The human pancreatic polypeptide receptor Y4

Genetic and functional variation

KATERYNA SHEBANITS
Humans are evolutionarily adapted to an environment where food is scarce, but today many live in a world of food abundance. Paired with low physical activity, this may lead to weight gain and obesity. Efficient anti-obesity treatments require understanding of the mechanisms that control hunger, satiety, energy metabolism and body weight. This thesis investigates possible genetic and physiological mechanisms behind these processes.

Genetic correlation between body-mass index (BMI) and a highly polymorphic region on chromosome 10 was analysed with regard to single nucleotide polymorphisms (SNPs) and gene copy number variation (CNV). This region contains the gene $\text{NPY4R}$ encoding the pancreatic polypeptide (PP) receptor Y4, which has been reported to reduce appetite.

The results show that the $\text{NPY4R}$ gene was duplicated before the divergence of modern humans from the Neanderthals and the Denisovans (approximately to 400,000–800,000 years ago). The CNV of the $\text{NPY4R}$ gene region was investigated by read depth analysis based on genome sequences and droplet digital PCR (ddPCR). The read depth results revealed a CNV range of 3–7 copies per genome, while the ddPCR results demonstrated a range of 2–11. Most humans have a total of 4–5 copies, in contrast to the two copies presumed by previous studies.

Investigation of an association between the $\text{NPY4R}$ CNV and body mass index (BMI) led to interesting and ambiguous results. A study of 558 Swedish individuals with a wide range of BMI suggested, surprisingly, a positive correlation between $\text{NPY4R}$ copy number and BMI for women. On the other hand, a study of 1009 individuals from Northern Sweden found no correlation between BMI and $\text{NPY4R}$ copy number. These diverging findings may be due to geographical variation or lack of power in one of these studies.

Twelve naturally-occurring amino acid variants of the Y4 receptor were investigated pharmacologically in cell culture. Three of these showed no functional response, which may be explained by altered conformation of the receptors. For two receptor variants PP had a significantly decreased potency. A 3D model of the Y4 receptor was generated based on the crystal structure of the human Y1 receptor. The functional responses of the Y4 variants agree well with the 3D model and with the degree of evolutionary conservation of the positions.

In conclusion, these studies reveal unexpectedly large CNV as well as extensive SNP for the $\text{NPY4R}$ gene and a possible correlation with BMI that may be due to the differing responses of the naturally occurring receptor variants.

**Keywords:** $\text{NPY4R}$, Y4, obesity, CNV

Kateryna Shebanits, Department of Neuroscience, Pharmacology, Box 593, Uppsala University, SE-75124 Uppsala, Sweden.

© Kateryna Shebanits 2018

ISSN 1651-6206
urn:nbn:se:uu:diva-356573 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-356573)
To my family
List of Papers

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

I  Shebanits, K., Günther, T., Johansson, A., Maqbool, K., Feuk, L., Jakobsson, M., Larhammar, D. Copy number determination of the gene for the human pancreatic polypeptide receptor \textit{NPY4R} using read depth analysis and droplet digital PCR. (Submitted)


III Shebanits, K., Johansson, Å., Rafati, N., Feuk, L., Larhammar, D. \textit{NPY4R} copy number is not associated with body weight in Northern Swedes. (Manuscript)

IV Shebanits, K., Vasile, S., Xu, B., Gutiérrez-de-Terán, H., Larhammar, D. Functional characterization in vitro of twelve naturally occurring variants of the human pancreatic polypeptide receptor \textit{NPY4R}. (Submitted)

Reprints were made with permission from the publisher.
Contents

Introduction ........................................................................................................11
    Obesity ........................................................................................................11
    Appetite and energy metabolism regulation ............................................12
        Hunger control ..................................................................................13
        Satiety control ...............................................................................14
        Crosstalk between hypothalamic regions regulating appetite ..........18
    NPY family of peptides and their receptors ...........................................19
        Pancreatic polypeptide .................................................................19
        Y4 receptor ....................................................................................20
        NPY4R gene organization ..............................................................21
    Genetics of obesity ................................................................................22
        Single nucleotide polymorphisms and copy number variation in obesity ..................................................23
        Syndromic forms of obesity ..............................................................24
        Non-syndromic obesity .....................................................................25
        SNPs and CNV of NPY4R and its role in appetite regulation ..........26
Aims ...............................................................................................................28

Materials and methods ................................................................................29
    Copy number investigation ....................................................................29
    The duplication unit and the NPY4R gene organization investigation ....30
    NPY4R copy number, body mass index and energy metabolism investigation ..................................................31
    Pharmacological analysis of naturally-occurring Y4 variants ..........32
        PCR for SNP detection .................................................................32
        Pharmacological investigation .......................................................32

Results and discussion ...............................................................................33
    Methodological considerations in CNV determination ......................33
        Read depth analysis and ddPCR ......................................................33
    Comparison and interpretation of the results ..................................34
    Duplication unit size and organisation .............................................35
NPY4R gene: organization, most common copy number state and CNV range (Papers I, II and III) ................................................................. 36
Splice variants .......................................................................................... 36
NPY4R CNV characterisation ..................................................................... 36
NPY4R copy number, body mass index and energy metabolism
(Papers II and III) .................................................................................... 37
NPY4R mutations, Y4 functions and body weight (Papers III and IV) .... 39
Inconsistent results of genetic association studies in obesity – FTO
andAMY .................................................................................................... 40
FTO gene .................................................................................................. 40
Amylase gene ............................................................................................ 41
Concluding remarks and future perspectives .......................................... 42
Acknowledgements .................................................................................. 44
References ............................................................................................... 45
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNV</td>
<td>Copy number variation</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index</td>
</tr>
<tr>
<td>WC</td>
<td>Waist circumference</td>
</tr>
<tr>
<td>NPY</td>
<td>Neuropeptide Y</td>
</tr>
<tr>
<td>PP</td>
<td>Pancreatic polypeptide</td>
</tr>
<tr>
<td>AgRP</td>
<td>Agouti-related peptide</td>
</tr>
<tr>
<td>MCH</td>
<td>Melanin-concentrating hormone</td>
</tr>
<tr>
<td>CART</td>
<td>Cocaine- and amphetamine-regulated transcript</td>
</tr>
<tr>
<td>CCK</td>
<td>Cholecystokinin</td>
</tr>
<tr>
<td>GLP-1</td>
<td>Glucagon-like peptide 1</td>
</tr>
<tr>
<td>PYY</td>
<td>Peptide YY</td>
</tr>
<tr>
<td>α-MSH</td>
<td>α-Melanocyte stimulating hormone</td>
</tr>
<tr>
<td>BBB</td>
<td>Blood-brain barrier</td>
</tr>
<tr>
<td>DVC</td>
<td>Dorsal vagal complex</td>
</tr>
<tr>
<td>AP</td>
<td>Area postrema</td>
</tr>
<tr>
<td>NTS</td>
<td>Nucleus of the solitary tract</td>
</tr>
<tr>
<td>ME</td>
<td>Median eminence</td>
</tr>
<tr>
<td>ARC</td>
<td>Arcuate nucleus</td>
</tr>
<tr>
<td>PVN</td>
<td>Paraventricular nucleus</td>
</tr>
<tr>
<td>VMN</td>
<td>Ventromedial nucleus</td>
</tr>
<tr>
<td>DMN</td>
<td>Dorsomedial nucleus</td>
</tr>
<tr>
<td>LHA</td>
<td>Lateral hypothalamic area</td>
</tr>
<tr>
<td>TRH</td>
<td>Thyrotropin releasing hormone</td>
</tr>
<tr>
<td>CRH</td>
<td>Corticotropin-releasing hormone</td>
</tr>
<tr>
<td>GPCR</td>
<td>G-protein-coupled receptor</td>
</tr>
<tr>
<td>TM</td>
<td>Transmembrane helix</td>
</tr>
<tr>
<td>ECL</td>
<td>Extracellular loop</td>
</tr>
<tr>
<td>ICL</td>
<td>Intracellular loop</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-wide association studies</td>
</tr>
<tr>
<td>PWS</td>
<td>Prader-Willi syndrome</td>
</tr>
<tr>
<td>ddPCR</td>
<td>Droplet digital polymerase chain reaction</td>
</tr>
<tr>
<td>RT-qPCR</td>
<td>Quantitative polymerase chain reaction</td>
</tr>
<tr>
<td>BAC</td>
<td>Bacterial artificial chromosome</td>
</tr>
<tr>
<td>YAC</td>
<td>Yeast artificial chromosome</td>
</tr>
<tr>
<td>IP</td>
<td>Inositol phosphate</td>
</tr>
</tbody>
</table>
Introduction

Obesity

As a species we have evolved in a world of hunger, where food was scarce and came at irregular intervals. It was beneficial to be able to store the energy from the food quickly and efficiently. Now, hundreds of thousands of years later, our bodies are still hungry with the same prehistoric hunger, but the world we live in has changed dramatically. Cravings for fatty and sweet food that used to be good for providing our bodies with the most energy-rich diet and allowing our survival has become detrimental to our health. Our living conditions have changed: most of us do not need to do hard physical work to obtain food, to run away from predators and to move much in our daily lives. All those calories our hungry bodies keep accumulating have become hard to spend. In these circumstances of food abundance and lack of physical activity our bodies keep increasing their energy stores, often leading to the disease called obesity. In addition, various mutations in our genomes can cause the physiological mechanisms that previously used to be beneficial to become detrimental and lead to obesity.

Obesity has become one of the major health problems across the globe and its prevalence has grown during the past few decades (World Health Organization 2014). The World Health Organization (WHO) has decided to define obesity as a body mass index (BMI) of 30 kg/m² and higher. As mentioned above, excessive weight gain as a result of an imbalance between consumed and spent calories might be caused by both environmental factors (Maes et al. 1997) (i.e. sedentary life style, food availability, low cost of fat- and sugar-rich fast food) and genetic factors (Maes et al. 1997) (mutations of various kinds). Genetics plays a very important role in the disease panorama of obesity due to the fact that body weight has high heritability that ranges from 78% to 81% (Elks et al. 2012). Addressing the environmental factors without the knowledge of the genetics underlying the disease is like trying to win a game without knowing the rules. Not only is obesity a disease in itself, it is also a risk factor for cardiovascular diseases, diabetes, several types of cancer and metabolic syndrome (Aronne and Segal 2002; Prospective Studies Collaboration 2009).

