UPTEC STS08 005
Examensarbete 20 p
Februari 2008

Single nucleotide polymorphisms in genes involved in body weight homeostasis and their association with obesity related phenotypes

Markus Larsson

UPPSALA
UNIVERSITET

Teknisk- naturvetenskaplig fakultet
UTH-enheten

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

Abstract

Single nucleotide polymorphisms in genes involved in body weight homeostasis and their association with obesity related phenotypes

Markus Larsson

Obesity is a growing health problem, primarily in the industrialized countries, with many different and co-varying sources. This thesis report is devoted to the genetic reasons for childhood obesity, and is examining the correlations between variations in three genes coding for plausible weight-regulating proteins, and various obesity-related traits. A cohort of 420 obese and 512 normal-weight youths were included in the study, donating blood from which DNA was extracted. The obese individuals also provided information on weight, length, and BMI, as well as various biochemical variables. The DNA of the individuals in the sampled population was amplified and examined for single-nucleotide polymorphisms, mutations of one single nucleotide base in the DNA sequence. A total of 11 SNPs were genotyped in the study. The distributions of such mutations among the examined individuals constituted the basis for a statistical analysis, where several parameters demonstrated correlations with the different genotypes. The SNPs seem to demonstrate some differences with regard to the parameters they influence, depending on which of the three examined genes they reside in. They also appear to have gender-specific influences on several parameters. If studied further, it would be interesting to reveal the combined effects from these SNPs.

Handledare: Josefin Jacobsson
Ämnesgranskare: Helgi Schiöth
Examinator: Elísabeth Andrésdóttir
ISSN: 1650-8319, UPTEC STS08 005

Populärvetenskaplig sammanfattning

I denna rapport undersöks kopplingar mellan å ena sidan genetisk variation i tre gener som kodar för membranbundna proteiner, och å andra sidan fetma hos barn och ungdomar, samt olika parametrar som har kopplingar till fetma och övervikt. DNA och olika biokemiska mätvärden från 420 kraftigt överviktiga ungdomar utgjorde material för en statistisk undersökning av sådana kopplingar, tillsammans med en kontrollgrupp bestående av 512 normalviktiga ungdomar. För att hitta variationer i de tre undersökta generna åtgick ett laborativt arbete till att amplifiera det DNA som kodar för de undersökta proteinerna. Punktmutationer i den genetiska koden söktes sedan, och funna mutationer undersöktes statistiskt för korrelationer med fetmarelaterade variabler. Ett antal signifikanta samband uppmättes, och mutationernas inverkan på variablerna visade sig i flera fall vara könsspecifika och komplexa. Studier av punktmutationernas kombinerade effekter kan vara ett intressant nästa steg efter detta examensarbetes färdigställande.

Note on confidentiality

Due to aspects of secrecy, the names of the genes in focus for this report, as well as the examined single-nucleotide polymorphisms (SNPs), are all coded. For the same reasons, the primer sequences used to amplify the selected DNA-sequences are not enclosed in this report.

Table of contents

Introduction

Obesity is a trait whose sources are numerous and correlating in complex ways. One person's susceptibility to become overweight or obese is determined by environmental as well as heritable factors. At the most basic level, the amount of food eaten versus the amount of physical exercise, together with the individual genetic predispositions towards obesity, decides the amount of body fat in an individual. At an elevated level, factors such as social status and environment, interests, or the living area may be important (Tiwari and Allison, 2003). However, in this report, the genetic factors for childhood obesity are the ones in focus.

There are many ways of performing a study of the genetic implications on childhood obesity. The strategy used in this thesis was to analyse genetic variations in a number of chosen genes in sample populations of obese and non-obese children, and to compare groups with different genotypes with respect to obesity and various related traits. Comparisons were made within the group of obese children, as well as between obese and non-obese children.

This thesis was done within the scope of the group led by Helgi Schiöth, Department of Neuroscience, Unit of Pharmacology, at the Uppsala Biomedical Center (BMC) at Uppsala University. It is a small contribution to the work done on the genetic causes for obesity and on a few of the genes in focus at the department, mainly on the research conducted by Josefin Jacobsson. Being primarily a substudy in a large research project, the possibilities of presenting long-gone conclusions from this work are limited. However, hopefully, this project will contribute with some data at a later stage in the work of the Schiöth research group.

Context and limitations

The framework for this thesis project provided by the Schiöth group, and above all by Jacobsson, needs to be mentioned in order to clarify the conditions and limitations of the study. This framework consists of several components, such as the network of collaborating research laboratories, the amount and quality of data available for analyzing, as well as the time and materials at hand for the experimental and analytical work. Such components are briefly reviewed in the nearest following sections, as well as a formulation of the aim and a discussion of the limits of the project.

The study objects

At the National Childhood Obesity Centre at Karolinska Institutet in Stockholm, obese children are monitored and treated by physicians, psychologists, and nutritionists. For the statistical purposes of this study, 420 obese patients from this centre are represented through DNA samples, which make up the foundation for genotype and sequence analyses. The mean age of the children was 14 years with a standard deviation of ± 3.1 years, and their BMI was distributed as 37.2 ± 6.2 kg/m^2. The gender distribution was approximately 52% girls versus 48% boys. The analyses conducted on this material will be described later on.

In addition to the genetic data made available through the DNA samples, a database was set up in 2006 in order to collect further information on the patients. From this database, named BORIS (BarnObesitasRegister i Sverige), plenty of data can be extracted, and

some of it has been used as analytical data for this thesis project. In June 2007, approximately 1400 children were registered in BORIS (Svensson, 2008). It contains information on age, gender, BMI measurements during the time of treatment, measurements of insulin sensitivity, glucose effectiveness, hormones, blood lipids, and puberty status, among other parameters. These factors will be explained and discussed thoroughly.

The control objects

In order to perform comparative studies on the data from obese children, a control group is required. For this task, adolescents from 17 upper secondary schools in Stockholm were asked to participate in the study and asked to give a blood sample. Each student also filled in a form, providing information on her or his health, ethnicity, and medical drug use (Jacobsson, et al., 2008). In this study, the genetic information from 512 control individuals has been used. The control individuals had a mean age of 17.1 ± 0.8, an average BMI of 21.1 ± 2.6 and a 52% female versus 48% male gender distribution.

Collaborations

Some of the methods crucial for the results derived from the present project have been outsourced to other research centres, due to limitations in the technology available at BMC. The sequencing has been performed at the Genome Center in Riga, Latvia. The genotyping was done at Uppsala Genome Center. Both methods will be described in the *Background* chapter.

Limitations

There are always aspects of, or prerequisites for, a project, which could be improved if it was to be redone. As is usually the case with thesis projects, time is an important limiting factor, setting the basic limits for what may be included in the survey. The material available is another limiting factor of importance. The populations of obese children and controls, 420 and 512 respectively, together constitute a firm basis for the statistical analyses. However, there are cases where statistical significance could not be achieved, but where trends still can be seen in the statistical material. I will try to be clear about the distinction between the two. For example, in the case of rare genetic variations, significances might be hard to attain, but it can still be of interest to acknowledge tendencies in the material.

Aim of study

The aim of this thesis project is to investigate the plausible associations between genetic variations in a number of chosen candidate genes on the one hand, and childhood obesity and related traits on the other. A staging post towards this goal is also to detect known and novel variations in the investigated DNA sequences and to evaluate their value for further analysis.

Theoretical background

In this section, I will briefly review the theoretical motivations for the methods used and the postulations made in the study. Initially, however, I will devote some space to the molecular genetics and biology of the central elements in this study, the chosen candidate genes, and obesity itself.

Single-nucleotide polymorphisms

The most common variant of *polymorphisms* – changes in the nucleotide structure in the DNA – is the variation known as a single-nucleotide polymorphism (SNP). Such variations constitute the most frequent form of genetic variation and are important potential targets for identifying phenotypic traits. A SNP is by definition a variation in a single nucleotide that can be found in more than one percent of the population. No insertions or deletions, but only shifts, are considered to be SNPs (Brookes, 1999). As will be evident later on in this thesis report, SNPs are far more common in non-coding parts of the genome than in coding parts. SNPs in coding regions of the DNA may alter the amino acid translated from the nucleotide sequence, and thus change the encoded protein or its stability. However, SNPs that do not change the encoded amino acid, so-called synonymous SNPs, as well as intronic SNPs, may also change the structure and stability of the encoded protein, which in turn may become the cause of a pathological condition, i.e. a disease They may also change the protein's interaction with a drug or toxin (Kim and Misra, 2007). A SNP in a non-coding region can also affect the splicing of the nucleotide sequence (Chang, et al., 2008).

Perspectives on obesity

Obesity is a major public health problem, primarily in the developed countries, and is also a source of stigmata, affecting relationships between individuals. While in the so-called "traditional" societies high social status was linked to obesity, the correlation has now turned to the opposite. The notions of thinness/health and fatness/illness have become concept pairs with contradictory meanings. In an English study, lower socio-economical status showed to be connected with a higher risk for obesity in both men and women, while lower occupational status was connected to obesity only in women (Wardle, et al., 2002). An association between socio-economic status and the risk of developing obesity has been detected also in a Swedish study with data from 1991 (Rahkonen, et al., 1998). In Sweden, the prevalence of overweight and obesity has shown to increase during the 1980s, and data from Official Statistics of Sweden shows a continuing increase in the amount of overweight and obese during the 1990s and 2000s, until 2005 (Kuskowska-Wolk and Bergstrom, 1993a) (Kuskowska-Wolk and Bergstrom, 1993b) (Statistiska Centralbyrån, 2007).

From another perspective, a growing body of work in endocrinology and neurology is deciphering the complex mechanisms that together form the sensations of hunger and satiety. These researchers are initially concerned with identifying the various organs, substances, and pathways involved in the process of developing obesity. A coarse description would involve the hypothalamus, the primary regulating centre for body weight and food intake in the brain. Appetite and metabolic rate is regulated by a range of different hormones, peptides, and neurotransmitters, acting on hypothalamus in different ways, inhibitory and stimulatory. Signals from the peripheral nerve and blood systems induce sensations of satiety or hunger in response to the amount of energy present in the body. Feeding stimulatory signals are produced in the brain and gut, in the form of hormones, neuropeptides, and neurotransmitters such as ghrelin and cholescystokinin (CCK), while inhibitory hormones include leptin, insulin, serotonin, and peptide YY (PYY). Multiple pathways make up an intricate pattern of signals affecting body weight homeostasis, and it is thus an impossible task to find one unique genetic body weight-regulating factor (Schwartz, et al., 2000) (Konturek, et al., 2003). In this study, genetic variations in genes involved in some of these pathways are

investigated for associations with obesity and related phenotypes. In the next section, the included parameters will briefly be described.

Association study parameters

The study population was scanned for genotype-phenotype associations, through the information provided from the BORIS database. A short review of the general, metabolic, and developmental parameters included in the association study is provided below.

General parameters

The population data was inspected for associations between genotype and birth weight, birth length, BMI, and the age of onset of obesity. BMI is a common measure of adiposity, and is calculated as the weight in kilograms divided by the square of the height in meters. Conventionally, an individual with a BMI of 25 or higher is considered overweight, while a BMI of 30 or higher corresponds to obesity. The BMI, however, is quite a crude measurement of body mass, especially among children, so in order to obtain a better estimate of this parameter, a variant of the BMI measurement was used, that is adjusted for the average changes related to growth, gender and ageing. The adjusted BMI Standard Deviation Score (BMI SDS) developed by Rolland-Cachera et al. will from here onwards be referred to as BMI SDS (RC). The measurements for this adjusted BMI were based on data taken from French individuals, from newborn up to the age of 87, and were gender specific (Rolland-Cachera, et al., 1991). The age of onset of obesity was calculated by the aid of a framework developed by Cole et al., and corresponds to the age when a child's BMI exceeds a level corresponding to a BMI of 30 kg/m^2 in an adult. These levels are also gender specific and are based on a large cohort of individuals (Cole, et al., 2000). The mean BMI changes during growth are depicted in *Figure 1* below.

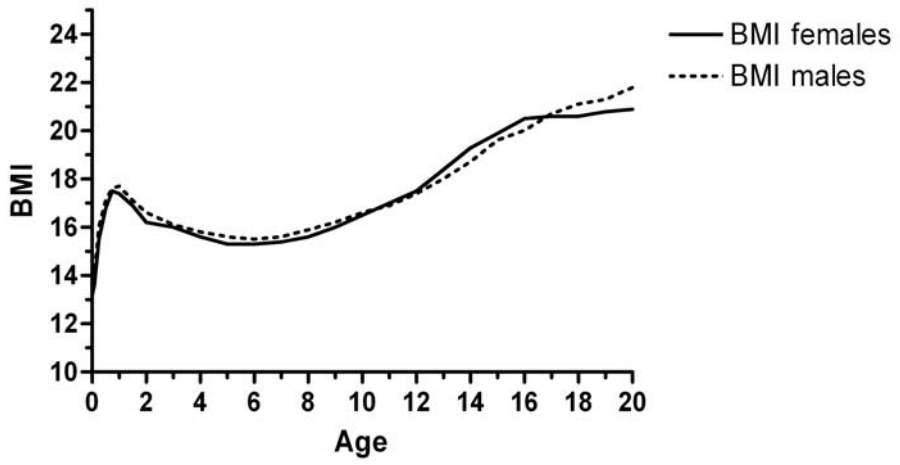

Figure 1. The changes in BMI during the ages in focus for this study. Values are derived from Rolland-Cachera, Cole et al. 1991.

Cholesterol and triglycerides

Cholesterol is an important component in cell membranes and does also have a vital function in the synthesis of steroid hormones. Triglycerides are responsible for storing energy as fat which can be taken up by the muscles during exercise. Both lipids are carried by the aid of apoproteins, which have structural functions and act as ligands. The whole compound is what constitutes a lipoprotein and is categorized by its size and the structural apoprotein that is determining the amount of cholesterol or triglyceride incorporated into it. The different variants of lipoproteins have different functions and are products in a chain of metabolic events. Initially, the so-called very low-density lipoproteins, with a 80% triglyceride content, are synthesized in the liver and thereafter secreted into the blood plasma. The concentrations of these compounds are measured in the blood serum of the patients (hence the "S" in the abbreviated form) and referred to as S-VLDL triglycerides in the tables in the appendices. VLDLs are primarily responsible for transport of triglycerides and are eventually transformed into a category lipoproteins known as low-density lipoproteins (LDL), in which triglycerides have been hydrolyzed in an intermediate step. The LDLs are responsible for the carrying of the largest share of serum cholesterol in the blood, around 60%, and are taken up by receptors on the cell surfaces of the liver and other tissues.

The high-density lipoproteins (HDLs), finally, are found in the plasma and carry a small amount of cholesterol. They are responsible for a pathway that is the opposite of the one of LDLs, referred to as "the reverse cholesterol transport system", where superfluous lipids are cleared off VLDLs and other large compounds known as chylomicrons, and cholesterol is removed from tissues (Lodish, et al., 2004, pp. 767-771).

