UPPSALA
UNIVERSITET

*Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 921

# Obesity Genetics

## *Functional Aspects of Four Genetic Loci Associated with Obesity and Body Mass*

MATHIAS RASK-ANDERSEN



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2013

ISSN 1651-6206
ISBN 978-91-554-8713-3
urn:nbn:se:uu:diva-204449

Dissertation presented at Uppsala University to be publicly examined in A1:111, BMC, Husargatan 3, Uppsala, Thursday, September 19, 2013 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English.

**Abstract**
Rask-Andersen, M. 2013. Obesity Genetics: Functional Aspects of Four Genetic Loci Associated with Obesity and Body Mass. Acta Universitatis Upsaliensis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 921. 48 pp. Upssala. ISBN 978-91-554-8713-3.

Obesity is a complex disorder which has reached epidemic proportions in many parts of the world. Twin studies have demonstrated a high heritability for obesity. The subsequent application of genome wide association studies (GWAS) in the last decade have identified at least 32 genetic loci associated with body mass and obesity. Despite these great advances, these loci are almost exclusively completely naïve in a functional context. Genetic variations within the gene encoding the fat mass and obesity associated gene (FTO) are the strongest and most consistently observed genetic variants associated with obesity and body mass throughout various studied populations from all parts of the world. The identification of association of FTO with obesity has spurred immense interest in the function of the FTO protein and the functional consequences of its variants. However, the implications of genetic variants at other genetic loci on protein molecular function and body mass development remain undetermined. This thesis aims to examine more closely four of the genetic loci associated with obesity; in proximity of, or associated with: FTO, TMEM18, MAP2K5 and STK33, in two cohorts of children of European descent: a case-control of clinically obese children and normal weight controls from the Stockholm area; and a cross sectional cohort of Greek children. These smaller cohorts allow for studies of more specific effects of genetic variants as individuals in these cohorts can be more carefully studied. TMEM18 gene expression was also studied in the rat-brain where a positive correlation was observed between the body weight of the animal and TMEM18 expression. We also employed next generation sequencing to more carefully study obesity-associated genetic loci related to FTO and TMEM18. We utilized a novel strategy in this project to study genetic variation in the entire FTO- and TMEM18 genes, as well as in the GWAS-identified BMI-associated loci located downstream from TMEM18. This analysis was performed on a case-control cohort of Swedish children (n = ~1000). Through this analysis, we were able to observe genetic variants within intron 1 of the FTO gene to be the main genetic variants asso-ciated with obesity at this locus. We also observed, for the first time, obesity-associated genetic variants within the gene encoding TMEM18. To analyze the potential functional context of FTO we used an in silico approach, utilizing public information databases on mRNA co-expression and protein-protein interaction. Based on our findings, we speculate on a wider functional role of FTO in extracellular ligand-induced neuronal plasticity, possibly via interaction or modulation of the BDNF/NTRK2 signaling pathway.

*Mathias Rask-Andersen, Uppsala University, Department of Neuroscience, Box 593, SE-751 24 Uppsala, Sweden.*



ISSN 1651-6206
ISBN 978-91-554-8713-3
urn:nbn:se:uu:diva-204449 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-204449)

*Conclusion:*
*Big helix in several chains, phosphates on outside, phosphate-phosphate inter-helical bonds disrupted by water.*
*Phosphate links available to proteins.*
***Rosalind Franklin - 1951***

# List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Rask-Andersen, M., Jacobsson, J.A., Moschonis, G., Chavan, R.A., Sikder, M.A.N., Allzén, E., Alsiö, J., Chrousos, G.P., Manios, Y., Fredriksson, R., Schiöth, H.B. (2012) Association of TMEM18 variants with BMI and waist circumference in children and correlation of mRNA expression in the PFC with body weight in rats. *European Journal of Human Genetics. 20 (2), pp. 192-197.*

II Rask-Andersen, M., Jacobsson, J.A., Moschonis, G., Ek, A.E., Chrousos, G.P., Marcus, C., Manios, Y., Fredriksson, R., Schiöth, H.B. (2012) The MAP2K5-linked SNP rs2241423 is associated with BMI and obesity in two cohorts of Swedish and Greek children. *BMC Medical Genetics 13 , art. no. 36.*

III Rask-Andersen M., Moschonis G., Chrousos GP., Marcus C., Manios Y., Fredriksson R., Schiöth HB. The STK33-linked SNP rs4929949 is associated with obesity and BMI in two independent cohorts of Swedish and Greek children. *PLOS ONE, in production.*

IV Sällman Almén, M., Rask-Andersen, M., Jacobsson, J.A., Ameur, A., Kalnina, I., Moschonis, G., Juhlin, S., Bringeland, N., Hedberg, L.A., Ignatovica, V., Chrousos, G.P., Manios, Y., Klovins, J., Marcus, C., Gyllensten, U., Fredriksson, R., Schiöth, H.B. (2013) Determination of the obesity-associated gene variants within the entire FTO gene by ultra-deep targeted sequencing in obese and lean children. *Int J Obes (Lond).Mar;37(3):424-31.*

V Xavier M.J., Rask-Andersen M., Almén M.S., Jacobsson J.A., Ameur A., Moschonis G., Manios Y., Marcus C., Gyllensten U., Fredriksson R., Schiöth H.B., Ultra-deep targeted re-sequencing of TMEM18 in obese and lean European children detects new genetic variants associated with obesity. *Eur J Hum Gen. (revised manuscript under review).*

VI Rask-Andersen, M., Almén, M.S., Olausen, H.R., Olszewski, P.K., Eriksson, J., Chavan, R.A., Levine, A.S., Fredriksson, R., Schiöth, H.B. (2011) Functional coupling analysis suggests link between the obesity gene FTO and the BDNF-NTRK2 signaling pathway. *BMC Neuroscience 12 art. no. 117.*

# Contents

# Abbreviations

BMI – Body mass index
BMISDS – body mass index standard deviation score
GWAS – Genome-wide association study
GIANT consortium – The Genetic Investigation of Anthropomorphic Traits consortium
SNP – Single nucleotide polymorphism
ENCODE – Encyclopedia of DNA elements
KEGG – Kyoto Encyclopedia of Genes and Genomes
OECD - Organisation for Economic Co-operation and Development
SOLiD – Sequencing by Oligonucleotide Ligation and Detection
PCR – Polymerase chain reaction
UVAM – unique valid adjacent mismatches
PFC – Prefrontal cortex

# Introduction

Over the past 30 years, overweight and obesity has become one of the world's leading health concerns. A recent report by the Organisation for Economic Co-operation and development (OECD) observed through epidemiological studies that overweight and obesity have reached high enough proportions to be classified as a global epidemic [1]. High body mass carries with it an increased risk of the development of a number of serious cardiovascular and metabolic diseases, such as type 2 diabetes, hypertension, dyslipidemia, stroke, osteoarthritis as well as several different forms of cancer e.g. prostrate, endometrial, breast as well as colon cancer [2]. For several of these diseases such as dyslipidaemia and related cardiovascular diseases there is a clear link between a body adiposity and disease. For osteoarthritis the increased load on weight bearing joints is believed to be the culprit, however, non-weight bearing joints are also affected. For other diseases such as cancer, the links to high body mass are less clearly defined.

The body mass index (BMI) is the most commonly used variable for estimation of body size. It does not, however, distinguish between lean and adipose body mass, but does serve as an excellent proxy-marker for body adiposity as it is easily calculated from just two simple measurements: weight (in kilograms) and length (in meters). Under current definitions BMI $> 25 kg/m^2$ is considered as overweight, and BMI $> 30\ kg/m^2$ is considered as obese. According to this classification, two out of three adults living in the United States can today be classified as overweight, and one in three as obese [1, 3]. In Sweden the corresponding numbers are not as dramatic but still alarming: today, almost 50% of the adult population is overweight and ten percent obese [1]. These numbers have also risen dramatically in the U.S. and Europe in the last 30 years [4-6]. Even more alarming are the increases in prevalence of childhood obesity and overweight: 34% of children in the US are overweight and 16% obese. In Sweden, prevalence rates for overweight have risen from about 8% in 1980, to 18% in 2000 [7]. On a more positive note, recent studies from Australia, the United States, France and Sweden report that the rise in prevalence rates of childhood overweight and obesity from 1980 throughout the 1990s to have slowed down or even plateaued in the last decade [8-11].

## The obesogenic environment

The industrialization of the western world over the 20th century has led to an increased availability of cheap and calorie-dense palatable foods which often are very aggressively marketed to the population. The industrial revolution has also produced several labor-reducing devices and structures which have removed much of the necessity for physical activity. In an industrialized society we are no longer forced to hunt, gather or otherwise manually produce or acquire food. These environmental changes have also happened very quickly from an evolutionary perspective. This short time span does not allow for evolutionary mechanisms to produce adequate adaptations to the human physiology in response to some of the benefits of modern life: the abundance of food, the reduced requirements for physical activity and a prolonged lifespan [12]. According to the "Thrifty Gene" hypothesis, evolution has favored genetic variants that are better adapted to restricted food supplies, endowing individuals with the ability to store nutrients a better chance of survival. In the modern environment this ability instead becomes detrimental as evidenced by the recent large increases in prevalence of metabolic disorders such as diabetes and obesity and related co-morbidities.

## The genetics of obesity

BMI as a trait is determined not only by environmental factors but also largely by interaction with inherited genetic factors. A method to determine the proportion of natural variance in BMI caused by genetic factors is to to study twins and comparing the concordance, i.e. the degree of similarity between monozygotic and dizygotic twin pairs with regard to a specific trait. Using this method, 60-70% of the variance of BMI has been estimated to be caused by inherited factors [13-14].

Obesity is commonly not a monogenic trait, i.e. one that is defined by the variation of a single gene. Rare instances of monogenic obesity have been identified and these cases have offered great insights into feeding regulatory systems and the pathological potential of a dysfunctional energy balance regulation. Deficiencies in components of the leptin signaling pathway were discovered as the cause of rare cases of severe obesity in 1997 [15]. Leptin is a protein hormone secreted by adipose tissue. It inhibits food intake by binding to leptin receptors in the arcuate nucleus of the hypothalamus, stimulating the production and release of anorexigenic neuropeptides pro-opiomelanocortin (POMC) and cocaine and amphetamine-regulated transcript (CART). Leptin deficient patients display hyperphagia (excessive hunger or increased appetite) and severe weight gain that appears early in life and eventually develops into obesity. These symptoms reverse once the deficiency is pharmacologically compensated [16]. Deficiency of the gene

encoding the melanocortin 4 receptor (MC4R), which is activated by the anorexigenic neuropeptide α-MSH, have also been observed to cause severe obesity [17].

