Rare Mendelian Forms of Obesity and Diabetes and their Implications for Treatment Outcomes

Thesis submitted by

Sumaya Obaid Almansoori

For the degree of Doctor of Philosophy

Department of Metabolism, Digestion and Reproduction

Faculty of Medicine

Imperial College London

2021

Declaration of originality

I hereby declare that research presented and submitted here is my own work. I have acknowledged and referenced all work and sources (published and unpublished) that are related to others. Any work resulted from collaborations and assistance have been appropriately mentioned and acknowledged in the text.

Copyright

Abstract

Obesity and diabetes are becoming epidemic health issues worldwide. In recent decades, a considerable amount of work has been done to study the pathogenicity underlying those diseases, which has led to valuable insights into the genetic basis, treatment and prevention of obesity and diabetes. Nevertheless, despite our more detailed pathophysiological understanding of the rare forms of diabetes and obesity than of more common polygenic forms, we still know little about their prevalence and implications outside specialised genetics services.

In the present work, I have explored the contribution of Mendelian forms of obesity in individuals with severe obesity. Initial segregation analyses of families with an obese proband, led to the identification of an oligogenic mode of inheritance for obesity. This was followed by re-analysis of pre-existing whole exome sequencing data from 91 individuals with extreme obesity, which revealed an additional 21 possible causative variants in known monogenic/syndromic obesity genes and three further cases of oligogenic inheritance. In addition, 11 candidate variants were identified in genes suggested by rodent models of obesity and/or diabetes, but not previously reported in humans.

To further expand the analysis, a unique custom genotyping array focusing on, obesity and diabetes mellitus (T2D, and monogenic forms of diabetes) was designed to be applied to a larger number of samples (N=2068). Application of the array led to the identification of a total of 161 potential causative variants in 40 monogenic obesity/syndromic obesity genes, with a putative diagnostic yield of 11%. Initial analysis suggests that having one of these putative Mendelian forms of obesity resulted in no statistical difference in percentage weight loss at 2 years post-surgery and diabetes remission.

Our first analysis on obesity indicates that the use of a custom-designed genotyping array for specific rare diseases may be an advantageous first level screening strategy in terms of cost and time. The work presented here also suggests that the true prevalence of Mendelian forms of obesity among bariatric surgery patients is likely to be high - this presents a significant unmet need for genetic analysis and follow-up.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof Alex Blakemore, for all her support, inspiration, encouragement and for being a great mentor from the first day of my PhD. I would also like to sincerely thank my second supervisor Prof Kevin Murphy for his consistent support, motivation, effort and guidance. I am grateful beyond words for having you as my supervisors and for being part of your research group.

I also wish to thank all those who have provided valuable contributions to this work and to the development of the PMMO cohort through their voluntary participation, data collection, follow-up, guidance and discussion. Additionally I would like also to acknowledge and thank the various collaboratives of the different cohorts including Geronimo, EndoBarrier, GLP-1 Receptor Agonist interVention for poor responders after bariAtric Surgery (GRAVITAS), Obesity plus family, and the Imperial College London Diabetic Centre (ICLDC) in Abu Dhabi for the samples and data. My gratitude extends to Dr Nader Lessan and Dr Terry Dovey for their continuous help and guidance. I would like to also recognize and convey my gratitude to Dr Fotios Drenos and Dr Andrianos Yiorkas for their help and support in the project and for their advice and inspiration. Your help and contribution were essential in every way.

Also I would like to thank my colleagues and friends who contributed in various ways throughout my PhD including Dr Olivia Szepietowski, Mitra Sato, Dr Sanne Alster, Nikman Nor Hashim, Hanis Nurul Ramzi, Dale Handley, Erdal Ozdemir, and, Hasnat Amin.

Furthermore, I am extremely grateful to my parents and family for their support and for always believing in me. Last, but certainly not least, I wish to express my sincere thanks to the UAE cultural office in the UK and the Ministry of Higher Education in UAE.

Table of Contents

List of figures

List of tables

CHAPTER 1: INTRODUCTION

1.1 Overview of the Research and Scope of this Thesis

This thesis describes an investigation the contribution of Mendelian forms of obesity in individuals with severe obesity. This was studied through analysis of whole exome sequencing (WES) data, followed by design and application of a customised genotyping array, and exploration of the effects of rare, deleterious mutations on baseline phenotype and outcomes of bariatric surgery

The study started with segregation analyses in families with an obese proband, leading to the identification of an oligogenic mode of inheritance for obesity (Chapter 4). This was followed by analysis of WES data from bariatric surgery patients (Chapter 5). The analysis of these data identified further cases of oligogenic and monogenic obesity, as well as obesity caused by genomic structural variation on chromosome 16p11.2.

Following this, a customised array, specifically focusing on obesity and type II diabetes mellitus was designed and applied to individuals with an extreme obesity phenotype. All the steps for this customised array are described in this thesis, including designing the content of the array (Chapter 6), sample processing and preparation (Chapter 3), quality control (QC) for rare variants analysis (Chapter 3), analysis of rare variants for monogenic obesity (Chapter 7), analysis of rare genomic copy number variants (CNVs) associated with obesity (Chapter 7) and, implications of rare mendelian forms of obesity for treatment outcomes (Chapter 8).

The global COVID-19 pandemic has unfortunately prevented laboratory-based validation of all the mutations identified. Analysis of common obesity-associated genetic variants is also outside the scope of this thesis.

1.2 Obesity as a global epidemic

In recent decades, the prevalence of obesity has increased dramatically throughout the world in both developed and developing countries, resulting in a serious health problem (Figure 1.1). According to the World Health Organisation (WHO), approximately 650 million people are obese and around 1.9 billion are overweight globally with overweight and obesity being considered the strongest risk factors that lead to type 2 diabetes [1-3]. In 2016 it was estimated that approximately 38.3 million children under 5 years of age are overweight worldwide, with 35 million in developing countries. According to WHO, obesity kills around 2.8 million people yearly; around 8% of worldwide deaths in 2017 resulted from obesity [1].

This globally escalating frequency of obesity presents a serious public health issue as obesity is considered a risk factor for developing many of the noncommunicable diseases such as diabetes mellitus, cardiovascular diseases, non-alcoholic fatty liver disease, hypertension, osteoarthritis, cancer, and Alzheimer's disease (Figure 1.2). Additionally, obesity might have other socio-economic consequences such as disability, lower productivity, early retirement and depression. The majority of these health complications are long-term and may be difficult to treat, leading to a negative impact on the individual's socio-economic situation and quality of life. It has been

stated by the WHO that obesity and weight problems are responsible for around 35.8 million global disability-adjusted life years [1].

In addition, obesity is a burden for the health sector, where it is estimated that obese individuals require 30% higher healthcare costs than individuals with "normal" weight [4].

Despite of great efforts from researchers and healthcare providers worldwide to better understand these diseases, obesity is currently recognised as an epidemic with clinical and health issues that have impacts not only at the individual level, but also at societal and economic levels. Thus the problem of obesity remains complex and requires strategies at many levels to prevent, control and manage it [5, 6].