VLDL, LDL, and HDL triglyceride concentrations in the serum (with the prefix "S" in the data sheets) were measured, as well as VLDL, LDL and HDL cholesterol concentrations. The total concentrations of triglycerides and cholesterol were also estimated and analyzed in connection to the genotype data. These biochemical parameters are important as they, naturally, determine the levels of body fat, but also since raised levels of LDL and low levels of HDL have shown to be connected to cardiovascular disease (Heyman and Gunnarsson, 2007).

Glucose and insulin

Glucose is the most important external energy source of the cell, and its concentration in the blood needs to be well calibrated. During exercise or fasting, the glucose usage rises and plasma glucose levels need to match the increased demand. During meals, the situation is the opposite; there is an abundance of plasma glucose which needs to be stored or used properly. Insulin and glucagon are the mediators of the glucose homeostasis, insulin being its inhibitor and glucagon its promoter. Elevated plasma glucose levels stimulate insulin secretion which halts the production of glucose in the liver and its release into the blood.

The individuals in the sample population had their serum insulin measured after a 12 hour overnight period of fasting (fS-insulin), and at the same time measured their plasma glucose levels (P-glucose). Insulin sensitivity, glucose effectiveness, and insulin response were estimated through repeated measurements after an injection of glucose, and were calculated according to the Bergman minimal model approach, which will not

be further elaborated on here (Bergman, 2005) (Bergman, 1989). Insulin sensitivity measures the ability of insulin to clear the plasma from glucose after a glucose injection. Glucose effectiveness measures the capacity by which glucose disappears from the plasma without any aid by increased insulin concentration levels. Insulin response, finally, is the measurement of the insulin secretion in response to elevated plasma glucose levels after a glucose injection (Jacobsson, et al., 2008) (Bergman, et al., 1981). Insulin resistance, an impairment of the cellular response to insulin secretion, is a precursor to, as well as a characteristic of, Type II diabetes and causes a lack of glucose tolerance. This resistance can also be evaluated through the use of the homeostatic model assessment (HOMA), which is calculated from fasting plasma insulin and glucose concentrations. The model provides a coarse but working estimation and is a widely used tool (Wallace, et al., 2004).

Puberty

An estimation of the degree of physical development was done for each participating individual. During the visits to a physicist, this was measured by the so-called Tanner scale, where maturity is assessed on the basis of sex characteristics such as breast development, genitalia, and pubic hair growth. Each category is divided into five stages (Tanner, 1986). For reasons of simplicity, only the categories genitalia and pubic hair growth were used, and were collapsed into one "pubertal development" variable with three stages. Biochemical data was selected to be measured as close in time to the puberty assessment as possible, as this factor has a large impact on other variables.

Candidate genes

Many genes are potential as topics of investigation in studies of obesity. Being a piece of a larger research project, the present report is engaged in the pursuit of SNPs in some of the genes of interest included in the encompassing study. It is based upon the previous and present research on the two large gene super families of solute carriers (SLCs) and G-protein-coupled receptors (GPCRs) conducted at Department of Neuroscience at Uppsala University. These genes encode cell membrane proteins, carrying or regulating the flow of many of the hormones, neurotransmitters, amino acids, and peptides involved in body weight homeostasis. Having a "gatekeeper" function, they are interesting in a study like this since a mutation in the genes encoding such proteins could possibly alter the manner or amount of substance migrating between cells (Gyllensten, et al., 2005).

GPRA

G protein-coupled receptors (GPCRs) constitute a large superfamily of cell surface receptors, which recognize and mediate substances and signals through the cell membrane. GPCRs are common cell membrane proteins, found in low amounts in most tissues in the body, and demonstrate a high variability as well as a high specificity, which makes them popular as drug targets (Bjarnadottir, et al., 2006). They are also important cell activity regulators, activating G proteins associated to the cell membrane or intracellular space, involved in the sensing of taste, smell, and vision, as well as in intracellular signalling due to external stimuli (Yeagle and Albert, 2007). It has been stated that about 50% of all recently produced drugs are targeted at GPCRs (Klabunde and Hessler, 2002). The superfamily is commonly divided into five families, Glutamate, Rhodopsin (which is in turn divided into four subgroups; α, β, γ, and δ, depending on

the type of ligand bound to the receptor), Frizzled/Taste2, and Secretin; the so-called GRAFS classification system (Bjarnadottir, et al., 2006).

The main structural characteristic of the GPCR-group are the seven antiparallel transmembrane (TM) α-helices that form the gate between the inside and outside of the lipid bilayer of the cell. The helices are connected by six loops, three extracellular and three intracellular. An N-terminus is located on the extracellular side and contains sites for glycosylation, and a C-terminus on the cytoplasmic side contains phosphorylation sites (Perez and Karnik, 2005). An early view of the GPCR functionality gave that the receptors functioned as "switches" with only two possible states, on and off, which resulted in ligands being either agonists or antagonists, stimulating or inhibiting the receptor. However, a newer, perhaps more nuanced, opinion is that the GPCRs change their conformation in different ways for different ligands, frequently referred to as "agonist-specific signalling states". Thus, GPCRs seem to demonstrate a broad range of conformational states, depending on the type of agonist or antagonist connected to it (Hoffmann, et al., 2007). The conformation change of the GPCR on the extracellular side of the cell membrane as a result of the contact with a certain ligand is generating a corresponding change on the cytoplasmic side, enabling a G-protein (short for GTP-binding protein) to couple with the receptor. The G-protein, in turn, functions as a intracellular signal, stimulating or inhibiting enzymes in the cytoplasm (Yeagle and Albert, 2007).

The Rhodopsin family is the largest of the families within the GRAFS system, and the receptor in focus here, GPRA, belongs to the Rhodopsin α subfamily. It is classified as an orphan receptor, implying that its function is still unknown. GPRA has a long N-terminus and third intracellular loop, a characteristic shared with one other GPCR of the same subfamily. Expressed Sequence Tag (EST) searches have however shown a higher rate of expression in GPRA than in its homolog, although their areas of expression follow the same pattern. GPRA is mainly expressed in the central nervous system, but is also present in the eye, lungs, and in the sexual organs. The protein also demonstrates a high degree of conservation, having a 95-96% amino acid identity with the corresponding receptor in rodents, as well as high overall sequence similarity in fish and chicken. This implies that the protein achieved an important role early in vertebrate evolution (Unpublished).

GPCRs have proved to be important in connection to obesity. One example of its significance in this context is the quite recent discovery of ghrelin, a small peptide synthesized primarily in the stomach. Its receptor, a GPCR, is found in the hypothalamus and in the pituitary in the brain and is called the growth hormone secretagogue (GHS) receptor. The activation of the GHS receptor sparks the secretion of growth hormone in the cell (Xu, et al., 2004). It has been shown, among others by Wren, that ghrelin increases hunger and appetite (Wren, et al., 2001). Analogously, the introduction of a ghrelin receptor antagonist has shown to reduce food intake in mice (Asakawa, et al., 2003). Similar connections between other GPCRs and potentially obesity related hormones has been reported in several cases (Xu, et al., 2004).

SLCA and SLCB

Proteins encoded by genes belonging to the solute carrier (SLC) super family constitute a diverse set of transporters, responsible for the carrying of a large number of different neurotransmitters, sugars, amino acids, and other compounds across the cell membrane

or intracellular compartment membranes (Hediger, et al., 2004). SLCs are responsible for sending solutes across the synaptic cleft, and for maintaining a certain concentration of compounds in different partitions of the cell or extracellular space. The family in focus for this report, the SLC6 family, consists of genes coding for transporters responsible for the carrying of neurotransmitters, osmolytes, and amino acids such as GABA, norepinephrine, dopamine, serotonin, glycin, taurine, L-proline, creatine, and betaine. The proteins encoded by the SLC6 family typically consist of twelve helices crossing the membrane of the cell or intracellular compartments, such as mitochondria or vesicles. The N- and C-termini are both located on the cytoplasmic side of the cell membrane (Chen, et al., 2004). The energy for the transportation conducted by SLC6 proteins is created from the electro-chemical gradient of sodium, as Na^+ is co-transported along with the solute and thus carries it against its concentration gradient. The SLCA and SLCB genes are currently categorized as orphans, lacking a known function and connection to pathological conditions. They show a similar but slightly diverging pattern of expression in the brain (Hoglund, et al., 2005).

Some of the proteins in the SLC6 family have shown to be involved in different diseases and pathological conditions. Conditions such as depression, parkinsonism, Tourettes syndrome, and certain kinds of mental retardation have been linked to various SLC6-encoded proteins (Hahn and Blakely, 2007). Being responsible for the homeostasis of various solutes in the cell and in the body, the proteins encoded by the SLC6 genes are plausible as proteins causing obesity-related phenotypes. Logically, obesity has already been connected to this family in several studies. Suviolahti et al have studied SLC6A14 in connection with obesity and found that several SNPs seem to dictate people's predisposition to become obese to some extent, through the regulation of receptor mechanisms involved in the synthesis of serotonin, a neurotransmitter thought to control appetite and body weight (Suviolahti, et al., 2003). This gender specific, and apparently ethnically specific, influence on obesity by the SLC6A14 gene has been examined by Tiwari and Allison (Tiwari and Allison, 2003).

Techniques

A number of different methods have been employed in order to produce the results of this thesis project. I will briefly review them in the following sections.

PCR

As useful as it is simple, the Polymerase Chain Reaction (PCR) method is an effective technique for amplifying fragments of genetic code for later use in genotyping or sequencing. The principles of the method are quite well known. A purified DNA solution is repeatedly heated and cooled down and thus, the DNA double helix is separated and realigned again. To the solution is added a polymerase and an abundance of deoxynucleoside triphosphates (dNTPs) A, T, G, and C. Added are also short fragments of single-stranded DNA, *primers*, which attach to the separated DNA as the temperature drops, faster than the corresponding helix due to their short length and high abundance. As temperature reaches a level of around 55° C, the polymerase elongates the attached primers along the helix, thus creating a corresponding strand. Next round of heating and cooling generates a DNA segment of the desired length, and each time the cycle is repeated, the amount of segments duplicates (Lodish, et al., 2004, p. 375f).

Gel electrophoresis

After the DNA-segments have been duplicated during the PCR-amplification, the samples are run through the microscopic pores of a gel, for the purpose of determining the existence and the length of the desired sequences. A heated solution containing the polysaccharide agarose and ethidium bromide, a fluorescent dye, is poured into a tray and is solidified as it cools off. The samples are loaded in wells on one side of the gel and an electric current is attached. The amplified negatively charged DNA-segments migrate towards the positive end of the gel, at a speed inversely proportional to the length of the segments. Under an ultraviolet lamp, the amplified DNA appears as bands on the gel. If the primers have annealed correctly, one unique segment should have been produced during the PCR-amplification, and one band representing the desired sequence should appear on the gel, as in *Figure 2* (Lodish, et al., 2004, p. 371f).

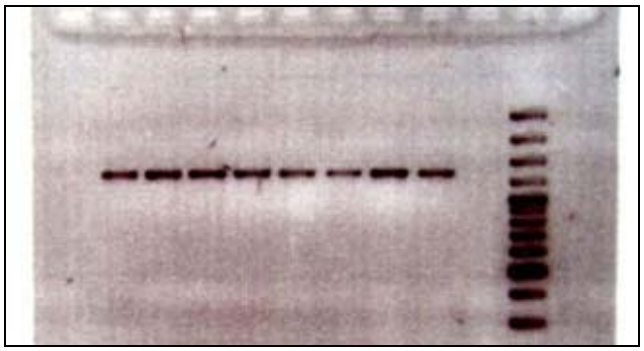

Figure 2. A gel with some amplified samples. The "ladder" to the right is used as a reference for determining the length of the amplified segments.

Sequencing

The sequencing for this project was conducted at the Genome Center in Riga. As the term implies, the technique is employed in order to produce information on the nucleotide sequence. The most common sequencing method, and the one used in this study, is the Sanger method. In short, it involves adding the four nucleotide bases A, T, G, and C, which constitutes the fundaments of the DNA-sequence, to a solution with the purified single-strand DNA-sample that is to be studied. To this solution, a low concentration of dideoxynucleoside triphosphates (ddNTPs), bases lacking its 3'-hydroxyl end, is added together with a DNA polymerase. These bases make further polymerization of the DNA-strand impossible, and terminate the elongation. One polymerization is conducted with ddNTPs corresponding one of the four bases at the time, resulting in a range of randomly terminated strands of different lengths. By marking each ddNTP with a base-specific tag, its identity can be detected. Strands are sorted by length and their ending ddNTP is identified; thus, the sequence of the DNA segment is revealed (Kim, et al., 2002) (Lodish, et al., 2004, p. 372f).

Genotyping

A substantial amount of samples is needed to statistically establish the presence of SNPs, especially rare ones, and to obtain a somewhat clear view of the proportions of different genotypes. A suitable method for the genotyping of the 900 samples of cases and controls included in this study is the so-called "TaqMan approach", a quite fast and labour non-intensive genotyping technique. One probe is designed for each allele, and is

made to be attached right above the area where the SNP of interest is located. On the 5'-end of the probe is put a fluorescent reporter, and on its 3'-end, a quencher is attached, inhibiting the reporter's fluorescence and blocking elongation of the primer along the complementary DNA strand. As the forward primer extends during polymerization, the reporter and quencher are cleaved off by the polymerase and the reporter turns fluorescent. The fluorescence emitted is measured by a detector, and alleles are carrying probes with reporters of different wavelengths depending on the character of the SNP (McGuigan and Ralston, 2002). Based on this information, the samples are clustered according to genotype, and no other post-PCR processing is needed (Livak, et al., 1995). *Figure 3* shows a schematic representation of TaqMan genotyping.

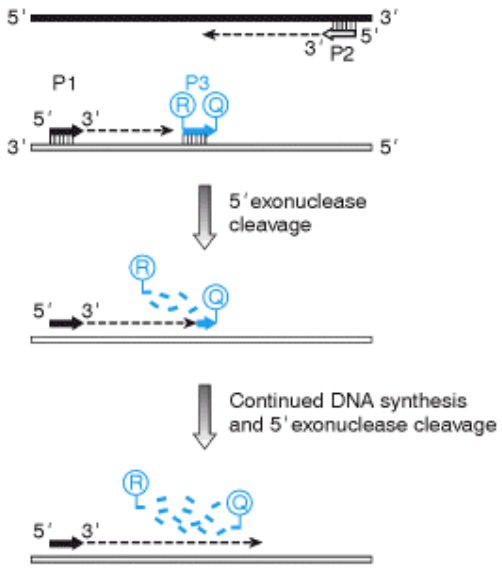

Figure 3. TaqMan genotyping. The reporter is marked "R" in the images, and the quencher with a "Q" (Strachan and Read, 1999).

Study design

Historically, two main approaches have been the most popular ones in order to establish the relationship between genotype and phenotype in populations. One approach utilizes prior knowledge on genes in order to present candidates to be examined more closely in connection to a certain phenotype. These genes may then be searched for genetic variation of some sort, for example SNPs, which might explain their putative involvement in a disease or trait. This strategy is commonly referred to as the *candidate gene approach* (Giallourakis, et al., 2005). In an *association study*, a population divided into cases, affected subjects, and controls, unaffected subjects, may then be used to investigate statistically whether the genotype distribution seems to have any connection to the phenotype in question (Long and Langley, 1999).