Human genetic studies aimed at identifying genetic elements associated with obesity started to emerge in the late 90s. The genome-wide scans of this era were commonly performed through tag SNP genotyping or microsatellite sequencing. These types of studies have very low resolution compared to modern techniques, analyzing ~400 genetic markers across the entire human genome vs. today's modern microarrays which can genotype several million genetic variants. The 'candidate-gene' approach was also employed to explore genetic variations in genes hypothesized to be involved in body weight regulation or certain related phenotypes based on the available literature. Genotyping of gene-associated SNPs or genetic sequencing can be utilized for this type of analysis. Due to the often low number of subjects, and the low coverage, early results were rarely replicated [18-24].

As the technology developed, genotyping of genetic variants became inexpensive enough to allow screening on a genome-level, using several million genetic markers. This allowed for a hypothesis-free screening but in turn brought about problems concerning statistical power, as the extremely high number of statistical tests required to analyze this type of data demands strict adjustments of significance thresholds to account for multiple testing. In order to increase statistical power, larger cohorts were needed, and in the last decade we have begun to see the results of ambitious collaborative genotyping-projects.

In 2007, the first results from genome wide association studies (GWAs) for BMI were published. Frayling and colleagues performed genotyping on the Affymetric Genechip 500k Mapping Array Set, containing 490 032 SNPs, in 4 862 individuals, and confirmed their findings in 13 cohorts totaling 29 596 individuals [25]. Scuteri and colleagues performed genotyping on the Affymetrix 500k and 10k arrays in 6 148 individuals and confirmed their findings in 3 205 individuals [26]. Both studies identified single nucleotide polymorphisms within the first intron of the gene encoding what was then termed the "fat mass and obesity associated gene" (FTO), to be associated to a higher BMI. Homozygotes for the risk allele, the A-allele, were observed to weigh on average 3 kilograms more compared to non-risk allele carriers [25]. The effects of this variant were furthermore already observable at age seven [25]. FTO will be further discussed in a later section.

A subsequent GWA-study in 16 876 individuals genotyped on the Affymetrix genechip 500k array, confirmed the original findings on FTO and also indentified common variants in proximity of MC4R to be associated to fat mass, weight and obesity [27]. In 2009, the Genetic Investigation of Anthropometric Traits consortium (GIANT) published their first study on body weight, a meta analysis on about 32 000 individuals using 2.4 million genotyped or imputed SNPs combined with follow-up analysis in 14 cohorts, a

total of about 59 000 individuals. Six new loci were reported in this study to be associated with BMI. These were in the proximity of the genes encoding transmembrane protein 18 (TMEM18), glucosamine-6-phosphate isomerase 2 (GNPDA2), BTB/POZ domain-containing protein KCTD15 (KCTD15) and Neuronal growth factor 1 (NEGR1) and in two gene-dense regions, with strong signals for association in close proximity to the genes: mitochondrial carrier homolog 2 (MTCH2) and SH2B adapter protein 1 (SH2B1) [28]. The associations of FTO and MC4R were confirmed to be strongly associated with BMI in this study. Variants in the proximity of TMEM18 were also observed to have an influence on BMI in the same order of magnitude as MC4R. At the same time, an independent group published a study confirming many of these new findings: the strong association of variants related to FTO, TMEM18 and MC4R and association of SH2B1, NEGR1 and KCTD15; as well as identifying other new loci not observed in the study by the GIANT-consortium [29]. Most of the genes related to the loci identified in these two studies are expressed in the brain and involved in neuronal function. Therefore, disturbances/alterations in the central regulation of feeding were suggested as major factors in the development of obesity, playing a larger role than peripheral metabolic factors.

The GIANT consortium further expanded their GWAs meta-analysis to include 34 cohorts totaling nearly a quarter of a million individuals (n = 249 796) [30]. All ten previously associated loci were confirmed in this study and an additional 22 were identified. However, the combined effects of the 32 loci only accounted for 1.45% of the inter-individual variability in BMI or about 2-4% of the genetic variance assuming a heritability of 40-70%. In addition, according to estimates from the GIANT consortium, more than 250 loci of similar effect size remain to be identified [30].

## The fat mass and obesity associated protein, FTO

When genetic variants within intron 1 of FTO were discovered to be associated with obesity in 2007, it was a fairly uncharacterized protein, apart from publications describing a mutant mouse-line carrying a large 1.6 Mb deletion containing several genes, among them *Fto* [31-32]. Mice homozygous for this mutation die during gestation and their embryos display severely altered brain- and craniofacial morphogenesis [33]. Heterozygous mice survived but were observed to have fused digits on their anterior limbs, giving the name of the mouse mutant “Fused toes” or “Ft”, and also displayed altered development of the thymus [32]. The *Fto* gene was subsequently cloned and found to encode a 58 kDa protein widely expressed in the adult mouse [34]. *Fto* was, serendipitously enough, referred to as “Fatso” in this paper due to the large size of the *Fto* gene, it was estimated at the time to be > 280 000 base pairs. No *Fto*-driven effects on body weight or adiposity were yet to be

published for several years. After the association with obesity and body weight was established, the *Fto* gene-product was officially termed "the fat mass and obesity associated protein" but is most commonly referred to as "FTO". FTO has since generated substantial interest. At the time of the licentiate defense for the author of this thesis in the spring of 2012, a PubMed search returned 664 hits for "FTO". Today (July 2013), that number has increased to 1000, of which 906 were published after 2007, when the first association to obesity was published by Frayling and associates [25]. While several replications of the original association have been published (reviewed by Jacobson and colleagues [35]), relatively few papers have been published on the molecular function of FTO.

Phylogenetic analyses found FTO to be highly conserved and present in vertebrates, missing in invertebrates, fungi and green plants but present in algae [36-37]. Bioinformatic analysis revealed sequence similarity to the Fe(II)-dependent *E-coli* de-alkylating enzyme AlkB. Substrate screening also showed a preferential de-methylating activity towards single-stranded 3-methylthymine ($m^3T$) and 3-methyluracil ($m^3U$) containing oligonucleotides suggesting a role for FTO in DNA and RNA de-methylation [38-39]. Expression analysis using *in situ* hybridization, immunohistochemistry and quantitative real-time polymerase chain reaction (qRT-PCR) on mice brain sections revealed high expression of FTO in the hypothalamus, particularly in feeding regulatory sites: the arcuate nucleus of the hypothalamus, ventromedial hypothalamus and the supraoptic nuclei [36, 39]. Double labeling also showed FTO to be predominantly expressed in neurons [36]. Gene expression studies on food restricted and starved rats using qRT-PCR showed hypothalamic FTO to be upregulated in starved and food-restricted rats [36]. However, subsequent expression studies have been inconsistent, reporting FTO to be both up- and down regulated, as well as not regulated at all, in response to food restriction and fasting [39-43].

Studies in transgenic mice have generated some interesting leads into the molecular function of FTO. Mice homozygous for a loss-of-function mutation in FTO display a leaner phenotype but also severe growth retardation and malformations as well as a higher metabolic rate. These mice also have an increased postnatal mortality rate [44]. A study on mice carrying a point mutation which inactivates FTO, I367F, reported a lean phenotype with reduced fat mass and a higher metabolic rate as well, despite normal food intake [45]. A case-report of a rare FTO-mutation in humans has also been published, describing individuals homozygous for a G to A variation at position 947, predicting an arginine to glutamine substitution at position 316, which was suggested to negatively affect enzymatic activity. These patients presented with severe developmental disorders, growth retardation, functional brain deficits as well as early lethality (1-30 months) [46]. Studies on transgenic mice carrying one or two additional copies of the FTO gene showed a dose dependant effect on weight gain by FTO expression [47].

This study, combined with observations in humans heterozygous for the rs9939609 obesity-associated SNP, that the risk allele is preferably transcribed [48-49], indicate the obesity risk allele to be gain-of-function, promoting weight gain through increased expression of FTO.

The crystal structure of FTO was published in 2010 and confirmed the similarity of FTO to AlkB and its human homologs ABH1 and ABH3. The active site of FTO was also found to be structurally hindered from interacting with double stranded DNA [50]. Consequently, the substrate for FTO is single stranded DNA, or perhaps more preferentially RNA. The role of FTO as a RNA-demethylase was further strengthened by immunohistochemical studies showing FTO to co-localize with nuclear speckle-specific markers SC35 [51] and MALAT1 [52]. The nuclear speckles are interchromatic structures enriched in the pre-mRNA splicing protein complex, the spliceosome.

FTO was subsequently shown to have a higher de-methylating activity for single stranded oligonucleotides and N6-methyladenine ($m^6A$) was identified as a major substrate for FTO [51]. This finding has sparked intense interest in the functional implications of $m^6A$ modifications. However, as $m^6A$ does not alter complementary sequences, analysis of $m^6A$ prevalence and distribution in the transcriptome was not possible using oligo hybridization technology. The development of $m^6A$-specific antibodies has however enabled immunoprecipitation protocols to be adapted for subsequent massive parallel sequencing platforms. Utilizing this technology, recent reports have revealed $m^6A$ to be a widespread RNA-specific modification; it has been detected in > 7 600 gene encoding transcripts and > 300 noncoding RNAs [53-54]. N6-methyladenosine sites were also found to be highly enriched near stop codons and in 3' untranslated regions. Together with the evolutionary conservation of $m^6A$ distribution patterns between mouse and humans, this observation indicates a specific regulatory role for $m^6A$ [53-54] that is dynamically regulated by FTO.

More recently, the obesity-associated SNP rs9939609 in intron one of FTO has been linked to both postprandial suppression of hunger circulating levels of the orexigenic peptide hormone acyl-ghrelin [49]. Homozygous carriers of the rs9939609 obesity risk allele (A) display impaired postprandial appetite suppression and higher acyl-ghrelin levels compared to C-homozygotes. Functional magnetic resonance imaging also revealed differential responses to food image cues in brain regions known to regulate appetite, reward and motivation such as the hypothalamus, insula, nucleus accumbens and orbitofrontal cortex [49]. FTO was also recently observed to influence dopamine signaling in dopaminergic neurons of the mouse midbrain. $FTO^{-/-}$ mice displayed blunted electrophysiological and behavioral responses to cocaine and the selective dopamine receptor $D_2$- and $D_3$-agonist quinpirole [55]. FTO was also observed to interact with aminoacyl-tRNA

synthesases, the enzymes catalyzing the tethering of amino acids to their corresponding transfer RNAs, a process known as tRNA-charging [56].

## TMEM18, transmembrane protein 18

Similar to FTO, TMEM18 was also relatively uncharacterized when it was originally found to be associated with obesity in 2009 [28]. It was first annotated as a transmembrane protein in the TMbase database of transmembrane proteins in 1993 [57]. In 2008, TMEM18 was identified from a library of tumor cDNA as a positive modulator of cell migration, promoting the migratory function of neural stem cells, specifically toward glioblastoma cells. TMEM18 was also found to localize to the nuclear membrane [58]. A recent paper also observed TMEM18 to bind DNA in a sequence-specific manner and, in effect, sequester DNA close to the nuclear membrane. A role in chromatin organizing or gene silencing was suggested based on this observation [59].