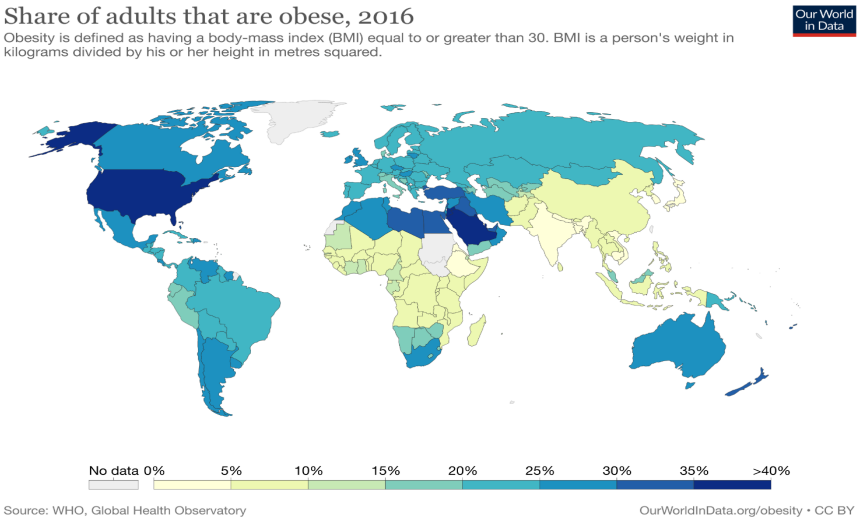
1.3 Definition of obesity

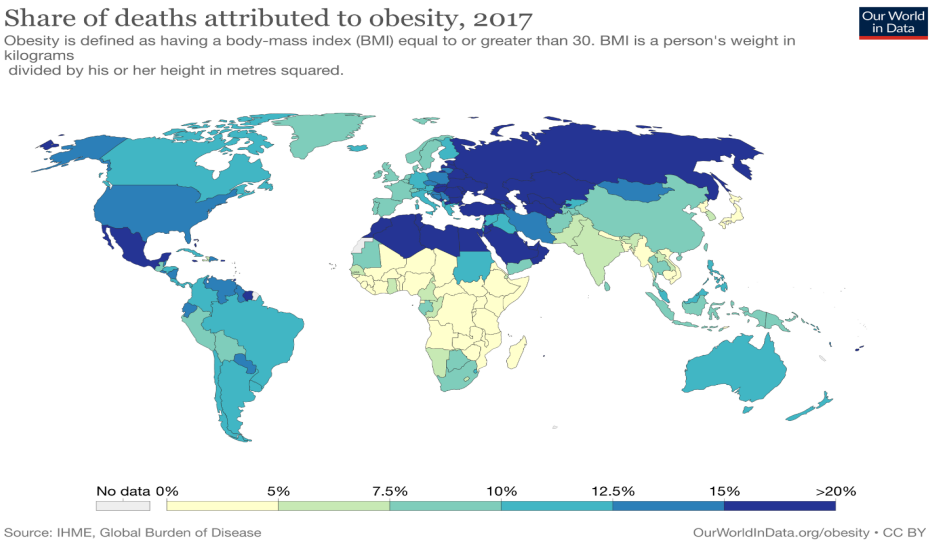
Obesity is a medical condition characterised by increased body weight that is due to the excessive accumulation of fat. Various measurements are used to assess obesity: the most commonly used is body mass index (BMI), a value that is calculated by dividing the person's weight in kg by the square of their height in metres. According to the WHO, weight can be classified into categories according to the BMI value as shown in Table 1.1. Individuals with BMI between 25 and 29.9 kg/m^2 are considered overweight while those with a BMI greater than 29.9 are considered obese. BMI greater than 40 kg/m^2 is defined as severe obesity (obese class III) [1]. Other measurements to assess fat quantity and distribution include waist circumference (WC) and waist-to-hip ratio (WHR).

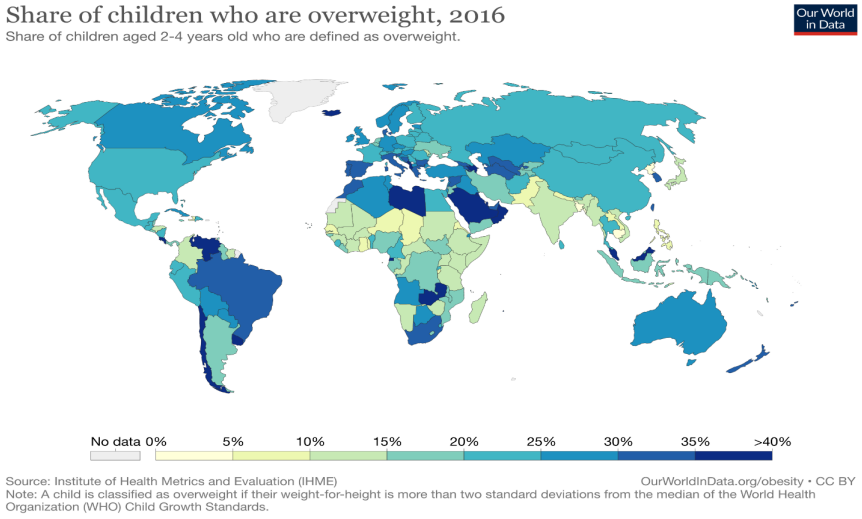
Table 1.1: BMI Classification of adult weight status according to WHO.

BMI, KG/M^2	WEIGHT CATEGORY FOR ADULTS
< 18.5	Underweight
18.5 - 24.9	Normal weight
≥ 25.0	Overweight
≥ 30.0	Obese
30.0 - 34.9	Class I
35.0 - 39.9	Class II
≥ 40.0	Class III (morbidly/extreme/severe obese)

Source: [1]







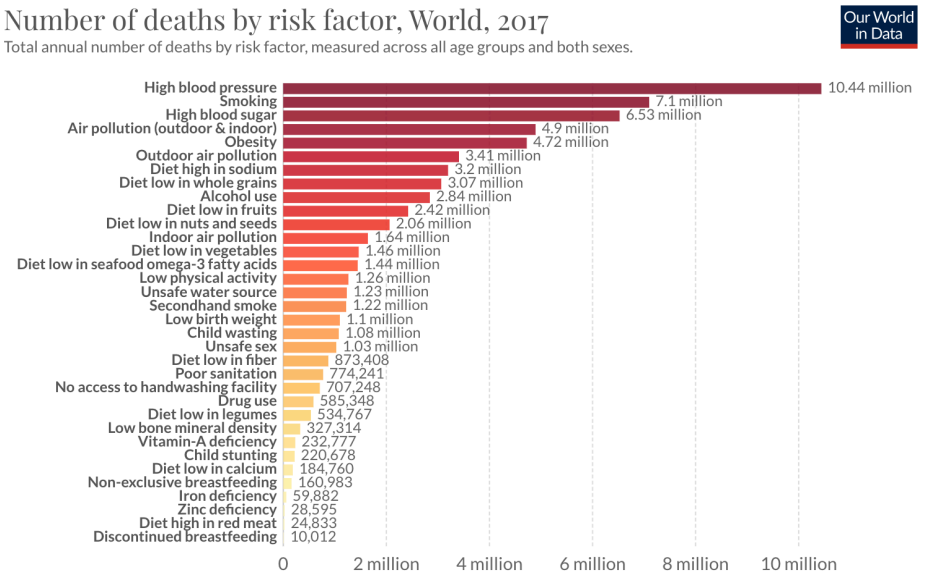


Figure 1.1: Obesity Worldwide. (A) Estimation of obese adults worldwide in 2016 according to the WHO (B) Estimation of deaths that resulted in obesity globally. (C) Estimation of overweight children worldwide in 2016. (D) Number of deaths by risk factor; as obesity contributed to around 4.72 million deaths worldwide in 2017. (Source: www.ourworldindata.org) [7]