Another common strategy is *linkage studies*, where the recombination occurring during meiosis is examined in order to make statements about the genetics and the genetic expression in populations. As maternal and paternal genetic information is combined, certain traits may or may not be passed on from either parent to the offspring. When recombination occurs, DNA segments in close proximity to each other are likely to be inherited together, and such stretches might contain several loci of polymorphisms.

Thus, one can search for easy-targeted markers in such blocks, in order to find links between a trait and a genotype. Such markers might consist of SNPs or other type of polymorphisms, and segments containing multiple markers that are inherited together are usually referred to as *haplotypes*. By investigating the probabilities by which markers are linked to a disease, it is possible to narrow down the area of investigation to a small region. While the candidate gene approach is built on hypotheses of genes that are causal of a certain phenotype, the linkage studies are often done without a preceding hypothesis and may be conducted as whole-genome scans (Lodish, et al., 2004, pp. 394-400). In this thesis, the candidate gene approach has been the prevailing strategy.

Materials and methods

In this section, I will present the work done during this thesis project in a chronological fashion and on a less elevated level.

Collection of samples

DNA samples were extracted and purified from the patients' blood samples at the Faculty of Medicine at Lund University. The information on the respective children available in the BORIS database was evaluated, and 48 samples were chosen for the sequencing analysis. When selecting the samples for the sequence analysis, attention was given to the severity of the patients' obesity, the most obese were included in the sequencing selection. The rationale of this decision was that if the patients' obesity is affected by the SNP under examination, the most obese fraction of the children should demonstrate an overrepresentation of the mutation. As one aim of the sequence analysis was to find candidate SNPs, this was deemed an important criterion in the selection of the cohort. Another decisive factor was to ensure that the patients were healthy apart from their obesity, that no other known pathological conditions, such as Type II diabetes, were causing or interacting with the obesity or even constituted the phenotype actually correlating with the point mutation.

For the genotyping analysis, samples from a total of 420 obese individuals and 512 control individuals were used as the test population. The genotyped cohort constitutes the basis for the primary statistical analysis in this thesis.

Primer design and PCR

Primers were designed for the exons of the candidate genes, SLCA, SLCB, and GPRA. The genes were looked up in Entrez Gene (NCBIa, 2008) and the protein sequences of the longer isoforms were chosen, in order to obtain only the coding exons in the longer transcripts. In FASTA format, the sequences were inserted into BLAT (http://genome.brc.mcw.edu/cgi-bin/hgBlat), and the exon and intron areas of the nucleotide sequences were thus identified. Segments were chosen that had a few hundred nucleotide bases outside the exon on each side, in order to obtain some space where primers could be attached. SNPs located in the areas in close proximity to the exons may affect the splicing of the synthesized protein, and also from this point of view, it was beneficial to include the areas close to the exons (Chang, et al., 2008). The chosen nucleotide sequence for each primer were then inserted into the Primer Premier 5 software, and a pair of fitting primers were selected on the basis of various parameters (Singh, et al., 1998). These parameters included the optimal polymerization

temperatures of the primers, their propensity to anneal to themselves or each other and hence form hairpins or dimers, and the risk for false priming to occur.

Primers were ordered from Thermo Scientific and arrived in dehydrated form. The primer samples were diluted to 100 pmol/µl with autoclaved water and stored in the freezer. A master mix solution was prepared, consisting of autoclaved water, buffer, 0,2 mM dNTPs, 1U/50µl Taq polymerase, and a concentration of $MgCl_2$ ranging from 1,0 to 2,0 mM. The $MgCl_2$ increases the affinity of the nucleotides to attach to each other, and is a variable factor which needs to be tested for each individual primer pair. For each sample that was to be amplified, a 30 µl solution was put together with 24 µl master mix, 3 µl primer mix with a concentration of 2,5 pmol/µl, and 3 µl DNA sample solution.

Initially, the amplifying program was tested in a series of optimization runs, where a combination of different annealing temperatures, lengths of the polymerization step, and number of PCR-cycles were evaluated in a small selection of DNA samples. The amplified samples were detected in a gel electrophoresis and, if the optimization had been successful, the rest of the samples were amplified with the optimized settings. A total of 48 sequence samples were amplified for each primer pair. When all samples had been amplified for one specific sequence, plates were put on heating blocks at 70°C and dehydrated.

Sequencing

There are two main options to choose between when designing primers for sequencing. One approach is to place the primers at the centre of the segment which is to be sequenced and progress outwards, towards the respective ends of the sequence. The sequencing in this project was however not conducted in this manner, but by using primers placed close to the PCR amplification primers with their 3'-end towards the centre of the exon, one primer on either DNA-strand. Thus, the same sequence is read twice, on both complementary strands, and a form of proof-reading is achieved (*Figure 4*).

Figure 4. Placement of primers for Sanger sequencing in this study. Thick grey lines indicate DNA-strands, thin black lines with arrows represent sequencing primers.

Most sequencing primers were designed at the Genome Center, Riga. The sequencing was performed with a solution containing a sequencing buffer with 80 mM Tris, pH 9,0 and 2 mM $MgCl_2$. Included was also 1 mM sequencing primer, 1 ml Big Dye v. 3.1 terminator mix (Applied Biosystems, Foster City, USA), 1 ml purified PCR product, and water to a final volume of 10 ml. This mix was subjected to a sequencing program with an initial denaturation phase at 95° C for 1 minute, followed by 30 repetitions of 95° C denaturation for 10 seconds, 50° C annealing for 15 seconds, and a 60° C extension phase for 4 minutes. Subsequently, the products from the sequencing were purified with G-50 sephadex (Sigma), and were evaporated and dissolved in formamide.

An ABI Prism 3741 (Applied Biosystems, Foster City, USA) machine was finally used for the analyzing of the products.

Sequence data analysis

When sequencing was finished, the results were analysed in the Staden Package software (Staden, et al., 2000). Initially, all sequences were aligned in the Staden Pregap4-software, and subsequently collected in a database created in Gap4. All sequences were first analyzed automatically by the Gap4-software, and highlighted areas in the genetic code were manually inspected for SNPs.

The detected SNPs were located using dbSNP BLAST (NCBIc, 2008), and their respective identification codes, the so-called rs-numbers, were identified. For known SNPs, the positions of the mutation, in the gene and in the codon, respectively, were found using the GeneView function on the dbSNP-site (NCBIc, 2008). These SNPs were also searched for on the HapMap-site, a database where common patterns of human genetic variation are mapped through the genotyping of individuals of diverse ethnicities. The genotype frequencies for a CEU-population (a population of Utah residents with ancestry from northern and western Europe) were registered in the cases where they had been measured for that particular SNP (International HapMap Project Consortium, 2003). In the case of novel SNPs, a nomenclature was used where the SNP was described by the distance, measured by the number of nucleotide bases, from the beginning of the first exon. In the case of intronic SNPs, a "+" was added, and after that, the distance from the end of the nearest exon. The prefix "C" for "coding" was also added, indicating that only sequences with exons coding for a protein were included. Further, a suffix was used, where the wild-type nucleotide, an arrow ">", and the mutated nucleotide followed the nucleotide sequence position (e.g. C.300+100A>G indicates a polymorphic locus in an intron, 100 bases from the end of the preceding exon with a ending position at base 300, and where the wild-type A has changed into a G). For novel SNPs, these locations were identified using BLAT and the DNASTAR EditSeq-software (DNASTAR, 2007).

The SNPs located in exons, which were found to change the amino acid translated from the genetic code, were examined more closely in order to obtain clues concerning their potential role. The protein sequence was inserted into the transmembrane topology predictor Phobius (Käll, 2008) so as to see its predicted location in the protein and thus be able to speculate on its function (Kall, et al., 2004).

Genotyping

A selection of SNPs was subjected for genotyping, performed at the Uppsala Genome Center at the Rudbeck Laboratory in Uppsala. The so-called TaqMan approach, described earlier, was used to genotype the 420 case and the 512 control samples. A total of 11 SNPs were genotyped during the course of this project. The resulting genotype data was initially analysed with the SDS software (*Figure 5* below contains a screenshot from SDS), which is described in more detail below, in order to separate the samples into three genotype clusters.

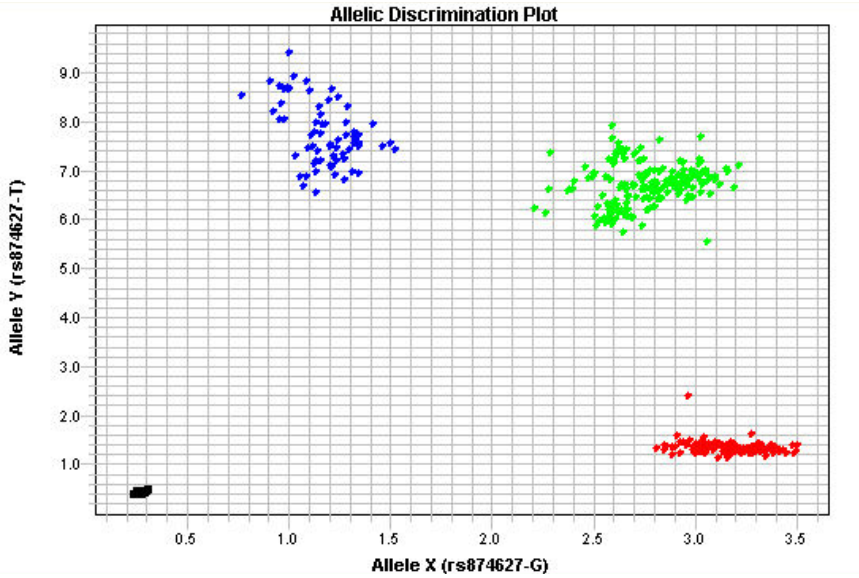

Figure 5. A view of the clustering of genotype data performed in the SDS software.

Genotype data analysis

The genotype data was initially analyzed using the Applied Biosystems SDS 2.2 software, a program used to partition the raw data into subpopulations for the statistical analysis (Applied Biosystems, 2003). The association analysis for this project was done by the aid of various programs. Initially, Pearson's chi 2-test (2 d.f) was applied to test for deviation from Hardy–Weinberg equilibrium (HWE) in both the case and control group, using R, a software for statistical computing. HWE states that the genotype frequencies in a population should remain constant over generations (R Development Core Team 2007). The HWE is in principle valid only during ideal conditions, such as an infinitely large population, and with no individuals leaving or entering the population. However, a population in approximate HWE can be considered quite uniform, and an HWE analysis is therefore performed on the genotype data in this study. Only the genotype distributions adhering to the HWE within a confidence interval of 95% were analyzed further. For the purpose of conducting the HWE-analysis, the supplementary R-packages Adegenet and Genetics were employed, providing algorithms specifically for calculating HWE (Jombart, 2007) (Warnes, et al., 2007). Genotype and allele frequencies among case and control subjects were determined, and odds ratio (OR) for the two groups was calculated in order to investigate the data for association with obesity. The genotype and allele frequencies, odds ratio and corresponding p-values as well as the p values for the HWE test are given in *Appendix 1*.

Statistical analysis of association between genotype and phenotype was done using the MINITAB software, another program for statistical calculations (Minitab Inc. 2005). All biochemical parameters were initially inspected for outliers that could disturb the results of the analysis. An outlier calculator created by GraphPad was used, employing the so-called Grubb's test with a significance level of 0.05 (GraphPad Software, 2002-2005). The parameters were then normalized in MINITAB using Johnson Transformation, a function which finds and applies the optimal transformation equation for the chosen data from three families of distributions. A p-value of 0,05 was accepted

as the threshold value for the normal distribution of the data. If no transformation could be done the data was instead ranked, with the downside that no multivariate analysis could be performed on that parameter. For such parameters, the Kruskal-Wallis test was performed.

The genotype data was divided into three groups if possible, but in cases where heterozygotes and/or homozygotes for the mutated allele were rare, the three groups were collapsed; the wild-type homozygotes in one group and the heterozygotes and mutated allele homozygotes in the other group. An individual evaluation for each genotype data set was done to decide whether a "three group analysis" should be performed. The data for each parameter that had not undergone Johnson Transformation was used to calculate the mean and standard deviation values for each parameter and genotype. This information was used for the purpose of identifying the trends inherent in the parameter values, on the basis of genotype.

The normally distributed data was subjected to a one-way analysis of variance (ANOVA) where the genotype of each individual was used as factor and the association study parameters were used as responses, with the default confidence level of 95%. This analysis was performed for all included individuals as well as for boys and girls separated. If a significant trend was found for any parameter, a multivariate analysis was carried out using a General Linear Model in MINITAB. Interesting parameters were analyzed, using BMI, age, gender, and "pubertal development" as covariates, if they had a significant effect on the response variable themselves. In some cases, where a trend was observable in the data but the association was not significant, a multivariate analysis was performed in order to explore the plausible effects of covariates. Additional analyses were performed in some cases in order to classify the type of influence the polymorphisms have on the phenotype parameters. The genotype categories could be grouped by combining wild-type homozygotes with heterozygotes or mutation homozygotes with heterozygotes in order to see if the polymorphism effect was recessive, dominant, or additive. Another, less explored, possible way of influence could be such, that the heterozygote is significantly deviating from the two homozygote groups, a phenomenon referred to in literature as *heterosis* (Hochholdinger and Hoecker, 2007). The different modes of influence from polymorphisms are depicted in *Figure 6* on the next page.

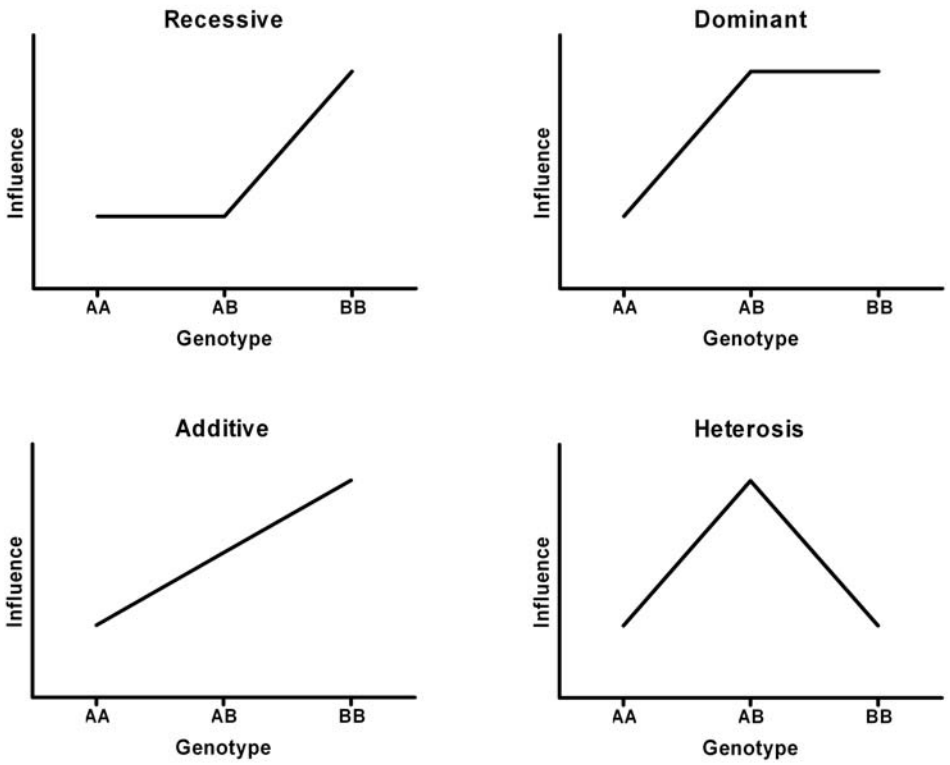

Figure 6. The four possible ways of influence from three genotypes. Genotype AA in the graphs corresponds to the wild-type allele; AB corresponds to the heterozygote and BB to the homozygote for the mutation.

