Vi utvärderade också huruvida den genetiska riskscoren påverkade viktuppgång vid olika tidpunkter i livet. Slutligen studerade vi om det finns ett orsakssamband mellan kardiometabola egenskaper och mortalitet, genom att använda en så kallad Mendeliansk randomiseringsmetod. Vi använde oss av data från två olika befolkningsbaserade prospektiva studier, MDCS som består av ca.30 000 individer från södra Sverige och GLACIER som består av ca.19 000 individer från norra Sverige.

I studie I såg vi att energigivande näringsämnen, fibrer och sammantaget energiintag inte påverkar styrkan i kopplingen mellan genetiska riskfaktorer och högre BMI och fetma. Å andra sidan observerade vi att vissa individuella genetiska varianter kan ha en effekt på mat- och energiintag och att effekten av vissa genetiska markörer kan förändras genom kosten. I studie II kunde vi se att individer som har en hög genetisk risk för fetma och dessutom konsumerar stora mängder av socker-sötade

drycker hade högre BMI. I studie III, studerade vi genen för enzymet amylas som förekommer i saliv (*AMY1*) i förhållande till mängd stärkelse i kosten och fetma. Genen förekommer i flera kopior i vår arvsmassa och antalet kopior varierar mellan individer. Vi kunde konstatera att kvinnor med färre kopior av genen *AMY1* hade högre BMI om deras kost innehöll lite stärkelse och *vice versa*, vilket tyder på att mängden stärkelse i kosten kan påverka den genetiska risken för fetma hos kvinnor. I studie IV kunde vi se att individer med högre genetisk risk för fetma ökar mer i vikt i tidig vuxenålder men därefter är förhållandet det motsatta. I studie V, observerade vi ett orsakssamband mellan HDL-kolesterol (”det goda kolesterolet”), triglycerider (en särskild typ av blodfetter) och mortalitet. Individer med högre genetisk risk för höga triglyceridnivåer och låga HDL-kolesterolnivåer har högre risk att dö i förtid, speciellt av hjärt-kärlsjukdomar.

Sammanfattningsvis så visar våra resultat att även om vi inte kan ändra våra gener så kan vi ändra deras effekt med våra livsstilsval. Genom att t.ex. minska konsumtion av sötade drycker kan man minska risken för fetma även hos individer med hög genetisk risk. Våra resultat tyder på att effekten av våra gener förändras under vår livstid och livsstilsanpassnig kan därför krävas. Speciellt individer med hög genetisk risk för fetma bör undvika stor viktuppgång tidigt i vuxenlivet för att inte drabbas av negativa konsekvenser senare i livet. Våra resultat pekar också på ett orsakssamband mellan högre nivåer av triglycerider, lägre nivåer av HDL-kolesterol och högre risk för ökad dödlighet, vilket kan hjälpa att utveckla nya behandlingsmetoder i kliniken.

Acknowledgements

I would like to thank everyone who contributed to this thesis and all those who have supported me during my time as a PhD student.

I would like to express my deepest gratitude to my supervisor **Marju Orho-Melander,** for your continuous support during my Masters and PhD thesis and for your sincerity, patience, understanding, motivation and immense knowledge. Marju, you have been a true inspiration both as a scientist and as a human, I cannot imagine having a better supervisor and mentor. I would also like to thank you outside science for all the summer lunches, X-mas dinners and gifts as well as for arranging our annual retreats at beautiful places in Sweden.

I would also like to thank my co-supervisors **Emily Sonestedt** and **Ulrika Ericson**. Emily, I was inspired by your friendly way of talking and easy going personality the very first day we met which strengthen over the years. Ulrika, for always listening to my questions carefully and giving your best possible advice. I thank you both for all your help, guidance, discussions and feedback whenever I needed. I have learned a lot about diet from you.

My sincere thanks also goes to **Olle Melander**, **Paul W Franks** and **Frida Renström**, for your valuable comments and suggestions in our collaborative projects. , **Elisabet Wirfält,** I would like to thank you for your feedback in my paper I, for the discussions during Nutrition brown bag lunches and for sharing your knowledge from International nutrition conferences. I would also like to thank **Peter Almgren,** for your advice in statistical analyses and to all the co-authors, for your contributions in my so far published and unpublished work.

Malin Svensson, my teacher in the lab and my dearest friend, I owe you a great deal of thanks for teaching me everything from pipetting to genotyping and also for your care, support and help all the times. **Marketa Sjögren**, for creating a lively atmosphere in the lab, for answering all my questions related to SPSS in the beginning and also for translating popular summary of my thesis in Swedish. I have always enjoyed talking to you. **Celine Fernandez** and **Widet Tas**, my friends in the lab, thank you for all the friendly discussions. I would also like to thank all four of you for the surprise of celebrating my birthday that is a great memory, the nice time we spend together in Poland, and for your always loving and caring attitude.

Joana Alves Dias, my sweet friend, office mate and my neighbor in student house, I only came to know you during the last year of my PhD and wish I had known you before. I would like to thank you for all the discussions about everything and for all the hugs**. Yan Borné**, my Chinese friend, I always enjoyed talking to you. **Faiza Siddiqui** and **Ayesha Fawad**, my fellow PhD students from Pakistan and my dear friends. Just talking with you two in Urdu makes me happy.

I would like to thank **George Hindy**, for all your help and guidance in the Mendelian randomization paper and **Gunilla Hughes Wulkan**, for all your help in making my life easy by taking care of all the administrative stuff. I would also like to thank rest of my group members **Ivana Stojkovic**, **Sophie Hellstrand**, **Christina-Alexandra Schulz**, **Louise Brunkwall** and **Tanja Stocks**, I learned a lot from all of you especially during our bi-weekly group meetings.

Ulrika Blom-Nilsson, you are among the first persons I met in CRC and am very much thankful to you for all the administrative assistance in the beginning and thereafter taking care of our progress reports and journal lunches. I would also like to thank all the faculty members, post-docs, fellow PhD students, it-service and CRC-service in CRC for making it a nice working environment. I would also like to thank **Linda**, **Anna** and **Carina,** my friends from daycare, for your understanding and support and making me feel that my child is in good hands.

My mother **Saleha** and my father **Habib**, both of you died during my PhD and I never had the opportunity to truly thank you and there are not enough words in the world to express my love for you. I thank you for giving me a life full of love and friendship, for believing in me and for supporting me in the best way possible. My parents-in-law **Fateh Alam** and **Amna**, I am truly grateful to you for your continuous support and love for all of us. I would also like to thank my beloved family members in Pakistan for being there for me in all the happy and difficult times**, Mujeeb, Haseeb, Mahrukh, Farukh, Ameena, Mehak, Rasheed, Nadeem, Ghazanfar, Ishtiaq, Shoaib, Rukhsana, Shazia, Nimra, Adeela, Ansar, Asiya, Sadia** and **Uzmah**.

Last but not least, **Shafqat**, the shelter and anchor of my life, I want to thank you for your encouragement when I felt low, for always seeing the bright side of things and cheering me up, for never letting me down, for providing me with all the respect and love and for managing the household activities while I completed my work. My kids, **Talha** and **Saad**, for your laughter and hugs that fill my heart with joy. I love you more than anything and have no meaning of life without you three