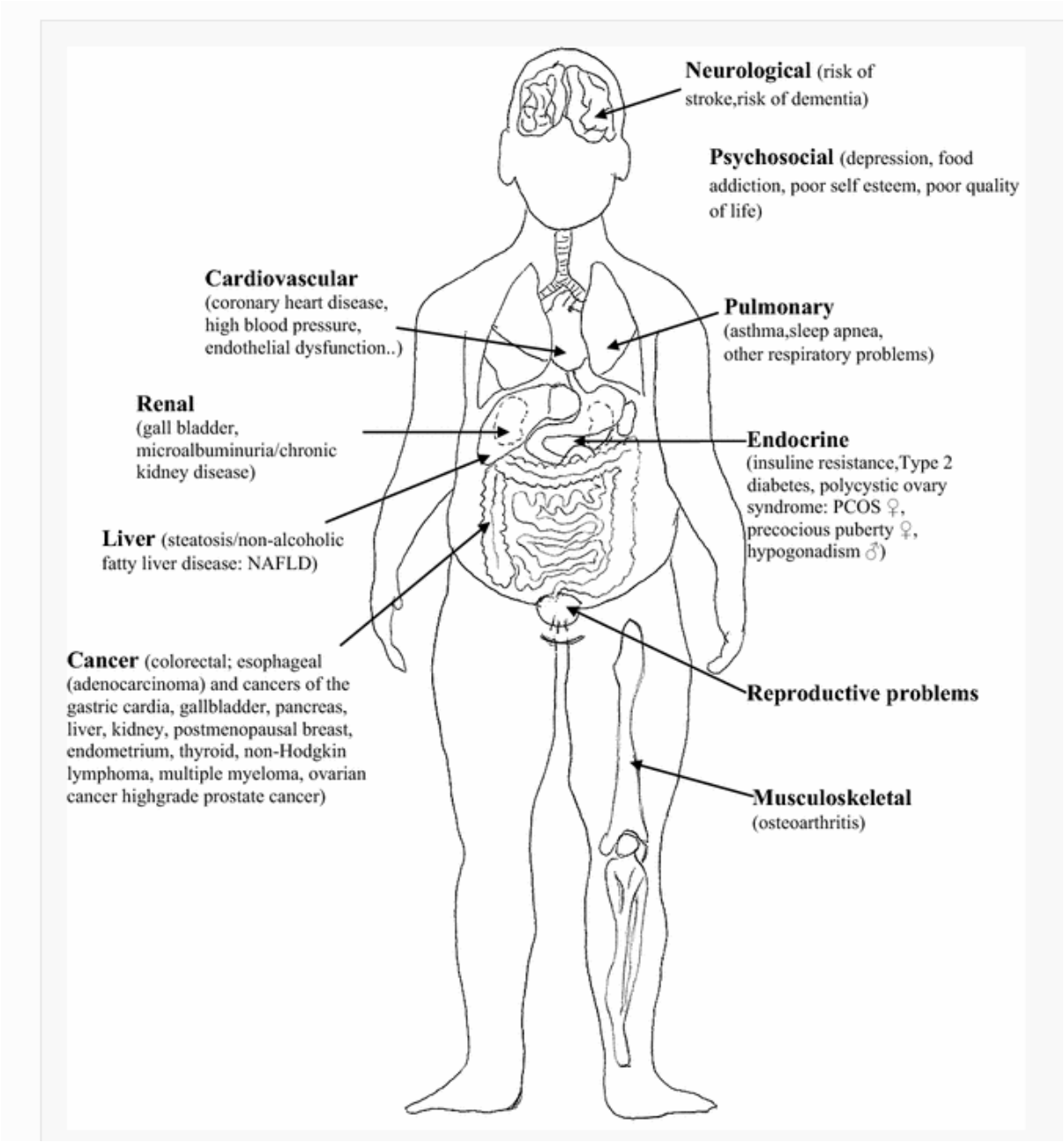


Figure 1.2: Consequences and correlates of obesity. (Source: Burgio *et al*. 2015 [8])

1.4 Causes of obesity

Obesity results from an imbalance between energy intake and energy expenditure, over an extended period of time, which drives an accumulation of body fat. In humans, energy intake is defined as the total energy consumed from food and beverages [9, 10]. Energy expenditure, on the other hand, represents the amount of energy a person needs to maintain body function while at complete rest, as well as the energy utilized during physical activity. The human body achieves energy balance when energy intake is equivalent to energy expenditure over a certain time period [9, 10]. Thus, obesity results when the energy intake exceeds expenditure, which is mainly due to factors such as insufficient physical activity and/or high food intake [9, 10]. In a modern environment, energy balance is believed to be harder to achieve compared to life in past environments [11]. In epidemiology, measuring energy expenditure and food intake is a challenge and the mechanisms regulating food intake are complex and involve many pathways, systems and organs [9, 10].

The energy status of an organism is regulated by the central nervous system (CNS), particularly the hypothalamus; a small region of the brain involved in regulating various homeostatic functions of the body, including appetite, temperature and the reproductive axis. The hypothalamus is divided into several nuclei that involve in controlling energy balance and food intake namely: the arcuate nucleus (ARC), the paraventricular nucleus (PVN), the dorsomedial nucleus (DMN), the lateral hypothalamic area (LHA), and the ventromedial nucleus (VMN).

Furthermore, additional work has revealed the importance of the brainstem in controlling energy expenditure and food intake. The dorsal vagal complex (DVC) in the brainstem is

considered a key module for interpretation and relaying peripheral signals from the gut to the hypothalamus. Neurons of vagal afferent have been found to express a variety of receptors that are involved in controlling food intake [12, 13]. This area has been reviewed by Schneeberger *et al.* (2014) [12].

There are a variety of factors that, individually and in combination, affect a person's energy homeostasis and food consumption. These include environmental factors, hormones, genetics and epigenetics. Many environmental and behavioral factors that contribute to the development of obesity have been well identified, including high-calorie food and beverages, high food intake, less physical activity, sleep alterations, sedentary lifestyle and urbanization. Details of the social and environmental factors are reviewed by Lee *et al*. (2019) [14]. On the other hand, there is considerable variation in individual body weight, shape and diet between people living within a given environment. Some of these inter-individual differences are due to genetics and epigenetics or a combination of environmental with genetic factors. The work described in this thesis focuses on the genetic aspects of severe obesity.

1.5 Genetics of obesity

There is no doubt that genetics play a substantial role in human variation, development, physiology and adaptation. Variation in weight among individuals who share the same environment is to a large extent attributable to genetic factors. Investigation of the underlying pathophysiological mechanisms has provided considerable evidence that supports and demonstrates the importance of genetics in human obesity.

The proportion of variation between individuals in a particular population that is due to genetics is known as heritability. This can be assessed in a number of ways, including twin studies or adoption studies.

Twin studies are a valuable tool to study the relative contribution of genetics and environmental risk factors in the development of a disease or a trait through comparing the degree of similarities of trait between monozygotic (MZ) twins, who are considered to be genetically identical, and dizygotic (DZ) twin pairs, which share only 50% genetic identity and are exposed to a similar environment. If the trait concordance within the MZ twins is higher than it is within the DZ twin, this indicates that the genetic contribution to trait is high. If the two type of twins show similar concordance for the trait, this indicates that genetic differences are having little impact[15, 16].

The other approach is the adoption analysis, which involves studying and comparing the proportion of concordance and discordance between an adoptive child and his adoptive parents or biological parent. If the phenotype in the adopted individual more closely resembles that of his/her biological parents rather than that of his/her adoptive parents, this indicates that the genetic influence is stronger than the environmental effects. If there is no difference in the phenotype concordance rate between the individual and his biological parents, then this indicates larger environmental component and less influence of genetics on the trait [17, 18]. Family segregation analysis is applied to determine the mode of inheritance of a trait by studying members of a family [15] and can also be used to estimate heritability – the more distant the biological relationship, the lower the rate of trait concordance is expected to be. It is estimated by these methods the heritability of obesity to be in the range of 40%-70% [19, 20]. The highest estimation of heritability, of 70%, is derived

from twin studies [21, 22] compared to both adoption studies and family segregation analysis which estimate the heritability of obesity to be around 20%.