Results

The first section in this chapter will primarily be concerned with presenting and describing the SNPs located in the sequence data. The second section will deal with the results from the genotyped individuals. Here, the most interesting findings from the analysis of the phenotype data in correlation with the genotype data will be presented. Additionally, some known SNPs that have been regarded as interesting to study have been chosen from databases, and this genotyping is thus not preceded by any sequencing.

Sequencing data results

Most of the sequences in the three candidate genes have been amplified and sequenced during the course of this thesis. Below, the sequenced exons in each gene are presented in *Table 1*.

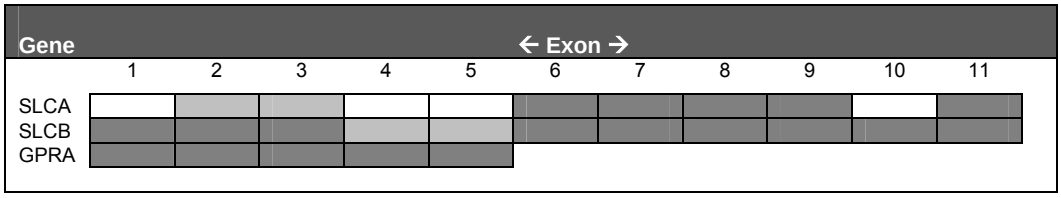

Table 1. The sequencing performed for each gene. Dark grey squares indicate that sequencing has been done on DNA that has been amplified during this project, light grey squares indicate that amplification was done earlier, and white squares indicate that amplification has not yet been performed on that exon.

The sequencing has resulted in the identification of 18 SNPs, whereof six are novel polymorphisms and 12 are known. Five of the SNPs are located in coding exons, all of these in SLCB. Additionally, four more SNPs are located in untranslated regions (UTRs), the last nine SNPs are intronic. The full results from the sequencing are presented in *Table 2* below.

Some of The SNPs in the coding exons are reported to change the amino acid sequence during the translation phase. It might thus be interesting to reveal their location in the encoded protein. Therefore, their codon positions were identified by the use of the Ensembl web site (EMBL-EBI, 2008) and the EditSeq software, and their positions in relation to the cell membrane was predicted using the Phobius web tool, and are included in *Table 2*.

SNP	Gene	Location	Alleles	Codon position	Function	Membrane position
C.11158+159T>A	SLCA	Intron	T>A			
SLCA420	SLCA	Intron	A>G			
C.5559+32T>C	SLCA	Intron	T>C			
SLCB364	SLCB	Exon 1	G>A	3	Synonymous	Cytoplasm, ligand
SLCB375	SLCB	Exon 1	T>C	3	Synonymous	Cytoplasm, ligand
SLCB742	SLCB	Exon 1	G>A	1	Missense	Cytoplasm, ligand
C.25738G>A	SLCB	Exon 7	G>A	3	Synonymous	Transmembrane
SLCB692	SLCB	Intron	T>A			
C.27840+97T>C	SLCB	Intron	T>C			
SLCB140	SLCB	Exon 9	A>G	3	Synonymous	Transmembrane
C.27840+627C>G	SLCB	Intron	G>C			
SLCB581	SLCB	Intron	C>A			
SLCB175	SLCB	3' UTR	G>A			
SLCB446	SLCB	3' UTR	C>G			
SLCB445	SLCB	Intron	G>A			
C.30669+445G>A	SLCB	Intron	G>A			
GPRA427	GPRA	5' UTR	A>G			
GPRA080	GPRA	5' UTR	A>G			

Table 2. The SNPs found in the sequence data. Only one of these polymorphisms causes a change in the encoded amino acid sequence. UTR *is short for "untranslated region",* 3' *implies a location downstreams, on the 3'-side, of the gene. Analogously,* 5' *refers to a location upstreams of the gene.*

For four SNPs, the genotype distribution for a CEU population was available on the HapMap web site (The International HapMap Consortium, 2008). A presentation of this data together with the genotype frequencies from the sequencing is provided in *Appendix 2*. Interestingly, none of the four distributions from the sequencing are similar to the HapMap genotype distributions. Thus, one might speculate on whether any of these differentiating results are mirroring the characteristics of the population chosen for sequencing, i.e. that the obesity phenotype could be connected to the genotype distribution. This data can perhaps serve as a source of material to point to further studies.

Genotyping data results

Eleven SNPs were genotyped in the three genes, whereof three were found to reside in SLCA, four in SLCB, and four in GPRA. As mentioned earlier, four SNPs were found using the sequencing data, the remaining SNPs were chosen from internet databases on the basis of their being potentially interesting as objects of study. Of the 11 SNPs, four are located in coding exons, four in UTRs, and three in introns. All genotyped SNPs are listed in *Table 3* below.

SNP	Gene	Location	Alleles	Codon position	Function	Membrane position	Amino acid change
SLCA654	SLCA	Promotor	T>C				
SLCA296	SLCA	Promotor	T>C				
SLCA577	SLCA	Intron	T>A				
SLCB140	SLCB	Exon 9	A>G	3	Synonymous	Transmembrane	
SLCB009	SLCB	Exon 12	G>A	3	Synonymous	Cytoplasm (C-termini)	
SLCB364	SLCB	Exon 1	C>T	3	Synonymous	Cytoplasm (N-termini)	
SLCB742	SLCB	Exon 1	G>A	1	Missense	Cytoplasm (N-termini)	Ala[A]>Thr[T]
GPRA627	GPRA	Intron	A>C				
GPRA261	GPRA	5' UTR	T>C				
GPRA081	GPRA	Intron	A>C				
GPRA427	GPRA	Exon 5	A>G	1	Missense	Cytoplasm (C-termini)	ARG[R]>Gly[G]

Table 3. Basic information on the genotyped SNPs. The rows marked light grey indicate SNPs that have also been identified through sequencing during this project. The text 5' UTR *indicates that the SNP is located "upstreams", on the 5' side, of the gene.*

Appendix 1 contains information on the genotype and allele distributions among the case and control populations, together with additional information on the odds ratios and p-values as well as HWE analyses. Two of the analyzed SNPs caused deviation from HWE, SLCB009 and SLCB140; the former for both case and control groups, and the latter only for the control group. It can be noted that no significant difference in allele or genotype frequencies can be identified between the case and control populations. Thus, no direct connection between the allele or genotype frequency and obesity can be traced from this material.

In the three following sections, a broad outline of the results obtained from the association studies between the genotype and phenotype data will be done, categorized by the genes in which they reside.

Statistical significances in SNP-data in SLCA

The examined variants in SLCA, SLCA654, SLCA296, and SLCA577, gave quite similar patterns of influence on the measured parameters. All SNPs showed some connection with birth length in girls, although in SLCA654, the correlation was not significant on a 95% confidence level. In all cases, wild-type homozygotes were taller at birth than individuals of the other genotypes.

SLCA577 showed significantly elevated plasma glucose levels and a lessened glucose effectiveness for heterozygous girls and girls homozygous for the mutation compared to girls homozygous for the reference allele, but not for boys. A trend towards higher plasma glucose levels could also be recognized for the SNP SLCA654, in girls. However, none of these parameters could be normalized. The variance analysis was thus conducted using the Kruskal-Wallis method with ranked data, by which follows that information of the distances between measurements is lost as the data is treated as ordinal numbers, and no multivariate analysis can be performed.

Information on the performed ANOVA for the three SNPs residing in SLCA is given in *Appendices 3 – 5*.

Statistical significances in SNP-data in SLCB

In SLCB, the genotype material of SLCB009 and SLCB364 contains merely two individuals homozygous for the mutation each, and an allele analysis was thus conducted. Here, the SLCB009 shows significantly elevated levels of serum cholesterol in the whole population and in girls analyzed separately for the heterozygotes and homozygotes for the mutated allele. When stratified for gender the analyze also showed increased HOMA and LDL-cholesterol values in boys. When adjusted for puberty, only the elevated serum cholesterol levels in girls were still significantly higher, with a p-value of 0,029. However, one should take precaution when analyzing the results from SLCB009, as it does not adhere to a HWE (see *Appendix 1* for details).

In the more evenly distributed genotype groups of SLCB140 and SLCB742, it was possible to perform a genotype comparison. SLCB140 was recessive in its negative influence on glucose effectiveness. The SNP also demonstrated a significant association with insulin sensitivity in girls in an univariate analysis, where the heterozygotes deviated from the homozygotes. This significance however disappeared when adjusted for puberty.

The SLCB742 provided significant associations with plasma glucose and HOMA in a univariate analysis, which both remained significant in a multivariate analysis adjusted for puberty. Both parameters demonstrated significant associations for boys only. Neither age nor BMI were correlated to these parameters in separate univariate analyses, and no multivariate analyses with these as covariates were performed. HDL triglycerides and VLDL triglycerides were also associated with this SNP, HDL triglycerides for girls only, and VLDL triglycerides for boys only. These two parameters were not normalized, and thus, no multivariate analysis could be conducted. Interestingly, and perhaps contra-intuitively, in three of the four parameters correlated with SLCB742, the heterozygote group was the one expressing the significant deviation from the other two groups, as is displayed in *Table 4* below. The genotype data thus

does not seem to follow any additional, nor dominant or recessive models, but rather adhere to the distribution of heterosis.

SLCB742 Parameter	GG	AG	AA	P	p adj. for puberty
P-glucose ♂	4.710 ± 0.587	4.819 ± 0.607	4.792 ± 0.498	**0.045**	**0.025**
HOMA ♂	3.375 ± 1.807	3.887 ± 2.302	3.507 ± 1.903	**0.006**	**0.008**
S-HDL triglycerides ♀	0.219 ± 0.084	0.207 ± 0.071	0.226 ± 0.079	**0.018**	
S-VLDL triglycerides ♂	0.812 ± 0.520	0.779 ± 0.450	0.775 ± 0.450	**0.044**	

Table 4. The p-values and the mean values for each genotype, for parameters significantly correlated with SLCB742. Gender specific parameters are marked with "♂" and "♀" signs.

Tables of ANOVA for the SNPs residing in SLCB can be found in *Appendices 6 – 9*.

Statistical significances in SNP data in GPRA

GPRA427 is one of the two genotyped SNPs examined in this thesis that alter the amino acid sequence in the translated protein. However, it showed no association with the phenotype data. The remaining three SNPs in GPRA had a more uniform pattern of associations. They all associated with fasting insulin and HOMA. The GPRA261 and GPRA081 were analyzed under a recessive model due to that homozygotes for the mutation were rare. Hence, the possibilities of analyzing different modes of genotype influence on the phenotype parameters became somewhat limited. What could be seen in a comparison between these SNPs was that while GPRA627 was connected to lowered levels of fasting insulin and HOMA, primarily in girls, GPRA261 and GPRA081 associated with elevated levels of these parameters. Both boys' and girls' fasting insulin and HOMA were significantly or close to significantly associated with these three SNPs. Thus, the effect does not seem to be gender specific.

What may also be noted from these data is that GPRA627 seems to have a recessive influence on elevated BMI for boys. When adjusted for puberty, the p-value for boys was calculated to 0,019, with an adjusted R^2-value of 25,42%. This is interesting, but should be considered with precaution, as no significance was seen for BMI SDS. GPRA081 showed a significant correlation with elevated BMI SDS for girls, but this significance disappeared when adjusted for puberty.

Appendices 10 – 13 contain ANOVA for the SNPs residing in GPRA.

Discussion

In this study, a mapping of some genetic variations in SLCA, SLCB, and GPRA in connection with obesity related phenotypes has been conducted. Among the resulting associations between the SNPs and phenotype parameters, some general trends have emerged, although it is difficult to explain the effects of genetic variations in such a complex context as obesity. Moreover, as the proteins in focus are membrane proteins, they are themselves not the products of, nor the matter that constitutes, obesity, although still a crucial component in complex metabolic pathways. Hopefully, this thesis can provide some advice for further studies in this area. This chapter will draw on the findings in the preceding *Results* chapter.

Gender specific SNPs

Although different SNPs naturally have diverging functionalities, some themes can be deduced by glancing through the results. One such theme is the gender specificity of SNPs. The SNPs in the SLCA gene are all demonstrating significances for girls exclusively. When boys are analyzed separately, the significances and trends vanish. In SLCB, significant results are observable in both genders, but a significance in boys is never accompanied by a significance or trend in girls, or vice versa. In GPRA, however, the SNPs are not demonstrating a clear gender specific impact on the phenotype parameters. In some cases, the SNPs seem to influence only one gender, but most often, significances or trends are observable for all individuals.

Trends in functional specificity of SNPs

Another theme, which is of great interest for further research in the search for gene function, is which type of phenotype the SNPs are affecting. Some speculations can be done on behalf of the present statistical material.

The three SNPs residing in SLCA do not affect a great many phenotype parameters. However, SLCA577 may affect the permeability of the cell surface for glucose to some extent, as glucose effectiveness is lessened and plasma glucose levels are elevated in individuals heterozygous or homozygous for the mutation. In SLCB, many parameters are influenced by the examined SNPs, mainly biochemical variables included in the homeostatic model assessment, HOMA. Various glucose and insulin parameters are affected in different ways, alternating between boys and girls. Another set or parameters to which the SNPs in SLCB seem to be relevant, are the various forms of triglycerides and cholesterols. In addition, one SNP, SLCB742, is affecting multiple parameters in quite complex ways. The consequences of SNPs in the GPRA gene, finally, are clearly associated with fasting insulin and HOMA. These significances are strong and hold for multivariate adjustments.

SNPs causing amino acid changes

The SNP SLCB742 is intriguing as it transforms the amino acid from the neutral nonpolar alanine into the, still neutral, but polar, threonine. Alanine also has a hydropathy index of 1.8, while threonine has an index of -0.7. This index is a measurement of the hydrophilic or hydrophobic properties of an amino acid, a positive number indicating a hydrophobic amino acid and a negative number indicating a hydrophilic one (Kyte and Doolittle, 1982). Thus, the SNP changes the properties of the translated amino acid quite drastically. According to the Phobius transmembrane topology predictor, the alanine/threonine amino acid resides in the cytoplasmic N-terminal of the protein, which indicates that it could have consequences for the binding of ligands to the membrane protein. SLCB742 affects a broad range of parameters, as seen in the *Results* chapter.