The TMEM18 gene encodes a 140 amino acid protein with three transmembrane alpha-helices with a nuclear localization signal at the C-terminal [60]. Expression profiling has suggested TMEM18 to be ubiquitously expressed with some differences in expression levels between tissues. TMEM18 is strongly expressed in feeding regulatory regions of the brain such as the brainstem and hypothalamus [28, 60].

More consistent effects of the TMEM18 risk-allele have been observed in child cohorts when compared with results from studies on adult populations. A study on 4923 adults from northern Sweden was unable to detect an association of a TMEM18-related SNP, rs6548238, with obesity, type 2 diabetes or measures of body composition and adipose distribution [61]. Another study on a cohort of 6013 Chinese type 2 diabetes patients, 1087 healthy adolescents and 605 healthy adults was also unable to detect an association of the TMEM18-related SNP rs7561317 to type 2 diabetes, waist circumference and waist-to-hip ratio. There was a trend toward an association with BMI but this did not reach statistical significance [62]. However, in a Japanese adult cohort of 1129 obese and 1736 normal weight controls, the TMEM18-associated SNPs, rs2867125, rs6548238, rs4854344 and rs7561317 were all observed to be associated with obesity [63].

Results from replication studies in child cohorts have been more consistent in showing the effects of TMEM18-related SNPs to obesity. SNPs in proximity of TMEM18 were observed to confer the strongest effect on pediatric BMI out of 25 obesity-associated variants studied in a cohort of 6078 children of European descent [64]. A similar effect was observed in a cohort of 2042 children and adolescents from four European countries when the effects of 17 variants were studied. The strongest effect size for BMI and sum of skin folds observed in this study was for the near-TMEM18 SNP

rs6548238 [65]. Results from our own group also showed an association of rs6548238 and rs7561317 with obesity in a case control cohort study of 1027 children from the Stockholm area, but no association with anthropomorphic traits was detected in this study [60].

## MAP2K5

Variants within and in the proximity of the immediate downstream region of the gene encoding MAP2K5 on chromosome 15 were identified in meta-analysis by the GIANT consortium in 2010 [30]. The strongest association in this locus was observed for the rs2241423 SNP, an A to G substitution within the last intron of MAP2K5 and approximately 30kb upstream from the gene encoding SKI family transcriptional co-repressor 1 (SKOR1), a loci that has been previously observed to be associated with restless legs syndrome [66]. A subsequent analysis in the cohorts containing children that were included in the meta-analysis revealed directionally consistent effects of rs2241423 on body mass [30]. Variations in the gene encoding the dual specificity mitogen-activated protein kinase kinase 5 (MAP2K5) were subsequently identified as risk factors for obesity in a meta-analysis of data from five cohorts of Chinese, Malay and Indian descent, with a total study population of 10 482 [67].

MAP2K5 has been extensively studied as it is a component of the mitogen activated protein kinase (MAPK) intracellular signaling pathways. According to the Kyoto Encyclopedia of Genes and Genomes (KEGG) (www.genome.jp/kegg), an online resource for understanding high level functions of biological systems, MAP2K5 is a component of the cellular response to extracellular growth factors such as brain derived neurotrophic factor (BDNF), nerve growth factor, insulin-like growth factor 2 [68], granulocyte colony-stimulating factor [69] as well as epidermal growth factor [70]. This pathway also responds to shear stress [71] and osmotic stress (KEGG). MAP2K5 is a 20 exon gene spanning about 280 000 bp. It encodes two main splice variants: MEKα, which is expressed in the liver and the brain to a higher degree [72] where it acts to phosphorylate and activate extracellular signal regulated kinase 5 (ERK5) which in turn activates transcription factors also via phosphorylation leading to cell differentiation and proliferation [73]; MEKβ, on the other hand, is ubiquitously expressed but does not activate ERK5. MEKβ has instead been suggested to act as an inhibitor of ERK5-activation via competitive binding, and to inhibit cell growth in terminally differentiated cell types.

# STK33

Genetic variants within, and in the proximity of, serine/threonine-protein kinase 33 (STK33) were found to be associated with body weight by the GIANT consortium in 2010 [30]. A single nucleotide polymorphism (SNP), rs4929949, located within intron 1 of STK33, produced the strongest signal in tests for association with body mass, but several other SNPs spanning a locus of ~200 000 bp, including the entire STK33 gene as well as the proximal upstream region, were also observed to be strongly associated [30]. Rs4929949 is a high-frequency SNP with a reported minor allele frequency of about 46% (http://www.1000genomes.org), which makes it suitable for replication in smaller independent cohorts. STK33 is located in a gene-rich region on chromosome 11p15.4 in close proximity to the genes tripartite motif-containing protein 66 (TRIM66), 60S ribosomal protein L27a (RPL27A) and suppressor of tumorigenicity 5 (ST5), which are all located within 200 000 bp of rs4929949 (Figure 1). Rs4929949 is also located ~500 000 bp downstream from the gene encoding TUB (Tubby protein homolog), which has been linked to body weight and obesity in mouse studies [74] as well as early genetic studies [75-76]. As of yet, no functional data exists linking any of the genes in closer proximity of rs4929949, or rs4929949 itself, to body weight.

STK33 is a relatively unstudied gene but has received some attention due to its potential involvement in GTPase KRas (KRAS) driven cancers. The potential role of STK33 as a target for pharmacological intervention has also led to the development of pharmacological STK33-inhibitors. STK33 was first discovered and classified as a serine/threonine-protein kinase putatively related to the $Ca^{2+}$/calmodulin-dependent kinase-family (CAMK) in 2001 [77]. It was later observed to preferentially be expressed in testes and lung [78] and to target the cytoskeletal protein vimentin for phosphorylation [79]. Although the normal functional role of STK33 has yet to be determined, a synthetic lethality RNA interference-screen identified a dependency between STK33 and the KRAS oncogene [80]. KRAS is one of the most frequently activated oncogenes, with mutated forms found in about 17-25% of tumor cells [81]. Despite this high prevalence pharmacological means to inhibit KRAS have yet to emerge. Small molecule inhibition of STK33 *in vitro* did not show sufficient effect on cancer cell-viability, leading the authors to speculate on an interaction between mutant KRAS and STK33 independent of its kinase activity [82].

# Aims

## Paper I

In Paper I, we aimed to more precisely determine the association of TMEM18 with obesity related traits in a cohort of Greek children. Anthropomorphic data on body weight and adiposity, as well as dietary intake data, such as energy and macronutrient intake as determined by 24-h recall interviews was available. We genotyped two TMEM18-related SNPs previously observed to be associated with BMI: rs6548238 and rs4854344. We also tested for correlation of TMEM18 expression to food intake, food preference and body weight in rat brain tissues involved in different aspects of food intake such as the hypothalamus and components of the mesolimbic and mesocortical pathways: the amygdala, nucleus accumbens and hippocampus; as well as the prefrontal cortex.

## Paper II & III

In papers II and III we aimed to examine the association of MAP2K5 and STK33-related SNPs to body mass in children and adolescents. For this purpose we utilized two unique cohorts: a case control cohort of obese Swedish children recruited at the pediatric department at the Karolinska hospital in Huddinge which were compared with a control population of adolescents recruited from high schools in the nearby region; and a cross-sectional cohort of about 2200 Greek children and adolescents recruited from the vicinity of Athens, Greece, as part of the Healthy Growth Study [83-84] conducted by the department of Nutrition and Dietetics, University of Athens, Greece. We also aimed to better determine the association of STK33- and MAP2K5-related variants with obesity by examining BMI-related traits such as body adiposity, dietary intake and metabolic traits such as insulin resistance and serum lipids.

## Paper IV & V

In papers IV and V we aimed to explore more deeply the obesity associated FTO and TMEM18 loci through genetic sequencing. Despite the immense

technological advances over the last two decades, genetic sequencing is still a costly enterprise. To overcome this we employed a novel strategy: sequencing on the massive parallel sequencing system SOLiD (Sequencing by Oligonucleotide Ligation and Detection) on pooled DNA samples. This sequencing approach was performed on a case control cohort of 524 severely obese Swedish children from the Stockholm area and a control population of 527 Swedish adolescents from neighboring high schools. Through this approach we were able to generate genetic variation maps of the obesity associated FTO- and TMEM18-loci. Polymorphism frequencies were compared between the obese and control population samples to identify novel SNPs associated with obesity.

## Paper VI

Paper VI aimed to explore the function of the FTO gene in regards to its molecular context. To this end, we utilized data from two co-expression and protein-protein interaction databases, as well as the "functional coupling" prediction database FunCoup. We identified nine candidate genes as functionally coupled to FTO, the expression of which were characterized with qRT-PCR along with the expression of FTO in an animal model known to affect hypothalamic FTO expression.

# Materials and methods

## Huddinge cohort

The Huddinge cohort consists of obese children and adolescents recruited at the National Childhood Obesity Centre at the Karolinska University Hospital, Huddinge, Sweden. A control population of normal weight Swedish adolescents aged 15–20 years were recruited from 17 upper secondary schools around Stockholm and matched to the obese group with respect to ethnicity and socioeconomic status. BMI was calculated from height and weight whereas body mass index standard deviation score (BMISDS) was calculated from weight and height standardized for age and gender. Subjects with overweight/obesity or chronic diseases were excluded from the group of normal weight adolescents, and subjects with type 2 diabetes were excluded from the obese group. Informed consent was provided by all participants or by their legal guardians.

## Greek children and adolescents – The Healthy Growth Study.

The cohort of Greek children comprised 2658 schoolchildren, attending the 5$^{th}$ and 6$^{th}$ grades of primary schools. This cohort was part of the 'Healthy Growth Study', a large-scale cross-sectional epidemiological study initiated in May 2007 [83-84]. Our access to this cohort was granted as part of collaboration with the department of pediatrics at the Harokopio University in Athens, Greece. An extended letter explaining the aims of the current study and a consent form were provided to each parent who had a child in one of the primary schools participating in the study. Those parents who agreed to participate in the study gave their informed consent by signing the consent form, and provided their contact details. Bodyweight and height were measured in all study participants using standard procedures and equipment. Body weight was measured to the nearest 10 g and height was measured to the nearest 0.1 cm in standing position. BMI z-score was calculated relative to the International Obesity Task Force (IOTF) definitions [85]. Waist circumference was measured to the nearest 0.1 cm with the use of a non-elastic tape (Hoechstmass, Sulzback, Germany) around the trunk, at the level of umbilicus midway between the lower rib margin and the iliac crest, and with the

subject at a standing position. Dietary intake data were obtained by trained dieticians and nutritionists conducting morning interviews with the children at their respective schools, for two consecutive weekdays and one weekend day, using the 24-h recall technique. All study participants were asked to describe the type and amount of different foods, as well as all beverages consumed during the previous day, provided that it was a usual day according to the participant's perception. To improve the accuracy of food descriptions, standard household measures (cups, table-spoons, etc) and food models were used to define amounts when appropriate. At the end of each interview, the interviewers, who were dieticians rigorously trained to minimize interviewer effect, reviewed the collected food intake data with the respondent in order to clarify entries, servings and possible forgotten foods. The ratio of reported energy intake and predicted basal metabolic rate was used to asses underreporting of calorie intake. Basal metabolic rate was estimated according to Schofield equations, [86] taking into account age, sex, and body weight and with cut-off limits developed by Goldberg and colleagues [87]. Food intake data were analyzed using the Nutritionist V diet analysis software (version 2.1, 1999, First Databank, San Bruno, CA, USA), which was extensively amended to include traditional Greek recipes, as described in Food Composition Tables of Greek Cooked Foods and Dishes [88]. Furthermore, the database was updated with nutritional information of processed foods provided by independent research institutes, food companies and fast-food chains. DNA for genotyping was available for 2352 subjects (1311 girls and1064 boys). Approval of the consent procedure, and to conduct the study, was granted by the Greek Ministry of National Education and the Ethical Committee of Harokopio, University of Athens.