Another important consideration is disease heterogeneity. Individuals can have different causes for their obesity. In this thesis we will consider two loose subcategories of obesity: Mendelian and polygenic (or common) obesity. Monogenic obesity refers to rare forms of the disease caused by a mutation, i.e. a change in the DNA sequence, in a single gene. Sometimes, however, the causative mutation involves a number of genes (such as the deletions on chromosome 16p11.2, discussed later in this thesis), and so we prefer the term Mendelian obesity (which would include both monogenic obesity and syndromic obesity).

Polygenic refers to the most common forms of obesity and results from interaction of multiple genes and environmental factors. The two forms are considered genetically distinct, but clinically similar. The proportion of people with extreme obesity who have Mendelian forms of disease, and the degree of overlap between Mendelian and common obesity are currently unknown. One of the major aims of this thesis is to address that knowledge gap.

1.5.1 Monogenic Obesity

1.5.1.1 Leptin (*LEP*) and Leptin Receptor (*LEPR*)

The first evidence of monogenic forms of human obesity relied heavily on early studies with mouse models. The pivotal discoveries were made in obese (ob) mice, which were characterised by obesity and hyperphagia, diabetes, infertility and neuroendocrine

abnormalities by a team in the Jackson Laboratory in 1949 [23, 24]. Joining the circulatory systems of ob/ob at wild-type mice revealed that lack of a blood-borne factor was responsible for the phenotype, and breeding experiments narrowed down the location of the causative gene. The positional cloning of the responsible variant in ob/ob mice led to the identification of the *LEP* gene (which encodes the leptin protein) in 1994 [25]. Figure 1.3-A shows an ob/ob mouse as compared to a wild type mouse. Another variant in a different obese mouse strain (db), whose phenotype was characterised by obesity and severe diabetes, was cloned and mapped to the *LEPR* gene, which encodes the leptin receptor [26].

Following the discovery of *LEPR* and *LEP* mutations in the mouse, the first human family with a mutation in *LEP* was identified in two related individuals [27]. Both carried a homozygous frameshift deletion that resulted in leptin deficiency. Since then, around 25 individuals have been identified with homozygous variants in the *LEP* gene that result in leptin deficiency [27-31]

The clinical characteristics of patients with LEP or LEPR deficiencies are considered similar, and include significant weight gain from the first months of being born, hyperphagia, ,and hypothalamic hypothyroidism. Figure 1.3-B shows a child with leptin deficiency who was treated with recombinant leptin therapy which is the current available treatment that has shown remarkable success in patients.

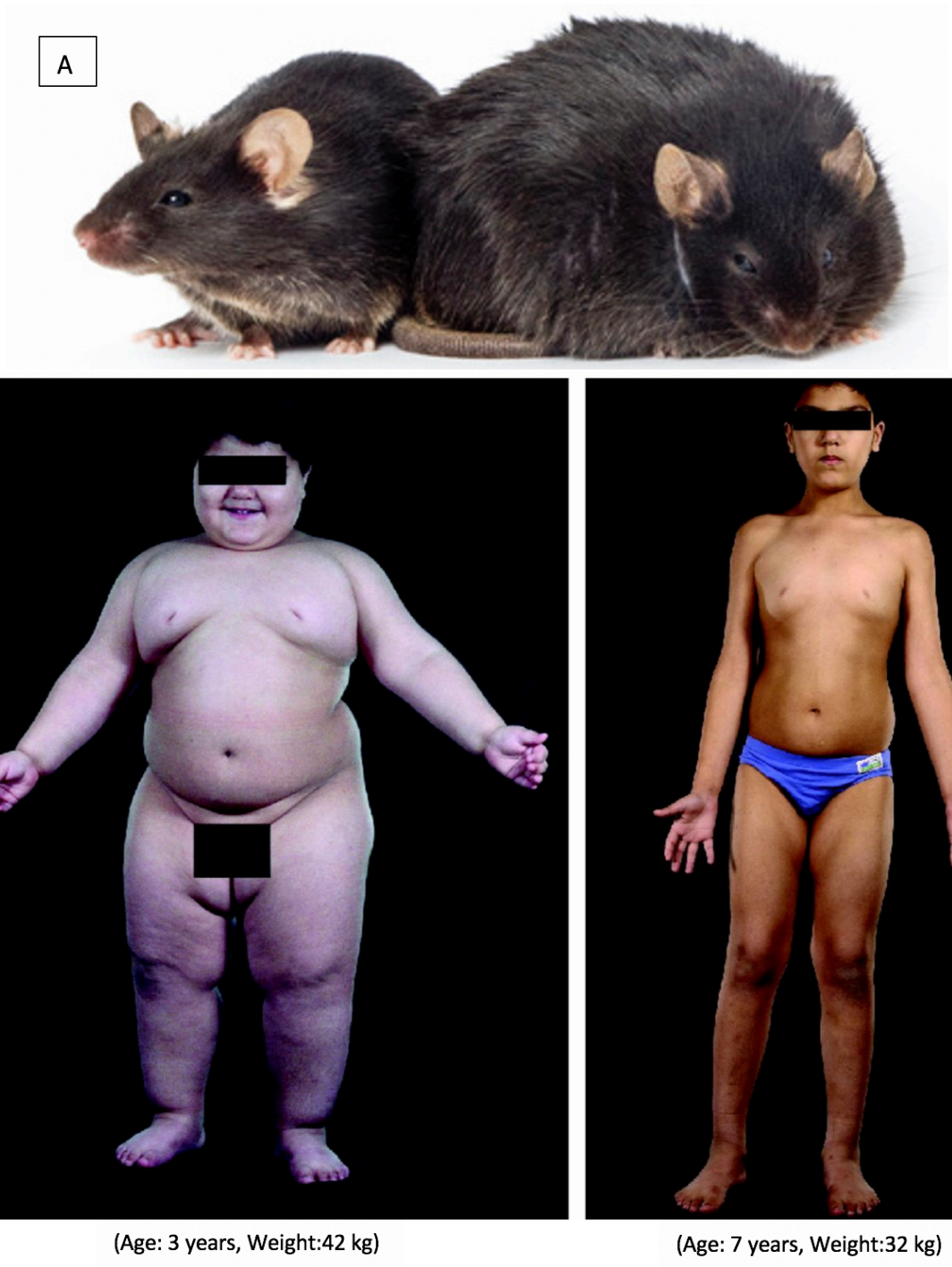


Figure 1.3 Mouse model and human with leptin deficiency. A: ob/ob mice with homozygous mutation for *LEP* gene as compared to wild type mice. (Source: https://www.jax.org/strain/000632) **B:** a child with leptin deficiency before and after recombinant leptin therapy. (Source: Farooqi *et al.* 2014 [32])

The *LEP* gene encodes the leptin protein, which consists of 167 amino acids and acts as a regulatory hormone. It is produced primarily in the adipose tissue (white and brown adipose tissue) and circulates in the blood to be transported to the brain, specifically the hypothalamus [33]. Circulating levels of leptin correlate with the amount of fat stored in the body [34]. The effect of and transport of leptin is regulated by the activation of leptin receptors, which are found in the brain and also in peripheral tissues.

In the hypothalamus, the leptin acts on the longest isoform of the leptin receptor LepRB/ObRb, which is one of six different isoforms: LEPR/ObRb, ObRa, ObRc, ObRd, ObRf and ObRe [35-37]. LepRB has identical extracellular ligand binding domain as the other isoforms but it has a longer transmembrane domain than the short isoforms (ObRa, ObRc, ObRd, ObRf). The long sequence of the transmembrane domain of the ObRB contains two identical domains to the short isoforms, the Janus tyrosine kinase (JAK) and box 1 isoform domains, along with two additional domains: the suppressor of cytokine signaling (SOCS) and box 2 motif domains.