Conversely, the other SNP studied here which causes an amino acid change, GPRA427, does not correlate with any parameter changes at all. This could imply that its area of influence is not among the parameters included in this study. It could also be an example of how many factors other than the ability to alter the encoded protein are determining the impact of one SNP. Lastly, it could be a piece of data supporting the

notion that a SNP would disappear due to evolutionary pressure, would its effects be too significant. All explanations probably hold some truth, I believe.

Heterosis

Besides the apparent gender distinction seen in the influence of SNPs on the phenotype parameters, the significant distinction of the heterozygotes for various phenotypes and SNPs is an interesting finding. The phenomenon of heterosis is an old concept that has experienced a renaissance in recent years, and that has gained some attention in association studies, but its causes are still largely unravelled and are being disputed. Interactions between superior dominant alleles of two different parental lines at several loci, or epistatic; e.g. non-allelic, interactions, have served as explanations for heterosis (Melchinger, et al., 2007) (Hochholdinger and Hoecker, 2007). While such ponderings lay beyond the scope of this thesis, it is evident that SNPs may have this kind of influence on phenotype parameters. Here, primarily SLCB742 demonstrates a complex heterosis-ish behaviour. For example, it seems to be connected to a lower amount of HDL-compounds in heterozygous girls, as both triglycerides and cholesterol in this form are of a lower concentration in these individuals. As mentioned in the *Theoretical background* chapter, low levels of HDL has shown a connection with cardiovascular disease, and trends like these may contribute to such pathological states.

Prospective perspectives

This thesis has hopefully provided some information on the doings of a number of SNPs, as well as some clues towards the functionalities of the examined genes. However, it is merely a small contribution to the research that leads up to the determination of gene function. Also, the project has perhaps given some indications of genetic influences on obesity-related phenotypes, although this phenomenon is obviously too complex to be captured in any work of this size, if any.

Some future tasks are quite easy to identify. The influences of SNPs on the phenotypes included in this study may be modest when considered individually, but an examination of their combined impact may reveal interesting effects. Thus, a haplotype analysis or linkage study could be appropriate as the next step.

There are also other phenotypes that could be considered in connection to the SNPs in this thesis. Together with the general and biochemical variables analyzed here, statistics for various sex hormones, blood lipids, enzymes, and other parameters have been generated. They are beyond the scope of this thesis, but are definitely interesting parameters for future studies. Naturally, there are also other plausible obesity-related genes that are interesting objects of study.

The “themes” proposed in the *Discussion* chapter of this report propose some possible routes which to direct further research. Gender specificity and the phenomenon of heterosis in some SNPs may be coincidental, but may also provide alternative ways of analyzing genetic variation.

Finally, there is much to be done cross-disciplinary in the field of obesity. A broadened set of parameters, such as area of residence, measurements of social class etcetera, may show to correlate more drastically with obesity and related traits in combination with genetic parameters than a, in this sense, randomly selected population does.

Acknowledgements

Many thanks to Josefin Jacobsson, my supervisor, who have had the patience to explain aspects of the work (in some cases over and over again), always being positive and supportive. Good luck with your house! Thanks also to Anna, Markus and Chris for many a good advice and many a cup of coffee. Thanks, finally, to Helgi Schiöth for accepting me for this thesis period.

Bibliography

Journal articles and books

Asakawa, A., Inui, A., Kaga, T., Katsuura, G., Fujimiya, M., Fujino, M.A. and Kasuga, M., (2003). 'Antagonism of ghrelin receptor reduces food intake and body weight gain in mice'. *Gut*, 52 (7):947-952.

Bergman, R.N., (1989). 'Lilly lecture 1989. Toward physiological understanding of glucose tolerance. Minimal-model approach'. *Diabetes*, 38 (12):1512-1527.

Bergman, R.N., (2005). 'Minimal model: perspective from 2005'. *Horm Res*, 64 Suppl 3:8-15.

Bergman, R.N., Phillips, L.S. and Cobelli, C., (1981). 'Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose'. *J Clin Invest*, 68 (6):1456-1467.

Bjarnadottir, T.K., Gloriam, D.E., Hellstrand, S.H., Kristiansson, H., Fredriksson, R. and Schioth, H.B., (2006). 'Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse'. *Genomics*, 88 (3):263-273.

Brookes, A.J., (1999). 'The essence of SNPs'. *Gene*, 234 (2):177-186.

Chang, W.C., Chen, H.H. and Tarn, W.Y., (2008). 'Alternative splicing of U12-type introns'. *Front Biosci*, 13:1681-1690.

Chen, N.H., Reith, M.E. and Quick, M.W., (2004). 'Synaptic uptake and beyond: the sodium- and chloride-dependent neurotransmitter transporter family SLC6'. *Pflugers Arch*, 447 (5):519-531.

Cole, T.J., Bellizzi, M.C., Flegal, K.M. and Dietz, W.H., (2000). 'Establishing a standard definition for child overweight and obesity worldwide: international survey'. *Bmj*, 320 (7244):1240-1243.

Consortium, T.I.H., (2003). 'The International HapMap Project'. *Nature*, 426 (6968):789-796.

Giallourakis, C., Henson, C., Reich, M., Xie, X. and Mootha, V.K., (2005). 'Disease gene discovery through integrative genomics'. *Annu Rev Genomics Hum Genet*, 6:381-406.

Gyllensten, U., Marcus, C., Lundeberg, J., Klovins, J., Fredriksson, R. and Schioth, H.B., (2005). 'Research plan - Obesity: The molecular foundation for individual variation in body weight regulation'. *Unpublished*.

Hahn, M.K. and Blakely, R.D., (2007). 'The functional impact of SLC6 transporter genetic variation'. *Annu Rev Pharmacol Toxicol*, 47:401-441.

Hediger, M.A., Romero, M.F., Peng, J.B., Rolfs, A., Takanaga, H. and Bruford, E.A., (2004). 'The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteinsIntroduction'. *Pflugers Arch*, 447 (5):465-468.

Heyman, C. and Gunnarsson, R., (2007). 'Normal S-cholesterol indicates unchanged S-LDL, S-HDL and S-triglycerides in patients with previously acceptable blood lipids'. *Scand J Clin Lab Invest*, 67 (5):498-506.

Hochholdinger, F. and Hoecker, N., (2007). 'Towards the molecular basis of heterosis'. *Trends Plant Sci*, 12 (9):427-432.

Hoffmann, C., Zurn, A., Bunemann, M. and Lohse, M.J., (2007). 'Conformational changes in G-protein-coupled receptors-the quest for functionally selective conformations is open'. *Br J Pharmacol*.

Hoglund, P.J., Adzic, D., Scicluna, S.J., Lindblom, J. and Fredriksson, R., (2005). 'The repertoire of solute carriers of family 6: identification of new human and rodent genes'. *Biochem Biophys Res Commun*, 336 (1):175-189.

Jacobsson, J.A., Danielsson, P., Svensson, V., Klovins, J., Gyllensten, U., Marcus, C., Schioth, H.B. and Fredriksson, R., (2008). 'Major gender difference in association of FTO gene variant among severely obese children with obesity and obesity related phenotypes'. *Biochem Biophys Res Commun*.

Jombart, T., (2007). 'The package adegenet for the R software: genetic data handling for multivariate analysis'. Lyon, France.

Kall, L., Krogh, A. and Sonnhammer, E.L., (2004). 'A combined transmembrane topology and signal peptide prediction method'. *J Mol Biol*, 338 (5):1027-1036.

Kim, S. and Misra, A., (2007). 'SNP genotyping: technologies and biomedical applications'. *Annu Rev Biomed Eng*, 9:289-320.

Kim, Y., Donoff, R.B., Wong, D.T. and Todd, R., (2002). 'The nucleotide: DNA sequencing and its clinical application'. *J Oral Maxillofac Surg*, 60 (8):924-930.

Klabunde, T. and Hessler, G., (2002). 'Drug design strategies for targeting G-protein-coupled receptors'. *Chembiochem*, 3 (10):928-944.

Konturek, S.J., Pepera, J., Zabielski, K., Konturek, P.C., Pawlik, T., Szlachcic, A. and Hahn, E.G., (2003). 'Brain-gut axis in pancreatic secretion and appetite control'. *J Physiol Pharmacol*, 54 (3):293-317.

Kuskowska-Wolk, A. and Bergstrom, R., (1993a). 'Trends in body mass index and prevalence of obesity in Swedish men 1980-89'. *J Epidemiol Community Health*, 47 (2):103-108.

Kuskowska-Wolk, A. and Bergstrom, R., (1993b). 'Trends in body mass index and prevalence of obesity in Swedish women 1980-89'. *J Epidemiol Community Health*, 47 (3):195-199.

Kyte, J. and Doolittle, R.F., (1982). 'A simple method for displaying the hydropathic character of a protein'. *J Mol Biol*, 157 (1):105-132.

Livak, K.J., Marmaro, J. and Todd, J.A., (1995). 'Towards fully automated genome-wide polymorphism screening'. *Nat Genet*, 9 (4):341-342.

Lodish, H., Berk, A., Matsudaira, P., Kaiser, C.A., Krieger, M., Scott, M.P., Zipursky, S.L. and Darnell, J., (2004). *Molecular Cell Biology*: W.H. Freeman and Company.

Long, A.D. and Langley, C.H., (1999). 'The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits'. *Genome Res*, 9 (8):720-731.

McGuigan, F.E. and Ralston, S.H., (2002). 'Single nucleotide polymorphism detection: allelic discrimination using TaqMan'. *Psychiatr Genet*, 12 (3):133-136.

Melchinger, A.E., Utz, H.F., Piepho, H.P., Zeng, Z.B. and Schon, C.C., (2007). 'The role of epistasis in the manifestation of heterosis: a systems-oriented approach'. *Genetics*, 177 (3):1815-1825.

Minitab_Inc., (2005). 'MINITAB Release 14.20'. Minitab Inc.

Perez, D.M. and Karnik, S.S., (2005). 'Multiple signaling states of G-protein-coupled receptors'. *Pharmacol Rev*, 57 (2):147-161.

Rahkonen, O., Lundberg, O., Lahelma, E. and Huuhka, M., (1998). 'Body mass and social class: a comparison of Finland and Sweden in the 1990s'. *J Public Health Policy*, 19 (1):88-105.

Rolland-Cachera, M.F., Cole, T.J., Sempe, M., Tichet, J., Rossignol, C. and Charraud, A., (1991). 'Body Mass Index variations: centiles from birth to 87 years'. *Eur J Clin Nutr*, 45 (1):13-21.

Schwartz, M.W., Woods, S.C., Porte, D., Jr., Seeley, R.J. and Baskin, D.G., (2000). 'Central nervous system control of food intake'. *Nature*, 404 (6778):661-671.

Singh, V.K., Mangalam, A.K., Dwivedi, S. and Naik, S., (1998). 'Primer premier: program for design of degenerate primers from a protein sequence'. *Biotechniques*, 24 (2):318-319.

Staden, R., Beal, K.F. and Bonfield, J.K., (2000). 'The Staden package, 1998'. *Methods Mol Biol*, 132:115-130.

Strachan, T. and Read, A., (1999). *Human Molecular Genetics 2*: BIOS Scientific Publishers Ltd.

Suviolahti, E., Oksanen, L.J., Ohman, M., Cantor, R.M., Ridderstrale, M., Tuomi, T., Kaprio, J., Rissanen, A., Mustajoki, P., Jousilahti, P., Vartiainen, E., Silander, K., Kilpikari, R., Salomaa, V., Groop, L., Kontula, K., Peltonen, L. and Pajukanta, P., (2003). 'The SLC6A14 gene shows evidence of association with obesity'. *J Clin Invest*, 112 (11):1762-1772.

Svensson, V., (2008). 'Severity of Obesity in Children is Related to Parental Weight Status and Early Onset of Obesity'. *Unpublished*.

Tanner, J.M., (1986). 'Normal growth and techniques of growth assessment'. *Clin Endocrinol Metab*, 15 (3):411-451.

Team, R.D.C., (2007). 'R'. Vienna, Austria: R Foundation for Statistical Computing, R: A language and environment for statistical computing.

Tiwari, H.K. and Allison, D.B., (2003). 'Do allelic variants of SLC6A14 predispose to obesity?' *J Clin Invest*, 112 (11):1633-1636.

Wallace, T.M., Levy, J.C. and Matthews, D.R., (2004). 'Use and abuse of HOMA modeling'. *Diabetes Care*, 27 (6):1487-1495.

Wardle, J., Waller, J. and Jarvis, M.J., (2002). 'Sex differences in the association of socioeconomic status with obesity'. *Am J Public Health*, 92 (8):1299-1304.

Warnes, G., Gorjanc, G., Leisch, F. and Man, M., (2007). 'genetics: Population Genetics'.

Wren, A.M., Seal, L.J., Cohen, M.A., Brynes, A.E., Frost, G.S., Murphy, K.G., Dhillo, W.S., Ghatei, M.A. and Bloom, S.R., (2001). 'Ghrelin enhances appetite and increases food intake in humans'. *J Clin Endocrinol Metab*, 86 (12):5992.

Xu, Y.L., Jackson, V.R. and Civelli, O., (2004). 'Orphan G protein-coupled receptors and obesity'. *Eur J Pharmacol*, 500 (1-3):243-253.

Yeagle, P.L. and Albert, A.D., (2007). 'G-protein coupled receptor structure'. *Biochim Biophys Acta*, 1768 (4):808-824.

Internet references

EMBL-EBI, Sanger Institute (2008). *Ensembl Genome Browser*, http://www.ensembl.org (2008-02-03).

GraphPad Software (2002-2005). *Outlier Calculator*, http://www.graphpad.com/quickcalcs/Grubbs1.cfm (2008-02-06).

International HapMap Consortium, the (2002). *The International HapMap Project*, http://www.hapmap.org/ (2008-02-03).

Käll, Lukas (2008). *Phobius*, http://phobius.sbc.su.se/ (2008-01-28).

NCBI (2008a). National Center for Biotechnology Information, *Entrez*, http://www.ncbi.nlm.nih.gov/sites/gquery (2008-02-01).

NCBI (2008b). National Center for Biotechnology Information, *BLAST*, http://www.ncbi.nlm.nih.gov/blast/Blast.cgi (2008-02-02).

NCBI (2008c). National Center for Biotechnology Information, *dbSNP*, http://www.ncbi.nlm.nih.gov/projects/SNP/ (2008-02-03).

Statistiska Centralbyrån (2007). 'Tabell HA 13. Överviktiga eller feta.' http://www.scb.se/statistik/LE/LE0101/2007A02/HA1305.xls (2008-02-01).

UCSC (2008). University of California Santa Cruz, *BLAT*, http://genome.brc.mcw.edu/cgi-bin/hgBlat (2008-01-25).

Software

Applied Biosystems (2003). *SDS 2.2*, http://www.appliedbiosystems.com (2008-02-08).