## Genotyping

Genotyping was performed with pre-designed TaqMan single-nucleotide polymorphism genotyping assays (Applied Biosystems, Foster City, USA) and an ABI7900 genetic analyzer with SDS 2.2 software at the Uppsala Genome Center. This method uses two allele-specific oligonucleotide probes in each assay which hybridize to the allele being studied. Each probe is covalently attached to a fluorophore and quencher molecule. The quencher molecule inhibits detection of the fluorophore when it is connected to the probe. Two primers surrounding the probe then lead to the probe being displaced and cleaved during a polymerase chain reaction, enabling detection of the fluorophore. Fluorescence is then measured in each cycle of the PCR program using a two-filter spectrophotometer. For our studies, this procedure was performed on 384-well plates, each well corresponding to one unique individual.

### Statistical analysis

Statistical analysis was performed with the PLINK software (http://pngu.mgh.harvard.edu/purcell/plink/). Test for deviation from Hardy–Weinberg equilibrium was performed using the Pearson's exact test (using 1 d.f). Association with obesity in the Swedish Cohort was analyzed with logistic regression. Associations with phenotypes were analyzed with linear regression in the Greek cohort, assuming an additive model. Quantitatively skewed variables were normalized by logarithmic transformation before analysis. The models were adjusted for age, gender, pubertal development and BMI z-score when needed.

## Next generation sequencing, SOLiD

The gold standard for genetic sequencing has, until recently, been the Sanger sequencing method developed by British biochemist and two-time Nobel price recipient, Frederick Sanger and colleagues [89]. A technical variant of the Sanger sequencing method using dye-terminators is in fact still the main technique used for small-scale genetic sequencing. This technique is however limited when applied to larger scale projects for economic as well as logistic reasons. The cost-per-base makes large scale projects such as whole-genome sequencing incredibly expensive and too labor-intensive to be practically useful. Sanger sequencing is however a valid approach for targeted genomic resequencing of a gene or region of interest in a small number of subjects but quickly becomes unfeasible for larger genes and genetic regions or for doing population genetics in larger cohorts.

The last decade has seen the emergence of a number of new sequencing techniques, termed next-generation sequencing (NGS), such as the Roche/454, Illumina/Solexa and the Applied biosystems' SOLiD (Sequencing by Oligonucleotide Ligation Detection) platforms. These techniques differ somewhat in their technical and biochemical approaches and sample preparation protocols but are similar in the aspect that they simultaneously sequence millions of substrate molecules in parallel, in a flow-cell setup [90-92]. For our targeted resequencing projects (Papers IV and V) we utilized the SOLiD sequencing platform in collaboration with the Uppsala Genome Center. With this method the DNA used as template for the sequencing reaction is sheared using sound waves in a liquid medium. This creates microscopic bubbles which then collapse, producing turbulence that denatures the DNA molecules. This process can be finely tuned to produce DNA fragments of suitable length with a relatively small degree of variation. For SOLiD sequencing a template length of 35-50 base pairs is commonly utilized. These short DNA molecules are then used as template to produce clonal bead populations, i.e. microscopic magnetic beads fused with universal adapter-

oligonucleotides fused to copies of a single species of DNA template fragment. This is achieved through an emulsion PCR where reagents are sequestered in microscopic lipid droplets that serve as reaction chambers for the PCR reaction. The beads are then deposited unto glass slides and sequencing reactions are performed in a flow cell.

SOLiD sequencing differs from other sequencing techniques in that it utilizes a ligase reaction rather than the DNA polymerase reaction. Fluorescently marked interrogation nucleotides specific for two bases in sequence are ligated complementary to the bead-fused template fragments in each reaction 'round'. Fluorescence is then measured before the fluorescent molecule is enzymatically cleaved from the oligonucleotide and washed from the reaction chamber. As the beads are deposited individually and given sufficient space on the plate, fluorescence is measured from each individual bead simultaneously in parallel in each round of ligation. Ligation is then repeated for several rounds and this can be said to constitute one 'cycle' of the sequencing reaction. In the next cycle the universal primer is replaced by a one-base offset primer, i.e. one base shorter (n-1), and the sequencing rounds are repeated. Due to the structure of the interrogation oligonucleotides a total of five cycles are required to generate a full complementary sequence of the interrogated template fragment and each base is interrogated twice as a result.

## Targeted resequencing pooling strategy

Despite the high efficiency of next generation sequencing and the low cost-per-base, the relatively high cost of reagents and library preparation still limit the application of these techniques. For population genetics projects such as ours (Paper IV and V) where we aimed to investigate and identify new genetic variants in large regions in a cohort of about 1000 children and adolescents, individual sequencing would be too costly. We instead utilized a pooled-DNA strategy using 6 template mixtures of purified DNA from about 150-200 individuals. We then enriched for the genetic regions we wanted to investigate by using traditional PCR on these template mixes with a high-fidelity DNA polymerase enzyme. A high-fidelity enzyme was used to reduce the risk of errors in DNA replication which is more common when using a less expensive polymerase enzyme. Enrichment was performed by amplification of overlapping fragments of about 2000 base pairs. The size of PCR products were then validated with electrophoresis on 2% agarose gels before purification. Fragments were then pooled in equal amounts and delivered to the Uppsala Genome Center for library preparation and sequencing. As we utilized the Huddinge case-control cohort, we delivered one combined pool for obese subjects and one for the control.

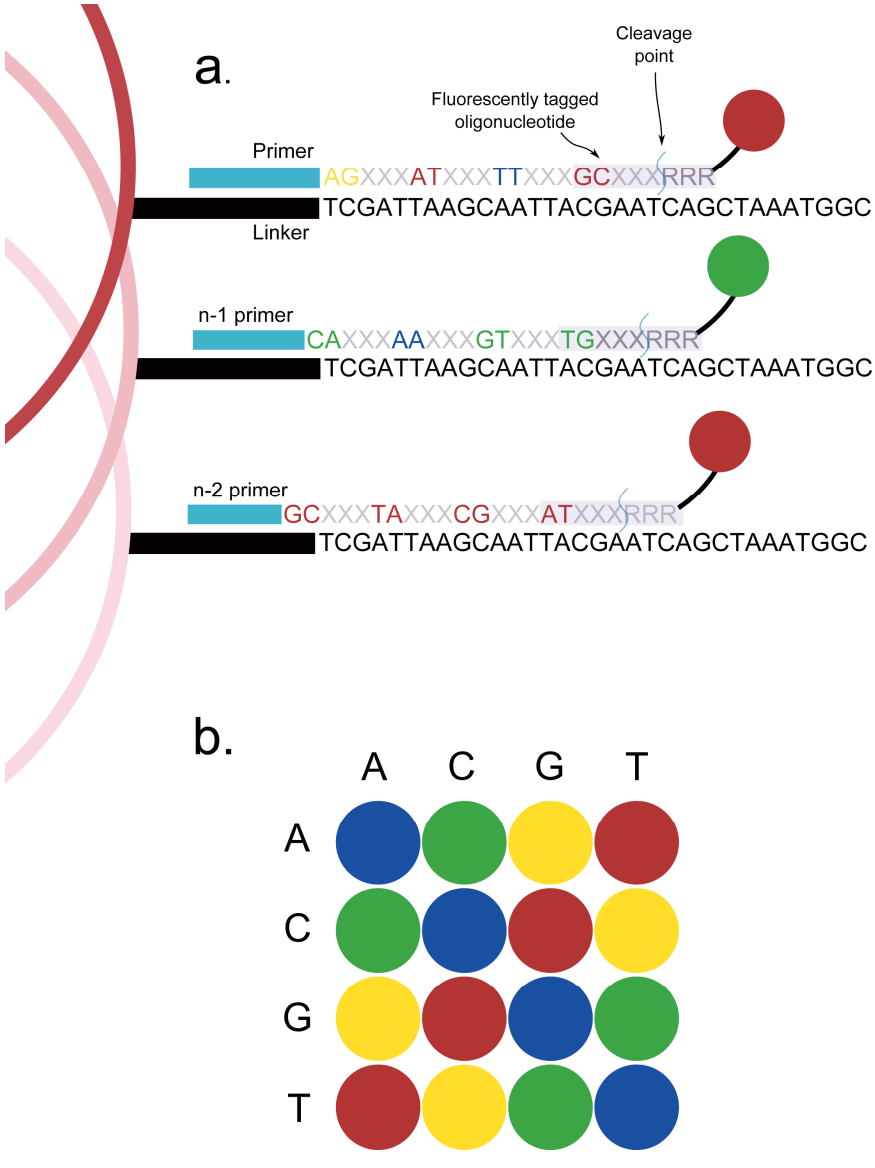


***Figure 1.*** **Schematic overview of the SOLiD sequencing oligonucleotide ligation process.** (a) Each interrogation oligonucleotide consists of eight bases connected to a fluorophore. The fluorophore corresponds to two bases in sequence according to the schematic (b). Each cycle of sequencing consists of several rounds of ligation. In the next cycle a new primer offset by one base is used (n-1 primer). Due to the structure of the oligonucleotides, a total of five cycles is required for a full sequencing and each base is interrogated twice.

## Analysis of sequencing data

The SOLiD sequencing method produces a 'color-space' representation of the template DNA sequence. As 16 different two-base sequences are possible (TA, AT, GC, CG, etc.) but only four fluorophores are used (red, green, yellow and blue) each color represents four unique two-base combinations (Figure 1b). The template sequence is then computationally inferred from the color-space sequence. Once collected and inferred, the total sequence data generated from one SOLiD sequencing run can amount to about 40-50 gigabases. This massive amount of data can then be assembled by computational methods for *de novo* sequencing, e.g. for sequencing of a newly identified pathogen or species. In our resequencing projects, sequences were aligned to a reference sequence (NCBI36/hg18) for the gene or region being investigated.