References

1. Pi-Sunyer, X., *The Medical Risks of Obesity.* Postgraduate medicine, 2009. **121**(6): p. 21-33.
2. Pischon, T., et al., *General and abdominal adiposity and risk of death in Europe.* N Engl J Med, 2008. **359**(20): p. 2105-20.
3. Stunkard, A.J., T.T. Foch, and Z. Hrubec, *A twin study of human obesity.* Jama, 1986. **256**(1): p. 51-4.
4. Maes, H.H., M.C. Neale, and L.J. Eaves, *Genetic and environmental factors in relative body weight and human adiposity.* Behav Genet, 1997. **27**(4): p. 325-51.
5. Allison, D.B., et al., *The heritability of body mass index among an international sample of monozygotic twins reared apart.* Int J Obes Relat Metab Disord, 1996. **20**(6): p. 501-6.
6. Locke, A.E., et al., *Genetic studies of body mass index yield new insights for obesity biology.* Nature, 2015. **518**(7538): p. 197-206.
7. Bray, G.A., *Risks of obesity.* Endocrinology and Metabolism Clinics of North America, 2003. **32**(4): p. 787-804.
8. Ogden, C.L., M.D. Carroll, and K.M. Flegal, *Epidemiologic trends in overweight and obesity.* Endocrinology and Metabolism Clinics of North America, 2003. **32**(4): p. 741-760.
9. Dietz, W.H. and T.N. Robinson, *Use of the body mass index (BMI) as a measure of overweight in children and adolescents.* J Pediatr, 1998. **132**(2): p. 191-3.
10. Keys, A., et al., *Indices of relative weight and obesity.* J Chronic Dis, 1972. **25**(6): p. 329-43.
11. *Obesity: preventing and managing the global epidemic. Report of a WHO consultation.* World Health Organ Tech Rep Ser, 2000. **894**: p. i-xii, 1-253.
12. Ng, M., et al., *Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013.* Lancet, 2014. **384**(9945): p. 766-81.
13. Wang, J., et al., *Anthropometry in body composition. An overview.* Ann N Y Acad Sci, 2000. **904**: p. 317-26.
14. Ulijaszek, S.J. and D.A. Kerr, *Anthropometric measurement error and the assessment of nutritional status.* Br J Nutr, 1999. **82**(3): p. 165-77.
15. Gallagher, D., et al., *How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups?* Am J Epidemiol, 1996. **143**(3): p. 228-39.
16. Blew, R.M., et al., *Assessing the validity of body mass index standards in early postmenopausal women.* Obes Res, 2002. **10**(8): p. 799-808.
17. Garn, S.M., W.R. Leonard, and V.M. Hawthorne, *Three limitations of the body mass index.* Am J Clin Nutr, 1986. **44**(6): p. 996-7.

18. Kamel, E.G., et al., *Measurement of abdominal fat by magnetic resonance imaging, dual-energy X-ray absorptiometry and anthropometry in non-obese men and women.* Int J Obes Relat Metab Disord, 1999. **23**(7): p. 686-92.
19. Clasey, J.L., et al., *The use of anthropometric and dual-energy X-ray absorptiometry (DXA) measures to estimate total abdominal and abdominal visceral fat in men and women.* Obes Res, 1999. **7**(3): p. 256-64.
20. *The Practical Guide: Identification, Evaluation, and Treatment of Overweight and Obesity in Adults.* 2000, National Institutes of Health, National Heart, Lung, and Blood Institute, North American Association for the Study of Obesity p. www.nhlbi.nih.gov.
21. Misra, A., J.S. Wasir, and N.K. Vikram, *Waist circumference criteria for the diagnosis of abdominal obesity are not applicable uniformly to all populations and ethnic groups.* Nutrition, 2005. **21**(9): p. 969-76.
22. Marks, G.C., J.P. Habicht, and W.H. Mueller, *Reliability, dependability, and precision of anthropometric measurements. The Second National Health and Nutrition Examination Survey 1976-1980.* Am J Epidemiol, 1989. **130**(3): p. 578-87.
23. Ellis, K.J., *Human body composition: in vivo methods.* Physiol Rev, 2000. **80**(2): p. 649-80.
24. Hu, F.B., *Measurements of Adiposity and Body Composition*, in *Obesity Epidemiology*. 2008, Oxford University Press: United States of America. p. 53-83.
25. Dempster, P. and S. Aitkens, *A new air displacement method for the determination of human body composition.* Med Sci Sports Exerc, 1995. **27**(12): p. 1692-7.
26. Chumlea WC, S.S., *Bioelectrical impedance analysis*, in *Human Body Composition*, L.T. Heymsfield SB, Wang Z, Going S, Editor. 2005, Human Kinetics, Champaign: IL. p. 79-88.
27. Kyle, U.G., et al., *Bioelectrical impedance analysis-part II: utilization in clinical practice.* Clin Nutr, 2004. **23**(6): p. 1430-53.
28. Sun, G., et al., *Comparison of multifrequency bioelectrical impedance analysis with dual-energy X-ray absorptiometry for assessment of percentage body fat in a large, healthy population.* Am J Clin Nutr, 2005. **81**(1): p. 74-8.
29. Klein, S., et al., *Clinical implications of obesity with specific focus on cardiovascular disease: a statement for professionals from the American Heart Association Council on Nutrition, Physical Activity, and Metabolism: endorsed by the American College of Cardiology Foundation.* Circulation, 2004. **110**(18): p. 2952-67.
30. Must, A., et al., *The disease burden associated with overweight and obesity.* Jama, 1999. **282**(16): p. 1523-9.
31. Burke, G.L., et al., *The Impact of Obesity on Cardiovascular Disease Risk Factors and Subclinical Vascular Disease: The Multi-Ethnic Study of Atherosclerosis.* Archives of internal medicine, 2008. **168**(9): p. 928-935.
32. Wilson, P.W., et al., *Overweight and obesity as determinants of cardiovascular risk: the Framingham experience.* Arch Intern Med, 2002. **162**(16): p. 1867-72.
33. Fox, C.S., et al., *Lifetime risk of cardiovascular disease among individuals with and without diabetes stratified by obesity status in the Framingham heart study.* Diabetes Care, 2008. **31**(8): p. 1582-4.
34. Wilson, P.W., et al., *Prediction of coronary heart disease using risk factor categories.* Circulation, 1998. **97**(18): p. 1837-47.
35. The Global Burden of Metabolic Risk Factors for Chronic Diseases, C., *Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and*

stroke: a pooled analysis of 97 prospective cohorts with 1·8 million participants. Lancet, 2014. **383**(9921): p. 970-983.
36. Lu, Y., et al., *Mediators of the effect of body mass index on coronary heart disease: decomposing direct and indirect effects.* Epidemiology, 2015. **26**(2): p. 153-62.
37. Kenchaiah, S., et al., *Obesity and the risk of heart failure.* N Engl J Med, 2002. **347**(5): p. 305-13.
38. Curtis, J.P., et al., *The obesity paradox: body mass index and outcomes in patients with heart failure.* Arch Intern Med, 2005. **165**(1): p. 55-61.
39. Sharma, A., et al., *Meta-analysis of the relation of body mass index to all-cause and cardiovascular mortality and hospitalization in patients with chronic heart failure.* Am J Cardiol, 2015. **115**(10): p. 1428-34.
40. Walker, S.P., et al., *Body size and fat distribution as predictors of stroke among US men.* Am J Epidemiol, 1996. **144**(12): p. 1143-50.
41. Folsom, A.R., et al., *Incidence of hypertension and stroke in relation to body fat distribution and other risk factors in older women.* Stroke, 1990. **21**(5): p. 701-6.
42. Wang, Z. and W.E. Hoy, *Waist circumference, body mass index, hip circumference and waist-to-hip ratio as predictors of cardiovascular disease in Aboriginal people.* Eur J Clin Nutr, 2004. **58**(6): p. 888-93.
43. Folsom, A.R., et al., *Body mass index, waist/hip ratio, and coronary heart disease incidence in African Americans and whites. Atherosclerosis Risk in Communities Study Investigators.* Am J Epidemiol, 1998. **148**(12): p. 1187-94.
44. Willett, W.C., et al., *Weight, weight change, and coronary heart disease in women. Risk within the 'normal' weight range.* Jama, 1995. **273**(6): p. 461-5.
45. Rimm, E.B., et al., *Body size and fat distribution as predictors of coronary heart disease among middle-aged and older US men.* Am J Epidemiol, 1995. **141**(12): p. 1117-27.
46. Garrison, R.J., et al., *Incidence and precursors of hypertension in young adults: the Framingham Offspring Study.* Prev Med, 1987. **16**(2): p. 235-51.
47. Nguyen, N.T., et al., *Association of Hypertension, Diabetes, Dyslipidemia, and Metabolic Syndrome with Obesity: Findings from the National Health and Nutrition Examination Survey, 1999 to 2004.* Journal of the American College of Surgeons, 2008. **207**(6): p. 928-934.
48. Droyvold, W.B., et al., *Change in body mass index and its impact on blood pressure: a prospective population study.* Int J Obes (Lond), 2005. **29**(6): p. 650-5.
49. Colditz, G.A., et al., *Weight gain as a risk factor for clinical diabetes mellitus in women.* Ann Intern Med, 1995. **122**(7): p. 481-6.
50. Abraham, T.M., et al., *Trends in diabetes incidence: the Framingham Heart Study.* Diabetes Care, 2015. **38**(3): p. 482-7.
51. Carey, V.J., et al., *Body fat distribution and risk of non-insulin-dependent diabetes mellitus in women. The Nurses' Health Study.* Am J Epidemiol, 1997. **145**(7): p. 614-9.
52. Chan, J.M., et al., *Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men.* Diabetes Care, 1994. **17**(9): p. 961-9.
53. Knowler, W.C., et al., *Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin.* N Engl J Med, 2002. **346**(6): p. 393-403.
54. Tuomilehto, J., et al., *Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance.* N Engl J Med, 2001. **344**(18): p. 1343-50.