DNASTAR (2007). *EditSeq*, http://www.dnastar.com/products/editseq.php (2008-02-07).

Appendix 1. Genotype data.

					Case				Controls				All
Gene	SNP	Position	Locus	Genotype	Genotype distribution, n (%)	Allele frequence, %	Sequencing genotype distribution	Genotype distribution, n (%)	Allele frequence, (%)	OR (95% CI)	HWE case / control	p, (95% CI)
SLCA	SLCA654	Promotor	T>C	TT	256 (62)	T C		288 (56)	T C		0.07 / 0.53	0.172
				TC	131 (31)	77 23		187 (37)	75 25	0.792 (0.598-1.048)		
				CC	28 (7)			36 (7)		0.878 (0.521-1.480)		
SLCA	SLCA296	Promotor	T>C	TT	172 (42)	T C		196 (38)	T C		0.16 / 0.41	0.425
				TC	179 (43)	63 37		233 (46)	61 39	1.029 (0.771-1.372)		
				CC	64 (15)			82 (16)		0.897 (0.610-1.321)		
SLCA	SLCA577	Intron	T>A	TT	297 (72)	T A		348 (68)	T A		0.62 / 0.61	0.369
				TA	106 (25)	84 16		150 (29)	83 17	0.829 (0.618-1.111)		
				AA	12(3)			13 (3)		1.096 (0.492-2.442)		
SLCB	SLCB140	Exon 9	A>G	AA	117 (29)	A G	4	123 (25)	A G		0.38 / 0.03	0.039
			Synonymous	AG	192 (48)	53 47	21	218 (45)	48 52	0.926 (0.673-1.274)		
			Codon position 3	GG	94 (23)		23	145 (30)		1.467 (1.021-2.108)		
SLCB	SLCB009	Exon 12	G>A	GG	423 (94)	G A		478 (94)	G A		0.00 / 0.00	0.993
			Synonymous	AG	26 (6)	97 3		30 (6)	94 6	0.979 (0.570-1.683)		
			Codon position 3	AA	2 (0)			2 (0)		0.885 (0.124-6.593)		
SLCB	SLCB364	Exon 1	C>T	CC	408 (94)	C T	44	480 (94)	C T		0.07 / 0.09	0.993
			Codon position 3	CT	25 (6)	97 3	4	29 (6)	97 3	1.014 (0.585-1.760)		
			Synonym	TT	0 (0)		0	2 (0)		4.251 (0.203-88.795)		
SLCB	SLCB742	Exon 1	G>A	GG	150 (33)	G A	15	167 (33)	G A		0.20 / 0.71	0.814
			Codon position 1	AG	208 (46)	56 44	22	245 (48)	57 43	0.945 (0.709-1.260)		
			Missense Ala[A]>Thr[T]	AA	92 (20)		11	96 (19)		0.937 (0.659-1.248)		

Appendix 1. Genotype data, continuation.

					Case				Controls					All
Gene	SNP	Position	Locus	Genotype	Genotype distribution, n (%)	Allele frequence, %		Sequencing genotype distribution	Genotype distribution, n (%)	Allele frequence, %		OR (95% CI)	HWE case / control	p, (95% CI)
GPRA	GPRA627	Intron	A>C	CC	147 (34)	C	A		176 (34)	C	A		0.41 / 0.64	0.672
				AC	205 (47)	57	43		244 (48)	58	42	1.006 (0.755-1.340)		
				AA	84 (19)				92 (18)			0.915 (0.633-1.322)		
GPRA	GPRA261	5' UTR	T>C	TT	300 (70)	T	C		349 (68)	T	C		0.59 / 0.15	0.285
				CT	121 (28)	84	16		144 (28)	82	18	0.978 (0.734-1.302)		
				CC	10 (2)				22 (4)			1.891 (0.882-4.057)		
GPRA	GPRA081	Intron	A>C	AA	251 (58)	A	C		320 (62)	A	C		0.30 / 0.66	0.260
				AC	164 (38)	77	23		169 (33)	79	21	1.237 (0.943-1.622)		
				CC	20 (5)				25 (5)			0.980 (0.532-1.806)		
GPRA	GPRA427	Exon 5	A>G	AA	217 (50)	A	G	22	264 (51)	A	G		0.68 / 0.34	0.493
			Missense	AG	183 (42)	71	29	23	214 (42)	72	28	1.040 (0.797-1.359)		
			Codon position 1 Arg [R] > Gly [G]	GG	35 (8)			3	35 (7)			0.822 (0.498-1.358)		

Appendix 2. Sequencing data and HapMap CEU-distributions.

Gene	SNP	Function	Position	Locus	Genotype	Genotype distribution, HapMap CEU	Distributions, sequencing samples (N=48)	Genotyped
SLCA	C.11158+159T>A	---	Intron	T>A	TT	---	15	No
					AT		21	
					AA		12	
SLCA	SLCA420	---	Intron	A>G	AA	---	44	No
					AG		4	
					GG		0	
SLCA	C.5559+32T>C	---	Intron	T>C	TT	---	45	No
					CT		3	
					CC		0	
SLCB	SLCB364	Synonymous	Exon 1	G>A	GG	---	44	Yes
		Codon position 3			AG		4	
					AA		0	
SLCB	SLCB375	Synonymous	Exon 1	T>C	TT	---	12	No
		Codon position 3			CT		23	
					CC		13	
SLCB	SLCB742	Missense	Exon 1	G>A	GG	---	15	Yes
		Codon position 1			AG		22	
		Ala[A]>Thr[T]			AA		11	
SLCB	C.25738G>A	Synonymous	Exon 7	G>A	GG	---	46	No
		Codon position 3			AG		2	
					AA		0	

Appendix 2. Sequencing data and HapMap CEU-distributions, continuation.

Gene	SNP	Function	Position	Locus	Genotype	Genotype distribution, HapMap CEU	Distributions, sequencing samples (N=48)	Genotyped
SLCB	SLCB692	---	Intron	T>A	TT	0.15	19	No
					AT	0.45	20	
					AA	0.40	9	
SLCB	C.27840+97T>C	---	Intron	T>C	TT	---	47	No
					CT		1	
					TT		0	
SLCB	SLCB140	Synonymous	Exon 9	A>G	AA	0.350	4	Yes
		Codon position 3			AG	0.467	21	
					GG	0.183	23	
SLCB	C.27840+627C>G	---	Intron	G>C	GG	---	47	No
					CG		1	
					CC		0	
SLCB	SLCB581	---	Intron	C>A	CC	---	23	No
					AC		21	
					AA		4	
SLCB	SLCB175	---	Exon 11	G>A	GG	---	38	No
			3' UTR		AG		10	
					AA		0	

Appendix 2. Sequencing data and HapMap CEU-distributions, continuation.

Gene	SNP	Function	Position	Locus	Genotype	Genotype distribution, HapMap CEU	Distributions, sequencing samples (N=48)	Genotyped
SLCB	SLCB446	---	Exon 11	C>G	CC	1.00	45	No
			3' UTR		CG	0.0	3	
					GG	0.0	0	
SLCB	SLCB445	---	Intron	G>A	GG	---	25	No
					AG		17	
					AA		6	
SLCB	30669+445G>A	---	Intron	G>A	GG	---	47	No
					AG		1	
					AA		0	
GPRA	GPRA427	Missense	Exon 5	A>G	AA	0.603	22	Yes
		Codon position 1	5' near gene		AG	0.328	23	
		Arg[R]>Gly[G]			GG	0.069	3	
GPRA	GPRA080	Synonymous	Exon 5	A>G	AA	---	38	No
		Codon position 3	5' near gene		AG		9	
					GG		1	

Appendix 3. ANOVA for SLCA654 (two pages).

Legend

Appendices 3 – 13 are all tables of ANOVA for the SNPs genotyped in this thesis. Hopefully, the interpretation of the following tables should not pose any problems. The top section contains information on the gene and SNP, the amount of samples in each genotype, and the number of examined individuals. The middle section contains mean values and standard deviation values for the different phenotypes according to genotype group, as well as p-values; gender specific and for the whole cohort. If adjusted R^2-values are given, they belong to the significant p-value, marked in bold. If several p-values are significant for any parameter, the R^2-value belongs to the gender non-specific p-value, if not accompanied by a "♀" or "♂" sign. P-values adjusted for significantly associated parameters, such as BMI, age, gender, or puberty, are given for normalized parameters that present a significant, or in some cases non-significant, trend. The bottom section contains gender specific information on mean and standard deviation values, and in some cases explorations of alternative partitions of data, recessive or dominant. Parameters that are not normalized and that have undergone a Kruskal-Wallis analysis are marked in light grey. For these parameters, no multivariate analysis could be conducted. In some cases, where trends were observable but not significant, multivariate analyses have been conducted to explore possible effects from covariates.

The parameters in the ANOVA tables are measured in units according to the following list.

Birth weight	g	S-cholesterol	mmol/l
Birth length	cm	Glucose effectiveness	/min
BMI SDS (RC)	m	P-glucose	mmol/l
BMI	kg/m^2	Insulin sensitivity	$Si \cdot 10^{-5}$/pM/min
Age obese	year	Insulin response	pM 0-10 min
S-LDL triglycerides	mmol/l	fS-insulin	pmol/l
S-HDL triglycerides	mmol/l	HOMA	---
S-VLDL triglycerides	mmol/l	LDL-cholesterol	mmol/l
Triglycerides	mmol/l	HDL-cholesterol	mmol/l
S-VLDL cholesterol	mmol/l		

Appendix 3. ANOVA for SLCA654, continuation.

Univariate test								Multivariate test		
SLCA	256	131	28	422	225	197				
SLCA654	**Homozygote**	**Heterozygote**	**Homozygote**					**P adjusted**		
	TT	CT + CC		♀/♂ **p**	♀, **p**	♂, **p**	**R^2 (adj)**	♀/♂	♀	♂
Birth weight	3579 ± 655	3507 ± 642		0.226	**0.041**	0.875	1.60%		0.986	
Birth length	50.42 ± 2.51	50.06 ± 2.32		0.073	0.086	0.257				
BMI SDS (RC)	6.067 ± 1.348	6.014 ± 1.236		0.908	0.856	0.681				
BMI	37.55 ± 6.12	37.65 ± 6.55		0.905	0.938	0.972				
Age obese	4.339 ± 2.619	4.615 ± 2.974		0.609	0.882	0.438				
S-LDL triglycerides	0.286 ± 0.102	0.284 ± 0.092		0.968	0.982	0.974				
S-HDL triglycerides	0.212 ± 0.078	0.214 ± 0.079		0.896	0.430	0.540				
S-VLDL triglycerides	0.821 ± 0.493	0.754 ± 0.500		0.154	0.665	0.115				
Triglycerides	1.199 ± 0.589	1.197 ± 0.607		0.780	0.891	0.465				
S-VLDL cholesterol	0.394 ± 0.218	0.379 ± 0.258		0.167	0.540	0.173				
S-cholesterol	4.302 ± 0.849	4.340 ± 0.833		0.669	0.945	0.536				
Glucose effectiveness	0.019 ± 0.011	0.018 ± 0.009		0.365	0.258	0.858				
P-glucose	4.788 ± 0.618	4.856 ± 0.569		0.227	0.087	0.988				
Insulin sensitivity	0.389 ± 0.262	0.347 ± 0.239		0.189	0.199	0.529				
Insulin response	6737 ± 4294	6434 ± 4071		0.557	0.505	0.907				
fS-insulin	110.6 ± 62.2	108.1 ± 52.6		0.875	0.723	0.869				
HOMA	3.257 ± 1.874	3.249 ± 1.693		0.693	0.391	0.714				
LDL-cholesterol	2.836 ± 0.797	2.779 ± 0.869		0.522	0.366	0.907				
HDL-cholesterol	0.998 ± 0.314	1.033 ± 0.303		0.367	0.254	0.986				
Significances	TT	CT + CC		♀/♂	♀	♂	**R^2 (adj)**	♀/♂	♀	♂
Birth weight ♀	3528 ± 712	3343 ± 590		0.226	**0.041**	0.875	1.60%		0.986	

Appendix 4. ANOVA for SLCA296.

Univariate test								Multivariate test	
SLCA	172	169	64	422	225	197			
SLCA296	**Homozygote**	**Heterozygote**	**Homozygote**					**P adjusted**	
	TT	CT	CC	♀/♂ **p**	♀, **p**	♂, **p**	**R^2 (adj)**	♀/♂	♀ ♂
Birth weight	3612 ± 661	3519 ± 671	3472 ± 544	0.229	0.152	0.790			
Birth length	50.61 ± 2.58	50.13 ± 2.29	49.78 ± 2.38	**0.017**	**0.039**	0.193			
BMI SDS (RC)	6.051 ± 1.402	6.037 ± 1.271	6.076 ± 1.158	0.959	0.811	0.799			
BMI	37.81 ± 6.60	37.30 ± 5.88	37.66 ± 6.31	0.949	0.775	0.750			
Age obese	4.373 ± 2.258	4.489 ± 2.994	4.522 ± 2.898	0.835	0.456	0.527			
S-LDL triglycerides	0.294 ± 0.107	0.276 ± 0.091	0.283 ± 0.092	0.572	0.386	0.427			
S-HDL triglycerides	0.212 ± 0.074	0.219 ± 0.084	0.200 ± 0.072	0.593	0.415	0.859			
S-VLDL triglycerides	0.874 ± 0.529	0.750 ± 0.446	0.707 ± 0.514	0.061	0.331	0.055			
Triglycerides	1.238 ± 0.623	1.171 ± 0.546	1.162 ± 0.657	0.440	0.417	0.576			
S-VLDL cholesterol	0.412 ± 0.234	0.360 ± 0.200	0.409 ± 0.306	0.228	0.694	0.061			
S-cholesterol	4.321 ± 0.892	4.282 ± 0.777	4.396 ± 0.888	0.817	0.680	0.935			
Glucose effectiveness	0.018 ± 0.010	0.018 ± 0.010	0.019 ± 0.008	0.642	0.760	0.361			
P-glucose	4.795 ± 0.653	4.817 ± 0.577	4.853 ± 0.520	0.608	0.408	0.989			
Insulin sensitivity	0.382 ± 0.252	0.366 ± 0.258	0.367 ± 0.245	0.615	0.213	0.747			
Insulin response	6520 ± 3962	6749 ± 4231	6514 ± 4730	0.889	0.889	0.985			
fS-insulin	109.3 ± 62.2	112.9 ± 57.7	101.9 ± 51.3	0.454	0.475	0.202			
HOMA	3.223 ± 1.903	3.316 ± 1.756	3.166 ± 1.695	0.654	0.275	0.492			
LDL-cholesterol	2.838 ± 0.859	2.794 ± 0.776	2.801 ± 0.885	0.938	0.854	0.949			
HDL-cholesterol	0.980 ± 0.298	1.022 ± 0.293	1.069 ± 0.382	0.355	0.517	0.694			
Significances	TT	CT	CC	♀/♂	♀	♂	**R^2 (adj)**	♀/♂	♀ ♂
Birth length ♀	50.21 ± 2.45	49.61 ± 2.30	49.20 ± 2.44		**0.039**				

Appendix 5. ANOVA for SLCA577.