So what is it that makes us so hungry? How are hunger and satiety regulated? What regulates our energy expenditure? Complex physiological processes in the central nervous system regulate both energy metabolism
homeostasis and behaviour. The key roles in this process are played by brain stem and hypothalamic nuclei, which receive and interpret multiple signals from peripheral systems (Sainsbury et al. 2002; Gao and Horvath 2008; Ladenheim 2012). Numerous hormones like insulin, leptin, ghrelin and NPY-family peptides are involved in appetite regulation and energy metabolism by neuronal and endocrine mechanisms (Karra and Batterham 2010; Dockray 2014).

**Appetite and energy metabolism regulation**

Multiple neuroendocrine peptides (Table 1) regulate appetite through various pathways. This chapter is a brief overview of the regulation of food intake and energy metabolism with focus on the most well-studied pathways and key regulators (Figure 1). It is primarily based on multiple reviews citing the original experimental work.

**Table 1. Neuroendocrine peptides involved in regulation of appetite and energy metabolism**

<table>
<thead>
<tr>
<th>Appetite stimulating or orexigenic neuroendocrine peptides</th>
<th>Appetite inhibiting or anorexigenic neuroendocrine peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>ghrelin</td>
<td>cholecystokinin (CCK),</td>
</tr>
<tr>
<td>gastric inhibitory polypeptide (GIP)</td>
<td>insulin</td>
</tr>
<tr>
<td>neuropeptide Y (NPY)</td>
<td>glucagon-like peptide 1 (GLP-1)</td>
</tr>
<tr>
<td>agouti-related peptide (AgRP)</td>
<td>leptin</td>
</tr>
<tr>
<td>melanin-concentrating hormone (MCH)</td>
<td>peptide YY (PYY)</td>
</tr>
<tr>
<td>orexin</td>
<td>pancreatic polypeptide (PP)</td>
</tr>
<tr>
<td>opioids</td>
<td>α-melanocyte stimulating hormone (α-MSH)</td>
</tr>
<tr>
<td></td>
<td>cocaine- and amphetamine-regulated transcript (CART)</td>
</tr>
<tr>
<td></td>
<td>corticotropin-releasing hormone (CRH)</td>
</tr>
<tr>
<td></td>
<td>oxyntomodulin (OXM)</td>
</tr>
<tr>
<td></td>
<td>bombesin/bombesin-related peptides (BNRP)</td>
</tr>
</tbody>
</table>

The peptide hormones are released from endocrine tissues in response to the nutritional and physical state of the organism and reach their receptors in the peripheral nervous system (vagus nerve) and parts of central nervous system that have an impaired blood-brain barrier (BBB), like area postrema (AP) of
the brain stem and the median eminence (ME) (Rodríguez et al. 2010) and arcuate nucleus (ARC) (Murphy and Bloom 2004; Faouzi et al. 2007; Rodriguez et al. 2010) of the hypothalamus (Figure 1). Signals coming from the receptors on the vagus nerve go through AP and other regions of the dorsal vagal complex (DVC), and ultimately reach the appetite-regulating nuclei of the hypothalamus. Some hormones, like insulin, PYY3-36 and leptin, are transported across the BBB (Simpson and Bloom 2010) and thus, can affect multiple central targets. The key hypothalamic nuclei involved in appetite regulation and energy homeostasis are the ARC, ventromedial nucleus (VMN), lateral hypothalamic area (LHA), dorsomedial nucleus (DMN) and paraventricular nucleus (PVN) (Figure 1). The hypothalamus is the key coordinator of the neuronal and endocrine signals that report the state of the body. It receives both long-term endocrine signals indicating general health and level of energy stored (leptin from the adipose tissue), and short-term signals regarding the physical activity level, levels of nutrients such as blood glucose, stretch of the stomach, smell and sight, etc.

Hunger control

Orexigenic pathways are summarised in Figure 1 B and F. Ghrelin is released by the stomach and is a potent appetite-stimulating hormone. Its plasma level has a negative correlation with body weight and increases after weight loss (Murphy and Bloom 2004; Sumithran and Prendergast 2011). Ghrelin plasma levels are increased during fasting and decrease rapidly following a meal. Ghrelin’s signaling is mediated through the GHS receptor (GHSR) which is expressed both in the periphery (vagus nerve, adipose tissue) and in the central nervous system (brain stem, hypothalamus) (Murphy and Bloom 2004; Huda et al. 2006). There is evidence suggesting that ghrelin is able to cross the BBB in the ARC and that small amounts of ghrelin are produced in the brain (Huda et al. 2006). Ghrelin neurons have terminals on NPY/AgRP, POMC (containing alpha-MSH) and CRH neurons and orexin fibers in the LHA (Huda et al. 2006; Simpson and Bloom 2010).

Orexin-A stimulates appetite through vagal afferents by inhibiting the CCK-dependent activation (Dockray 2014).

NPY is a potent orexigenic neuropeptide that decreases energy expenditure. It is expressed in multiple regions of the central nervous system that are involved in appetite and energy homeostasis control (brain stem, hypothalamic nuclei) (Gao and Horvath 2008).

AgRP is an endogenous antagonist of the MC3 and MC4 receptors. It exerts its orexigenic action by counteracting the anorexigenic α-MSH signal (Gao and Horvath 2008).
Satiety control

Long-term and short-term signals regulate satiety and energy metabolism. Leptin and insulin levels are indicators of long-term adiposity state, while acute signal for hunger (like ghrelin) and satiety (like CCK and PYY) regulate meal start, size and termination. Anorexigenic pathways are summarised in Figure 1 C and G.

Leptin is secreted from white adipose tissue in proportion to fat mass. Both peripheral and central administration of leptin inhibits appetite. Leptin is able to cross the BBB and directly inhibit orexigenic NPY/AgRP neurons and stimulate the anorexigenic POMC neurons of the ARC (Murphy and Bloom 2004; Simpson and Bloom 2010). It also indirectly acts through DMN, VMN and LHA (Faouzi et al. 2007).
Figure 1. Simplified schematic representation of some of the most well studied appetite control mechanisms. Coronal section of a human brain (A) with orexigenic pathways in orange (B) and anorexigenic pathways in blue (C). Sagittal view of a human brain (D) with a line indicating a brainstem cross section at the obex level (E). Orexigenic pathways are labelled with orange (F) and anorexigenic pathways with blue (G). Original experimental work was performed in rodents and the results were transferred to a human model.
Leptin decreases hypothalamic expression levels of orexin and MCH, increases CRH expression, and modulates ghrelin signalling in the ARC (Simpson and Bloom 2010). It also potentiates CCK signalling through the vagal afferents (Dockray 2014).

Insulin is released from pancreatic β-cells. Its basal circulating level correlates with adiposity (Lau and Herzog 2014). Insulin can cross the BBB (Gray et al. 2017) and it promotes satiety and increases energy expenditure by inhibiting the orexigenic NPY/AgRP neurons of the ARC (Murphy and Bloom 2004) and increasing POMC expression in POMC neurons of the ARC (Simpson and Bloom 2010).

Postprandial satiety signals come from the periphery and originate from mechanical receptors indicating stomach distension, from chemical receptors of the intestine that indicate nutrient content, and from the endocrine system of the intestine that releases satiety hormones in response to the nutrient content (Murphy and Bloom 2004).

CCK is released from the small intestine in response to products of digestion. It reduces the meal size in a dose-dependent manner by acting through CCK1 receptors (CCK1R) on the vagus nerve (Murphy and Bloom 2004; Dockray 2014), NTS and DMN (Huda et al. 2006). The CCK signal from vagal afferent fibers (through the DVC) activates DMN and CART neurons of the PVN (Simpson and Bloom 2010). NTS has CCK neurons and small amounts of CCK are produced within the brain (D’Agostino et al. 2016).

PP is an anorexigenic hormone secreted by the pancreas in response to food intake and proportionally to caloric intake (Murphy and Bloom 2004; Simpson and Bloom 2010). It might be able to cross the incomplete BBB of AP, NTS and ARC and its receptors (Y4) are expressed in these regions (Whitcomb et al. 1997). Both PP and Y4 will be discussed in detail further below.

PYY is secreted postprandially by the endocrine cells of the small and large intestine in proportion to energy intake and nutrient composition. It is cleaved to produce a truncated form of the peptide, PYY3-36, which is able to cross the BBB (Huda et al. 2006; Simpson and Bloom 2010). Peripheral PYY3-36 acts on presynaptic inhibitory Y2 receptors on the vagus nerve, in the brain stem and in the ARC. NPY/AgRP neurons of the ARC are inhibited by PYY3-36, which in turn relieves the tonic inhibition of anorexigenic POMC/CART neurons and promotes satiety (Murphy and Bloom 2004; Huda et al. 2006; Simpson and Bloom 2010).

GLP-1 and OXM are both products of posttranslational processing of proglucagon. Both are synthesised in the central nervous system, intestine and colon, and released into circulation in response to food intake (Murphy and Bloom 2004; Simpson and Bloom 2010). GLP-1 and OXM inhibit gastric acid secretion and food intake (Huda et al. 2006). GLP-1 also stimulates insulin secretion (Simpson and Bloom 2010) and slows down gastric emptying (Murphy and Bloom 2004). Both peptides act through GLP-1 receptors, but via different neural pathways.
The neuropeptide precursor POMC is cleaved in ARC neurons to produce α-MSH which is a potent anorexigenic hormone. α-MSH inhibits food intake, reduces body weight and increases energy expenditure by acting through MC3 and MC4 receptors that are expressed in such appetite regulating hypothalamic nuclei as ARC, DMN, PVN and LHA (Gao and Horvath 2008).