Thus, the long isoform of leptin receptor is known as a fully active receptor for activation of Janus kinase 2 (JAK2) and signal transducer activators of transcription (STAT). It has high expression in a region that plays a critical role in the regulation of feeding behavior; the arcuate nucleus (ARC) of the hypothalamus. Within the ARC it is specifically expressed in two different groups of neurons; proopiomelanocortin (POMC)-expressing neurons, which are anorexigenic i.e. they reduce food intake and appetite, and agouti-related peptide (AgRP)-expressing neurons, which are orexigenic. i.e. they promote food intake and appetite. These two neuronal populations receive signals from peripheral organs about the body's energy and nutritional status and work together (in opposite directions) to modulate food intake and control energy homeostasis as shown in Figure 1.4. Binding of leptin activates the

anorexigenic POMC neurons, stimulating the expression of the POMC precursor and inhibiting the activity of neuropeptide (NPY) /(AgRP) neurons and increasing the expression of their secreted neuropeptides. On activation of the LEPR on POMC neurons, the ligand precursor, POMC, is converted into α-melanocyte stimulating hormone (α-MSH), which is released from the neurons and acts on the melanocortin 4 receptor (MC4R), driving a downstream cascade of neuronal activation which results in decreased appetite and food intake, and increased energy expenditure [38-40]. In contrast, leptin will inhibit NPY/AgRP neurons, and therefore a fall in leptin will lead to greater activation of these neurons, increasing food intake and reducing energy expenditure [38, 41-43].

In addition leptin is involved in other pathways in the hypothalamus such as JAK2-signal transducer and activator of transcription 3 (STAT3) [44, 45], mitogen-activated protein kinase (MAPK) [46], phosphatidylinositol 3-kinase (PI3K) [47], 5' adenosine monophosphate-activated protein kinase (AMPK) [48] and many others [44, 49, 50]. External to the hypothalamus, leptin also has further effects on other peripheral tissues, such as neuroendocrine functions, regulation of male and female reproductive organs and of the immune system [50].

The discovery of the mouse and human *LEP* and *LEPR* deficiency provided the initial gateway into the investigation of further genes that relate to the leptin-melanocortin pathway in bodyweight regulation. This led to the discovery of loss-of-function mutations in different genes which can be categorised in two groups: genes that are related to the leptin-melanocortin pathways including pro-opiomelanocortin (*POMC*), melanocortin-4 receptor (*MC4R*), Src homology 2 B adapter protein 1 (*SH2B1*), pro-hormone convertase subtilisin/kexin type 1 (*PCSK1*), and genes that related to the development of the

hypothalamus: neurotrophic receptor tyrosine kinase 2 (*NTRK2)*, brain-derived neurotrophic factor *(BDNF*), kinase suppressor of ras 2 *(KSR2)* and single-minded homolog 1 (*SIM1).* These are considered to be the most prominent genes implicated in monogenic obesity and a number will be discussed further below.

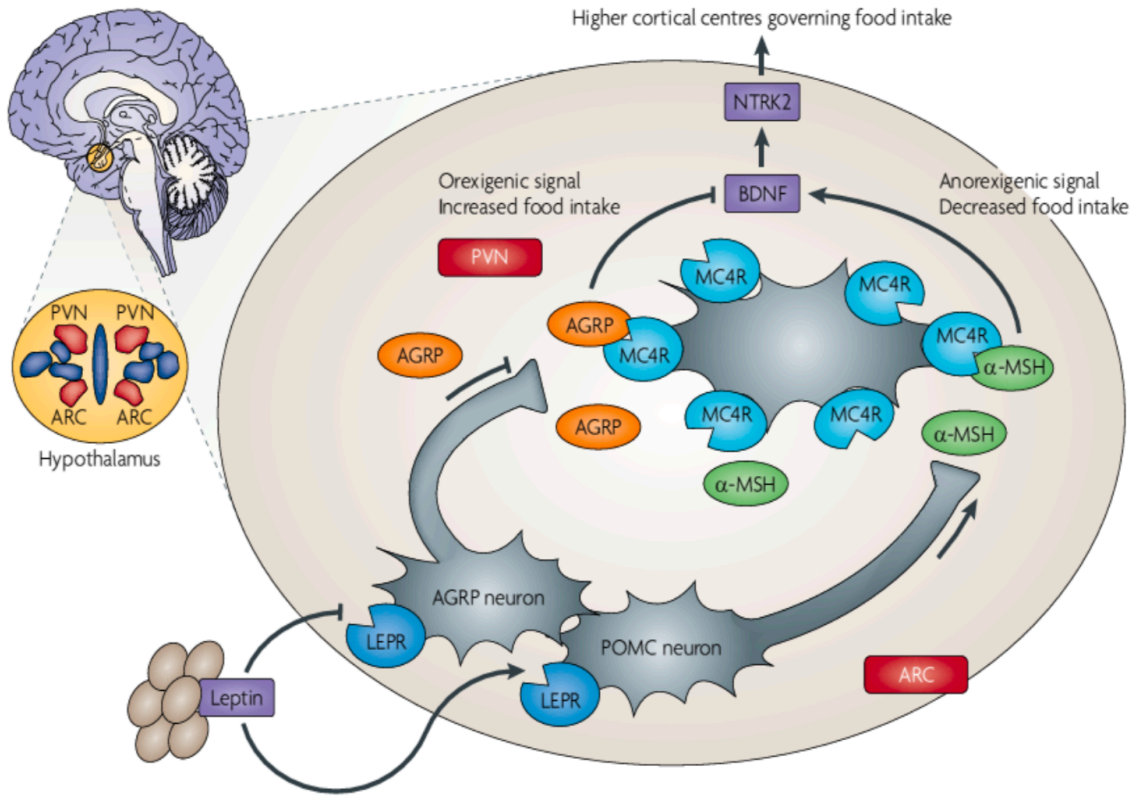


Figure 1.4 The leptin–melanocortin pathway.

An overview of hypothalamic regulation of the food appetite. Food intake is regulated by a complex mechanism known as the leptin–melanocortin pathway. Adipose tissues release leptin that acts on leptin receptor (LEPR) on neurons. There are two relevant types of arcuate nucleus neurons: (1) POMC neurons which releases a range of peptides, including α-, β- and γ-melanocyte stimulating hormone (MSH). α-MSH activates MC4R, increasing energy expenditure and decreasing appetite and (2) (NPY)/(AgRP) neurons which act to decrease energy expenditure and increase food intake [51].

1.5.1.2 *MC4R*

MC4R encodes a membrane-bound receptor that belongs to a family of seven G-protein coupled receptors (GPCRs) [52]. *MC4R* mutations cause autosomal dominant obesity. *MC4R* is highly expressed throughout the CNS, particularly in the hypothalamic paraventricular nucleus (PVN) [53, 54]. It plays an important role in many biological processes, including the maintenance of energy balance, cardiovascular function and, glucose homeostasis [55]. A–MSH acts as an agonist of the MC4R to suppress food intake, whereas AgRP acts as an endogenous antagonist, increasing food intake [54].

Knockout of murine Mc4r- produces mice with obesity, hyperinsulinemia and increased linear growth [56, 57]. Heterozygous mice exhibited a milder phenotype, but developed obesity on a high-fat diet [56, 57], [58]. Following the discovery of MC4R deficiency in rodents, several people heterozygous for deleterious *MC4R* variants have been reported in different populations. *MC4R* deficiency causes a clinical phenotype characterised by severe early-onset obesity, increased height and lean mass, hyperphagia and hyperinsulinemia at a young age. A more severe phenotype is observed in homozygous individuals compared to those who are heterozygous [59, 60].