55. Fall, T., et al., *The role of adiposity in cardiometabolic traits: a Mendelian randomization analysis.* PLoS Med, 2013. **10**(6): p. e1001474.
56. Holmes, M.V., et al., *Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis.* Am J Hum Genet, 2014. **94**(2): p. 198-208.
57. Howard, B.V., G. Ruotolo, and D.C. Robbins, *Obesity and dyslipidemia.* Endocrinol Metab Clin North Am, 2003. **32**(4): p. 855-67.
58. Ginsberg, H.N., Y.-L. Zhang, and A. Hernandez-Ono, *Metabolic Syndrome: Focus on Dyslipidemia.* Obesity, 2006. **14**(S2): p. 41S-49S.
59. Troiano, R.P., et al., *The relationship between body weight and mortality: a quantitative analysis of combined information from existing studies.* Int J Obes Relat Metab Disord, 1996. **20**(1): p. 63-75.
60. Manson, J.E., et al., *Estimating the number of deaths due to obesity: can the divergent findings be reconciled?* J Womens Health (Larchmt), 2007. **16**(2): p. 168-76.
61. Manson, J.E., et al., *Body weight and longevity. A reassessment.* Jama, 1987. **257**(3): p. 353-8.
62. Whitlock, G., et al., *Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies.* Lancet, 2009. **373**(9669): p. 1083-96.
63. Flegal, K.M., et al., *Association of all-cause mortality with overweight and obesity using standard body mass index categories: A systematic review and meta-analysis.* JAMA, 2013. **309**(1): p. 71-82.
64. Flegal, K.M., et al., *Excess deaths associated with underweight, overweight, and obesity.* Jama, 2005. **293**(15): p. 1861-7.
65. Orpana, H.M., et al., *BMI and mortality: results from a national longitudinal study of Canadian adults.* Obesity (Silver Spring), 2010. **18**(1): p. 214-8.
66. Allison, D.B., et al., *Annual deaths attributable to obesity in the United States.* Jama, 1999. **282**(16): p. 1530-8.
67. Calle, E.E., et al., *Body-mass index and mortality in a prospective cohort of U.S. adults.* N Engl J Med, 1999. **341**(15): p. 1097-105.
68. Gillum, R.F. and C.T. Sempos, *Ethnic variation in validity of classification of overweight and obesity using self-reported weight and height in American women and men: the Third National Health and Nutrition Examination Survey.* Nutr J, 2005. **4**: p. 27.
69. Fontaine, K.R., et al., *Years of life lost due to obesity.* Jama, 2003. **289**(2): p. 187-93.
70. Gregg, E.W., et al., *Intentional weight loss and death in overweight and obese U.S. adults 35 years of age and older.* Ann Intern Med, 2003. **138**(5): p. 383-9.
71. Zhang, C., et al., *Abdominal obesity and the risk of all-cause, cardiovascular, and cancer mortality: sixteen years of follow-up in US women.* Circulation, 2008. **117**(13): p. 1658-67.
72. Weinsier, R.L., et al., *The etiology of obesity: relative contribution of metabolic factors, diet, and physical activity.* The American Journal of Medicine, 1998. **105**(2): p. 145-150.
73. Rolls, B.J., E.L. Morris, and L.S. Roe, *Portion size of food affects energy intake in normal-weight and overweight men and women.* American Journal of Clinical Nutrition, 2002. **76**(6): p. 1207-1213.
74. Lissner, L. and B.L. Heitmann, *Dietary fat and obesity: evidence from epidemiology.* Eur J Clin Nutr, 1995. **49**(2): p. 79-90.
75. Summerbell, C.D., C. Cameron, and P.P. Glasziou, *WITHDRAWN: Advice on low-fat diets for obesity.* Cochrane Database Syst Rev, 2008(3): p. Cd003640.

76. Ludwig, D.S., *Clinical update: the low-glycaemic-index diet.* The Lancet. **369**(9565): p. 890-892.
77. Malik, V.S. and F.B. Hu, *Popular weight-loss diets: from evidence to practice.* Nat Clin Pract Cardiovasc Med, 2007. **4**(1): p. 34-41.
78. Nordmann, A.J., et al., *Effects of low-carbohydrate vs low-fat diets on weight loss and cardiovascular risk factors: a meta-analysis of randomized controlled trials.* Arch Intern Med, 2006. **166**(3): p. 285-93.
79. Hu, T., et al., *The Effects of a Low-Carbohydrate Diet vs. a Low-Fat Diet on Novel Cardiovascular Risk Factors: A Randomized Controlled Trial.* Nutrients, 2015. **7**(9): p. 7978-94.
80. Halton, T.L. and F.B. Hu, *The effects of high protein diets on thermogenesis, satiety and weight loss: a critical review.* J Am Coll Nutr, 2004. **23**(5): p. 373-85.
81. Merchant, A.T., et al., *Protein intake is inversely associated with abdominal obesity in a multi-ethnic population.* J Nutr, 2005. **135**(5): p. 1196-201.
82. Skov, A.R., et al., *Randomized trial on protein vs carbohydrate in ad libitum fat reduced diet for the treatment of obesity.* Int J Obes Relat Metab Disord, 1999. **23**(5): p. 528-36.
83. Malik, V.S., M.B. Schulze, and F.B. Hu, *Intake of sugar-sweetened beverages and weight gain: A systematic review.* American Journal of Clinical Nutrition, 2006. **84**(2): p. 274-288.
84. Bray, G.A., S.J. Nielsen, and B.M. Popkin, *Consumption of high-fructose corn syrup in beverages may play a role in the epidemic of obesity.* American Journal of Clinical Nutrition, 2004. **79**(4): p. 537-543.
85. Bray, G.A., *How do we get fat? An epidemiologic and metabolic approach.* Clinics in Dermatology, 2004. **22**(4 SPEC. ISS.): p. 281-288.
86. Hunter, G.R., et al., *Intra-abdominal adipose tissue, physical activity and cardiovascular risk in pre- and post-menopausal women.* International Journal of Obesity, 1996. **20**(9): p. 860-865.
87. Brownson, R.C., T.K. Boehmer, and D.A. Luke, *Declining rates of physical activity in the United States: what are the contributors?* Annu Rev Public Health, 2005. **26**: p. 421-43.
88. Chau, J.Y., et al., *Cross-sectional associations between occupational and leisure-time sitting, physical activity and obesity in working adults.* Preventive Medicine, 2012. **54**(3–4): p. 195-200.
89. Van Dyck, D., et al., *International study of objectively measured physical activity and sedentary time with body mass index and obesity: IPEN adult study.* Int J Obes (Lond), 2015. **39**(2): p. 199-207.
90. Hemmingsson, E. and U. Ekelund, *Is the association between physical activity and body mass index obesity dependent?* Int J Obes (Lond), 2007. **31**(4): p. 663-8.
91. Dombrowski, S.U., et al., *Long term maintenance of weight loss with non-surgical interventions in obese adults: systematic review and meta-analyses of randomised controlled trials.* Bmj, 2014. **348**: p. g2646.
92. Flegal, K.M., et al., *The Influence of Smoking Cessation on the Prevalence of Overweight in the United States.* New England Journal of Medicine, 1995. **333**(18): p. 1165-1170.
93. Shimokata, H., D.C. Muller, and R. Andres, *Studies in the distribution of body fat. III. Effects of cigarette smoking.* Jama, 1989. **261**(8): p. 1169-73.
94. Hofstetter, A., et al., *Increased 24-hour energy expenditure in cigarette smokers.* N Engl J Med, 1986. **314**(2): p. 79-82.