A SNP is identified in the color-space sequence as a 'valid adjacent mismatch'. As the interrogation oligonucleotides are specific for two-base pairs in sequence this entails that a variation in a single base will produce a change in two subsequent fluorescence-signals and a corresponding change in the color-space sequence. For a SNP, only three resulting changes are possible and considered valid: e.g. a substitution of the center base in CAT (green-red) to CGT (red-green), CCT (blue-yellow) and CTT (yellow-blue). All other color-space sequences at this position should then not be possible and can be filtered out as an invalid read. Another factor to consider is the coverage, i.e. the number of times each genetic position is read during sequencing. In our analysis (Papers IV and V) we observed a very high coverage averaging at about 18 000 reads per base for the ~400 000 bp FTO gene (Paper IV) and about 30 000 - 60 000 reads per base for the ~70 000 bp region associated with TMEM18 analyzed in paper V. However, we also observed positions of extremely high coverage ranging up to several hundred thousand reads. These high-coverage peaks preferentially occurred at amplicon borders near primer positions from the originally sequenced amplicons. This was considered a methodological artifact produced by the biochemical methods employed during library preparation, e.g. preferential shearing of amplicons at certain positions or near amplicon ends due to biophysical properties of the specific sequence. We therefore removed these regions from the analyses. To avoid similar effects of pileup of identical reads, we utilized the UVAM ('unique valid adjacent mismatches') score developed by Zaboli and colleagues [93-94]. The UVAM score takes into account the number of uniquely placed reads at a position with a valid mismatch. For reads with a 50 bp length the mismatch can theoretically be read by sequences starting at 50 unique positions. A UVAM score of 100 means that the position has been read by 50 sequences with unique starting points on each DNA-strand. Cut-off UVAM values for valid SNPs were inferred from the distribution of UVAM-scores for known SNPs reported in dbSNP

(www.ncbi.nlm.nih.gov/SNP) and by the 1000 Genomes Project (www.1000genomes.org). It was evident when comparing these distributions that most novel candidate SNPs identified through SOLiD sequencing were false positives, however, false positive SNPs were observed to have a lower UVAM score and using a binomial hypergeometric distribution to model the probability of false positive SNPs to have a certain UVAM score, cut-off UVAM values corresponding to a false discovery rate of 1% could be generated.

A SNP is identified in the color-space sequence as a 'valid adjacent mismatch'. As the interrogation oligonucleotides are specific for two-base pairs in sequence this entails that a variation in a single base will produce a change in two subsequent fluorescence-signals and a corresponding change in the color-space sequence. For a SNP, only three resulting changes are possible and considered valid: e.g. a substitution of the center base in CAT (green-red) to CGT (red-green), CCT (blue-yellow) and CTT (yellow-blue). All other color-space sequences at this position should then not be possible and can be filtered out as an invalid read. Another factor to consider is the coverage, i.e. the number of times each genetic position is read during sequencing. In our analysis (Papers IV and V) we observed a very high coverage averaging at about 18 000 reads per base for the ~400 000 bp FTO gene (Paper IV) and about 30 000 - 60 000 reads per base for the ~70 000 bp region associated with TMEM18 analyzed in paper V. However, we also observed positions of extremely high coverage ranging up to several hundred thousand reads. These high-coverage peaks preferentially occurred at amplicon borders near primer positions from the originally sequenced amplicons. This was considered a methodological artifact produced by the biochemical methods employed during library preparation, e.g. preferential shearing of amplicons at certain positions or near amplicon ends due to biophysical properties of the specific sequence. We therefore removed these regions from the analyses. To avoid similar effects of pileup of identical reads, we utilized the UVAM ('unique valid adjacent mismatches') score developed by Zaboli and colleagues [93-94]. The UVAM score takes into account the number of uniquely placed reads at a position with a valid mismatch. For reads with a 50 bp length the mismatch can theoretically be read by sequences starting at 50 unique positions. A UVAM score of 100 means that the position has been read by 50 sequences with unique starting points on each DNA-strand. Cut-off UVAM values for valid SNPs were inferred from the distribution of UVAM-scores for known SNPs reported in dbSNP (www.ncbi.nlm.nih.gov/SNP) and by the 1000 Genomes Project (www.1000genomes.org). It was evident when comparing these distributions that most novel candidate SNPs identified through SOLiD sequencing were false positives, however, false positive SNPs were observed to have a lower UVAM score and using a binomial hypergeometric distribution to model the probability of false positive SNPs to have a certain UVAM score, cut-off

UVAM values corresponding to a false discovery rate of 1% could be generated.

In addition, for a SNP to be considered valid it had to be detected in both template DNA pools and the number of reads were required to be distributed equally between the DNA-strands. Identified valid SNPs were then tested for association with obesity by using Fischer's exact test based on the allele frequencies for each SNP in the obese and control populations.

## Mouse feeding models

In the gene expression analysis, we utilized the tissue collected in the experiments described previously [41, 95]. Briefly, male C57BL/6J mice (Scanbur, Sweden) were housed individually or in groups of two in macrolon cages with LD 12:12 (lights on at 07.00). Animals were twelve weeks old at the beginning of the experiment. Water and standard chow (Lactamin, Sweden) were available *ad libitum* unless specified otherwise. All animal procedures were approved by the Uppsala Animal Ethical Committee (ID: C228/7 & C262/7) and followed the guidelines of Swedish legislation on animal experimentation (Animal Welfare Act SFS1998:56) and European Union legislation (Convention ETS123 and Directive 86/609/EEC).

### Food choice model

Thirty-six outbred male Wistar rats (Scanbur BK AB, Sollentuna, Sweden) were housed in standard macrolon cages at constant temperature (22 ± 1°C) and humidity (50 ± 5%). Rats were ten weeks old at the start of the study, and housed individually for 8 days during the food preference paradigm. Water and standard food chow (R36, Lactamin, Lidköping, Sweden) was supplied *ad libitum* during the entire experiment except for the 5-day food preference test, during which the animals had free access to three diets: a palatable high-fat diet and two reference diets with high amounts of either casein (high-protein diet) or maize starch (high-carbohydrate diet). Compositions of the different diets were described by Alsiö and colleagues [95]. The food was provided in bowls and weighed each day to measure the ingested amount. This model allowed the determination of the following outcomes: body weight at endpoint, total food intake during palatable diet presentation, high-fat diet preference (high-fat diet consumption divided by total intake) and consumption of standard chow. Subsequent to the 5-day food choice, the animals were kept on standard chow for 12 days; this 'washout' period was inserted to allow effects of the food choice diets on gene expression to subside. Animals were then killed by decapitation. All samples were collected immediately following decapitation. Dissection of the brain is described by

Alsiö and colleagues [95]. Samples were then immersed in RNAlater (Applied Biosystems/Ambion, Austin, TX, USA) and stored at room temperature for 2 h before being stored at -20°C until preparation.

### Sixteen-hr food deprivation

Chow was removed just before the onset of darkness and mice were sacrificed 16 hours later. Control mice had *ad libitum* access to chow. Each group contained eight animals. Hypothalami were dissected and immersed in RNAlater. A larger group of *ad libitum* chow fed animals was used to better study co-regulation of FTO with other genes. Ten group-housed mice had *ad libitum* access to chow for 48 hours before sacrifice and dissection of hypothalami.

## Generation of cDNA

Animal tissues were kept in RNAlater at room temperature for two hours and then stored at -80°C until further processing. Tissues were homogenized by sonification (Branson sonifier B15) and RNA was purified from the samples using the TRIzol method (Sigma-Aldrich, Sweden) [96]. Samples were treated with Dnase I (Roche Diagnostics, Scandinavia) to remove residual DNA contamination. DNA contamination after Dnase I treatment was checked with PCR. Complementary DNA, cDNA, was then generated using reverse MLV reverse transcriptase (Invitrogen, Sweden) according to the manufacturer's specifications.

## Quantitative real time PCR

PCR reactions were run in a total volume of 20 µl using Taq polymerase kits (Biotools, Madrid, Spain). Each reaction was performed in duplicate according to the manufacturer's specifications, and contained 75 mM Tris/HCl, 50 mM KCl and 20 mM $(NH_4)_2SO_4$, 4mM $MgCl_2$, 0.25mM dNTPs, 1:20 DMSO, 20 mU/ml Taq polymerase, 50 mM forward and reverse primer and 1:4 SYBR-green (Invitrogen). Reactions were run on iCycler temperature cyclers, and fluorescence was measured using MyiQ single color real-time PCR detection system. Data were analyzed using iQ5 software (BioRad, Sundbyberg, Sweden). Primer temperatures were optimized for specificity using temperature gradients and through analyzing PCR product melting curves.

## Statistical analysis

Student's t-test was used to test the difference in gene expression between *ad libitum* chow fed, and food restricted mice. Linear regression was used to determine the relationship between expression of FTO and candidate genes. FTO was considered to correlate to candidate genes if the slope of the trend-line significantly deviated from zero. For all tests nominal p-values $< 0.05$ were considered significant. The statistical analysis was performed with Prism v5.02 (GraphPad Software, San Diego, California, USA, http://www.graphpad.com).

# Results

## Paper I

The major alleles of the near-TMEM18 SNPs rs6548238 and rs4854344 were significantly associated with obesity in the studied cohort of 2 352 Greek children. These results are in line with the previous results on a cohort of children from the Stockholm area [60], as well as with the results from Zhao et al [64] who observed strong associations of three TMEM18-associated SNPs with obesity in a cohort of 6078 children of European descent. We also observed associations of rs6548238 and rs4854344 with anthropomorphic traits: waist circumference, body weight and BMI z-score. These associations were however not detected in the Stockholm cohort. We did not observe any association to dietary energy and macronutrient intake in any of the cohorts.

We also present for the first time an association of body weight in rats with gene expression of TMEM18. This association was neuroanatomically localized to the prefrontal cortex (PFC). The PFC has been proposed to have an important role in integrating behavioral cues from multiple sources in the brain (reviewed by Miller and Cohen [97]). We also investigated the expression of several proteins known to be involved in cortical signaling. Body weight was significantly correlated to expression of GABA receptor subunit α3 (GABRA3), as well as 5-HT2A receptor (HTR2A), α-1B adrenergic receptor (ADRA1B) and GABA B receptor 2 (GABBR2). Regulation of monoaminergic- and GABA-ergic receptor transcripts concomitantly with TMEM18 could prompt further studies for a role on TMEM18 in synaptic plasticity.