CART is a potent appetite inhibitor and it is also involved in energy expenditure control in brown adipose tissue. It is expressed in ARC, PVN and LHA. It stimulates CRH and thyrotropin releasing hormone (TRH) release in the PVN (Lau and Herzog 2014).

Crosstalk between hypothalamic regions regulating appetite

The ARC is one of the most critical regions in energy homeostasis control. Its proximity to fenestrated capillaries at the base of hypothalamus allows it to respond to multiple hormones (Huda et al. 2006; Gao and Horvath 2008). Furthermore, ARC neurons receive neuronal projections from the brainstem. The ARC neurons project to multiple central targets (Gao and Horvath 2008). The ARC regulates both hunger and satiety through two distinct populations of neurons: orexigenic NPY/AgRP neurons and anorexigenic POMC/CART neurons (Huda et al. 2006; Gao and Horvath 2008). GABA-ergic NPY/AgRP neurons exert tonic inhibition of POMC/CART neurons, inhibit satiety signals in VMN, influence orexin and MCH neurons of LHA, and inhibit α-MSH signalling in PVN (Gao and Horvath 2008). POMC/CART neurons project to DMN, PVN and LHA, where they promote satiety signal through MC3 and MC4 receptors (Gao and Horvath 2008).

The VMN is an important satiety centre. It expresses the leptin receptor LEPR and is crucial in leptin-mediated regulation of energy metabolism. Activation of LEPR directly increases brain-derived neurotrophic factor (BDNF) expression in VMN, which promotes synaptic plasticity and leads to anorexigenic effects. Direct neuronal projections from VMN activate POMC neurons of the ARC (Gao and Horvath 2008).

The LHA is a “hunger center” in the hypothalamus. It contains orexin and MCH neurons, which receive input from NPY and POMC neurons of the ARC (Huda et al. 2006; Gao and Horvath 2008). The activity of these neuronal populations is regulated by different hormones and multiple neuronal projections. They interact with each other as well as with various brain regions involved in behavioural responses related to memory, learning and emotions and also to motor responses to changes in energy balance. Orexin neurons of LHA and NPY neurons of ARC project to each other and have synergistic action. MCH neurons act via an independent pathway and counteract the anorexigenic action of α-MSH (Gao and Horvath 2008).

The DMN is important for entrainment of the circadian rhythms to feeding schedules. It receives input from ARC and brain stem regions involved in appetite control and projects to PVN and LHA (Gao and Horvath 2008).
The PVN is critical for control of food intake and energy expenditure. It receives projections from the brain stem, ARC (both POMC/CART and NPY/AgRP neurons), DMN, VMN and LHA (Murphy and Bloom 2004). ARC projections to PVN regulate CRH and TRH release (Sainsbury et al. 2002; Simpson and Bloom 2010) and DMN projections are involved in regulation of circadian rhythm and temperature (Gao and Horvath 2008). Projections from DMN containing α-MSH terminate on TRH neurons (Simpson and Bloom 2010).

NPY family of peptides and their receptors

The NPY family of peptides and their receptors comprise one of the important neuroendocrine systems that regulate energy metabolism, hunger and satiety (Herzog 2003), lipid metabolism, insulin secretion (Renshaw and Batterham 2005; Huda et al. 2006) and many other physiological functions. The NPY family of peptides consists of NPY, PYY and PP, all of which are 36 amino acids long, have an amidated C-terminal tyrosine and share considerable sequence identity (Larhammar 1996; Cerdá-Reverter and Larhammar 2000; Sundström et al. 2008). As was described above, NPY stimulates appetite (Edwards et al. 1999) while PYY and PP inhibit appetite (Ueno et al. 1999; Batterham et al. 2002, 2003; Kojima et al. 2007; Akerberg et al. 2010). These three peptides are ligands that bind to G-protein-coupled receptors belonging to the NPY-receptor family. Humans have four functional receptor subtypes: Y1, Y2, Y4 and Y5 (Larhammar and Salaneck 2004), and a pseudogenised Y6 receptor. All Y-receptors are expressed in the brain, especially in hypothalamic nuclei where they are involved in control of appetite and energy metabolism (Whitcomb et al. 1997; Fetissov et al. 2004; Murphy and Bloom 2004; Huda et al. 2006; Simpson and Bloom 2010).

NPY exhibits orexigenic action through the Y1 and Y5 receptors in the paraventricular nucleus of the hypothalamus (Lecklin et al. 2002). Both Y2 and Y4 are considered to promote satiety. PYY acts primarily on Y2 in the arcuate nucleus of the hypothalamus (Batterham et al. 2002; Murphy and Bloom 2004; Huda et al. 2006; Simpson and Bloom 2010), and PP acts on Y4 receptors (Kojima et al. 2007). Mutations in these peptides and their receptors have been associated with obesity and are discussed below (Bray et al. 2000; Karvonen et al. 2001; Ding et al. 2005; Siddiq et al. 2007).

Pancreatic polypeptide

PP affects appetite and energy balance by acting through the Y4 receptor. PP is released from pancreatic PP cells (previously called F cells) (Adrian 1978; Katsuura et al. 2002) postprandially and in proportion to caloric intake (Sive et al. 1979; Track et al. 1980; Inui et al. 1993; Simpson and Bloom 2010) as
well as in response to other stimuli such as hypoglycaemia (Havel et al. 1993). Its plasma half-life is approximately 6 min (Adrian 1978). The plasma concentration of PP in fasting individuals displays a circadian rhythm (Track et al. 1980). Both postprandial and circadian releases of the hormone are regulated by vagal tone (Schwartz et al. 1978; Taylor et al. 1978). PP inhibits pancreatic exocrine secretion, gall bladder contraction and stimulates gastrointestinal motility and gastric acid secretion (McTigue and Rogers 1995a, b; Kojima et al. 2007). Vagotomy completely abolishes the ability of PP to decrease food intake (Asakawa et al. 2003). Hepatic vagotomy in rats abolished PP-induced increase of efferent activity in sympathetic nerves innervating white and brown adipose tissue and adrenal medulla (Asakawa et al. 2003).

Acute peripheral administration of physiological doses of PP leads to rapid and lasting reduction of food intake in fasted normal weight mice (Asakawa et al. 1999, 2003; Balasubramaniam et al. 2006; Lin et al. 2009) by decreasing the hypothalamic expression of the potent hunger stimulants NPY, ghrelin and orexin, and increases anorexigenic urocortin (Asakawa et al. 2003). Intra-peritoneal injections of PP in genetically obese ob/ob mice and fatty liver Shionogi-ob/ob mice leads to reduction of food intake (Asakawa et al. 2003). Repeated PP administration leads not only to food intake reduction, but also to reduction of body weight gain and ameliorated insulin responsiveness in ob/ob mice (Asakawa et al. 2003).

Intravenous (i.v.) administration of PP (10 pmol/kg/min) to healthy men reduced energy intake by 22% at a buffet setting provided 2 hours after infusion. It also decreased the cumulative food intake for next 24 hours by 25% (Batterham et al. 2003). I.v. administration of 5 pmol/kg/min to fasted lean individuals reduced energy intake in a buffet setting 1 hour after infusion by 11% (Jesudason et al. 2007). Prader-Willi syndrome (PWS) patients, who are characterised by childhood-onset hyperphagia and morbid obesity, had reduced food intake after i.v. infusion of PP (Berntson et al. 1993).

Y4 receptor

The Y4 receptor is encoded by NPY4R gene, which in human is located on chromosome 10q11.22. The receptor was cloned in our group in 1995 (Lundell et al. 1995) and independently by three other groups (Bard et al. 1995; Gregor et al. 1996; Yan et al. 1996). It is a part of the NPY-receptor family, which belongs to the rhodopsin-like superfamily of G protein-coupled receptors (GPCRs of class A). Among the NPY family receptors, Y4 is the fastest evolving functional member (Wraith et al. 2000).

The human Y4 receptor is 375 amino acids long and, like all GPCRs, is comprised by seven transmembrane helices (TM) connected with intracellular (ICL) and extracellular loops (ECL) (Figure 2). Like the other NPYR family receptors, Y4 can couple via Gi and Go as well as Gq (Misra et al. 2004). It has been found to be expressed in the regions of central nervous system that
play important roles in energy balance, like the brain stem (Parker and Herzog 1999; Kojima et al. 2007; Lin et al. 2009) and the hypothalamus (ARC, PVN, DMN, VMN and LHA) (Parker and Herzog 1999; Campbell et al. 2003; Lin et al. 2009). Animal studies have demonstrated that it is also expressed in several other organs including coronary artery, heart, stomach, ileum (Bard et al. 1995), colon (Cox and Tough 2002), pancreas and prostate (Lundell et al. 1995). Y4 is involved in the regulation of multiple physiological processes such as food intake (Lin et al. 2009; Sainsbury et al. 2010; Li et al. 2010), energy metabolism and physical activity (Zhang et al. 2010), colonic anion transport (Tough et al. 2006), adipose tissue and bone formation synergistically with Y2 (Sainsbury et al. 2003; Lee et al. 2011).

A crystal structural model of the human Y4 receptor does not exist up to date, but a computational model has been proposed based on crystal structures for other class A receptors, followed by mutagenesis and measurements of functional response to PP stimulation (Pedragosa-Badia et al. 2014). Several residues in the outer parts of the transmembrane regions were identified to be important for receptor activation by PP.

![Figure 2. Schematic structure of the Y4 receptor. Y4 is comprised by an extracellular N-terminus, seven transmembrane regions, separated by intracellular loops (ICL) and extracellular loops (ECL), and an intracellular C-terminus. The overall structure is characteristic for GPCRs.](image)

**NPy4R gene organization**

Gene organization, features in the genomic neighbourhood and genetic variation are important for understanding gene expression. *NPy4R* is 8 kb long and
contains three exons, but only one of the exons is protein coding (Figure 3). Our phylogenetic analysis of gene organization by multiple species alignment revealed the presence of a potential adipocyte enhancer element partially composed of a primate-specific Alu repeat (Figure 3). According to the previous analysis of the splice variants performed by our group in adipose mRNA, either of the 5’UTR exons could be spliced onto the coding 3’ exon (Figure 2B, mRNAs 2 and 4).