95. Filozof, C., M.C. Fernández Pinilla, and A. Fernández-Cruz, *Smoking cessation and weight gain.* Obesity Reviews, 2004. **5**(2): p. 95-103.
96. Manson, J.E., et al., *A Prospective Study of Obesity and Risk of Coronary Heart Disease in Women.* New England Journal of Medicine, 1990. **322**(13): p. 882-889.
97. Katzmarzyk, P.T., et al., *Familial risk of obesity and central adipose tissue distribution in the general Canadian population.* Am J Epidemiol, 1999. **149**(10): p. 933-42.
98. Lee, J.H., D.R. Reed, and R.A. Price, *Familial risk ratios for extreme obesity: implications for mapping human obesity genes.* Int J Obes Relat Metab Disord, 1997. **21**(10): p. 935-40.
99. Ziegler, A., H. Schafer, and J. Hebebrand, *Risch's lambda values for human obesity estimated from segregation analysis.* Int J Obes Relat Metab Disord, 1997. **21**(10): p. 952-3.
100. Allison, D.B., M.S. Faith, and J.S. Nathan, *Risch's lambda values for human obesity.* Int J Obes Relat Metab Disord, 1996. **20**(11): p. 990-9.
101. Knowler, W.C., et al., *Obesity in the Pima Indians: its magnitude and relationship with diabetes.* Am J Clin Nutr, 1991. **53**(6 Suppl): p. 1543s-1551s.
102. Ravussin, E., et al., *Effects of a traditional lifestyle on obesity in Pima Indians.* Diabetes Care, 1994. **17**(9): p. 1067-74.
103. Elks, C.E., et al., *Variability in the heritability of body mass index: a systematic review and meta-regression.* Front Endocrinol (Lausanne), 2012. **3**: p. 29.
104. Silventoinen, K., et al., *The genetic and environmental influences on childhood obesity: a systematic review of twin and adoption studies.* Int J Obes (Lond), 2010. **34**(1): p. 29-40.
105. Montague, C.T., et al., *Congenital leptin deficiency is associated with severe early-onset obesity in humans*. Nature, 1997. **387**(6636): p. 903-908.
106. Farooqi, I.S., et al., *Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency.* J Clin Invest, 2002. **110**(8): p. 1093-103.
107. Farooqi, I.S. and S. O'Rahilly, *Monogenic obesity in humans.* Annu Rev Med, 2005. **56**: p. 443-58.
108. Hinney, A., et al., *Prevalence, Spectrum, and Functional Characterization of Melanocortin-4 Receptor Gene Mutations in a Representative Population-Based Sample and Obese Adults from Germany.* The Journal of Clinical Endocrinology & Metabolism, 2006. **91**(5): p. 1761-1769.
109. Clement, K., *Genetics of human obesity.* Proc Nutr Soc, 2005. **64**(2): p. 133-42.
110. Newton-Cheh, C. and J.N. Hirschhorn, *Genetic association studies of complex traits: design and analysis issues.* Mutat Res, 2005. **573**(1-2): p. 54-69.
111. Loos, R.F., *The Genetic Determinants of Common Obesity-Susceptibility*, in *Adipose Tissue Biology*, M.E. Symonds, Editor. 2012, Springer New York. p. 317-378.
112. Norman, R.A., et al., *Genomewide search for genes influencing percent body fat in Pima Indians: suggestive linkage at chromosome 11q21-q22. Pima Diabetes Gene Group.* Am J Hum Genet, 1997. **60**(1): p. 166-73.
113. Rankinen, T., et al., *The human obesity gene map: the 2005 update.* Obesity (Silver Spring), 2006. **14**(4): p. 529-644.
114. Saunders, C.L., et al., *Meta-analysis of genome-wide linkage studies in BMI and obesity.* Obesity (Silver Spring), 2007. **15**(9): p. 2263-75.

115. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
116. *A haplotype map of the human genome.* Nature, 2005. **437**(7063): p. 1299-320.
117. Abecasis, G.R., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
118. Fan, J.B., M.S. Chee, and K.L. Gunderson, *Highly parallel genomic assays.* Nat Rev Genet, 2006. **7**(8): p. 632-44.
119. Visscher, Peter M., et al., *Five Years of GWAS Discovery.* American Journal of Human Genetics, 2012. **90**(1): p. 7-24.
120. de Bakker, P.I.W., et al., *Practical aspects of imputation-driven meta-analysis of genome-wide association studies.* Human Molecular Genetics, 2008. **17**(R2): p. R122-R128.
121. Klein, R.J., et al., *Complement factor H polymorphism in age-related macular degeneration.* Science, 2005. **308**(5720): p. 385-9.
122. Herbert, A., et al., *A common genetic variant is associated with adult and childhood obesity.* Science, 2006. **312**(5771): p. 279-83.
123. Scuteri, A., et al., *Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits.* PLoS Genet, 2007. **3**(7): p. e115.
124. Frayling, T.M., et al., *A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity.* Science, 2007. **316**(5826): p. 889-94.
125. Dina, C., et al., *Variation in FTO contributes to childhood obesity and severe adult obesity.* Nat Genet, 2007. **39**(6): p. 724-6.
126. Jacobsson, J.A., H.B. Schiöth, and R. Fredriksson, *The impact of intronic single nucleotide polymorphisms and ethnic diversity for studies on the obesity gene FTO.* Obesity Reviews, 2012. **13**(12): p. 1096-1109.
127. Fawcett, K.A. and I. Barroso, *The genetics of obesity: FTO leads the way.* Trends in Genetics, 2010. **26**(6): p. 266-274.
128. Loos, R.J., et al., *Common variants near MC4R are associated with fat mass, weight and risk of obesity.* Nat Genet, 2008. **40**(6): p. 768-75.
129. Chambers, J.C., et al., *Common genetic variation near MC4R is associated with waist circumference and insulin resistance.* Nat Genet, 2008. **40**(6): p. 716-8.
130. Willer, C.J., et al., *Six new loci associated with body mass index highlight a neuronal influence on body weight regulation.* Nat Genet, 2009. **41**(1): p. 25-34.
131. Thorleifsson, G., et al., *Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity.* Nat Genet, 2009. **41**(1): p. 18-24.
132. Speliotes, E.K., et al., *Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index.* Nat Genet, 2010. **42**(11): p. 937-48.
133. Lindgren, C.M., et al., *Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution.* PLoS Genet, 2009. **5**(6): p. e1000508.
134. Heard-Costa, N.L., et al., *NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium.* PLoS Genet, 2009. **5**(6): p. e1000539.
135. Kilpelainen, T.O., et al., *Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile.* Nat Genet, 2011. **43**(8): p. 753-60.
136. Bradfield, J.P., et al., *A genome-wide association meta-analysis identifies new childhood obesity loci.* Nat Genet, 2012. **44**(5): p. 526-31.