## Paper II & III

We were able to observe associations of the MAP2K5 and STK33-realated variants rs2241423 and rs4929949 with obesity in the case-control cohort of Swedish children and adolescents, as well as with body mass in the cross-sectional cohort of Greek children and adolescents. The effects we observed were directionally consistent with previous reports from GWAS [28, 30]. We also performed secondary analyses in the cohort of Greek children ado-

lescents with the aim to better determine the specific obesity-associated effects of these variants. We were able to demonstrate associations of rs4929949 with several measurements of body adiposity. We also observed a trend towards association with fasting insulin that was directionally consistent with previous reports from GWAS [28, 30]. Rs2241423 was observed to be associated with hip circumference in the cohort of Greek children and adolescents.

## Paper IV & V

Through targeted resequencing of the FTO gene utilizing the SOLiD platform on enriched pooled DNA samples from a case control cohort of obese and control children and adolescents, we were able to provide a high resolution variation pattern of the FTO gene. We also observed 19 obesity-associated SNPs within intron one of FTO. In total we sequenced 389 501 bases with a mean coverage of about 18 000 reads per base. 348 557 bases remained after removing pileup-regions near primer positions. The three strongest associated SNPs identified in this study have not previously been associated with obesity. Cross reference with the ENCODE database (ENCyclopedia Of DNA Elements) revealed these SNPs to be located close to potential regulatory sites.

About 70 000 bases across the TMEM18 locus were successfully sequenced with a very high sequencing depth of about 30 000 and 66 000 reads in the obese and control pools, respectively. 41 820 bases were available for variant calling after removing pileup regions near primer positions. A total of 239 SNPs were identified and tested for association with obesity. Due to the large number of tests, the threshold for significance was adjusted, after which 23 obesity-associated SNPs could be observed. Seven obesity-associated SNPs were selected for validation through TaqMan genotyping in the case-control cohort of Swedish children and adolescents as well as the cross sectional cohort of Greek children. Although one SNP within intron 3 of the TMEM18 gene could be observed to be associated with obesity in the Swedish cohort, no association with BMI could be observed in the cohort of Greek Children. More consistent results were observed for genetic variants within the TMEM18 downstream region.

## Paper VI

We confirmed increased levels of FTO mRNA in the hypothalamus of food restricted mice and observed simultaneous up-regulation of four out of the nine genes in the hypothalamus during food deprivation: PFN2, PRKACB,

TRKB and STAT3; as well as down-regulation of BDNF. The genes BDNF, NTRK2 and STAT3 have been previously implicated in the regulation of energy homeostasis. Two of the genes, PRKACB and PFN2, have not previously been implicated in feeding/obesity paradigms. Intriguingly, we observed the expression of FTO to correlate with expression of NTRK2 and BDNF in the hypothalamus of chow-fed mice. We also observed a strong linearity between FTO expression and that of NTRK2 and BDNF. Based on our findings and recent publications of the interaction between FTO and transcription factor CCAAT-enhancer binding protein beta (C/EBPβ), a hypothetical model for the functional coupling between FTO and the BDNF/NTRK2 signalling pathway is presented.

# Perspectives

The genome wide association studies performed for detecting the association of genetic variants to higher BMI and obesity have generated a number of interesting leads into the functional mechanisms regulating body mass. A few of the genes within or in the vicinity of body mass-associated loci have defined functional roles in body mass regulation. The most clearly defined models have been established for the melanocortin 4 receptor, MC4R [27], and proopiomelanocortin, POMC [30]. POMC encodes the precursor to several biologically active agents, among these the anorexigenic peptide hormone α-MSH. These proteins have essential roles in satiety-signaling within the arcuate nucleus of the hypothalamus. Transgenic animal models have demonstrated functional links between SH2B1 and central regulation of energy balance and glucose homeostasis [98]. BDNF has also been observed to produce anorexigenic effects and increase energy expenditure when administered centrally [99-100].

SNPs within intron one of FTO remain the strongest associated genetic variants to higher BMI on a population level, and functional characterization of this gene has come a long way since the original discovery and cloning of the gene in 1994 and 1999, respectively [32, 34]. Case reports and animal studies have shown FTO to affect embryological development and development of the central nervous system. The revelation of the molecular function of FTO as an RNA demethylase makes this one of the more interesting, but nevertheless complex, obesity-associated proteins to study more closely. RNA-methylation is poorly understood from its regulation to its molecular function and consequences.

Some functional aspects of TMEM18 have been characterized but its role in body weight maintenance is as of yet unknown. Its function regarding neuronal migration of neural stem cells indicates that it is related to differentiation and neural development. Animal studies may be informative in exploring the role of TMEM18. At the moment, no knockout studies for TMEM18 in transgenic rodents have been published. Functional assays using transgenic cell-lines may also be informative to study how the TMEM18 affects the response of neuronal cells to various stimuli.

## Opportunities for pharmacological intervention

Although FTO is highly expressed in central regions regulation feeding behavior, it is also expressed in most other tissues of the human body. FTO function is also critical for normal development of the human body as seen by the increased lethality observed in FTO-deficient mice and humans [34, 46]. While pharmacological manipulation of FTO for the specific purpose of treating obesity seems to be unlikely to be a successful approach, recent observations of the interaction of FTO-genotype with circulating acyl-ghrelin and central responses to food cues [49] is of particular interest. These are the first observations of links between FTO and a feeding regulatory system which is relatively well understood. Ghrelin is a circulating peptide hormone released from P/D1 cells within the fundus of the human stomach and from epsilon cells in the pancreas. It functions to stimulate appetite by stimulating ghrelin receptors in the arcuate nucleus of the hypothalamus. The ghrelin-axis is currently under exploration as an intervention point for treatment of obesity, cachexia and gastroparesis [101]. As genotyping technology becomes more inexpensive, individualized obesity treatments based on genotype are likely to emerge in the coming decades. These programs may be able to exploit links like the one between FTO and ghrelin or potential similar links between other body mass-associated loci and feeding regulatory systems.

Similarly to FTO, the role of TMEM18 in body mass regulation is also poorly understood at present. Its function as a nuclear membrane bound protein which sequesters DNA and suppresses transcription [59] also makes this unlikely as a potential drug target for treatment of obesity. Exploring the functional links of TMEM18 to more canonical feeding regulatory paradigms, analogous to recent clinical experiments exploring the functional role of FTO, may lead to the emergence of potential opportunities for intervention, pharmacological or otherwise.

MAP2K5 and STK33 serve as examples of more likely points of intervention. Protein kinases are currently under intense investigation as drug targets in antineoplastic treatment which has led to a large number of experimental and clinical agents having been developed [101-102]. Protein kinases also have unique properties in their protein structure making them suitable for pharmacological intervention. However, protein kinases are commonly linked to more than one cellular function making them difficult to exploit for more specific biological systems. Protein kinases also have common structural elements, particularly in the nucleotide tri-phosphate binding region. This region is commonly exploited for pharmacological intervention making selective targeting very difficult [103-104]. For antineoplastic disease, relatively unspecific protein kinase inhibition is less likely to be a problem as the goal is to slow down cellular metabolism and proliferation [105]. Protein kinase-based intervention with the aim to affect specific biological functions,

such as regulation of feeding behavior and accumulation of body weight, may perhaps pose a greater challenge.

In conclusion, the genetic component of body weight regulation and the heritable risk factors predisposing increased susceptibility for obesity are slowly being elucidated. The great efforts of multi-center collaborations such as the GIANT consortium have greatly helped advance our knowledge of the genetic variations predisposing higher body mass. We now face the challenge of determining the functional implications of these loci. Advances in our understanding of regulatory DNA elements and how they interplay with trait-associated genetic variants are also essential for this understanding. International collaborative efforts such as those by the ENCODE consortium have created invaluable data resources for integration of this data with novel genetic findings [106]. This thesis presents some applicable approaches for determining the effects of body mass-associated loci through genotyping in specialized cohorts that are rich in phenotypic data measurements, targeted re-sequencing, and animal models. As we are now in a very expansive era of technological development in genetic sequencing, we can expect these methods to become less expensive and more efficient in the coming years. Whole genome sequencing in specialized cohorts is one of the more interesting prospects for future projects in population genetics and may help uncover causative genetic variants as well as many of the presumed SNPs/gene regions associated with obesity that so far remain undiscovered.

# Acknowledgements

I would like to thank all my colleagues and friends at HelgiLab and the department of Neuroscience for all the support and assistance during my time as a PhD-student. I would like to specially thank:

Professor **Helgi B. Schiöth**, my supervisor, for giving me the opportunity and privilege to work in his lab. Your guidance and confidence in me has been, and continues to be an inspiration. **Madeleine le Grevés**, my co-supervisor, for all her kind assistance and support. **Robert Fredriksson** for his invaluable technical assistance in all things under the sun. My dear friend **Markus Sällman Almén** for his invaluable mentoring, support, guidance and friendship. **Josefin Jacobsson** and **Johan Alsiö** for their supervision and patience and for their substantial efforts which have laid the groundwork for much of the work in this thesis. **Miguel Xavier** for his work on Paper V. **Pawel Olszewski** for assisting me with writing. **Jonathan Cedernaes** for proofing, expertise and movie nights. **Rohit Chavan** and **Rahul Sawat** for their invaluable assistance and hard work. **Sofie Juhlin, Nathalie Bringeland, Caroline Ingman, Abu Noman Sikder, Robert Olausen, Susann Goran, Dessi Edén, Vaman Tahir, Praveen Kumar, Jasmin Musleh, Rebecca Gutierrez Liljestrand, Colin Chapman** and **Lilia Hedberg** for all their hard work.

My gratitude also goes out to our collaborators at KI, especially **Claude Marcus**. I would also like to extend my gratitude to our collaborators in Greece, **George Moschonis**, **Yannis Manios** and **George P. Chrousos** for their assistance, support and valuable commentary. A great deal of thanks also goes out to **Ulf Gyllensten**, **Adam Ameur**, **Inger Jonasson** and **Joanna Hammer** at the dept. of Immunology, Genetics and Pathology and the Uppsala Genome Center.

My friends in the lab **Maria, Sahar, Karl, Smitha, Christian, Atieh, Emil, Pleunie, Mike, Anica, Samantha, Olga, Sofie, Anders, Linda, Anirudha, Emilie, Emilia,** and **Gaetan.**

I would go crazy if I didn't have a band to play in. My gratitude goes out to my bandmates, **Joel Borg**, **Jonas Eriksson Slove** and **Pontus Wallin** for

their camaraderie, talent and for keeping me sane over the years. This research thing is only until we make it big guys!

My thanks goes out to my Family: my brother **Niels** and my new sister-in-law **Hanna**. **Charlotta** and **Albert** and their wonderful children: my nephew **Elliot,** and nieces **Julia** and **Emilia**. My mother **Anna Rask-Andersen** for all her support and reverence for the academic life in Uppsala. My father **Helge Rask-Andersen** for all of his wonderful support and enthusiasm for research which has been an inspiration to me for as long as I can remember. My extended American family: in San Diego, My rediscovered friends the Hoffmans and especially **Ari** and **Micah Hoffman** for their gracious hospitality and support. In Chicago, min jättesnälla barnvakt **Barbie Markay** for listening, ranting and sending me American sweets.