Univariate test								Multivariate test		
SLCA	297	106	12	422	225	197				
SLCA577	**Homozygote**	**Heterozygote**	**Homozygote**					**P adjusted**		
	TT	AT + AT		♀/♂ **p**	♀, **p**	♂, **p**	R^2 **(adj)**	♀/♂	♀	♂
Birth weight	3572 ± 668	3499 ± 604		0.204	0.092	0.828				
Birth length	50.43 ± 2.53	49.93 ± 2.17		**0.038**	**0.045**	0.233				
BMI SDS (RC)	6.05 ± 1.31	6.05 ± 1.34		0.984	0.597	0.504				
BMI	37.60 ± 5.98	37.53 ± 7.16		0.646	0.728	0.374				
Age obese	4.407 ± 2.672	4.540 ± 2.981		0.845	0.992	0.830				
S-LDL triglycerides	0.288 ± 0.103	0.288 ± 0.103		0.601	0.979	0.373				
S-HDL triglycerides	0.215 ± 0.081	0.209 ± 0.069		0.760	0.418	0.173				
S-VLDL triglycerides	0.809 ± 0.492	0.763 ± 0.506		0.351	0.578	0.461				
Triglycerides	1.205 ± 0.604	1.179 ± 0.573		0.770	0.966	0.603				
S-VLDL cholesterol	0.392 ± 0.229	0.380 ± 0.245		0.438	0.955	0.292				
S-cholesterol	4.325 ± 0.847	4.293 ± 0.832		0.778	0.727	0.998				
Glucose effectiveness	0.019 ± 0.010	0.017 ± 0.009		0.180	**0.007**	0.424				
P-glucose	4.790 ± 0.608	4.869 ± 0.578		0.241	**0.043**	0.705				
Insulin sensitivity	0.385 ± 0.261	0.344 ± 0.232		0.247	0.117	0.854				
Insulin response	6778 ± 4357	6251 ± 3826		0.407	0.198	0.831				
fS-insulin	108.3 ± 60.6	112.5 ± 53.9		0.307	0.411	0.557				
HOMA	3.212 ± 1.849	3.350 ± 1.698		0.312	0.170	0.985				
LDL-cholesterol	2.843 ± 0.795	2.739 ± 0.895		0.287	0.296	0.696				
HDL-cholesterol	1.014 ± 0.325	1.006 ± 0.268		0.905	0.810	0.900				
Significances	TT	AT + AT		♀/♂	♀	♂	R^2 **(adj)**	♀/♂	♀	♂
Birthlength	50.43 ± 2.53	49.93 ± 2.17		**0.038**		0.233				
Birthlength	49.99 ± 2.48	49.37 ± 2.16			**0.045**	0.233				
Glucose effectiveness ♀	0.019 ± 0.010	0.015 ± 0.008		0.180	**0.007**	0.424				
P-glucose ♀	4.744 ± 0.589	4.937 ± 0.593		0.241	**0.043**	0.705				

Appendix 6. ANOVA for SLCB140.

Univariate test								Multivariate test		
SLCB	117	192	94	422	225	197				
SLCB140	**Homozygote**	**Heterozygote**	**Homozygote**					**P adjusted**		
	AA	AG	GG	♀/♂ **p**	♀, **p**	♂, **p**	**R^2 (adj)**	♀/♂	♀	♂
Birth weight	3573 ± 599	3543 ± 668	3568 ± 674	0.602	0.367	0.099				
Birth length	50.17 ± 2.44	50.14 ± 2.41	50.69 ± 2.52	0.148	0.914	**0.001**				
BMI SDS (RC)	6.009 ± 1.544	6.142 ± 1.323	6.058 ± 1.038	0.658	0.994	0.306				
BMI	37.25 ± 7.18	38.03 ± 6.06	37.67 ± 5.97	0.597	0.936	0.257				
Age obese	4.333 ± 2.879	4.349 ± 2.566	4.840 ± 2.921	0.304	0.731	0.210				
S-LDL triglycerides	0.297 ± 0.103	0.285 ± 0.103	0.271 ± 0.082	0.382	0.100	0.763				
S-HDL triglycerides	0.231 ± 0.090	0.208 ± 0.075	0.205 ± 0.066	0.258	0.100	0.977				
S-VLDL triglycerides	0.731 ± 0.483	0.838 ± 0.505	0.803 ± 0.509	0.266	0.649	0.266				
Triglycerides	1.205 ± 0.641	1.217 ± 0.594	1.153 ± 0.547	0.652	0.722	0.866				
S-VLDL cholesterol	0.366 ± 0.239	0.401 ± 0.240	0.395 ± 0.220	0.417	0.526	0.694				
S-cholesterol	4.355 ± 0.866	4.268 ± 0.837	4.356 ± 0.849	0.672	0.202	0.673				
Glucose effectiveness	0.020 ± 0.009	0.019 ± 0.010	0.016 ± 0.009	0.084	0.506	**0.032**				
P-glucose	4.770 ± 0.613	4.846 ± 0.630	4.788 ± 0.526	0.677	0.662	0.936				
Insulin sensitivity	0.339 ± 0.238	0.397 ± 0.262	0.344 ± 0.248	0.116	**0.049**	0.739	2.69%		0.138	
Insulin response	7402 ± 4574	6278 ± 3924	6712 ± 4355	0.216	0.283	0.452				
fS-insulin	119.9 ± 66.3	102.2 ± 50.6	115.7 ± 62.0	0.083	0.144	0.206				
HOMA	3.549 ± 2.087	3.038 ± 1.527	3.435 ± 1.919	0.144	0.234	0.313				
LDL-cholesterol	2.891 ± 0.815	2.705 ± 0.822	2.909 ± 0.814	0.132	0.105	0.342				
HDL-cholesterol	1.036 ± 0.362	1.004 ± 0.274	1.024 ± 0.320	0.927	0.723	0.709				
Significances	AA	AG	GG	♀/♂	♀	♂	**R^2 (adj)**	♀/♂	♀	♂
Glucose effectiveness ♂	0.022 ± 0.009	0.018 ± 0.011	0.016 ± 0.010	0.084	0.506	**0.032**				
Insulin sensitivity ♀	0.308 ± 0.166	0.409 ± 0.222	0.350 ± 0.262	0.116	**0.049**	0.739	2.69%		0.138	

Appendix 7. ANOVA for SLCB009.

Univariate test								Multivariate test			
SLCB	423	26	2	422	228	208					
SLCB009	**Homozygote**	**Heterozygote**	**Homozygote**					**P adjusted**			
	GG	AG + AA		♀/♂, **p**	♀, **p**	♂, **p**	**R^2 (adj)**	♀/♂	♀	♂	**R^2 (adj)**
Birth weight	3526 ± 654	3547 ± 630		0.864	0.355	0.438					
Birth length	50.25 ± 2.43	50.38 ± 2.12		0.800	0.937	0.652					
BMI SDS (RC)	5.990 ± 1.345	5.484 ± 0.916		0.332	0.494	0.495					
BMI	37.14 ± 6.32	37.79 ± 4.21		0.564	0.766	0.617					
Age obese	4.525 ± 2.930	4.429 ± 3.171		0.535	0.439	0.965					
S-LDL triglycerides	0.288 ± 0.099	0.327 ± 0.0961		0.141	0.311	0.285					
S-HDL triglycerides	0.215 ± 0.078	0.207 ± 0.056		0.849	0.986	0.763					
S-VLDL triglycerides	0.782 ± 0.455	0.925 ± 0.730		0.919	0.806	0.637					
Triglycerides	1.213 ± 0.596	1.288 ± 0.719		0.962	0.690	0.554					
S-VLDL cholesterol	0.390 ± 0.230	0.425 ± 0.286		0.816	0.908	0.606					
S-cholesterol	4.313 ± 0.797	4.775 ± 0.863		**0.034**	**0.031**	0.432	2.24% (♀)		**0.029**		2.69%
Glucose efficiency	0.019 ± 0.010	0.018 ± 0.009		0.999	0.367	0.254					
P-glucose	4.786 ± 0.582	4.683 ± 0.561		0.475	0.588	0.669					
Insulin sensitivity	0.354 ± 0.247	0.374 ± 0.240		0.633	0.838	0.394					
Insulin response	6577 ± 4083	7484 ± 6251		0.670	0.557	0.194					
fS-insulin	113.8 ± 64.9	126.9 ± 74.5		0.408	0.725	0.414					
HOMA	3.579 ± 2.025	4.289 ± 2.499		0.217	0.872	**0.017**	3.69%			0.109	
LDL-cholesterol	2.827 ± 0.798	2.994 ± 0.991		0.469	0.563	**0.049**	2.25%			0.099	
HDL-cholesterol	1.002 ± 0.304	0.994 ± 0.297		0.659	0.562	0.914					
Significances	GG	AG + AA		♀/♂	♀	♂	**R^2 (adj)**	♀/♂	♀	♂	**R^2 (adj)**
S-cholesterol	4.313 ± 0.797	4.775 ± 0.863		**0.034**			1.18%	0.064			
S-cholesterol ♀	4.563 ± 0.759	4.237 ± 0.617			**0.031**	0.432	2.24%		**0.029**		2.69%
HOMA ♂	3.743 ± 2.194	6.002 ± 2.173		0.217	0.872	**0.017**	3.69%			0.109	
LDL-cholesterol ♂	2.759 ± 0.815	3.480 ± 0.319		0.469	0.563	**0.049**	2.25%			0.099	

Appendix 8. ANOVA for SLCB364.

Univariate test								Multivariate test		
SLCB	408	23	2	422	228	208				
SLCB364	**Homozygote**	**Heterozygote**	**Homozygote**					**P adjusted**		
	CC	CT + TT		♀/♂ **p**	♀, p	♂, p	R^2 (adj)	♀/♂	♀	♂
Birth weight	3523 ± 656	3590 ± 583		0.661	0.434	0.814				
Birth length	50.29 ± 2.43	50.00 ± 2.28		0.525	0.436	0.761				
BMI SDS (RC)	5.973 ± 1.352	5.752 ± 0.488		0.841	0.883	0.818				
BMI	37.24 ± 6.28	35.25 ± 4.33		0.496	0.597	0.401				
Age obese	4.522 ± 2.943	4.750 ± 3.015		0.486	0.064	0.429				
S-LDL triglycerides	0.290 ± 0.098	0.289 ± 0.108		0.925	0.454	0.514				
S-HDL triglycerides	0.214 ± 0.078	0.222 ± 0.065		0.471	0.367	0.055				
S-VLDL triglycerides	0.798 ± 0.478	0.672 ± 0.397		0.323	0.396	0.567				
Triglycerides	1.229 ± 0.610	1.043 ± 0.457		0.182	0.340	0.293				
S-VLDL cholesterol	0.397 ± 0.235	0.316 ± 0.186		0.105	0.290	0.197				
S-cholesterol	4.340 ± 0.817	4.300 ± 0.647		0.967	0.862	0.896				
Glucose effectiveness	0.019 ± 0.010	0.018 ± 0.011		0.818	0.803	0.936				
P-glucose	4.765 ± 0.575	5.000 ± 0.626		0.090	0.194	0.289				
Insulin sensitivity	0.354 ± 0.250	0.377 ± 0.186		0.318	0.052	0.678			0.053	
Insulin response	6726 ± 4303	5320 ± 3209		0.239	0.317	0.484				
fS-insulin	115.9 ± 66.79	95.53 ± 36.22		0.375	0.263	0.883				
HOMA	3.635 ± 2.065	3.433 ± 2.045		0.733	0.275	0.568				
LDL-cholesterol	2.838 ± 0.819	2.794 ± 0.624		0.913	0.912	0.982				
HDL-cholesterol	1.010 ± 0.297	0.850 ± 0.378		0.129	0.246	0.284				

Significances	CC	CT + TT	♀/♂	♀	♂	R^2 (adj)	♀/♂	♀	♂
None.									

Appendix 9. ANOVA for SLCB742 (two pages).

Univariate test								Multivariate test			
SLCB	150	208	92	422	228	208					
SLCB742	**Homozygote**	**Heterozygote**	**Homozygote**					**P adjusted**			
	GG	AG	AA	♀/♂ **p**	♀, **p**	♂, **p**	R^2 **(adj)**	♀/♂	♀	♂	R^2 **(adj)**
Birth weight	3478 ± 660	3543 ± 668	3573 ± 599	0.497	0.843	0.565					
Birth length	50.14 ± 2.70	50.33 ± 2.18	50.36 ± 2.44	0.392	0.464	0.574					
BMI SDS (RC)	5.911 ± 1.486	6.038 ± 1.298	5.916 ± 1.170	0.788	0.814	0.484					
BMI	35.87 ± 6.59	37.72 ± 6.44	37.97 ± 4.99	0.070	0.063	0.717		0.287			
Age obese	4.413 ± 2.938	4.542 ± 2.820	4.674 ± 3.205	0.595	0.264	0.945					
S-LDL triglycerides	0.306 ± 0.104	0.284 ± 0.102	0.275 ± 0.079	0.172	0.078	0.783					
S-HDL triglycerides	0.219 ± 0.084	0.207 ± 0.071	0.226 ± 0.079	0.321	**0.018**	0.428					
S-VLDL triglycerides	0.812 ± 0.520	0.779 ± 0.450	0.775 ± 0.450	0.964	0.195	**0.044**					
Triglycerides	1.257 ± 0.626	1.190 ± 0.569	1.212 ± 0.636	0.678	0.228	0.157					
S-VLDL cholesterol	0.404 ± 0.226	0.386 ± 0.235	0.386 ± 0.237	0.636	0.156	0.138					
S-cholesterol	4.360 ± 0.737	4.346 ± 0.881	4.267 ± 0.729	0.790	0.672	0.543					
Glucose effectiveness	0.020 ± 0.011	0.017 ± 0.010	0.020 ± 0.010	0.093	0.605	0.064					
P-glucose	4.710 ± 0.587	4.819 ± 0.607	4.792 ± 0.498	0.334	0.749	**0.045**	3.04%			**0.010**	7.45%
Insulin sensitivity	0.390 ± 0.261	0.355 ± 0.249	0.310 ± 0.216	0.164	0.066	0.526			0.097		
Insulin response	6990 ± 4495	6374 ± 4020	6657 ± 4410	0.563	0.901	0.457					
fS-insulin	110.1 ± 63.9	117.2 ± 68.7	115.0 ± 61.3	0.690	0.533	0.333					
HOMA	3.375 ± 1.807	3.887 ± 2.302	3.507 ± 1.903	0.325	0.339	**0.006**	6.46%			**0.009**	10.42%
LDL-cholesterol	2.781 ± 0.784	2.890 ± 0.866	2.779 ± 0.714	0.592	0.492	0.901					
HDL-cholesterol	1.018 ± 0.266	0.958 ± 0.327	1.064 ± 0.304	0.093	0.074	0.627					

Appendix 9. ANOVA for SLCB742, continuation.