137. Berndt, S.I., et al., *Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture.* Nat Genet, 2013. **45**(5): p. 501-12.
138. Monda, K.L., et al., *A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry.* Nat Genet, 2013. **45**(6): p. 690-6.
139. Guo, Y., et al., *Gene-centric meta-analyses of 108 912 individuals confirm known body mass index loci and reveal three novel signals.* Hum Mol Genet, 2013. **22**(1): p. 184-201.
140. Tung, Y.C., et al., *Obesity and FTO: Changing Focus at a Complex Locus.* Cell Metab, 2014. **20**(5): p. 710-8.
141. Stratigopoulos, G., et al., *Hypomorphism for RPGRIP1L, a ciliary gene vicinal to the FTO locus, causes increased adiposity in mice.* Cell Metab, 2014. **19**(5): p. 767-79.
142. Smemo, S., et al., *Obesity-associated variants within FTO form long-range functional connections with IRX3.* Nature, 2014. **507**(7492): p. 371-375.
143. Tung, Y.C.L., et al., *Obesity and FTO: Changing Focus at a Complex Locus.* Cell Metabolism, 2014. **20**(5): p. 710-718.
144. Claussnitzer, M., C.C. Hui, and M. Kellis, *FTO Obesity Variant and Adipocyte Browning in Humans.* N Engl J Med, 2016. **374**(2): p. 192-3.
145. Herman, Mark A. and Evan D. Rosen, *Making Biological Sense of GWAS Data: Lessons from the FTO Locus.* Cell Metabolism, 2015. **22**(4): p. 538-539.
146. Julia Sarah El-Sayed Moustafa, P.F., *copy number variants and their contribution to the risk of obesity*, in *The Genetics of Obesity*, S.F.A. Grant, Editor. 2014, Springer: e-Book. p. 55-70.
147. Conrad, D.F., et al., *Origins and functional impact of copy number variation in the human genome.* Nature, 2010. **464**(7289): p. 704-12.
148. Handsaker, R.E., et al., *Large multiallelic copy number variations in humans.* Nat Genet, 2015. **47**(3): p. 296-303.
149. Sha, B.Y., et al., *Genome-wide association study suggested copy number variation may be associated with body mass index in the Chinese population.* J Hum Genet, 2009. **54**(4): p. 199-202.
150. Jarick, I., et al., *Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis.* Hum Mol Genet, 2011. **20**(4): p. 840-52.
151. El-Sayed Moustafa, J.S., et al., *Novel association approach for variable number tandem repeats (VNTRs) identifies DOCK5 as a susceptibility gene for severe obesity.* Hum Mol Genet, 2012. **21**(16): p. 3727-38.
152. Walters, R.G., et al., *A new highly penetrant form of obesity due to deletions on chromosome 16p11.2.* Nature, 2010. **463**(7281): p. 671-5.
153. Bochukova, E.G., et al., *Large, rare chromosomal deletions associated with severe early-onset obesity.* Nature, 2010. **463**(7281): p. 666-70.
154. Jacquemont, S., et al., *Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus.* Nature, 2011. **478**(7367): p. 97-102.
155. Bachmann-Gagescu, R., et al., *Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity.* Genet Med, 2010. **12**(10): p. 641-7.
156. Wang, K., et al., *Large copy-number variations are enriched in cases with moderate to extreme obesity.* Diabetes, 2010. **59**(10): p. 2690-4.

157. Ottman, R., *Gene-environment interaction: definitions and study designs.* Prev Med, 1996. **25**(6): p. 764-70.
158. Winkler, T.W., et al., *The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study.* PLoS Genet, 2015. **11**(10): p. e1005378.
159. Thomas, D., *Gene-Environment-Wide Association Studies: Emerging Approaches.* Nature reviews. Genetics, 2010. **11**(4): p. 259-272.
160. Huang, T. and F.B. Hu, *Gene-environment interactions and obesity: recent developments and future directions.* BMC Medical Genomics, 2015. **8**(Suppl 1): p. S2-S2.
161. Kraft, P. and D. Hunter, *Integrating epidemiology and genetic association: the challenge of gene–environment interaction.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2005. **360**(1460): p. 1609-1616.
162. Albert, P.S., et al., *Limitations of the case-only design for identifying gene-environment interactions.* Am J Epidemiol, 2001. **154**(8): p. 687-93.
163. Wareham, N.J., E.H. Young, and R.J. Loos, *Epidemiological study designs to investigate gene-behavior interactions in the context of human obesity.* Obesity (Silver Spring), 2008. **16 Suppl 3**: p. S66-71.
164. Franks, P.W., *Gene x environment interactions in type 2 diabetes.* Curr Diab Rep, 2011. **11**(6): p. 552-61.
165. Franks, P., *Gene × Environment Interactions in Type 2 Diabetes.* Current Diabetes Reports, 2011. **11**(6): p. 552-561.
166. Loos, R.J. and C. Bouchard, *FTO: the first gene contributing to common forms of human obesity.* Obes Rev, 2008. **9**(3): p. 246-50.
167. Andreasen, C.H., et al., *Low physical activity accentuates the effect of the FTO rs9939609 polymorphism on body fat accumulation.* Diabetes, 2008. **57**(1): p. 95-101.
168. Sonestedt, E., et al., *Fat and carbohydrate intake modify the association between genetic variation in the FTO genotype and obesity.* Am J Clin Nutr, 2009. **90**(5): p. 1418-25.
169. Sonestedt, E., et al., *Association between fat intake, physical activity and mortality depending on genetic variation in FTO.* Int J Obes (Lond), 2011. **35**(8): p. 1041-9.
170. Kilpelainen, T.O., et al., *Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children.* PLoS Med, 2011. **8**(11): p. e1001116.
171. Qi, Q., et al., *Dietary Intake, FTO Genetic Variants, and Adiposity: A Combined Analysis of Over 16,000 Children and Adolescents.* Diabetes, 2015. **64**(7): p. 2467-76.
172. Qi, Q., et al., *FTO genetic variants, dietary intake and body mass index: insights from 177,330 individuals.* Hum Mol Genet, 2014. **23**(25): p. 6961-72.
173. Vimaleswaran, K.S., et al., *Association between FTO variant and change in body weight and its interaction with dietary factors: the DiOGenes study.* Obesity (Silver Spring), 2012. **20**(8): p. 1669-74.
174. Roswall, N., et al., *Association between Mediterranean and Nordic diet scores and changes in weight and waist circumference: influence of FTO and TCF7L2 loci.* Am J Clin Nutr, 2014. **100**(4): p. 1188-97.
175. Corella, D., et al., *Statistical and Biological Gene-Lifestyle Interactions of MC4R and FTO with Diet and Physical Activity on Obesity: New Effects on Alcohol Consumption.* PLoS ONE, 2012. **7**(12): p. e52344.

176. Li, S., et al., *Physical activity attenuates the genetic predisposition to obesity in 20,000 men and women from EPIC-Norfolk prospective population study.* PLoS Med, 2010. **7**(8).
177. Ahmad, S., et al., *Gene x physical activity interactions in obesity: combined analysis of 111,421 individuals of European ancestry.* PLoS Genet, 2013. **9**(7): p. e1003607.
178. Sandholt, C.H., et al., *The effect of GWAS identified BMI loci on changes in body weight among middle-aged danes during a five-year period.* Obesity, 2014. **22**(3): p. 901-908.
179. Qi, Q., et al., *Sugar-sweetened beverages and genetic risk of obesity.* N Engl J Med, 2012. **367**(15): p. 1387-96.
180. Qi, Q., et al., *Fried food consumption, genetic risk, and body mass index: gene-diet interaction analysis in three US cohort studies.* Bmj, 2014. **348**: p. g1610.
181. Qi, Q., et al., *Television watching, leisure time physical activity, and the genetic predisposition in relation to body mass index in women and men.* Circulation, 2012. **126**(15): p. 1821-7.
182. Nettleton, J.A., et al., *Gene x dietary pattern interactions in obesity: analysis of up to 68 317 adults of European ancestry.* Hum Mol Genet, 2015. **24**(16): p. 4728-38.
183. Yoshida, T., et al., *Mutation of beta 3-adrenergic-receptor gene and response to treatment of obesity.* Lancet, 1995. **346**(8987): p. 1433-4.
184. Erez, G., et al., *Phenotypic and genetic variation in leptin as determinants of weight regain.* Int J Obes (Lond), 2011. **35**(6): p. 785-92.
185. Ruiz, J.R., et al., *Preliminary findings on the role of PLIN1 polymorphisms on body composition and energy metabolism response to energy restriction in obese women.* Br J Nutr, 2011. **106**(4): p. 486-90.
186. Santos, J.L., et al., *Allelic variants of melanocortin 3 receptor gene (MC3R) and weight loss in obesity: a randomised trial of hypo-energetic high- versus low-fat diets.* PLoS One, 2011. **6**(6): p. e19934.
187. Razquin, C., et al., *A 3-year intervention with a Mediterranean diet modified the association between the rs9939609 gene variant in FTO and body weight changes.* Int J Obes (Lond), 2010. **34**(2): p. 266-72.
188. Lappalainen, T.J., et al., *The common variant in the FTO gene did not modify the effect of lifestyle changes on body weight: the Finnish Diabetes Prevention Study.* Obesity (Silver Spring), 2009. **17**(4): p. 832-6.
189. Huang, T., et al., *FTO genotype, dietary protein, and change in appetite: the Preventing Overweight Using Novel Dietary Strategies trial.* Am J Clin Nutr, 2014. **99**(5): p. 1126-30.
190. Zhang, X., et al., *FTO genotype and 2-year change in body composition and fat distribution in response to weight-loss diets: the POUNDS LOST Trial.* Diabetes, 2012. **61**(11): p. 3005-11.
191. Reinehr, T., et al., *Aggravating effect of INSIG2 and FTO on overweight reduction in a one-year lifestyle intervention.* Arch Dis Child, 2009. **94**(12): p. 965-7.
192. Jaaskelainen, T., et al., *Genetic predisposition to obesity and lifestyle factors--the combined analyses of twenty-six known BMI- and fourteen known waist:hip ratio (WHR)-associated variants in the Finnish Diabetes Prevention Study.* Br J Nutr, 2013. **110**(10): p. 1856-65.
193. Papandonatos, G.D., et al., *Genetic Predisposition to Weight Loss and Regain With Lifestyle Intervention: Analyses From the Diabetes Prevention Program and the Look AHEAD Randomized Controlled Trials.* Diabetes, 2015. **64**(12): p. 4312-21.