# References


1. OECD, *Obesity and the Economics of Prevention: Fit not Fat.* 2010.
2. Malnick, S.D. and H. Knobler, *The medical complications of obesity.* QJM, 2006. **99**(9): p. 565-79.
3. Flegal, K.M., et al., *Prevalence and trends in obesity among US adults, 1999-2008.* JAMA, 2010. **303**(3): p. 235-41.
4. Sturm, R., *Increases in clinically severe obesity in the United States, 1986-2000.* Arch Intern Med, 2003. **163**(18): p. 2146-8.
5. Sturm, R., *Increases in morbid obesity in the USA: 2000-2005.* Public Health, 2007. **121**(7): p. 492-6.
6. Branca, F., H. Nikogosian, and T. Lobstein, eds. *The challenge of obesity in the WHO European Region and the strategies for response*. 2007, WHO Library Cataloguing in Publication Data.
7. Sjoberg, A., et al., *Overweight and obesity in a representative sample of schoolchildren - exploring the urban-rural gradient in Sweden.* Obes Rev, 2011. **12**(5): p. 305-14.
8. Olds, T.S., et al., *Trends in the prevalence of childhood overweight and obesity in Australia between 1985 and 2008.* Int J Obes (Lond), 2010. **34**(1): p. 57-66.
9. Ogden, C.L., et al., *Prevalence of obesity and trends in body mass index among US children and adolescents, 1999-2010.* JAMA, 2012. **307**(5): p. 483-90.
10. Sjoberg, A., et al., *Recent anthropometric trends among Swedish school children: evidence for decreasing prevalence of overweight in girls.* Acta Paediatr, 2008. **97**(1): p. 118-23.
11. Peneau, S., et al., *Prevalence of overweight in 6- to 15-year-old children in central/western France from 1996 to 2006: trends toward stabilization.* Int J Obes (Lond), 2009. **33**(4): p. 401-7.
12. Poston, W.S., 2nd and J.P. Foreyt, *Obesity is an environmental issue.* Atherosclerosis, 1999. **146**(2): p. 201-9.

13. Maes, H.H., M.C. Neale, and L.J. Eaves, *Genetic and environmental factors in relative body weight and human adiposity.* Behav Genet, 1997. **27**(4): p. 325-51.
14. Schousboe, K., et al., *Sex differences in heritability of BMI: a comparative study of results from twin studies in eight countries.* Twin Res, 2003. **6**(5): p. 409-21.
15. Montague, C.T., et al., *Congenital leptin deficiency is associated with severe early-onset obesity in humans.* Nature, 1997. **387**(6636): p. 903-8.
16. Farooqi, I.S., et al., *Effects of recombinant leptin therapy in a child with congenital leptin deficiency.* N Engl J Med, 1999. **341**(12): p. 879-84.
17. Farooqi, I.S., et al., *Dominant and recessive inheritance of morbid obesity associated with melanocortin 4 receptor deficiency.* J Clin Invest, 2000. **106**(2): p. 271-9.
18. Hager, J., et al., *A genome-wide scan for human obesity genes reveals a major susceptibility locus on chromosome 10.* Nat Genet, 1998. **20**(3): p. 304-8.
19. Ohman, M., et al., *Genome-wide scan of obesity in Finnish sibpairs reveals linkage to chromosome Xq24.* J Clin Endocrinol Metab, 2000. **85**(9): p. 3183-90.
20. Zhu, X., et al., *A genome-wide scan for obesity in African-Americans.* Diabetes, 2002. **51**(2): p. 541-4.
21. Loos, R.J. and C. Bouchard, *Obesity--is it a genetic disorder?* J Intern Med, 2003. **254**(5): p. 401-25.
22. Bell, C.G., et al., *Genome-wide linkage analysis for severe obesity in french caucasians finds significant susceptibility locus on chromosome 19q.* Diabetes, 2004. **53**(7): p. 1857-65.
23. Hsueh, W.C., et al., *Genome-wide scan of obesity in the Old Order Amish.* J Clin Endocrinol Metab, 2001. **86**(3): p. 1199-205.
24. Chen, G., et al., *A genome-wide scan for quantitative trait loci linked to obesity phenotypes among West Africans.* Int J Obes (Lond), 2005. **29**(3): p. 255-9.
25. Frayling, T.M., et al., *A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity.* Science, 2007. **316**(5826): p. 889-94.
26. Scuteri, A., et al., *Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits.* PLoS Genet, 2007. **3**(7): p. e115.

27. Loos, R.J., et al., *Common variants near MC4R are associated with fat mass, weight and risk of obesity.* Nat Genet, 2008. **40**(6): p. 768-75.
28. Willer, C.J., et al., *Six new loci associated with body mass index highlight a neuronal influence on body weight regulation.* Nat Genet, 2009. **41**(1): p. 25-34.
29. Thorleifsson, G., et al., *Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity.* Nat Genet, 2009. **41**(1): p. 18-24.
30. Speliotes, E.K., et al., *Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index.* Nat Genet, 2010. **42**(11): p. 937-48.
31. Peters, T., et al., *The mouse Fused toes (Ft) mutation is the result of a 1.6-Mb deletion including the entire Iroquois B gene cluster.* Mamm Genome, 2002. **13**(4): p. 186-8.
32. van der Hoeven, F., et al., *Programmed cell death is affected in the novel mouse mutant Fused toes (Ft).* Development, 1994. **120**(9): p. 2601-7.
33. Anselme, I., et al., *Defects in brain patterning and head morphogenesis in the mouse mutant Fused toes.* Dev Biol, 2007. **304**(1): p. 208-20.
34. Peters, T., K. Ausmeier, and U. Ruther, *Cloning of Fatso (Fto), a novel gene deleted by the Fused toes (Ft) mouse mutation.* Mamm Genome, 1999. **10**(10): p. 983-6.
35. Jacobsson, J.A., H.B. Schioth, and R. Fredriksson, *The impact of intronic single nucleotide polymorphisms and ethnic diversity for studies on the obesity gene FTO.* Obes Rev, 2012. **13**(12): p. 1096-109.
36. Fredriksson, R., et al., *The obesity gene, FTO, is of ancient origin, up-regulated during food deprivation and expressed in neurons of feeding-related nuclei of the brain.* Endocrinology, 2008. **149**(5): p. 2062-71.
37. Robbens, S., et al., *The FTO gene, implicated in human obesity, is found only in vertebrates and marine algae.* J Mol Evol, 2008. **66**(1): p. 80-4.
38. Jia, G., et al., *Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO.* FEBS Lett, 2008. **582**(23-24): p. 3313-9.
39. Gerken, T., et al., *The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase.* Science, 2007. **318**(5855): p. 1469-72.

40. McTaggart, J.S., et al., *FTO is expressed in neurones throughout the brain and its expression is unaltered by fasting.* PLoS One, 2011. **6**(11): p. e27968.
41. Olszewski, P.K., et al., *Hypothalamic FTO is associated with the regulation of energy intake not feeding reward.* BMC Neurosci, 2009. **10**: p. 129.
42. Stratigopoulos, G., et al., *Regulation of Fto/Ftm gene expression in mice and humans.* Am J Physiol Regul Integr Comp Physiol, 2008. **294**(4): p. R1185-96.
43. Wang, P., et al., *Involvement of leptin receptor long isoform (LepRb)-STAT3 signaling pathway in brain fat mass- and obesity-associated (FTO) downregulation during energy restriction.* Mol Med, 2011. **17**(5-6): p. 523-32.
44. Fischer, J., et al., *Inactivation of the Fto gene protects from obesity.* Nature, 2009. **458**(7240): p. 894-8.
45. Church, C., et al., *A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene.* PLoS Genet, 2009. **5**(8): p. e1000599.
46. Boissel, S., et al., *Loss-of-function mutation in the dioxygenase-encoding FTO gene causes severe growth retardation and multiple malformations.* Am J Hum Genet, 2009. **85**(1): p. 106-11.
47. Church, C., et al., *Overexpression of Fto leads to increased food intake and results in obesity.* Nat Genet, 2010. **42**(12): p. 1086-92.
48. Berulava, T. and B. Horsthemke, *The obesity-associated SNPs in intron 1 of the FTO gene affect primary transcript levels.* Eur J Hum Genet, 2010. **18**(9): p. 1054-6.
49. Karra, E., et al., *A link between FTO, ghrelin, and impaired brain food-cue responsivity.* J Clin Invest, 2013. **123**(7).
50. Han, Z., et al., *Crystal structure of the FTO protein reveals basis for its substrate specificity.* Nature, 2010. **464**(7292): p. 1205-9.
51. Jia, G., et al., *N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO.* Nat Chem Biol, 2011. **7**(12): p. 885-7.
52. Berulava, T., et al., *FTO levels affect RNA modification and the transcriptome.* Eur J Hum Genet, 2013. **21**(3): p. 317-23.
53. Dominissini, D., et al., *Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq.* Nature, 2012. **485**(7397): p. 201-6.

54. Meyer, K.D., et al., *Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons.* Cell, 2012. **149**(7): p. 1635-46.
55. Hess, M.E., et al., *The fat mass and obesity associated gene (Fto) regulates activity of the dopaminergic midbrain circuitry.* Nat Neurosci, 2013.
56. Gulati, P., et al., *Role for the obesity-related FTO gene in the cellular sensing of amino acids.* Proc Natl Acad Sci U S A, 2013. **110**(7): p. 2557-62.
57. Hofmann, K. and W. Stoffel, *TMBASE - A database of membrane spanning protein segments.* Biol. Chem. Hoppe-Seyler, 1993. **374**(166).
58. Jurvansuu, J., et al., *Transmembrane protein 18 enhances the tropism of neural stem cells for glioma cells.* Cancer Res, 2008. **68**(12): p. 4614-22.
59. Jurvansuu, J.M. and A. Goldman, *Obesity risk gene TMEM18 encodes a sequence-specific DNA-binding protein.* PLoS One, 2011. **6**(9): p. e25317.
60. Almen, M.S., et al., *The obesity gene, TMEM18, is of ancient origin, found in majority of neuronal cells in all major brain regions and associated with obesity in severely obese children.* BMC Med Genet, 2010. **11**: p. 58.
61. Renstrom, F., et al., *Replication and extension of genome-wide association study results for obesity in 4923 adults from northern Sweden.* Hum Mol Genet, 2009. **18**(8): p. 1489-96.
62. Truong, K., et al., *Measuring the physical food environment and its relationship with obesity: evidence from California.* Public Health, 2010. **124**(2): p. 115-8.
63. Hotta, K., et al., *Association between obesity and polymorphisms in SEC16B, TMEM18, GNPDA2, BDNF, FAIM2 and MC4R in a Japanese population.* J Hum Genet, 2009. **54**(12): p. 727-31.
64. Zhao, J., et al., *The role of obesity-associated loci identified in genome-wide association studies in the determination of pediatric BMI.* Obesity (Silver Spring), 2009. **17**(12): p. 2254-7.
65. den Hoed, M., et al., *Genetic susceptibility to obesity and related traits in childhood and adolescence: influence of loci identified by genome-wide association studies.* Diabetes, 2010. **59**(11): p. 2980-8.