Figure 3. NPY4R gene organisation (A) and possible mRNA variants (B).

Genetic variation

Human genomes are remarkably similar across individuals and the differences that account for the genetic heterogeneity and phenotypic diversity in human populations are comprised by multiple types of genetic variants. Variation can occur in both coding and non-coding parts of the genome. Most of such alterations (or mutations) are harmless (rarely beneficial) and constitute a pool of genetic variability. However, some can become detrimental, leading to increased susceptibility to certain diseases. Genetic variations can be divided by size into small- and large-scale variants. Small-scale structural variants are single nucleotide polymorphisms (SNPs), tandem nucleotide substitutions and repeated sequences (e.g. short and long interspersed elements and tandem repeats) (Feuk et al. 2006). Large-scale structural variants are usually defined as regions longer than 1 kb that constitute inversions, translocations, insertions and copy number variations (CNVs): deletions and duplications (Feuk et al. 2006; Sjödin and Jakobsson 2012). All these alterations can affect structural integrity, copy number, regulation and functions of genes.
Genetics of obesity

Due to the fact that obesity is a multifactorial and polygenic disease, it is difficult to study specific genetic factors that influence body weight. But despite the challenges, multiple single nucleotide polymorphisms (SNPs) and copy number variation (CNV) regions have been associated with obesity.

Single nucleotide polymorphisms and copy number variation in obesity

SNPs are defined as DNA variations in single position that occur in at least 1% of randomly selected individuals in a population (Keats and Sherman 2013). Most of the SNPs are functionally neutral, but many are implicated in various disorders, including obesity. More than 500 SNPs for various adiposity-associated traits have already been identified (Loos 2018). Genome-wide association studies (GWAS) have found multiple loci associated with adiposity and indicates that most of these SNPs are involved in CNS functions (Loos 2018). These data suggest that CNS is a key organ in the regulation of energy balance. Analysis of low-frequency and rare SNP variants has also confirmed that the CNS plays a primary role in body weight regulation (Turcot et al. 2018).

CNVs are defined as genomic regions of one kilobase or larger, up to several megabases in length that display different numbers of copies between individuals (Feuk et al. 2006). A growing body of evidence suggests that CNVs are associated with multiple disorders including obesity (Willer et al. 2009; Sha et al. 2009; Walters et al. 2010; Bochukova et al. 2010; Jarick et al. 2011; Gilman et al. 2011; Sanders et al. 2011; Artuso et al. 2011; Girirajan et al. 2011; Hitz et al. 2012; Kodama et al. 2014; Lee et al. 2014; Aerts et al. 2016). Large, rare CNVs have been overrepresented in obese individuals compared to the control cohort (Wang et al. 2010), and CNV in salivary amylase gene has been reported to predispose to obesity (Falchi et al. 2014).

Microdeletions and microduplications, sometimes accompanied by a translocation, are known to be associated with both monogenic and syndromic obesity (D’Angelo and Koiffmann 2012). Some of the examples are: a 43-kb deletion upstream of NEGR1 locus (Willer et al. 2009), two deletions in 16p11.2 region (Bochukova et al. 2010; Walters et al. 2010; Bachmann-Gagescu et al. 2010), and a de novo balanced translocation between chromosomes 1p22.1 and 6q16.2, which disrupts SIM1 (Holder et al. 2000). PWS (Prader-Willi syndrome) is associated with a deletion on 15q11.2-q13 (Driscoll et al. 1993). An unbalanced translocation involving a microdeletion of the distal part of 15q and a microduplication of the distal part of 18 has been reported to lead to Prader-Willi-like phenotype (Dello Russo et al. 2016) and a deletion at chromosome 11p13 is associated with Wilms tumour, anorexia, ambiguous
genitalia and mental retardation (WAGR) syndrome (Hingorani and Moore 1993).

Syndromic forms of obesity

PWS is an autosomal dominant disorder characterized by hyperphagia and obesity, mental retardation and muscular hypotonia. Its most frequent causes are a paternally inherited deletion, maternal uniparental disomy and, rarely, an imprinting defect in 15q11.2-q13 region (PWS/Angelman syndrome region) (Driscoll et al. 1993). The exact effect of the deletion is not clear, but the phenotype can be explained by impaired hypothalamic development and increased ghrelin production (Cummings et al. 2002), which, probably, overstimulates the NPY/AgRP neurons of the ARC.

Mutations leading to deletion or disruption of the \textit{SIM1} gene (encodes a transcription factor that plays an important role in neurogenesis) lead to hyperphagia (Holder et al. 2000). Mouse models link the hyperphagic phenotype to impaired development of the PVN, crucial for appetite and energy expenditure control (Michaud et al. 2001).

Pseudohypoparathyroidism type 1A (PHP1A) syndrome is caused by a maternally transmitted mutation in \textit{GNAS1} (encodes an \(\alpha\)-subunit of the Gs protein). The resulting variant of the Gs protein is, probably, affecting the functions of multiple GPCRs in the hypothalamic circuitry that controls appetite and energy balance (Spiegel and Weinstein 2004).

WAGR syndrome is caused by a deletion at chromosome 11p13 that encompasses the \textit{WT1} and \textit{PAX6} genes (Hingorani and Moore 1993) and can encompass \textit{BDNF} (Takada et al. 2017). This syndrome leads to obesity phenotype.

Cohen syndrome is an autosomal recessive disorder characterized by mild childhood obesity, mental retardation, small stature and microcephaly (Cohen et al. 1973). It has been demonstrated that \textit{COHI} gene locus is associated with Cohen syndrome (Chandler 2003).

Bardet–Biedl syndrome (BBS) is characterised by early-onset obesity, polydactyly, learning disabilities, dyslexia, progressive rod–cone dystrophy, hypogonadism and progressive renal abnormalities (O’Rahilly and Farooqi 2006). Multiple BBS loci have been identified: BBS1 on 11q13, BBS2 on 16q21, BBS3 on 3p13–p12, BBS4 on 15q22.3–q23, BBS5 on 2q31, BBS6 on 20p12, BBS7 on 4q27 and BBS8 on 14q32.11 (O’Rahilly and Farooqi 2006).

Fragile X syndrome is associated with mild obesity, severe mental retardation, macrocephaly, prominent jaws and high pitch in voice. It is caused by expansion of a polymorphic CGG trinucleotide repeat in the fragile X mental retardation gene (\textit{FMR1}) (Hagerman and Hagerman 2004).
Non-syndromic obesity

**Monogenic forms of obesity**

Monogenic forms of obesity are caused by mutations of a single gene and are rare, severe and manifest in early childhood. The genes causing this type of obesity are involved in regulation of food intake, energy metabolism and adipogenesis. Mutations in genes encoding leptin (Montague et al. 1997; O’Rahilly and Farooqi 2006) and its receptor (LEPR) (Clément et al. 1998), POMC (Krude et al. 1998; O’Rahilly and Farooqi 2006) and prohormone convertase 1 (PC1) that is involved in processing of POMC (Jackson et al. 1997) lead to rare recessive forms of obesity caused by increased energy intake and reduced expenditure. Multiple MC3R and MC4R mutations have been associated with obesity (Singh et al. 2017). Dominant autosomal mutations in MC4R (Yeo et al. 1998) are one of the most frequent causes of monogenic obesity. Ghrelin receptor mutations have been associated with obesity and short stature (Wang and Tao 2016). Multiple variants of the FTO gene locus have been implicated in development of obesity (Singh et al. 2017). Mutation in BDNF gene (Martínez-Ezquerro et al. 2017; Harcourt et al. 2018) and BDNF receptor TrkB (NTRK2) has been associated with extreme hyperphagia and obesity (Yeo et al. 2004).

**Polygenic (common) forms of obesity**

Association studies have identified hundreds of loci associated with obesity (Loos 2018). While the exact functions of some loci remain obscure, the joint effort of multiple research groups has already provided knowledge about many loci. Below are some examples of known loci contributing to non-syndromic polygenic obesity.

Mutations in α- and β-adrenergic receptor genes (ADRA2B, ADRB1, ADRB2 and ADRB3) lead to obesity (Bell et al. 2005), since these receptors are involved in energy and lipid metabolism (Singh et al. 2017). It has been demonstrated that genetic variation in uncoupling protein (UCP) genes have been associated with energy metabolism regulation (UCP1 and UCP2) and obesity (UCP2 and UCP3). Mutations in the SLC6A14 gene lead to obesity, probably by affecting appetite control and energy balance (Suviolahti et al. 2003), since this gene encodes a tryptophan transporter that is necessary for serotonin synthesis. The common SNP in the proximal promoter of the adiponectin gene (ADIPOQ) that is known to influence adiponectin levels, has been associated with severe childhood and adult obesity (Walley et al. 2009). Obesity has been associated with genes encoding cannabinoid receptor 1 (CNR1), dopamine receptor 2 (DRD2), and genes involved in regulating serotonin function, namely the serotonin transporter (SLC6A4) and serotonin receptor 2C (HTR2C) (Walley et al. 2009). Common sequence variants in coding and regulatory sequences of NPY (Bray et al. 2000; Karvonen et al. 2001;
Ding et al. 2005) as well as NPY2R and PYY (Ma et al. 2005; Siddiq et al. 2007) have been associated with obesity.

SNPs and CNV of NPY4R and its role in appetite regulation

A genetic screen of a cohort of obese German children carried out by the genetics company IntegraGen found an association between a genomic region on chromosome 10q11.22 and obesity. NPY4R was the strongest candidate for association with obesity among the genes located in the region, and was found to have nonsynonymous single nucleotide polymorphisms (SNPs) that segregated with childhood obesity. We have characterized two receptor variants pharmacologically and found that one of them displayed reduced signal transduction as compared to the most common receptor sequence, the wildtype (Sjödin 2005). The second receptor variant displayed a tendency towards reduced signalling, but did not reach statistical significance. Further studies in our group have revealed deviation of SNP frequencies, within the NPY4R region, from Hardy-Weinberg equilibrium (unpublished) and led to our studies of NPY4R copy number variation (CNV).