194. Davey Smith, G. and S. Ebrahim, *Epidemiology--is it time to call it a day?* Int J Epidemiol, 2001. **30**(1): p. 1-11.
195. Kivimäki, M., et al., *Lifetime body mass index and later atherosclerosis risk in young adults: examining causal links using Mendelian randomization in the Cardiovascular Risk in Young Finns study.* European Heart Journal, 2008. **29**(20): p. 2552-2560.
196. Lawlor, D.A., et al., *Mendelian randomization: using genes as instruments for making causal inferences in epidemiology.* Stat Med, 2008. **27**(8): p. 1133-63.
197. Davey Smith, G. and G. Hemani, *Mendelian randomization: genetic anchors for causal inference in epidemiological studies.* Human Molecular Genetics, 2014. **23**(R1): p. R89-R98.
198. Taylor, A.E., et al., *Mendelian randomization in health research: Using appropriate genetic variants and avoiding biased estimates().* Economics and Human Biology, 2014. **13**(100): p. 99-106.
199. Jansen, H., N.J. Samani, and H. Schunkert, *Mendelian randomization studies in coronary artery disease.* European Heart Journal, 2014. **35**(29): p. 1917-1924.
200. Burgess, S. and S.G. Thompson, *Use of allele scores as instrumental variables for Mendelian randomization.* International Journal of Epidemiology, 2013. **42**(4): p. 1134-1144.
201. Freathy, R.M., et al., *Common variation in the FTO gene alters diabetes-related metabolic traits to the extent expected given its effect on BMI.* Diabetes, 2008. **57**(5): p. 1419-26.
202. Timpson, N.J., et al., *Does greater adiposity increase blood pressure and hypertension risk?: Mendelian randomization using the FTO/MC4R genotype.* Hypertension, 2009. **54**(1): p. 84-90.
203. Nordestgaard, B.G., et al., *The effect of elevated body mass index on ischemic heart disease risk: causal estimates from a Mendelian randomisation approach.* PLoS Med, 2012. **9**(5): p. e1001212.
204. Berglund, G., et al., *The Malmo Diet and Cancer Study. Design and feasibility.* J Intern Med, 1993. **233**(1): p. 45-51.
205. Manjer, J., et al., *The Malmo Diet and Cancer Study: representativity, cancer incidence and mortality in participants and non-participants.* Eur J Cancer Prev, 2001. **10**(6): p. 489-99.
206. Riboli, E., *Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC).* Ann Oncol, 1992. **3**(10): p. 783-91.
207. Slimani, N., et al., *European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study: rationale, design and population characteristics.* Public Health Nutr, 2002. **5**(6b): p. 1125-45.
208. Manjer, J., et al., *Invitation to a population-based cohort study: differences between subjects recruited using various strategies.* Scand J Public Health, 2002. **30**(2): p. 103-12.
209. Pero, R.W., et al., *Quality control program for storage of biologically banked blood specimens in the Malmo Diet and Cancer Study.* Cancer Epidemiol Biomarkers Prev, 1998. **7**(9): p. 803-8.
210. Pero, R.W., et al., *The Malmo biological bank.* J Intern Med, 1993. **233**(1): p. 63-7.
211. Nilsson, P.M., G. Engström, and B. Hedblad, *The metabolic syndrome and incidence of cardiovascular disease in non-diabetic subjects—a population-based study comparing three different definitions.* Diabetic Medicine, 2007. **24**(5): p. 464-472.

212. Callmer, E., et al., *Dietary assessment methods evaluated in the Malmo food study.* J Intern Med, 1993. **233**(1): p. 53-7.
213. Riboli, E., et al., *The Malmo Food Study: validity of two dietary assessment methods for measuring nutrient intake.* Int J Epidemiol, 1997. **26 Suppl 1**: p. S161-73.
214. Elmstahl, S., et al., *The Malmo Food Study: the reproducibility of a novel diet history method and an extensive food frequency questionnaire.* Eur J Clin Nutr, 1996. **50**(3): p. 134-42.
215. Mattisson, I., et al., *Misreporting of energy: prevalence, characteristics of misreporters and influence on observed risk estimates in the Malmo Diet and Cancer cohort.* Br J Nutr, 2005. **94**(5): p. 832-42.
216. Hallmans, G., et al., *Cardiovascular disease and diabetes in the Northern Sweden Health and Disease Study Cohort - evaluation of risk factors and their interactions.* Scand J Public Health Suppl, 2003. **61**: p. 18-24.
217. Kurbasic, A., et al., *Gene-Lifestyle Interactions in Complex Diseases: Design and Description of the GLACIER and VIKING Studies.* Curr Nutr Rep, 2014. **3**(4): p. 400-411.
218. Taylor, H.L., et al., *A questionnaire for the assessment of leisure time physical activities.* J Chronic Dis, 1978. **31**(12): p. 741-55.
219. Meyre, D., et al., *Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations.* Nat Genet, 2009. **41**(2): p. 157-9.
220. Falchi, M., et al., *Low copy number of the salivary amylase gene predisposes to obesity.* Nat Genet, 2014. **46**(5): p. 492-7.
221. Newton-Cheh, C., et al., *Genome-wide association study identifies eight loci associated with blood pressure.* Nat Genet, 2009. **41**(6): p. 666-76.
222. Ehret, G.B., et al., *Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk.* Nature, 2011. **478**(7367): p. 103-9.
223. Wain, L.V., et al., *Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure.* Nat Genet, 2011. **43**(10): p. 1005-11.
224. Teslovich, T.M., et al., *Biological, clinical and population relevance of 95 loci for blood lipids.* Nature, 2010. **466**(7307): p. 707-13.
225. Dupuis, J., et al., *New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk.* Nat Genet, 2010. **42**(2): p. 105-16.
226. Usher, C.L. and S.A. McCarroll, *Complex and multi-allelic copy number variation in human disease.* Brief Funct Genomics, 2015. **14**(5): p. 329-38.
227. Usher, C.L., et al., *Structural forms of the human amylase locus and their relationships to SNPs, haplotypes and obesity.* Nat Genet, 2015. **47**(8): p. 921-5.
228. Carpenter, D., et al., *Obesity, starch digestion and amylase: association between copy number variants at human salivary (AMY1) and pancreatic (AMY2) amylase genes.* Hum Mol Genet, 2015. **24**(12): p. 3472-80.
229. Cornelis, M.C., et al., *Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry.* Ann Intern Med, 2009. **150**(8): p. 541-50.
230. Do, R., et al., *Common variants associated with plasma triglycerides and risk for coronary artery disease.* Nat Genet, 2013. **45**(11): p. 1345-52.
231. Burgess, S., F. Dudbridge, and S.G. Thompson, *Re: "Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects".* Am J Epidemiol, 2015. **181**(4): p. 290-1.