In recent decades, mutations in the MC4R have been recognised as one of the most common monogenetic causes of obesity, representing around 1–6% of cases depending on age and ethnicity [61].

1.5.1.3 *POMC*

As mentioned above, the *POMC* gene encodes a precursor molecule that is cleaved to produce multiple active peptides, namely melanocyte-stimulating hormones (MSH) (α, β and γ), β endorphin, and adrenocorticotropic hormone (ACTH), which are highly conserved across species and have important biological functions through interacting with different melanocortin receptors [62], including the regulation of energy homeostasis and food intake, pigmentation and adrenal function, as well as roles in other body systems such as the nervous and immune systems [62, 63].

Deficiency of *POMC* results in a lack of MC4R signalling that leads to a constant feeling of hunger, and autosomal recessive obesity. Knockout mice and humans exhibit early-onset obesity, tall stature and adrenal failure with altered skin and hair pigmentation [64, 65]. Evidence of loss-of-function mutations in *POMC* was initially identified in two individuals who were compound heterozygotes. The subjects presented with ACTH deficiency, red hair, pale skin and obesity [65] due to a lack of ACTH and α-MSH signalling.

Further investigations led to the identification of Heterozygous individuals with milder increases in body weight and adiposity [66]. To date , only nine cases with homozygous or compound heterozygous mutations have been reported: these are mainly found in consanguineous families [67]. In spite of the initial evidence of autosomal recessive inheritance of *POMC* deficiency, findings in affected heterozygous carriers suggest that carriage of just one mutated allele may also be involved in increasing obesity risk through interfering with the signaling mechanisms of the central melanocortin system. [67, 68]. In

addition, mice with heterozygous Pomc mutations have been reported to develop obesity and significant hyperphagia on a high fat diet but not when fed with a normal chow diet.

1.5.1.4 *NTRK2* and *BDNF*

Neurotrophic receptor tyrosine kinase 2 (*NTRK2),* also known as tropomyosin receptor kinase B (*TRKB*) encodes a kinase membrane bound receptor that is expressed in neuronal cells in the CNS and peripheral nervous system (PNS). It acts as a receptor for various neurotrophin ligands including NT3 and NT4 and high infinity for *BDNF*, which is considered its major endogenous ligand. Both *NTRK2* and *BDNF* are highly expressed in the hypothalamus and the hindbrain and, have crucial roles in energy balance, and neuronal development, function and maintenance [69, 70]. Mutations in *NTRK2* and *BDNF* have been reported to cause dominant forms of obesity in human and rodents [69, 71, 72]. Despite that fact that *NTRK2* and *BDNF* have a substantial influence on regulating energy and intake and hypothalamus development, the exact functions of these genes in obesity remain unclear [73, 74].

1.5.1.5 Heterozygous mutation carriers of the genes follow the AR.

For genes that follow the autosomal recessive mode of inheritance, such as *LEPR*, *LEP* and *POMC*, Heterozygous carriers have been reported in several studies to have an intermediate obesity phenotype. Heterozygous carriers of *LEP* and *LEPR* had higher BMI and body fat mass compared to the controls, but did not have severe obesity [75, 76]. On the other hand, carrier of Heterozygous *POMC* mutations had a significant higher BMI than controls, but did not manifest other phenotypes observed in patients homozygous for *POMC* mutations [66]. [77-79].

The identification of monogenic forms of obesity has provided a considerable evidence and progress in understanding molecular pathogenesis of the condition. Continued discovery and study of the new forms of obesity is vital in further clarifying unsuspected biological pathways of obesity.

1.5.2 Syndromic obesity

Another form of Mendelian obesity is syndromic obesity. Syndromic obesity refers to a Mendelian form of obesity with additional features that often include intellectual disability, cognitive delay, mental retardation and many other features**.** Cases of syndromic obesity were initially found long before the identification of the monogenic form of obesity based on animal models.

The genetics causes vary from a point mutation to chromosomal defects**.** To date there are around 79 forms of syndromic obesity reported; 54% have no assigned name and 13.9% have been assigned more than one name [80]. Some named examples are CHOPS syndrome, Cohen syndrome, Alstrom syndrome and Smith-Magenis syndrome.

There is considerable overlap in clinical features among the majority of syndromic obesity cases. However, it remains challenging to classify whether obesity is correlational or merely coincidental to the syndrome [80]. The first obesity syndrome identified, known as Bardet-Biedl syndrome (BBS), was recognised in 1966. It is an autosomal recessive condition that is genetically and clinically variable. The syndrome is characterised by obesity, intellectual disability, behavioural dysfunction, kidney dysfunction, retinal degeneration and polydactyly

[81, 82]. To date, 21 different genes have been reported to cause BBS, all of which are involve a cell structure, the cilium, that is involved in movement and signalling pathways, and are thus collectively described as ciliopathies [81, 82]. Diagnosis of BBS relies on the presence of at least 3–4 of the major BBS characteristics and 2 minor characteristics, conforming with diagnostics guidelines published by Beales *et al.* 1999 [83].

Another well-known type of syndromic obesity is Prader-Willi syndrome (PWS). It is estimated that around 70% of the PWS cases are caused by inactivation or deletion of number paternal genes in chromosome 15, considered as one of the two main human genomic regions which undergo imprinting. It is a multisystem disorder effects different parts of the body. It is characterised by feeding difficulties and severe hypotonia in infancy, followed by excessive eating at early childhood, leading to severe obesity. In addition, there are other characteristics which include short stature, cognitive impairment, hypogonadism, behavioural problems, developmental delay, failure to meet motor milestones and sleep disorder. Some of these features vary in the degree of severity, between low to mild to moderate [84, 85]. PWS is primarily caused by loss of expression of at least one of the paternally inherited genes at 15q11.2-q13. Around 65–75% of cases are caused by paternal deletion at chromosomes 15q11.2-q13, 20-30% are caused by maternal uniparental disomy (UPD) and 1-3% are due to imprinting defects [84, 85].

Another example of syndromic obesity is Smith-Magenis syndrome (SMS), which is caused by haploinsufficiency of the RAI1 gene resulting from either deletion of a chromosome 17p11.2 region or a pathogenic mutation in RAI1. The disease is characterised by variable features including intellectual disability, obesity, cognitive impairment and, behavioural deficits. Analysis of genotype-phenotype correlation in SMS showed that RAI1 haploinsufficiency is

the primary contributor to most of the SMS features, and the severity and variability of the phenotype varies as the deletion increases to encompass more genes. Patients with mutations in RAI1 alone tend to have less severe or absence of features of cognitive impairment, heart and rental defect, short stature, hearing loss and motor delay compared with the patients with the 17p11.2 deletion [86, 87]. However, patients harbouring mutations in RAI1 do exhibit other core features of SMS such sleep disturbance, intellectual disability and more frequently overeating and obesity.

One under-recognised rare cause of syndromic obesity is deletion of either one or other of (or, even more rarely, both) regions of 16p11.2. This copy number variant (CNV) region is strongly associated with obesity and morbid obesity, supported by replication findings in several independent studies [88] . Duplication of this region has been found to be associated with underweight, autism and schizophrenia. The subjects showed other features in addition to obesity, including developmental delay and intellectual disability. Further studies detected another CNV, 222 KB in size, within the 16p12.2 and encompassing the SH2B1 gene [89].