Later, a CNV region on chromosome 10q11.22 was described by several research groups (Sebat et al. 2004; Park et al. 2010; Wang et al. 2010; Sudmant et al. 2010; Jarick et al. 2011; Artuso et al. 2011; Aerts et al. 2016) and was reported to span approximately 194 kb (Sha et al. 2009) across NPY4R, SYT15 and GPRIN2 genes. SYT15 encodes the membrane trafficking protein synaptotagmin 15 that, unlike other members of the synaptotagmin family, is expressed outside of the nervous system (lungs and testis) (Fukuda 2003). GPRIN2 encodes G protein regulated inducer of neurite outgrowth 2. It interacts preferentially with activated forms of subunits of the Gi subfamily (Iida and Kozasa 2004), but its exact functions are unknown.

This genomic region contains multiple repeated elements and its organization has not been fully resolved and mapped in the current version of human genome assembly (Hg38). Most of the previous reports indicate that the normal copy number of NPY4R is 2 copies per genome (Sha et al. 2009; Park et al. 2010; Jarick et al. 2011; Sun et al. 2013; Aerts et al. 2016) and describe the CNV as either gain or loss, not always specifying the exact copy number.

The first study of CNV in NPY4R region and association with BMI reported an inverse correlation: a higher gene copy number was associated with reduced BMI in an elderly Chinese cohort (Sha et al. 2009). Subsequent studies had contradicting conclusions regarding whether weight gain is associated with copy number loss or copy number gain. An inverse correlation of the CNV in this region and BMI has been demonstrated in a German cohort (Jarick et al. 2011) and in a Belgian cohort of children and adolescents with obesity and healthy adults with normal weight (Aerts et al. 2016), while a study of Italian patients with Rett syndrome has demonstrated a positive
correlation (Artuso et al. 2011). Finally, a study of young Chinese individuals did not detect either such association or CNV of \textit{NPY4R} (Sun et al. 2013).
Aims

The overall aim of this thesis is to investigate the genetic and functional variation of the human pancreatic polypeptide receptor Y4, with focus on copy number variation and single nucleotide polymorphisms, their relationship to body weight and impact on functional properties of the receptor.

Specific aims for the different papers:

- Paper I. To confirm the \( NPY4R \) CNV and to investigate the \( NPY4R \) copy number in different populations, as well as to compare read depth analysis and droplet digital PCR as tools for CNV analysis.

- Paper II. To investigate the relationship between the \( NPY4R \) copy number and body weight as well as energy intake in a group of Swedish individuals with wide range of BMI.

- Paper III. To investigate the relationship between the \( NPY4R \) copy number and body weight in a cohort of Northern Swedes and to study the effect of the nonsynonymous Y4 variants on body weight.

- Paper IV. To functionally characterise naturally occurring variants of pancreatic polypeptide receptor Y4, encoded by the \( NPY4R \) gene.
Materials and methods

Copy number investigation

The NPY4R copy number was analysed in order to validate previous findings of the presence of the NPY4R CNV, possible geographical and age differences in the NPY4R CNV, and to investigate the range of the CNV. Two different methodological approaches were used for CNV detection: read depth analysis (Papers I and IV) and ddPCR (Papers I, II and IV). The two methods were compared. Read depth analysis was performed on:

- 66 genomes from the 1000 Genomes Project (Abecasis et al. 2010);
- three high-coverage genomes of archaic hominins (two Neanderthals and one Denisovan);
- 1009 genomes from the Northern Sweden Population Health Study (NSPHS) cohort.

Three different read depth methods were used: CNVrd2, Control-FREEC and CNVnator. First, the total read base count for the region containing the NPY4R was extracted. Then average read depth across the region was computed and compared to the average read depth across the reference region (chromosome 10 or autosomes).

Droplet digital PCR was performed on:

- 18 DNA samples from the 1000 Genomes Project;
- 558 DNA samples from the Swedish Obese Subjects (SOS) study (Sjöström et al. 2004), the SOS reference (SOS Ref) study (Larsson et al. 2004) and the SibPair cohort (Carlsson et al. 2009);
- 61 DNA samples from the NSPHS cohort.

The ddPCR is based on digesting and separating different copies of region of interest into 1 nL droplets, amplifying the region of interest by PCR and detecting a fluorescent signal from hydrolysis probes. One such probe was designed for a gene of interest (the NPY4R) and the other one for a reference gene (the H1 component of the RNaseP encoded by RPPH1). Final data evaluation was performed using Poisson statistics.

Comparisons of NPY4R copy number generated by different methods were performed. The degree of correlation between NPY4R copy number data generated by read depth and ddPCR was calculated using Spearman correlation. Statistical analysis was performed using SPSS version 22.0.
The duplication unit and the *NPY4R* gene organization investigation

Eight bacterial artificial chromosome (BAC) clones and seven fosmid clones from BACPAC Resource Center (Children's Hospital Oakland Research Institute, Oakland California, USA) along with four yeast artificial chromosome (YAC) clones (Table 2) from ECEPH B Mega YAC library (kindly provided to us by IntegraGen) were used in order to investigate the genomic neighbourhood of the *NPY4R* gene and duplication unit organization. BAC and fosmid clones were sequenced using IonTorrent technology (SciLifeLab, Uppsala, Sweden). YAC clones were first sequenced using SOLID sequencing technology and later with long-read PacBio method (SciLifeLab, Uppsala, Sweden).

Table 2. Clones used in the duplication unit organization investigation

<table>
<thead>
<tr>
<th>BAC clones</th>
<th>Fosmid clones</th>
<th>YAC clones</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP11-314P12</td>
<td>WI2-630M17 (G248P80524G9)</td>
<td>912-f-10</td>
</tr>
<tr>
<td>CH17-19K23</td>
<td>WI2-1834P9 (G248P85397H5)</td>
<td>917-c-2</td>
</tr>
<tr>
<td>CH17-252I16</td>
<td>WI2-1889A21 (G248P86916A11)</td>
<td>859-e-2</td>
</tr>
<tr>
<td>CH17-335B8</td>
<td>WI2-3004Q19 (G248P89816H10)</td>
<td>831-a-4</td>
</tr>
<tr>
<td>CH17-351H10</td>
<td>WI2-275218 (G248P8318E4)</td>
<td></td>
</tr>
<tr>
<td>RP11-1151E16</td>
<td>WI2-2743D4 (G248P8327B2)</td>
<td></td>
</tr>
<tr>
<td>RP11-1105L11</td>
<td>WI2-1243G6 (G248P85122D3)</td>
<td></td>
</tr>
<tr>
<td>RP11-319C16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Quantitative polymerase chain reaction (RT-qPCR) was used to explore the *NPY4R* gene organization in intestinal adenocarcinoma cell lines colony (col) 1, 6 and 24 (kindly provided by professor Helen Cox, King’s College, London). Total RNA from col1, col6 and col24 cell cultures was extracted with RNeasy Mini Kit (Qiagen) and treated with DNase I (Fermentas) according to manufacturer’s protocol. For all qPCRs iScript™cDNA synthesis kit (Bio-Rad) was used. Reverse transcriptase (RT) as well as non-RT samples were diluted to the equivalent of 10ng/μl and 1 μl of sample was used. RT-qPCRs were performed in a CFX96 RT-PCR detection system (BioRad) using IQ SYBR Green Supemix (BioRad) and exon-specific primers (Table 3). All reactions were run in triplicate for each primer pair and no template controls were included.
<table>
<thead>
<tr>
<th>Name</th>
<th>Primer</th>
<th>Sequence</th>
<th>Product length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y4 human Exon1/2 tr1</td>
<td>Forward</td>
<td>TCTCCTCACCTGAGACATGG</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CTTCCACCAGGAAGATG</td>
<td></td>
</tr>
<tr>
<td>Y4 human Exon2/3 tr1</td>
<td>Forward</td>
<td>TGAGAATCTGCAACCTGTG</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CAGGAGGTGAGGTTCA</td>
<td></td>
</tr>
<tr>
<td>Y4 human Exon1/2 tr2</td>
<td>Forward</td>
<td>AGAGAATCTGAGAGGCCTCA</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CAGGAGGTGAGGTTCA</td>
<td></td>
</tr>
<tr>
<td>Y4 human Exon 1/3</td>
<td>Forward</td>
<td>TCTCCTCACCTGAGACATGG</td>
<td>139</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CAGGAGGTGAGGTTCA</td>
<td></td>
</tr>
<tr>
<td>Y4 human Before STOP</td>
<td>Forward</td>
<td>TGGCTTTTCTCAACACAACT</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CCAACGCTCCTCCACTTAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCACGTCCCTCAGTGGAAATG</td>
<td></td>
</tr>
<tr>
<td>Y4 human 3'UTR</td>
<td>Forward</td>
<td>CCAGGCTCTTCCACTTAG</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CAGTCTTCCAGTGGAAATG</td>
<td></td>
</tr>
<tr>
<td>ACTB</td>
<td>Forward</td>
<td>AGAAATCTGGCAACACACC</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CCGGGGTGTTGAAGGTCTCAAA</td>
<td></td>
</tr>
<tr>
<td>TBP</td>
<td>Forward</td>
<td>TGCACAGGAGCCAGAGTGA</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CACATCACAGCTCCCCACCA</td>
<td></td>
</tr>
<tr>
<td>B2M</td>
<td>Forward</td>
<td>CTGTTCACAGGCCAGATAG</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CAATCCAAATGCGGCATCTTC</td>
<td></td>
</tr>
</tbody>
</table>

**NPY4R copy number, body mass index and energy metabolism investigation**

The relationships between the **NPY4R** gene and BMI (Papers II and III), waist circumference (WC) and energy intake (Paper II) were explored in a study sample of 558 individuals from SOS, SOS Ref and SibPair studies (Paper II) and in 1009 individuals from the NSPHS cohort (Paper III). Weight, height and WC of the participants were objectively measured, while food intake data was obtained from a semi-quantitative dietary questionnaire on habitual food and beverage intake covering the last three months (Lindroos et al. 1993; Lissner et al. 1998). The participants specified intake frequency from standard portions of different foods and photographs of portion sizes of cooked meals. These questionnaire data were used to calculate total energy intake, energy intake normalized to body weight, energy intake for carbohydrates, proteins and fat, energy percent from carbohydrates, proteins and fat and energy intake from different food groups.