232. Sandholt, C.H., et al., *Studies of metabolic phenotypic correlates of 15 obesity associated gene variants.* PLoS One, 2011. **6**(9): p. e23531.
233. den Hoed, M., et al., *Evaluation of common genetic variants identified by GWAS for early onset and morbid obesity in population-based samples.* Int J Obes (Lond), 2013. **37**(2): p. 191-6.
234. Kraft, P., E. Zeggini, and J.P.A. Ioannidis, *Replication in genome-wide association studies.* Statistical science : a review journal of the Institute of Mathematical Statistics, 2009. **24**(4): p. 561-573.
235. 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 Genet, 2010. **6**(4): p. e1000916.
236. Renstrom, F., et al., *Replication and extension of genome-wide association study results for obesity in 4923 adults from northern Sweden.* Hum Mol Genet, 2009. **18**(8): p. 1489-96.
237. Hotta, K., et al., *Association between obesity and polymorphisms in SEC16B, TMEM18, GNPDA2, BDNF, FAIM2 and MC4R in a Japanese population.* J Hum Genet, 2009. **54**(12): p. 727-31.
238. Holzapfel, C., et al., *Genes and lifestyle factors in obesity: results from 12,462 subjects from MONICA/KORA.* Int J Obes (Lond), 2010. **34**(10): p. 1538-45.
239. Haupt, A., et al., *Novel obesity risk loci do not determine distribution of body fat depots: a whole-body MRI/MRS study.* Obesity (Silver Spring), 2010. **18**(6): p. 1212-7.
240. Bauer, F., et al., *Obesity genes identified in genome-wide association studies are associated with adiposity measures and potentially with nutrient-specific food preference.* Am J Clin Nutr, 2009. **90**(4): p. 951-9.
241. Almen, M.S., et al., *The obesity gene, TMEM18, is of ancient origin, found in majority of neuronal cells in all major brain regions and associated with obesity in severely obese children.* BMC Med Genet, 2010. **11**: p. 58.
242. 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, 2010. **95**(5): p. 2418-25.
243. Hofker, M. and C. Wijmenga, *A supersized list of obesity genes.* Nat Genet, 2009. **41**(2): p. 139-40.
244. Hill, J.O., et al., *Obesity and the environment: where do we go from here?* Science, 2003. **299**(5608): p. 853-5.
245. Cecil, J.E., et al., *An obesity-associated FTO gene variant and increased energy intake in children.* N Engl J Med, 2008. **359**(24): p. 2558-66.
246. Wardle, J., et al., *The FTO gene and measured food intake in children.* Int J Obes (Lond), 2009. **33**(1): p. 42-5.
247. Haupt, A., et al., *Variation in the FTO gene influences food intake but not energy expenditure.* Exp Clin Endocrinol Diabetes, 2009. **117**(4): p. 194-7.
248. Lear, S.A., et al., *Associations of the FTO rs9939609 variant with discrete body fat depots and dietary intake in a multi-ethnic cohort.* Genet Res (Camb), 2011. **93**(6): p. 419-26.
249. Lappalainen, T., et al., *Association of the fat mass and obesity-associated (FTO) gene variant (rs9939609) with dietary intake in the Finnish Diabetes Prevention Study.* Br J Nutr, 2012. **108**(10): p. 1859-65.

250. Hakanen, M., et al., *FTO genotype is associated with body mass index after the age of seven years but not with energy intake or leisure-time physical activity.* J Clin Endocrinol Metab, 2009. **94**(4): p. 1281-7.
251. Liu, G., et al., *FTO variant rs9939609 is associated with body mass index and waist circumference, but not with energy intake or physical activity in European- and African-American youth.* BMC Med Genet, 2010. **11**: p. 57.
252. Livingstone, K.M., et al., *Associations between FTO genotype and total energy and macronutrient intake in adults: a systematic review and meta-analysis.* Obes Rev, 2015. **16**(8): p. 666-78.
253. Kipnis, V. and L.S. Freedman, *Impact of exposure measurement error in nutritional epidemiology.* J Natl Cancer Inst, 2008. **100**(23): p. 1658-9.
254. Paeratakul, S., et al., *Measurement error in dietary data: implications for the epidemiologic study of the diet-disease relationship.* Eur J Clin Nutr, 1998. **52**(10): p. 722-7.
255. Church, C., et al., *Overexpression of Fto leads to increased food intake and results in obesity.* Nat Genet, 2010. **42**(12): p. 1086-92.
256. McMurray, F., et al., *Adult onset global loss of the fto gene alters body composition and metabolism in the mouse.* PLoS Genet, 2013. **9**(1): p. e1003166.
257. Fischer, J., et al., *Inactivation of the Fto gene protects from obesity.* Nature, 2009. **458**(7240): p. 894-8.
258. Church, C., et al., *A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene.* PLoS Genet, 2009. **5**(8): p. e1000599.
259. Boender, A.J., et al., *The obesity-associated gene Negr1 regulates aspects of energy balance in rat hypothalamic areas.* Physiol Rep, 2014. **2**(7).
260. Lee, A.W., et al., *Functional inactivation of the genome-wide association study obesity gene neuronal growth regulator 1 in mice causes a body mass phenotype.* PLoS One, 2012. **7**(7): p. e41537.
261. Rosas-Vargas, H., J.D. Martinez-Ezquerro, and T. Bienvenu, *Brain-derived neurotrophic factor, food intake regulation, and obesity.* Arch Med Res, 2011. **42**(6): p. 482-94.
262. Ma, X.-Y., et al., *Association between BDNF rs6265 and Obesity in the Boston Puerto Rican Health Study.* Journal of Obesity, 2012. **2012**: p. 102942.
263. Cornelis, M.C. and F.B. Hu, *Gene-environment interactions in the development of type 2 diabetes: recent progress and continuing challenges.* Annu Rev Nutr, 2012. **32**: p. 245-59.
264. Wong, M.Y., et al., *The detection of gene-environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement?* Int J Epidemiol, 2003. **32**(1): p. 51-7.
265. Frazier-Wood, A.C., *Dietary Patterns, Genes, and Health: Challenges and Obstacles to be Overcome.* Curr Nutr Rep, 2015. **4**: p. 82-87.
266. Hu, F.B. and V.S. Malik, *Sugar-sweetened beverages and risk of obesity and type 2 diabetes: epidemiologic evidence.* Physiol Behav, 2010. **100**(1): p. 47-54.
267. Ebbeling, C.B., et al., *A randomized trial of sugar-sweetened beverages and adolescent body weight.* N Engl J Med, 2012. **367**(15): p. 1407-16.
268. de Ruyter, J.C., et al., *A trial of sugar-free or sugar-sweetened beverages and body weight in children.* N Engl J Med, 2012. **367**(15): p. 1397-406.

269. Trumbo, P.R. and C.R. Rivers, *Systematic review of the evidence for an association between sugar-sweetened beverage consumption and risk of obesity.* Nutrition Reviews, 2014. **72**(9): p. 566-574.
270. Keller, A. and S. Bucher Della Torre, *Sugar-Sweetened Beverages and Obesity among Children and Adolescents: A Review of Systematic Literature Reviews.* Child Obes, 2015. **11**(4): p. 338-46.
271. Belsky, D.W., et al., *Development and evaluation of a genetic risk score for obesity.* Biodemography Soc Biol, 2013. **59**(1): p. 85-100.
272. He, M., et al., *Obesity genotype score and cardiovascular risk in women with type 2 diabetes mellitus.* Arterioscler Thromb Vasc Biol, 2010. **30**(2): p. 327-32.
273. Hu, F.B., *Resolved: there is sufficient scientific evidence that decreasing sugar-sweetened beverage consumption will reduce the prevalence of obesity and obesity-related diseases.* Obes Rev, 2013. **14**(8): p. 606-19.
274. Perry, G.H., et al., *Diet and the evolution of human amylase gene copy number variation.* Nat Genet, 2007. **39**(10): p. 1256-60.
275. Groot, P.C., W.H. Mager, and R.R. Frants, *Interpretation of polymorphic DNA patterns in the human alpha-amylase multigene family.* Genomics, 1991. **10**(3): p. 779-85.
276. Mejia-Benitez, M.A., et al., *Beneficial effect of a high number of copies of salivary amylase AMY1 gene on obesity risk in Mexican children.* Diabetologia, 2015. **58**(2): p. 290-4.
277. Viljakainen, H., et al., *Low Copy Number of the AMY1 Locus Is Associated with Early-Onset Female Obesity in Finland.* PLoS One, 2015. **10**(7): p. e0131883.
278. Zanda, M., et al., *A Genome-Wide Assessment of the Role of Untagged Copy Number Variants in Type 1 Diabetes.* PLoS Genet, 2014. **10**(5): p. e1004367.
279. Cantsilieris, S. and S.J. White, *Correlating multiallelic copy number polymorphisms with disease susceptibility.* Hum Mutat, 2013. **34**(1): p. 1-13.
280. Elmstahl, S., et al., *The Malmo Food Study: the relative validity of a modified diet history method and an extensive food frequency questionnaire for measuring food intake.* Eur J Clin Nutr, 1996. **50**(3): p. 143-51.
281. Noble, R.E., *Salivary alpha-amylase and lysozyme levels: a non-invasive technique for measuring parotid vs submandibular/sublingual gland activity.* J Oral Sci, 2000. **42**(2): p. 83-6.
282. Fogel, M.R. and G.M. Gray, *Starch hydrolysis in man: an intraluminal process not requiring membrane digestion.* J Appl Physiol, 1973. **35**(2): p. 263-7.
283. Willis, H.J., et al., *Greater satiety response with resistant starch and corn bran in human subjects.* Nutr Res, 2009. **29**(2): p. 100-5.
284. Cust, A.E., et al., *Total dietary carbohydrate, sugar, starch and fibre intakes in the European Prospective Investigation into Cancer and Nutrition.* Eur J Clin Nutr, 2009. **63 Suppl 4**: p. S37-60.
285. Zhao, J., et al., *Role of BMI-associated loci identified in GWAS meta-analyses in the context of common childhood obesity in European Americans.* Obesity (Silver Spring), 2011. **19**(12): p. 2436-9.
286. den Hoed, M., et al., *Genetic susceptibility to obesity and related traits in childhood and adolescence: influence of loci identified by genome-wide association studies.* Diabetes, 2010. **59**(11): p. 2980-8.