The improvement and success in identification of the causes of syndromic obesity has led to improvement in patients' care, accurate diagnosis and insights into pathogenesis. Nevertheless, only around 24% of cases have been fully explained while around 13% have been partly elucidated [80]. A large fraction of obesity syndromes remain to be identified and characterised, specifically in terms of the casual genes. Thus, improvements in syndromic obesity research can be applied at different levels, including the analysis of larger cohorts, and the establishing of guidelines for classifying syndromic obesity. This may involve revising and reclassifying the known types based on molecular and clinical evidence.

1.5.3 The role of common genetic variants in obesity

Polygenic obesity is assumed to be the most common form of obesity, and is generally studied at the level of populations, rather than within family groups. It differs from the monogenic forms of obesity in its complex pattern of segregation, where more than one gene is involved. Early studies investigating the genetics of common obesity relied mainly on two methodologies: linkage analysis and candidate gene analysis, with each having its own advantages and disadvantages. Linkage analysis is a technique that is used to identify and map a chromosomal region/segment which co-segregates with a phenotype or disease of interest through related individuals [90]. Despite the success of linkages analyses in studying monogenic obesity, it has been less successful in studying the common type due to low statistic power and few consistent replications in subsequent studies [5, 91, 92]. Candidate gene studies, on the other hand, rely on detecting association between genomic markers within specific candidate genes. The candidate genes are chosen at either positional level (where they are located at important genomic regions based on previous linkage or association analysis), or functional level, based on current understanding of the pathogenesis [92-94]. A number of candidate gene studies have been performed, and at least 127 genes have been identified to be associated with obesity or obesity-related traits [95]. In the initial stage there was limited success with the candidate genes studies due to a low number of samples and different populations showing different results [91, 96]. This was followed by advances in testing for association within larger populations (N>5000) and performing meta-analyses [91]. The enhancement has led to the identification of robust associations between

obesity and many genes. Well-replicated candidate genes include *MC4R*, *BDNF*, *PCSK1* and β-adrenergic receptor 3 (*ADRB3*) genes [91, 97, 98].

In recent decades there has been noticeable progress in common obesity research using genome wide association studies (GWAS). GWAS is a non-hypothesis-driven methodology to explore the biology of a phenotype of interest without significant prior knowledge regarding the genes involved. The methodology of GWAS relies on the idea of common diseases being caused by common variants, where the minor allele frequency (MAF) of each genotyped SNP is tested for association with a particular phenotype of interest. The methodology has identified many SNPs that are associated with obesity and obesity-related traits. The strongest and most consistently-replicated associations were identified in the regions of the FTO [99-101] and MC4R genes [102-104].

The first important discovery was a strong association in a region of chromosome 16q12 that includes the *FTO* gene with obesity and BMI, which was reported separately by three different independent groups [99-101]. This association was later replicated in different populations [5]. FTO is highly expressed in tissues including the hypothalamus, adipose and skeletal muscles. Although little is known about the function of *FTO*, it is located in the cell nucleus and has a role in regulating gene expression through DNA demethylation. It acts as RNA demethylase which binds to different of RNAs including mRNA, tRNA and snRNA, where its most favourable nucleobase substrate is m6A [105]. The M6A has an important roles in post-transcriptional regulation including translation, RNA splicing, degradation and nuclear production [106].

Based on the identification of heterozygous loss of functions mutations in obese and lean individuals, *FTO* is not thought to be a cause of monogenic obesity. Although the exact molecular mechanism between *FTO* and obesity has yet to be fully elucidated, recent studies showed that animal models showed that *FTO* expression is regulated by fasting and feeding behaviour [106, 107]. In addition, several studies have suggested that polymorphisms located in the introns region of the *FTO* may contribute to obesity by acting and regulating the expression not only of the *FTO* but on other neighbourhood genes such as *IRX3, IRX5, RRBL2* and RPGRIP1L [106, 107]. Furthermore, it has been identified that *FTO* is involved in the development of obesity by regulating and affecting the m6A levels of hormones and molecules related to food intake and adipogenesis [106, 107].

The association of the *MC4R* with BMI was identified by the first meta-analysis, based on 16,786 individuals and performed by the Genetic Investigation of Anthropometric Traits (GIANT) consortium [108]. This was then confirmed in different meta-analyses [5]. Following the discovery of the association between *FTO* and *MC4R* SNPs and obesity, further SNPs were identified in different genes that include transmembrane protein 18 (*TMEM18*),mito-chondrial carrier homolog 2 (*MTCH2*), potassium channel tetramerization domain containing 15 (*KCTD15*), neuronal growth regulator 1 (*NEGR1*), glucosamine-6-phosphate deaminase 2 (*GNPDA2*) and *SH2B1*, *BDNF* and *ETV5*. In addition, multiple large meta-analyses identified associations of SNPs with obesity-related traits, such as neurexin 3 (*NRXN3*) with waist circumference, hip circumference (HIP) and waist-to-hip ratio (WHR) [109]. To date, large-scale GWAS from well-powered studies and research consortiums have identified 870 SNPs that are strongly associated with BMI [110-112].

Despite the efficiency of GWAS in the investigation of complex diseases through clearly defining the region, as well as the international collaborations that have enabled the inclusion of large numbers of study participants, many challenges and limitations have emerged. These challenges include those identified when using GWAS to investigate other complex diseases, such as interpretation of findings, a lack of functional characterisation of relevant genes, the facts that the genes identified to date account for only a small proportion (17-27%) of the total heritability, and lack of power to confer risk for obesity [113]. Much of the heritability of obesity thus remains unexplained.

1.5.4 Common genomic copy number variation (CNV)

Another form of gene alteration that results in human variability is genomic copy number variation. Copy number variation is a type of structural variation when a section of the genome, of 1 kb or larger, is duplicated, amplified or deleted [114, 115]. CNVS are classified into two categories: common CNVs, which are present in more than 5% of the general population, and rare CNVs, which are present in less than 1%. Some CNVs are harmless when they present by themselves but can cause disease when they occur in combination with other variants such as SNPs, CNVs or inversions [114-116]. There has been a much interest in studying the relationship between structural variants and different diseases [116].

The first two common CNVs identified were a deletion near the neuronal growth regulator 1 gene (*NEGR1*) and a deletion at 10q11.22 which encompasses the pancreatic polypeptide receptor 1 (*PPYR1*) gene [117, 118]. Supporting these findings, another study by Jarick *et al*.

2011 identified another common CNV, on chromosome 11q11, associated with extreme obesity of early onset, followed by two meta-analyses that identified two further CNVs; G-protein-coupled receptors (GPCRs) and one within *NEGR1* that is associated with obesity and regulation of body weight [119]. These two CNVs were identified by linkage disequilibrium with BMI SNPs.

Rare CNVs, on the other hand, have been suggested to make a larger contribution to the heritability of obesity and BMI compared to the common CNVs that are described above. One of the well-known rare CNVs with a potential large effect is the large deletion of 16p11.2. which has been described in section 1.5.2.

1.6 Missing heritability

As described above, heritability refers to the proportion of phenotypic variation among individuals in a particular population that is due to genetic variation. Our understanding of the monogenic forms of obesity and diabetes in terms of pathophysiology is further advanced than that of the common form of those diseases. As it has been mentioned earlier, the effect size of the identified common loci is relatively small and altogether they have been estimated to account for 17% to 27% of the variation in BMI [113]. In addition, the causal genes and mechanisms of the majority of these common variants remain unexplained and the "missing"genetic components have not been found yet [120].

Many possible explanations have been suggested to account for this missing heritability of human obesity, including limitations of current methodologies in exploring genetic variants at large scale. For example, the potential contribution of rare variants, which is addressed in this current study [120, 121], has been little explored. Another possibility that many SNPs remain undetected as their effects do not reach statistical significance or lack of complete linkage disequilibrium between causal variants and genotyped SNPs [122].