Univariate test							Multivariate test				
SLCB	150	208	92	422	228	208					
SLCB742	**Homozygote**	**Heterozygote**	**Homozygote**					**P adjusted**			
Significances	GG	AG	AA	♀/♂	♀	♂	**R^2 (adj)**	♀/♂	♀	♂	**R^2 (adj)**
S-HDL triglycerides ♀	0.225 ± 0.095	0.199 ± 0.071	0.247 ± 0.086	0.321	**0.018**	0.428					
S-VDLDL triglycerides ♂	0.775 ± 0.498	0.867 ± 0.439	0.646 ± 0.387	0.964	0.195	**0.044**					
P-glucose ♂	4.626 ± 0.529	4.878 ± 0.669	4.943 ± 0.508	0.334	0.749	**0.045**	3.04%			**0.010**	7.45%
HOMA ♂	3.335 ± 1.903	4.563 ± 2.482	3.166 ± 1.758	0.325	0.339	**0.006**	6.46%			**0.009**	10.42%

Appendix 10. ANOVA for GPRA627 (two pages).

Univariate test						Multivariate test					
GPRA	147	205	84		228	208					
GPRA627	**Homozygote**	**Heterozygote**	**Homozygote**					**P adjusted**			
	GG	GT	TT	♀/♂ **p**	♀, **p**	♂, **p**	**R^2 (adj)**	♀/♂	♀	♂	**R^2 (adj)**
Birth weight	3467 ± 596	3530 ± 673	3624 ± 682	0.253	0.349	0.170					
Birth length	50.22 ± 2.34	50.20 ± 2.39	50.54 ± 2.59	0.783	0.868	0.449					
BMI SDS (RC)	6.029 ± 1.273	5.914 ± 1.445	6.014 ± 1.034	0.755	0.288	0.393					
BMI	37.62 ± 6.87	36.65 ± 6.04	38.00 ± 5.19	0.439	0.183	**0.040**				**0.019**	25.42%
Age obese	4.804 ± 3.194	4.441 ± 2.779	4.170 ± 2.775	0.248	0.314	0.671					
S-LDL triglycerides	0.298 ± 0.098	0.285 ± 0.104	0.285 ± 0.085	0.604	0.880	0.264					
S-HDL triglycerides	0.216 ± 0.077	0.217 ± 0.077	0.209 ± 0.079	0.774	0.359	0.814					
S-VLDL triglycerides	0.798 ± 0.445	0.794 ± 0.492	0.764 ± 0.473	0.750	0.368	0.574					
Triglycerides	1.291 ± 0.611	1.197 ± 0.584	1.141 ± 0.624	0.062	**0.027**	0.307					
S-VLDL cholesterol	0.406 ± 0.252	0.386 ± 0.227	0.380 ± 0.211	0.939	0.785	0.873					
S-cholesterol	4.329 ± 0.829	4.328 ± 0.732	4.360 ± 0.943	0.983	0.655	0.791					
Glucose effectiveness	0.019 ± 0.010	0.018 ± 0.011	0.019 ± 0.009	0.333	0.211	0.846					
P-glucose	4.858 ± 0.577	4.694 ± 0.541	4.830 ± 0.657	0.070	0.166	0.207					
Insulin sensitivity	0.313 ± 0.213	0.371 ± 0.259	0.398 ± 0.267	0.112	0.068	0.107					
Insulin response	6991 ± 4221	6714 ± 4371	5705 ± 3933	0.080	0.130	0.521					
fS-insulin	127.6 ± 72.1	111.3 ± 62.3	97.3 ± 55.6	**0.022**	**0.016**	0.051	1.75%	0.123	0.054		2.85%
HOMA	4.180 ± 2.274	3.328 ± 1.781	3.336 ± 2.142	**0.005**	**0.002**	0.066	2.99%	**0.012**	**0.009**	**0.046**	3.52%
LDL-cholesterol	2.867 ± 0.809	2.831 ± 0.812	2.775 ± 0.813	0.779	0.572	0.921					
HDL-cholesterol	0.997 ± 0.333	0.993 ± 0.283	1.036 ± 0.295	0.828	0.768	0.994					

Appendix 10. ANOVA for GPRA627, continuation.

Univariate test								Multivariate test			
GPRA	147	205	84		228	208					
GPRA627	**Homozygote**	**Heterozygote**	**Homozygote**					**P adjusted**			
Significances	GG	GT	TT	♀/♂	♀	♂	R^2 (adj)	♀/♂	♀	♂	R^2 (adj)
BMI ♂	34.02 ± 4.16	35.85 ± 6.36	40.79 ± 6.98	0.439	0.183	**0.040**	6.84%			**0.019**	25.42%
Triglycerides ♀	1.235 ± 0.553	1.205 ± 0.599	1.014 ± 0.514	0.062	**0.027**	0.307					
Triglycerides ♀ *recessive*	1.218 ± 0.578		1.014 ± 0.514		**0.012**						
fS-insulin ♀	121.8 ± 68.9	119.5 ± 61.8	86.5 ± 39.5	0.022	**0.016**	0.051	3.57%		0.054		2.94%
fS-insulin ♀ *recessive*	120.6 ± 64.9		86.5 ± 39.5		**0.004**		4.11%		**0.017**		3.46%
HOMA	4.180 ± 2.274	3.328 ± 1.781	3.336 ± 2.142	**0.005**	0.002	0.066	2.99%	**0.012**			3.52%
HOMA *dominant*	4.180 ± 2.274		3.330 ± 1.879	**0.001**			3.30%	**0.004**			3.76%
HOMA ♀	4.103 ± 2.239	3.263 ± 1.603	2.666 ± 1.442	0.005	**0.002**	0.066	6.42%		**0.009**		5.37%
HOMA ♀ *recessive*	3.632 ± 1.946		2.666 ± 1.442		**0.006**		4.16%		**0.043**		2.19%
HOMA ♂	4.278 ± 2.340	3.397 ± 1.964	4.429 ± 2.645			0.066	2.76%			**0.009**	7.63%

Appendix 11. ANOVA for GPRA261.

Univariate test							Multivariate test			
GPRA	300	121	10	228	208					
GPRA261	**Homozygote**	**Heterozygote**	**Homozygote**				**P adjusted**			
	TT	CT + CC		♀/♂ **p**	♀, p	♂, p	R^2 (adj)	♀/♂	♀	♂	R^2 (adj)
Birth weight	3539 ± 659	3503 ± 639		0.691	0.843	0.652					
Birth length	50.31 ± 2.44	50.18 ± 2.39		0.994	0.716	0.594					
BMI SDS (RC)	5.955 ± 1.399	5.998 ± 1.165		0.683	0.367	0.590					
BMI	37.24 ± 6.46	37.10 ± 5.70		0.928	0.834	0.819					
Age obese	4.502 ± 2.799	4.480 ± 3.110		0.643	0.299	0.633					
S-LDL triglycerides	0.286 ± 0.099	0.296 ± 0.099		0.364	**0.045**	0.274					
S-HDL triglycerides	0.214 ± 0.077	0.216 ± 0.077		0.767	0.818	0.878					
S-VLDL triglycerides	0.808 ± 0.481	0.750 ± 0.452		0.335	0.878	0.106					
Triglycerides	1.228 ± 0.621	1.203 ± 0.560		0.955	0.397	0.418					
S-VLDL cholesterol	0.403 ± 0.240	0.368 ± 0.211		0.337	0.948	0.190					
S-cholesterol	4.293 ± 0.822	4.426 ± 0.747		0.145	0.224	0.460					
Glucose effectiveness	0.019 ± 0.011	0.018 ± 0.009		0.505	0.436	0.087					
P-glucose	4.748 ± 0.559	4.848 ± 0.626		0.205	0.161	0.713					
Insulin sensitivity	0.375 ± 0.255	0.295 ± 0.204		**0.017**	0.236	**0.036**	1.55%	0.074			2.54%
Insulin response	6759 ± 4333	6340 ± 4049		0.450	0.604	0.598					
fS-insulin	110.1 ± 66.4	127.5 ± 62.4		**0.010**	0.207	**0.018**	1.77%	**0.045**			3.35%
HOMA	3.504 ± 2.097	3.977 ± 1.944		**0.036**	0.081	0.249	1.21%				
LDL-cholesterol	2.828 ± 0.827	2.846 ± 0.769		0.807	0.915	0.781					
HDL-cholesterol	0.992 ± 0.297	1.026 ± 0.321		0.281	0.933	0.121					
Significances	TT	CT + CC		♀/♂	♀	♂	R^2 (adj)	♀/♂	♀	♂	R^2 (adj)
S-LDL triglycerides ♀	0.280 ± 0.104	0.313 ± 0.104		0.364	**0.045**	0.274					
Insulin sensitivity ♂	0.387 ± 0.288	0.279 ± 0.219		0.017	0.236	**0.036**	2.51%			0.066	8.30%
fS-insulin ♂	108.6 ± 72.4	132.4 ± 58.7		0.010	0.207	**0.018**	1.61%			0.065	7.08%

Appendix 12. ANOVA for GPRA081.

Univariate test						Multivariate test				
GPRA	251	164	20		228	208				
GPRA081	**Homozygote**	**Heterozygote**	**Homozygote**				**P adjusted**			
	AA	AC + CC		♀/♂ **p**	♀, **p**	♂, **p**	**R^2 (adj)**	♀/♂	♀	♂	**R^2 (adj)**
Birth weight	3538 ± 652	3514 ± 651		0.682	0.616	0.976					
Birth length	50.19 ± 2.41	50.39 ± 2.42		0.232	0.722	0.222					
BMI SDS (RC)	5.833 ± 1.283	6.132 ± 1.366		0.165	**0.032**	0.584	3.78%		0.081		8.84%
BMI	36.81 ± 5.56	37.65 ± 6.92		0.601	0.843	0.351					
Age obese	4.611 ± 2.949	4.411 ± 2.926		0.365	**0.050**	0.407					
S-LDL triglycerides	0.284 ± 0.101	0.296 ± 0.096		0.200	0.576	0.176					
S-HDL triglycerides	0.210 ± 0.073	0.221 ± 0.081		0.320	0.507	0.456					
S-VLDL triglycerides	0.805 ± 0.510	0.770 ± 0.422		0.992	0.635	0.581					
Triglycerides	1.232 ± 0.640	1.198 ± 0.549		0.868	0.802	0.604					
S-VLDL cholesterol	0.408 ± 0.253	0.371 ± 0.201		0.560	0.278	0.717					
S-cholesterol	4.284 ± 0.865	0.397 ± 0.720		0.129	0.459	0.157					
Glucose effectiveness	0.019 ± 0.011	0.018 ± 0.009		0.098	0.562	0.104					
P-glucose	4.766 ± 0.527	4.796 ± 0.645		0.834	0.458	0.647					
Insulin sensitivity	0.378 ± 0.263	0.326 ± 0.222		0.129	0.181	0.396					
Insulin response	6436 ± 4289	6885 ± 4201		0.273	0.200	0.777					
fS-insulin	105.7 ± 62.8	126.3 ± 67.5		**0.003**	0.067	**0.016**	2.50%	**0.004**			4.42%
HOMA	3.330 ± 1.894	3.981 ± 2.201		**0.006**	**0.041**	0.069	2.35%	**0.023**			2.45%
LDL-cholesterol	2.791 ± 0.835	2.885 ± 0.773		0.281	0.186	0.829					
HDL-cholesterol	0.991 ± 0.296	1.020 ± 0.314		0.499	0.849	0.270					
Significances	AA	AC + CC		♀/♂	♀	♂	**R^2 (adj)**	♀/♂	♀	♂	**R^2 (adj)**
BMI SDS (RC) ♀	5.663 ± 1.264	6.260 ± 1.359		0.165	**0.032**	0.584	3.78%		0.081		8.84%
Age obese ♀	4.825 ± 3.159	4.085 ± 2.854		0.365	**0.050**	0.407					
fS-insulin ♂	106.2 ± 71.0	126.8 ± 65.8		0.003	0.067	**0.016**	3.27%			**0.034**	7.79%
HOMA ♀	3.133 ± 1.591	3.829 ± 2.166		0.006	**0.041**	0.069	2.06%		**0.062**		1.73%

Appendix 13. ANOVA for GPRA427.

Univariate test								Multivariate test			
GPRA	217	183	35		228	208					
GPRA427	**Homozygote**	**Heterozygote**	**Homozygote**					**P adjusted**			
	AA	AG	GG	♀/♂ **p**	♀, **p**	♂, **p**	**R^2 (adj)**	♀/♂	♀	♂	**R^2 (adj)**
Birth weight	3536 ± 636	3531 ± 685	3460 ± 570	0.846	0.885	0.897					
Birth length	50.39 ± 2.44	50.18 ± 2.36	50.06 ± 2.60	0.635	0.914	0.539					
BMI SDS (RC)	5.950 ± 1.217	5.961 ± 1.383	6.111 ± 1.702	0.977	0.781	0.600					
BMI	36.76 ± 5.46	37.52 ± 6.70	38.10 ± 7.96	0.847	0.580	0.918					
Age obese	4.627 ± 2.999	4.338 ± 2.659	4.808 ± 3.803	0.715	0.320	0.859					
S-LDL triglycerides	0.299 ± 0.097	0.280 ± 0.100	0.280 ± 0.100	0.220	0.155	0.889					
S-HDL triglycerides	0.213 ± 0.073	0.216 ± 0.085	0.219 ± 0.057	0.822	0.335	0.316					
S-VLDL triglycerides	0.806 ± 0.499	0.771 ± 0.444	0.785 ± 0.460	0.840	0.851	0.671					
Triglycerides	0.194 ± 0.583	1.229 ± 0.603	1.307 ± 0.715	0.644	0.184	0.817					
S-VLDL cholesterol	0.405 ± 0.243	0.376 ± 0.214	0.392 ± 0.250	0.677	0.956	0.314					
S-cholesterol	4.331 ± 0.816	4.329 ± 0.782	4.381 ± 0.876	0.977	0.490	0.616					
Glucose effectiveness	0.019 ± 0.011	0.019 ± 0.009	0.018 ± 0.010	0.924	0.788	0.438					
P-glucose	5.764 ± 0.598	4.833 ± 0.582	4.579 ± 0.383	0.148	0.562	0.069					
Insulin sensitivity	0.363 ± 0.259	0.351 ± 0.232	0.337 ± 0.243	0.849	0.992	0.808					
Insulin response	6317 ± 4357	7062 ± 4192	6332 ± 2841	0.173	0.481	0.373					
fS-insulin	114.2 ± 68.4	114.2 ± 64.2	116.9 ± 56.0	0.848	0.736	0.998					
HOMA	3.582 ± 2.139	3.659 ± 2.050	3.709 ± 1.620	0.781	0.203	0.575					
LDL-cholesterol	2.833 ± 0.827	2.816 ± 0.792	2.914 ± 0.810	0.874	0.480	0.539					
HDL-cholesterol	1.008 ± 0.291	1.005 ± 0.322	0.950 ± 27.1	0.537	0.442	0.490					

Significances	AA	AG	GG	♀/♂	♀	♂	**R^2 (adj)**	♀/♂	♀	♂	**R^2 (adj)**
None.											