66. Winkelmann, J., et al., *Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions.* Nat Genet, 2007. **39**(8): p. 1000-6.
67. Dorajoo, R., et al., *Replication of 13 obesity loci among Singaporean Chinese, Malay and Asian-Indian populations.* Int J Obes (Lond), 2011.
68. Carter, E.J., et al., *MEK5 and ERK5 are mediators of the pro-myogenic actions of IGF-2.* J Cell Sci, 2009. **122**(Pt 17): p. 3104-12.
69. Dong, F., J.S. Gutkind, and A.C. Larner, *Granulocyte colony-stimulating factor induces ERK5 activation, which is differentially regulated by protein-tyrosine kinases and protein kinase C. Regulation of cell proliferation and survival.* J Biol Chem, 2001. **276**(14): p. 10811-6.
70. Kato, Y., et al., *Bmk1/Erk5 is required for cell proliferation induced by epidermal growth factor.* Nature, 1998. **395**(6703): p. 713-6.
71. Clark, P.R., et al., *MEK5 is activated by shear stress, activates ERK5 and induces KLF4 to modulate TNF responses in human dermal microvascular endothelial cells.* Microcirculation, 2011. **18**(2): p. 102-17.
72. English, J.M., et al., *Isolation of MEK5 and differential expression of alternatively spliced forms.* J Biol Chem, 1995. **270**(48): p. 28897-902.
73. Drew, B.A., M.E. Burow, and B.S. Beckman, *MEK5/ERK5 pathway: The first fifteen years.* Biochim Biophys Acta, 2012. **1825**(1): p. 37-48.
74. Prada, P.O., et al., *Tub has a key role in insulin and leptin signaling and action in vivo in hypothalamic nuclei.* Diabetes, 2013. **62**(1): p. 137-48.
75. Shiri-Sverdlov, R., et al., *Identification of TUB as a novel candidate gene influencing body weight in humans.* Diabetes, 2006. **55**(2): p. 385-9.
76. Snieder, H., et al., *TUB is a candidate gene for late-onset obesity in women.* Diabetologia, 2008. **51**(1): p. 54-61.
77. Mujica, A.O., T. Hankeln, and E.R. Schmidt, *A novel serine/threonine kinase gene, STK33, on human chromosome 11p15.3.* Gene, 2001. **280**(1-2): p. 175-81.
78. Mujica, A.O., et al., *Differential expression pattern of the novel serine/threonine kinase, STK33, in mice and men.* FEBS J, 2005. **272**(19): p. 4884-98.

79. Brauksiepe, B., et al., *The Serine/threonine kinase Stk33 exhibits autophosphorylation and phosphorylates the intermediate filament protein Vimentin.* BMC Biochem, 2008. **9**: p. 25.
80. Scholl, C., et al., *Synthetic lethal interaction between oncogenic KRAS dependency and STK33 suppression in human cancer cells.* Cell, 2009. **137**(5): p. 821-34.
81. Kranenburg, O., *The KRAS oncogene: past, present, and future.* Biochim Biophys Acta, 2005. **1756**(2): p. 81-2.
82. Luo, T., et al., *STK33 kinase inhibitor BRD-8899 has no effect on KRAS-dependent cancer cell viability.* Proc Natl Acad Sci U S A, 2012. **109**(8): p. 2860-5.
83. Moschonis, G., et al., *Social, economic and demographic correlates of overweight and obesity in primary-school children: preliminary data from the Healthy Growth Study.* Public Health Nutr, 2010. **13**(10A): p. 1693-700.
84. Jacobsson, J.A., et al., *Genetic variants near the MGAT1 gene are associated with body weight, BMI and fatty acid metabolism among adults and children.* Int J Obes, 2011.
85. Cole, T.J., et al., *Establishing a standard definition for child overweight and obesity worldwide: international survey.* BMJ, 2000. **320**(7244): p. 1240-3.
86. Schofield, W.N., *Predicting basal metabolic rate, new standards and review of previous work.* Hum Nutr Clin Nutr, 1985. **39 Suppl 1**: p. 5-41.
87. Goldberg, G.R., et al., *Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording.* Eur J Clin Nutr, 1991. **45**(12): p. 569-81.
88. Trichopoulou, A., *Composition tables of Foods and Greek dishes.* 2004, Athens, Greece: Parisianau Publications.
89. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.
90. Metzker, M.L., *Sequencing technologies - the next generation.* Nat Rev Genet, 2010. **11**(1): p. 31-46.
91. Voelkerding, K.V., S.A. Dames, and J.D. Durtschi, *Next-generation sequencing: from basic research to diagnostics.* Clin Chem, 2009. **55**(4): p. 641-58.
92. McKernan, K.J., et al., *Sequence and structural variation in a human genome uncovered by short-read, massively parallel*

*ligation sequencing using two-base encoding.* Genome Res, 2009. **19**(9): p. 1527-41.
93. Zaboli, G., et al., *Sequencing of high-complexity DNA pools for identification of nucleotide and structural variants in regions associated with complex traits.* Eur J Hum Genet, 2012. **20**(1): p. 77-83.
94. Sallman Almen, M., et al., *Determination of the obesity-associated gene variants within the entire FTO gene by ultra-deep targeted sequencing in obese and lean children.* Int J Obes (Lond), 2013. **37**(3): p. 424-31.
95. Alsio, J., et al., *Inverse association of high-fat diet preference and anxiety-like behavior: a putative role for urocortin 2.* Genes Brain Behav, 2009. **8**(2): p. 193-202.
96. Rio, D.C., et al., *Purification of RNA using TRIzol (TRI reagent).* Cold Spring Harb Protoc, 2010. **2010**(6): p. pdb prot5439.
97. Miller, E.K. and J.D. Cohen, *An integrative theory of prefrontal cortex function.* Annu Rev Neurosci, 2001. **24**: p. 167-202.
98. Ren, D., et al., *Neuronal SH2B1 is essential for controlling energy and glucose homeostasis.* J Clin Invest, 2007. **117**(2): p. 397-406.
99. Wang, C., et al., *Brain-derived neurotrophic factor (BDNF) in the hypothalamic ventromedial nucleus increases energy expenditure.* Brain Res, 2010. **1336**: p. 66-77.
100. Noble, E.E., et al., *The lighter side of BDNF.* Am J Physiol Regul Integr Comp Physiol, 2011. **300**(5): p. R1053-69.
101. Rask-Andersen, M., S. Masuram, and H.B. Schioth, *The Druggable Genome: Evaluation of Drug Targets in Clinical Trials Suggests Major Shifts in Molecular Class and Indication.* Ann Rev Pharm Tox, 2014. **54**.
102. Rask-Andersen, M., J. Zhang, and H.B. Schioth, *Protein kinases as therapeutic drug targets: Current use and Clinical Trials*. in production.
103. Fabian, M.A., et al., *A small molecule-kinase interaction map for clinical kinase inhibitors.* Nat Biotechnol, 2005. **23**(3): p. 329-36.
104. Karaman, M.W., et al., *A quantitative analysis of kinase inhibitor selectivity.* Nat Biotechnol, 2008. **26**(1): p. 127-32.

105. Zhang, J., P.L. Yang, and N.S. Gray, *Targeting cancer with small molecule kinase inhibitors.* Nat Rev Cancer, 2009. **9**(1): p. 28-39.
106. Bernstein, B.E., et al., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

# Additional Papers

Rask-Andersen M, Masuram S, Fredriksson R, Schiöth HB. Solute carriers as drug targets: Current use, clinical trials and prospective. Mol Aspects Med. 2013;34(2-3):702-10.

Carlini VP, Poretti MB, Rask-Andersen M, Chavan RA, Ponzio MF, Sawant RS, de Barioglio SR, Schiöth HB, de Cuneo MF. Differential effects of fluoxetine and venlafaxine on memory recognition: Possible mechanisms of action. Prog Neuro-Psychopharmacol Biol Psychiatry. 2012;38(2):159-67.

Brooks SJ, Rask-Andersen M, Benedict C, Schiöth HB. A debate on current eating disorder diagnoses in light of neurobiological findings: Is it time for a spectrum model? BMC Psychiatry. 2012;12

Jacobsson JA, Rask-Andersen M, Risérus U, Moschonis G, Koumpitski A, Chrousos GP, Lannfelt L, Marcus C, Gyllensten U,Schiöth HB, Fredriksson R. Genetic variants near the MGAT1 gene are associated with body weight, BMI and fatty acid metabolism among adults and children. Int J Obes. 2012;36(1):119-29.

Rask-Andersen M, Almén MS, Schiöth HB. Trends in the exploitation of novel drug targets. Nature Reviews Drug Discovery. 2011;10(8):579-90.

Zheleznyakova GY, Kiselev AV, Vakharlovsky VG, Rask-Andersen M, Chavan R, Egorova AA, Schiöth HB, Baranov VS. Genetic and expression studies of SMN2 gene in russian patients with spinal muscular atrophy type II and III. BMC Medical Genetics. 2011;12

Rask-Andersen M, Olszewski PK, Levine AS, Schiöth HB. Molecular mechanisms underlying anorexia nervosa: Focus on human gene association studies and systems controlling food intake. Brain Res Rev. 2010;62(2):147-64.

Rask-Andersen M, Masuram S, Schiöth HB. The Druggable Genome: Evaluation of Drug Targets in Clinical Trials Suggests Major Shifts in Molecular Class and Indication. ARPT. Manuscript in production.

Rask-Andersn M, Alsiö J, Chavan RA, Olszewski PK, Levine AS, Fredriksson R, Schiöth HB. Hindbrain proopiomelanocortin induction and differential regulation of dopamine D1 and D2 receptors in the hindbrain of diet-induced obese rats. Neuroscience Letters. Manuscript being revised for publication.

Rask-Andersen M, Zhang J, Schiöth HB. The Druggable Kinome: Protein kinases as therapeutic drug targets: Current use and Clinical Trials. Manuscript in production.

Rask-Andersen M, Masuram S, Schiöth HB. Novel GPCRs as Drug Targets in Clinical Trials. Manuscript in production.

Västermark Å, Rask-Andersen M, Reiter J, Williams M, Schiöth HB. Insulin receptor-like "ectodoomain only" genes and splice variants are found in both arthropods and human brain cDNA. Journal of Systematics and Evolution. Manuscript being revised for publication.

Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 921

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine.

Distribution: publications.uu.se
urn:nbn:se:uu:diva-204449

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2013