For normally distributed data, Pearson’s rho was used for correlation analysis and linear regression was used to investigate the relationship between the variables. Non-normally distributed dietary data was analysed using Mann-Whitney U-test. False discovery rate analysis was performed in order to correct for multiple testing.
Pharmacological analysis of naturally-occurring Y4 variants

PCR for SNP detection

Taq DNA Polymerase (Invitrogen) was used according to manufacturer’s instructions to perform PCRs on genomic DNA samples from 1000 Genomes. Briefly, all reactions were composed of 1X PCR buffer, 1.5 mM MgCl2, 0.4 mM dNTPs, 0.4 μM forward and reverse primers, 0.05% W1, 5U Taq DNA Polymerase (Invitrogen) and 100 ng gDNA. Running conditions were optimized for different primer combinations (see Supplementary Table 1 in Paper IV). All PCR products were sequenced using Sanger sequencing method (Eurofins, Germany).

Pharmacological investigation

Twelve naturally occurring amino acid changing SNPs were found in the samples from 1000 Genomes and the NSPHS cohort. These variants of the Y4 receptor were pharmacologically characterized in comparison to the wild-type receptor. Corresponding mutant Y4 plasmids were synthesized and used for transient transfection of human embryonic kidney (HEK293) cells. In order to measure the functional response of each receptor variants, the cells were cotransfected with a chimeric Gαq-protein (courtesy of E. Kostenis) that has the last four amino acids replaced by the corresponding amino acids from Gαi-protein and can change a Gαi signal transduction pathway to the Gαq pathway, leading to inositol phosphate (IP) generation (Kostenis 2002). Cells were incubated with 3H-myoinositol, harvested, stimulated with PP and the production of 3H-labeled IP was measured. Each assay was performed at least three times. Expression of each Y4 mutant was confirmed by confocal microscopy.

Non-linear regression was used to calculate the EC50 values of PP for wild-type Y4 and each mutant. The statistical analyses of pEC50 were performed using one-way ANOVA with Dunnett’s post hoc test.
Results and discussion

Methodological considerations in CNV determination

There is constant methodological development in the field of biotechnology and, doing research, we have to be aware of the limitations of different methods and learn by comparing them with one another. There are multiple methodological approaches to CNV determination: genome sequencing data analysis (read depth analysis), hybridisation array-based methods (SNP-arrays, aCGH), and PCR-based methods (RT-PCR, quantitative RT-PCR, various kinds of digital PCR). CNV determination by sequencing data analysis requires a high quality reference assembly, which can be problematic to achieve for complex repeated regions, and high sequencing coverage. Hybridisation- and PCR-based methods require a reference copy number, which is most commonly set at two copies per genome (Sha et al. 2009; Park et al. 2010, 2012; Jarick et al. 2011; Sun et al. 2013). Studies based on SNP arrays and aCGH heavily depend on relative fluorescence data quality (Chu et al. 2013) and often utilize different analysis algorithms with different sensitivity to the inherent variation in relative fluorescence between genomic loci on SNP and hybridization arrays (Talseth-Palmer et al. 2013). Another consideration is that current genotyping platforms have limited or no probe coverage for a large number of common CNVs (Cooper et al. 2008) and limited power to detect CNVs in repeat-rich and duplication-rich genome regions, such as chr10q11.22.

Read depth analysis and ddPCR

In our venture into copy number determination, we have considered different methods that were available during the course of this work. Read depth was our first choice, since the 1000 Genomes project phase I data became available. We have considered fluorescence in situ hybridisation (FISH) and qPCR, but these methods were extremely laborious. FISH also required human tissue, which is much more difficult to obtain and work with, than DNA. We have tried to use microsatellite analysis, but this approach has proven to not be suitable for CNV analysis. The most recent method we considered was ddPCR, which allows a direct molecular CNV analysis.

In this study we used two methods: read depth analysis that compares an average read count in the region of interest and in the reference region (whole
genome, chromosome of interest); and ddPCR that determines the copy number of region of interest compared to the reference gene. Each of these methods has its own advantages and disadvantages. As was mentioned above, read depth analysis heavily depends on sequencing data coverage and a reliable reference assembly, but it is relatively easy to perform, albeit costly and labour intense and therefore time consuming, and can be used to analyse genomes that have already been sequenced. Reads from genomic re-sequencing are mapped to a reference assembly and any comparison between individuals is relative to this assembly. Among the molecular methods for target nucleic acid quantification, ddPCR has demonstrated equal (Svobodová et al. 2015) or better precision and reliability in copy number determination (Bharuthram et al. 2014; Sillence et al. 2015) and was comparable to other digital PCR methods (Dong et al. 2015). The method utilizes a reference assay for a gene with known copy number (like RNaseP in this study) and normalizes the target assay data to this reference. Such an approach requires careful choice of the reference gene, but allows for more accurate determination of the absolute copy number.

Comparison and interpretation of the results

We have studied the NPY4R CNV using read depth analysis in 66 modern human genomes from the 1000 Genomes Project (Paper I) and the 1009 genomes from NSPHS cohort (Paper III). Then we compared the results from the read depth analysis with ddPCR in 18 individuals from the 1000 Genomes Project (Paper I) and 61 individuals from the NSPHS cohort (Paper III). First we have studied the NPY4R region using the CNVrd2 read depth method on phase 1 data from the 1000 Genome Project. However, these samples were only sequenced to a relatively low coverage. Then, we used three different read depth methods (CNVrd2, Control-FREEC and CNVnator) to analyse the genomes from 1000 Genome Project (phase 3) that had higher and more even coverage. Lastly, we have used single base-based read depth analysis and CNVnator to study the genomes from NSPHS cohort, with samples that were sequenced at a high (30x) coverage.

In the samples from 1000 Genomes Project, the CNV range estimates varied depending on the method: from three to six copies per genome according to CNVrd2 read depth of phase 1 data, from one to five copies per genome according to CNVrd2 read depth of phase 3 data and from three to seven copies according to Control-FREEC and CNVnator read depth analyses (phase 3 data) and ddPCR. Control-FREEC and CNVnator demonstrated a high correlation with each other (Spearman’s $\rho=0.734$, $p=3.52 \times 10^{-12}$) and with ddPCR results (Spearman’s $\rho=0.867$, $p=0.3 \times 10^{-2}$ and Spearman’s $\rho=0.8$, $p=0.68 \times 10^{-4}$ respectively). Among the NSPHS genomes read depth analysis demonstrated that the copy number for NPY4R ranged from three to seven copies and according to ddPCR it ranged from three to 11. Read depth results correlated
closely with ddPCR results (Spearman’s $\rho=0.854$, $p=2.1 \times 10^{-18}$), despite a 1.3 copy shift between the two methods. The chr10q11.22 is a complex genomic region with many repeats that are difficult to assemble. Hence, it is difficult to obtain accurate read depth results for such complex regions. Our results demonstrate that different read depth methods can produce different results for the same data set. It is important to keep in mind that both data quality and methodological differences can influence the results of the read depth analysis. Despite its limitations read depth is a useful tool for analysis of cohorts that have already been sequenced. However, determination of the absolute number of copies may be more problematic and require calibration. Based on previous studies showing that ddPCR is an accurate tool for nucleic acid quantification (Hindson et al. 2011; Pinheiro et al. 2012) and absolute copy number determination (Boettger et al. 2012; Beck et al. 2013; Tayoun et al. 2016) as well as our own observations, ddPCR can be used as a calibration tool for read depth analysis. It has been demonstrated that ddPCR is a more precise CNV determination method than read depth analysis, especially for high copy numbers (Eisfeldt et al. 2018).

**Duplication unit size and organisation**

After confirming the CNV for NPY4R, we were interested in the organisation of the genomic region surrounding the gene. Read depth analysis has confirmed that SYT15 and GPRIN2 are located within the same duplication unit, as has been previously described (Sha et al. 2009) and that the duplication unit is around 200 kb long. Subsequent YAC clone analysis indicated that the duplication unit might be much larger. In our efforts to investigate the breakpoints of the duplication unit and distinguish between copies, we determined the complete sequence of eight BAC clones and seven fosmid clones containing the human Y4 gene, using the Ion Torrent method offered by SciLifeLab at Uppsala University. The clones contained no duplicates of NPY4R and did not cover the break points between the duplication units as far as we have been able to tell. These results suggest that NPY4R duplicates are not in the immediate neighbourhood of one another and might be located some distance apart along the chromosome. YAC clones were first sequenced by SOLID sequencing, but could not be properly assembled *de novo* due to the short length of the reads. To achieve longer reads we have utilized PacBio sequencing method. One of the YAC clones (917-c-2) contained NPY4R but was chimeric, composed of chromosomes 10 and 14. We assume that the fusion might have happened naturally or during the construction of the YAC library. Clones 912-f-10 and 859-e-2 have been assembled and their analysis is in progress. We expect these *de novo* assemblies to help us resolve the structure of the NPY4R region, including the organisation, orientation and break points of the duplication units.
NPY4R gene: organization, most common copy number state and CNV range (Papers I, II and III)

Splice variants

According to the current version of the Ensembl data base (GRCh38.p5) NPY4R has three predicted splice variants – either all three exons are spliced or only the first 5’UTR exon and coding 3’ exon or only a truncated version of the second 5’UTR exon with the coding exon (Figure 3B, mRNAs 1, 3 and 4). As was mentioned above, our previous analysis of the splice variants in adipose mRNA suggested that either of the 5’UTR exons was spliced onto the coding 3’ exon (Figure 3B, transcripts 2 and 4). Using the human adenocarcinoma colony cell lines that express NPY4R we have identified mRNAs that contain the second 5’UTR and coding exons. It was not possible to say whether this mRNA contained full or a truncated exon 2 or if both exist. It is possible that different tissues express different transcripts and mRNAs and more studies are needed to identify them.