287. Hardy, R., et al., *Life course variations in the associations between FTO and MC4R gene variants and body size.* Human Molecular Genetics, 2010. **19**(3): p. 545-552.
288. Murphy, R.A., et al., *Candidate Gene Association Study of BMI-Related Loci, Weight, and Adiposity in Old Age.* The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 2013. **68**(6): p. 661-666.
289. Dahl, A.K., et al., *Multifactorial analysis of changes in body mass index across the adult life course: a study with 65 years of follow-up.* Int J Obes (Lond), 2014. **38**(8): p. 1133-41.
290. Hardy, R., et al., *Life course variations in the associations between FTO and MC4R gene variants and body size.* Hum Mol Genet, 2010. **19**(3): p. 545-52.
291. Kaji, H., *Linkage between muscle and bone: common catabolic signals resulting in osteoporosis and sarcopenia.* Curr Opin Clin Nutr Metab Care, 2013. **16**(3): p. 272-7.
292. Datta, N.S., *Muscle-bone and fat-bone interactions in regulating bone mass: do PTH and PTHrP play any role?* Endocrine, 2014. **47**(2): p. 389-400.
293. Han, T.S., A. Tajar, and M.E. Lean, *Obesity and weight management in the elderly.* Br Med Bull, 2011. **97**: p. 169-96.
294. Zajacova, A. and J. Ailshire, *Body mass trajectories and mortality among older adults: a joint growth mixture-discrete-time survival analysis.* Gerontologist, 2014. **54**(2): p. 221-31.
295. Lawlor, D.A. and G.D. Smith, *Cardiovascular risk and hormone replacement therapy.* Curr Opin Obstet Gynecol, 2006. **18**(6): p. 658-65.
296. Lawlor, D.A., et al., *Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence?* Lancet, 2004. **363**(9422): p. 1724-7.
297. Bochud, M. and V. Rousson, *Usefulness of Mendelian Randomization in Observational Epidemiology.* International Journal of Environmental Research and Public Health, 2010. **7**(3): p. 711-728.
298. James, W.P., et al., *Effect of sibutramine on cardiovascular outcomes in overweight and obese subjects.* N Engl J Med, 2010. **363**(10): p. 905-17.
299. Wing, R.R., et al., *Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes.* N Engl J Med, 2013. **369**(2): p. 145-54.
300. Hagg, S., et al., *Adiposity as a cause of cardiovascular disease: a Mendelian randomization study.* Int J Epidemiol, 2015. **44**(2): p. 578-86.
301. Murakami, Y., et al., *Relation of blood pressure and all-cause mortality in 180,000 Japanese participants: pooled analysis of 13 cohort studies.* Hypertension, 2008. **51**(6): p. 1483-91.
302. Lewington, S., et al., *Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies.* Lancet, 2002. **360**(9349): p. 1903-13.
303. Lieb, W., et al., *Genetic predisposition to higher blood pressure increases coronary artery disease risk.* Hypertension, 2013. **61**(5): p. 995-1001.
304. Staessen, J.A., et al., *Risks of untreated and treated isolated systolic hypertension in the elderly: meta-analysis of outcome trials.* Lancet, 2000. **355**(9207): p. 865-72.
305. Fulcher, J., et al., *Efficacy and safety of LDL-lowering therapy among men and women: meta-analysis of individual data from 174,000 participants in 27 randomised trials.* Lancet, 2015. **385**(9976): p. 1397-405.
306. Abramson, J.D., et al., *Should people at low risk of cardiovascular disease take a statin?* Bmj, 2013. **347**: p. f6123.

307. Ridker, P.M., et al., *Cardiovascular benefits and diabetes risks of statin therapy in primary prevention: an analysis from the JUPITER trial.* Lancet, 2012. **380**(9841): p. 565-71.
308. Preiss, D., et al., *Risk of incident diabetes with intensive-dose compared with moderate-dose statin therapy: a meta-analysis.* Jama, 2011. **305**(24): p. 2556-64.
309. Corrao, G., et al., *Statins and the risk of diabetes: evidence from a large population-based cohort study.* Diabetes Care, 2014. **37**(8): p. 2225-32.
310. Hindy G, R.G., Almgren P, Schulz C, Ericson U, Melander O, Orho-Melander M, *Causal effect of decreased LDL cholesterol and increased blood pressure on higher incidence of type 2 diabetes by Mendelian randomisation in the Malmo Diet and Cancer Study.* , in *Diabetologia*. 2014. p. S67-S67.
311. Fall, T., et al., *Using Genetic Variants to Assess the Relationship Between Circulating Lipids and Type 2 Diabetes.* Diabetes, 2015. **64**(7): p. 2676-84.
312. Voight, B.F., et al., *Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study.* Lancet, 2012. **380**(9841): p. 572-80.
313. Barter, P.J., et al., *Effects of torcetrapib in patients at high risk for coronary events.* N Engl J Med, 2007. **357**(21): p. 2109-22.
314. Varbo, A., et al., *Remnant cholesterol as a causal risk factor for ischemic heart disease.* J Am Coll Cardiol, 2013. **61**(4): p. 427-36.
315. Jorgensen, A.B., et al., *Genetically elevated non-fasting triglycerides and calculated remnant cholesterol as causal risk factors for myocardial infarction.* Eur Heart J, 2013. **34**(24): p. 1826-33.
316. Thomsen, M., et al., *Low nonfasting triglycerides and reduced all-cause mortality: a mendelian randomization study.* Clin Chem, 2014. **60**(5): p. 737-46.
317. Keene, D., et al., *Effect on cardiovascular risk of high density lipoprotein targeted drug treatments niacin, fibrates, and CETP inhibitors: meta-analysis of randomised controlled trials including 117,411 patients.* Bmj, 2014. **349**: p. g4379.
318. Lee, M., et al., *Efficacy of fibrates for cardiovascular risk reduction in persons with atherogenic dyslipidemia: a meta-analysis.* Atherosclerosis, 2011. **217**(2): p. 492-8.
319. Durrington, P.N., *With statin co-administration, drugs designed to increase HDL have no impact on cardiovascular outcomes.* Evid Based Med, 2015. **20**(1): p. 12.

Gull Rukh graduated in biochemistry from Punjab University, Lahore Pakistan. She obtained her M.Sc in Molecular biology from University of Skövde, Sweden.

In her doctoral thesis in genetic epidemiology she focused on:

- Gene-diet interactions in obesity
- Association between genetic risk for obesity and weight gain at different ages
- Causal associations between cardiometabolic traits and mortality

Printed by Media-Tryck, Lund University 2016

LUND UNIVERSITY
Faculty of Medicine

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2016:43
ISBN 978-91-7619-269-6
ISSN 1652-8220