NPY4R CNV characterisation

Our findings regarding the most common copy number of NPY4R and its range differ from previously published results. Several research groups have reported that the most common copy number of NPY4R is two copies per diploid genome (Sha et al. 2009; Park et al. 2010; Jarick et al. 2011; Sun et al. 2013) and a study of young Chinese cohort did not detect any CNV in NPY4R region (Sun et al. 2013). Based on observed Hardy-Weinberg disequilibrium of SNP frequencies observed in the NPY4R region it is clear that there are more than two copies of this gene in diploid genome. The results of read depth analyses and ddPCR (with two carefully chosen reference genes) described in this work demonstrate that four and five copies per diploid genome are the most common genotypes.

We have studied the NPY4R copy number in a Denisovan genome and two Neanderthal genomes (Paper I), and found that the archaic hominins had four copies of the NPY4R gene. This led us to the conclusion that the duplication of this genomic region happened before the split of modern humans from the Neanderthals and the Denisovans between 400,000 and 800,000 years ago (Langergraber et al. 2012). In all modern human genomes investigated (Papers I, II and III) four was the most frequent copy number. These results agree with the most recent human genome assembly (GRCh38), where two copies of NPY4R are placed on the same chromosome, indicating that four copies per genome is the most likely common copy number.

As was mentioned above the CNV range estimates depended on the method. In the samples from 1000 Genomes it ranged from three to seven copies per genome according to read depth (Control-FREEC and CNVnator)
and ddPCR (Paper I). In the mixed Swedish study sample with a wide range of BMI (558 individuals from SOS, SOS Ref and SibPair cohorts) the CNV was determined by ddPCR and ranged from two to eight copies (Paper II). Finally, in the NSPHS samples the copy number for NPY4R ranged from three to seven according to the read depth and three to as much as 11 according to ddPCR (Paper III).

Since there are two contradictory studies of Chinese cohorts, one demonstrating the CNV on chr10q11.22 in an elderly Chinese cohort (Sha et al. 2009) and one reporting no CNV in a young Chinese cohort (Sun et al. 2013). We have studied 9 CHB individuals (Han Chinese in Beijing, China) from the 1000 Genomes Project (Paper II). Both our read depth analysis and ddPCR showed that the NPY4R copy number varied between four and five copies per genome. More individuals are needed to make reliable conclusions at the populations level, but even with such a small sample we were able to demonstrate the presence of NPY4R CNV in individuals of Chinese origin.

NPY4R copy number, body mass index and energy metabolism (Papers II and III)

As it has been reported that Y4 promotes satiety (Batterham et al. 2003) and we had confirmed that the gene encoding Y4 displays CNV, we investigated the correlation of the CNV with body weight and energy metabolism. Several research groups reported that NPY4R CNV correlated with BMI (Sha et al. 2009; Jarick et al. 2011; Artuso et al. 2011; Aerts et al. 2016) or was identified in individuals with obesity only (Wang et al. 2010). The majority of the studies have demonstrated a negative correlation: low copy number correlated with higher BMI and low copy number high BMI (Sha et al. 2009; Jarick et al. 2011; Aerts et al. 2016). However, a study of a young Chinese cohort did not detect a CNV in NPY4R region and, hence, did not observe any correlation between CNV and BMI (Sun et al. 2013). The negative correlation between the CNV and BMI was expected and seems logical at first look: the fewer copies of the gene, the less of the receptor for satiety, the more weight gain. Therefore, it came as a big surprise when we see a positive correlation in a Swedish study sample (Paper II) and even more surprising when we did not find a correlation with BMI in a Northern Swedish cohort (Paper III).

First, we have analysed the effect of the NPY4R CNV on BMI and waist circumference in 558 Swedish individuals (216 men and 342 women) (Paper II). This study sample was comprised by individuals participating in the SOS, the SOS-ref and the SibPair studies, in order to cover a wide range of BMI. Copy number was determined by ddPCR. Our results suggest a strong positive correlation between NPY4R copy number and BMI for men and women together, and even stronger for women separately, but not significant for men.
alone. Every additional NPY4R copy increased BMI by 2.60 kg/m² in women. In comparison, each FTO risk-allele increases BMI by 0.20-0.66 kg/m² (Loos 2009; Willer et al. 2009). Waist circumference correlated with NPY4R copy number for men and women together and for women separately, but not for men alone. An increase of one copy was associated with 5.67 cm increase in waist circumference for women. These results indicate that the NPY4R copy number influences body weight in women but probably not, or only marginally, in men.

Since Y4 supposedly regulates satiety, we investigated the correlation of the NPY4R CNV with energy intake (Paper II). We observed no correlation between total energy intake, energy intake from specific food groups or energy percent from carbohydrates, proteins and fat and NPY4R copy number, neither for the whole group nor for men and women separately. Total energy intake adjusted to body weight had a strong and statistically significant negative correlation with the NPY4R copy number in men and women together and in women separately, but not in men. Each additional copy was associated with a decrease of 3.54 kcal/kg in women. These results may indicate that NPY4R influences body weight through metabolic pathways, rather than by influencing food intake itself. Even though dietary questionnaires provide reproducible data on food intake, the physical and psychological state of study participants as well as social desirability can influence self-reported dietary measures, which can result in specific underreporting or overestimation of the intake of certain food types. Another complication of energy intake analysis is that the observed large variation in energy intake from the specific food groups may obscure any correlations between CNV and total energy intake or energy from macronutrients.

We have also analysed the relationship between the NPY4R copy number and BMI in a study sample from Northern Sweden (Paper III). We have done read depth analysis in 1009 individuals (474 men and 535 women) from NSPHS cohort. Surprisingly, the NPY4R copy number displayed no correlation with BMI in this cohort, neither for the whole group nor for men and women separately.

Thus, our results differ between the studies described in Paper II and IV and are different from the previously published studies carried out by other groups. The differences might be explained by study design (including cohort selection and CNV detection methods), possible type II error (getting a false-positive result by chance) and true cohort differences. In order to give a fuller explanation, it is necessary to analyse NPY4R CNV in several large cohorts and to keep in mind that CNV and weight may correlate differently across geographical regions depending on genetic background. It is also important to mention that the role of Y4 and PP in appetite regulation is not very clear, since there is evidence of the receptor/ligand pair being involved in promoting satiety in animals and humans (Asakawa et al. 1999, 2003; Batterham et al. 2003; Balasubramaniam et al. 2006; Jesudason et al. 2007; Lin et al. 2009),
but there are human studies showing that PP might have a more complex role. A long-term study of diet-induced weight loss in adults has demonstrated that even a year after weight loss the hunger hormones were elevated (as if trying to get the body back to its original higher weight at the start of the study) and so was PP (Sumithran and Prendergast 2011), as if PP works more like a hunger hormone that a satiety hormone. In contrast, the satiety hormones such as leptin, CCK, insulin and amylin were reduced. Similar results of PP have also been demonstrated in children (Reinehr et al. 2006). Lastly, when trying to disentangle the effect of the NPY4R CNV on body weight and energy metabolism, it is important to keep in mind that it is not known how many and which of the copies are expressed, if any of the copies have mutations altering the functions of Y4 receptor and if there is tissue-specific expression of different copies. We already know that there is a potential adipocyte enhancer formed by a primate-specific Alu repeat between the first and the second 5’UTR exons (Figure 3). It might be possible that some NPY4R copies have accumulated other mutations that influence the gene expression. It is not yet possible to differentiate between the copies because their genomic surrounding has not been assembled with sufficient accuracy. There could also be regulatory mechanisms induced by the receptor protein. For instance, a cell with a high gene copy number may initially have a high level of receptor expression that autoregulates by negative feedback and eventually leads to a lowered expression level. Regarding SNP variation, see below.

**NPY4R mutations, Y4 functions and body weight (Papers III and IV)**

CNV is not the only type of genetic variation present in NPY4R. In order to investigate SNPs in NPY4R, we have studied 24 samples from the 1000 Genomes Project by performing a number of overlapping PCRs, followed by Sanger sequencing (Paper IV). These samples had between three and seven copies of the NPY4R gene. We found a total of 24 SNPs in the coding region, out of which seven were nonsynonymous (i.e., they changed the amino acid). We have also analysed the genotype frequencies from the sequence data of the NSPHS (Paper III). There we found 18 SNPs, out of which eight were nonsynonymous. The SNPs that change amino acids were introduced separately by mutagenesis of the wildtype Y4 receptor in an expression vector and the resulting constructs were functionally expressed in transfected cells in vitro. Three mutants, Cys201TyrECL2, Val271Leu6.41 and Asn318Asp7.49, showed no functional response, which may be explained by altered conformational state of the receptors making them unable to respond to the peptide agonist, since the mutant receptors were verified to be expressed by the transfected mammalian cells. We observed a statistically significant decrease in the potency of PP
on mutants Cys34SerNT and Val135Met3.46, while other mutants did not have a significant impact on receptor response. The Val271Leu6.41 mutant has been previously reported to have a decreased receptor activity in response to PP (Aerts et al. 2016) but showed no functional response in our study.

Most of the nonsynonymous variants in the NSPHS samples were rare, being present in single individuals or a few, and none of them exhibited any association with body weight in this study group (Paper III). We did not detect any cumulative effect of minor alleles on BMI. Since the rare alleles are present in individuals with variable NPY4R copy number, it is impossible to know how many gene copies carry the rare variants. Resolving the effect of rare variants of the NPY4R gene will be challenging because it is not known if and which of the rare variants are expressed. It is possible that the most common variants either compensate for, or dilute, any effect of the rare variants.

Inconsistent results of genetic association studies in obesity – FTO and AMY

The inconsistent results of NPY4R copy number association with BMI are not unique. There are other genes that exhibit complex genetic variation and whose role in obesity has been difficult to assess. Two such problematic genes are FTO and AMY1.

FTO gene

In 2007, a SNP in the fat mass and obesity-associated (FTO) gene was associated with obesity (Frayling et al. 2007; Dina et al. 2007). It is still debated whether the association with obesity is caused by FTO itself or the neighbouring genes and whether FTO expression levels in adipose tissue are positively or negatively associated with BMI.

FTO encodes an AlkB-like 2-oxoglutarate–dependent nucleic acid demethylase (Gerken et al. 2007) whose exact functions remained unknown until a recent study demonstrated that FTO regulates fat metabolism through demethylation. FTO has been shown to down-regulate m^6^A levels, decrease mitochondrial content, and increase triglyceride deposition in hepatocytes (Kang et al. 2018), which indicates functions in energy metabolism. FTO mRNA was shown to be expressed, among other places, in hypothalamic nuclei, suggesting roles in food intake and energy metabolism regulation (Gerken et al. 2007).

Three studies have demonstrated contradictory results; FTO mRNA levels in visceral adipose tissue in humans demonstrated both negative (Klöting et al. 2008) and positive association with obesity (Wåhlén et al. 2008; Grunnet et al. 2009). The later studies have also demonstrated that the FTO genotype
did not affect FTO mRNA levels in adipose tissue (Wåhlén et al. 2008; Grunnet et al. 2009). It has been suggested that the noncoding SNPs in introns 1 and 2 of FTO do not affect its expression, but rather interact directly with the promoter of a homeobox gene, IRX3, and are associated with the expression of IRX3 in human brain. IRX3 is expressed in multiple regions of the brain, including hypothalamus, which connects it to body weight regulation (Smemo et al. 2014). One of these intronic FTO SNPs has also been reported to disrupt a conserved motif for the ARID5B repressor, which leads to de-repression of a potent preadipocyte enhancer and doubles IRX3 and IRX5 expression during early adipocyte differentiation. This results in a developmental shift from energy-dissipating beige adipocytes to energy-storing white adipocytes, involved in lipid storage and increase of body weight (Claussnitzer et al. 2015).

Amylase gene

The amylase gene cluster encodes the salivary (AMY1) and pancreatic (AMY2) amylase. In 2014, a multi-allelic CNV encompassing the amylase gene cluster was associated with BMI and obesity, suggesting that lower copy number of AMY1 in adults and children predisposes to obesity (Falchi et al. 2014; Viljakainen et al. 2015; Mejia-Benítez et al. 2015; Marcovecchio et al. 2016). Later, combined AMY1 and AMY2 CNV studies have demonstrated that there is a difference between amylase haplotypes. According to these studies, the majority of amylase haplotypes worldwide contain odd numbers of AMY1 repeat units (an even copy number per diploid genome) (Carpenter et al. 2015; Usher et al. 2015; Eisfeldt et al. 2018). These haplotypes partially correlate with nearby SNPs, which do not exhibit association with BMI (Usher et al. 2015). A study of two East Asian cohorts also questions the role of this CNV, since it reports no association between AMY1 copy number and BMI (Yong et al. 2016). Interestingly, the haplotypes that carry even numbers of AMY1 were associated with CNV of the AMY2A and AMY2B genes (Carpenter et al. 2015). A later comparison of female carriers of low AMY1 copy number to carriers of high copy number explored the metabolic pathways affected by this CNV. It suggested an increased metabolic reliance on fatty acids in low carriers (reduced cellular glucose uptake and shift towards lipid-based energy production) (Arredouani et al. 2016), which strengthens the CNV role in energy metabolism regulation. One of the most recent studies reports a negative association of AMY1 and AMY2 enzyme activity and BMI, a negative association of AMY1 copy number and obesity risk in children, and no association between AMY2A copy number and BMI (Bonnefond et al. 2017). The other study demonstrates no direct effect of AMY1 copy number on BMI and an interaction with starch intake and BMI (Rukh et al. 2017).
Concluding remarks and future perspectives

Obesity is an extremely complex disease with high genetic heritability. Exception for a few rare monogenic forms, obesity is usually highly polygenic, which means that each contributing gene makes a very small contribution to the disease. But every drop of knowledge brings us closer to understanding and, hopefully, counteracting this global problem.

This thesis focuses on the *NPY4R* gene that encodes Y4, a receptor for human pancreatic polypeptide (PP). The *NPY4R* gene is located in a highly complex genomic region that has proven to be exceedingly difficult to sequence and assemble due to multiple repeats and duplications. The results of several different types of read depth analyses as well as ddPCR confirm previously controversial results regarding the presence of a CNV of this genomic region. They also demonstrate that the most common copy numbers of *NPY4R* are 4 or 5 copies per genome, in contrast to the two copies presumed by all previous studies by other investigators. The read depth analyses suggest that *NPY4R* is part of a large duplication unit (approximately 200 kb or more) that also includes *SYT15* and *GPRIN2* and demonstrated a duplication unit CNV range of three to seven copies. The ddPCR results found a range of two to eleven copies of the *NPY4R* per genome. Investigation of archaic hominins (Denisovan and Neanderthal individuals) revealed that, just as modern humans, they had four copies of *NPY4R*, indicating that the duplication happened before the divergence of modern humans from the common ancestor of Denisovans and Neanderthals.

The results of this work as well as previous research suggest that ddPCR is a robust, precise and reliable method for CNV analysis. In a study of 558 Swedish individuals from SOS-Ref, SOS and Sib-Pair cohorts, the most common copy number for the individuals with normal weight was found to be four copies per genome. The *NPY4R* copy number exhibited a positive correlation with BMI and waist circumference for men and women together and for women only. There was no correlation between the *NPY4R* copy number and BMI or weight circumference for men only. No correlation between dietary intake and the *NPY4R* copy number was found for men and women together or separately. On the other hand, a study of 1009 individuals from Northern Sweden did not find a correlation between BMI and *NPY4R* copy number. Different explanations for such diverging results are conceivable: there may exist a geographical variation in the correlation between *NPY4R* copy number and BMI, life style differences may influence BMI, lack of power in the study
of 558 individuals as well as studies by others may have led to false positive results, or lack of power in our study of 1009 individuals may have led to a false negative result.

The investigation of naturally-occurring amino acid variants of Y4 in 24 samples from the 1000 Genomes Project and 1009 samples from the Northern Sweden Population Health Study (NSPHS) has revealed 12 variants that were functionally expressed and studied pharmacologically. Three of the variants led to non-functional Y4 receptors, while two variants displayed decreased functional response to PP. Using the recently determined crystal structure of the human Y1 receptor, a 3D model of the Y4 receptor was generated. The functional responses of the receptor variants are in agreement with both the locations of the amino acid residues in the 3D model and with the degree of conservation shown by evolutionary sequence comparisons.

This work strengthens the position of the NPY4R gene as a factor that may influence body weight, and invites further studies to explore possible population differences. Several more cohorts of both similar and different origin would help to shed light upon the relationship between NPY4R copy number and body weight. A better understanding of the physiological action of Y4 is needed, such as receptor expression pattern, to which extent amino acid variants change the response to PP in vivo, whether the nodose ganglion of the vagus nerve is a primary site of synthesis of Y4, and if PP can reach brain regions that have a fenestrated blood-brain barrier. Conditional Y4 knockout and Y4-overexpressing animal models would greatly aid in our understanding of the role of NPY4R in the regulation of body weight and energy metabolism regulation.

Finally, this work has attempted to investigate differences between the duplication units of this CNV region, but despite the efforts of sequencing fosmids, BACs and YACs this task has proven to be extremely challenging. More extensive sequencing efforts with long-read methods and high coverage are necessary to resolve the structure of this region, distinguish the duplication units, assemble them correctly and start to investigate different haplotypes. Since the gene region displays both CNV and SNPs further studies are needed to resolve how SNPs and CNV are superimposed upon one another and how the different versions of the NPY4R gene may contribute to variation in body weight.
Acknowledgements

Many people have contributed to this work by guiding, helping, teaching, collaborating, inspiring and supporting me.

First, I would like to express my gratitude to my supervisor Dan Larhammar, who took me in and let me grow, and my co-supervisor Lars Feuk, who supported and advised me along the way. I also want to acknowledge my collaborators Johanna Andersson-Assarsson, Ingrid Larsson and Lena Carlsson for our work on Paper II. Torsten Günther, Anna Johansson, Mattias Jakobsson and Khurram Maqbool for their contributions to Paper I. Åsa Johansson and Nima Rafati for their major work on Paper III. Silvana Vasile and Hugo Gutiérrez-de-Terán for making the Y4 model in Paper IV.

I would like to thank all the present and past members of our group. Especially, to Christina Bergqvist for all her lab advice and our own “project talks”, which were the best. Daniel Ocampo Daza for all the moral support (including late night beverages) and practical advice. David Lagman for turning from a lab mate into a friend. Bo Xu for all the pharmacology we did together and for introducing me to the Golden China, which had a very big impact on my PhD life. Chus Abalo for our conversations and all the energy you were giving me. Julia Pedersen for the feeling of PhD comradery and making the last year a bit more fun.

I also want to give many thanks to all my teachers, colleagues, teaching mates, project students and friends.

And finally, a separate and very special acknowledgement of my family’s contribution. Thank you, Mom, Oleg and Bobby, for your support, help and for always inspiring me to be my best self.
References


Rodríguez EM, Blázquez JL, Guerra M (2010) The design of barriers in the hypothalamus allows the median eminence and the arcuate nucleus to enjoy private milieus: the former opens to the portal blood and the latter to the cerebrospinal fluid. Peptides 31:757–76.


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1479

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. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-356573