It is t has also been proposed that part of the missing heritability can be explained by genetic interaction or epistasis rather than additional unidentified genetic variants. Another suggestion is the contribution of epigenetics, whereby the variability of complex diseases could be due to variability in gene expression among individuals [120, 121]. Nevertheless, although a large proportion of missing heritability remains to be elucidated, the introduction of advanced techniques and increasingly well-powered studies may lead to explanations and promising outputs that improve our understanding of the complexity of those diseases [121].

1.7 Advanced technologies and databases

The development of the first generation sequencing, Sanger sequencing, undoubtedly initially dominated the field of genetic research, allowing identification of deleterious variants causing genetic diseases and led to the creation of accomplishments such as the human genome project [123]q.

Nonetheless, this technology has been superceded by the next generation of massive-parallel sequencers, which have overcome the limitations of the Sanger sequencing in terms of cost, time consuming and scalability [124]. Whole genome sequencing (WGS) allows an untargeted view of the genome to detect any possible genetic variations while whole exome sequencing (WES) enables a targeted view of the human exons and regulatory regions to detect genetic variants within those regions. Targeted gene sequencing, on the other hand, allows flexibility in the size of analysis from at least two genes to thousands of genes. The choice of sequencing approach varies according to the question and type of analyses to be used.

Another well-established approach to studying genetic variation is through use of array-based genotyping technologies. This allows investigation of a large number of variants across the genome. Generally, most commercial genotyping arrays are based on common variants, designed for conducting GWAS studies. The development of most of those arrays was based on validated SNPs, sourced from the International HapMap project. Nevertheless, in the last couple of years, more customised arrays have been developed to include more rare variants in addition to common SNPs. These include arrays based on the 1000 Genomes Project content, and exome arrays based on all possible variants from exome sequencing projects. Another more efficient approach to address specific traits or diseases is through a customised

genotyping array, for instance the metabochip [125].

The rapid advancement and availability of cost-effective methods for sequencing and genotyping that have emerged in the last few years have led to new strategies to investigate the genetic architecture of diseases. It has also enabled the development of very large-scale publicly available genetic and functional databases such as gnomAD (www. https://gnomad.broadinstitute.org/) and HGMD (www.http://www.hgmd.cf.ac.uk/) [126, 127]. The accessibility of those databases has brought numerous benefits for research, including population genetic analysis, clinical genetics, and functional interpretation.

1.8 Treatment and management of obesity

Treating and preventing obesity and diabetes has become one of the vital goals in public health. Obesity treatments are classified mainly into three groups: (1) lifestyle intervention, (2) pharmacotherapy and (3) surgery. Each of the treatments has some degree of risk; the treatment with the lowest risk is the lifestyle intervention. Thus, the treatment option depends largely on various factors including the patient's BMI, measurements of WHR and WC, the patient's health and other treatments [128].

1.8.1 Lifestyle interventions in obesity treatment

Lifestyle intervention is the fundamental treatment for obesity, essentially a reduction in the dietary energy intake (food intake) and/or increased physical activity (energy expenditure) to attain a negative energy balance. Several types of strategy are used, including caloric intake reduction such as counting calories, reducing/eliminating specific types of food or macronutrients or meal replacement [128]. Nevertheless, determining the absolute best diet is still challenging as all of these food-specific diets show comparable results of weight loss and there is still wide research going on this field.

Physical activity also plays an important part of obesity treatment. However, it is not considered as effective as food intake reduction for weight loss [129, 130], though it can be useful to prevent weight regain over time. Lifestyle intervention has some limitations which

include that although the mean body weight lost is estimated to be between 5–8% over the starting 6 months, most people experience weight regain, with around two thirds of this weight beingregained during the first year, [131].

1.8.2 Pharmacotherapy

Another approach to weight management is through drug therapy. Until 2012, there were only two weight loss drugs that were available in the market and approved by the US Food and Drug Administration (FDA): sibutramine (Meridia) and orlistat (Xenical). Orlistat is an inhibitor of pancreatic lipases, and thus inhibits the digestion and subsequent absorbtion of dietary fat. This is estimated to reduce energy intake into the body byaround 200 calories reduction per day based on a diet of 2,000 calories/day [132, 133]. However, Sibutramine was removed from the market in 2010 due to associtaion with cardiovascular complications [134]. In the UK, the only medication that is approved for treating obesity is orlistat .

With advances in drug development, further obesity medications have been approved by the FDA. One of the known treatments is the administration of leptin for patients who have loss of function mutations in LEP or LEPR genes that cause leptin deficiency [135]. The treatment is known as Myalept (metreleptin), recombinant methionyl human leptin [135]. The other recent approved drugs therapy includes liraglutide 3.0 (Saxenda) and, phentermine/Topamax (Qsymia)[131].

Liraglutide is a glucagon-like peptide-1 (GLP-1) receptor agonist that have been also approved a treatment for both obesity and diabetes. It mimics the function of GLP-1 hormone through binding to GLP1 recepts that results in stimulating insulin secretion, reducting postprandial glucagon and, reducting appetite and reducing foot intake [136, 137]. Another GLP-1 agonist has been recently approved by Semaglutide, which differs from liraglutide in its peptide structure by the inclusion of two modifications (amino acid substitution). The modifications provide significant protection for the peptide through preventing enzymatic degradation by DPP-4. Overall, the mechanism of both GLP-1 receptor agonists is similar however semaglutitde has shown to be more effective in weight loss [138].

The Qsymia on other hands, is a combination of phentermine (appetite suppressant) and topiramate (anticonvulsant: treatments for seizure and migraines) that promotes weight loss through reducing appetite and increasing satiety energy used by the body [139].

1.8.3 Bariatric surgery

Bariatric surgery is currently the most successful and reliable therapy for long-term weight loss and improvement or permanent remission of co-morbidities including T2D, hypertension and dyslipidaemia [140]. In addition, bariatric surgery is associated with a decrease in mortality rates related to various diseases and cancer incidence [141, 142]. In the UK, the National Institute for Clinical Excellence (NICE) recommends surgery as the first-line treatment for people with a BMI over 50, since other treatments are unlikely to have a durable effect in these extreme cases.

There are four common types of bariatric surgery including those shown in Figure 1.5 and described in details by Pories WJ *et al.* [143]. Roux-en-Y gastric bypass (RYGB) and laparoscopic sleeve gastrectomy (VSG) are the most frequently performed procedures in the United Kingdom. The RYGB has a greater effect in terms of weight loss and diabetic remission [144]. As with other surgeries, there are several risks and side effects of bariatric surgery including: high rate of re-operation for some surgery types, difficulties in adapting to the lifestyle changes in terms of type and amount of food, deficiencies in nutritional vitamins and minerals that require regular monitoring and weight regain. However, in terms of efficacy it significantly surpasses conventional lifestyle and medical approaches, and has a more durable effect.

Patients' responses to surgery vary in terms of weight loss and diabetes remission. It is estimated that around 20% of patients fail to achieve or maintain more than 50% excess weight loss. Several studies have investigated the role of genetics in the variability of weight loss after bariatric surgery. They have used evidence from twins and close relatives to

demonstrate that there is a genetic contribution to weight loss after bariatric surgery [145]. In addition, association studies, mainly based on candidate genes, identified that patients carrying a homozygous allele of the SNP rs490683, located at the promoter of the ghrelin receptor gene, also known as growth hormone secretagogue receptor (GHSR), tend to lose significantly more weight compared to patients carrying the heterozygous genotype. Furthermore, mutations in *MC4R* have been shown in some mice and human studies to be linked with poor surgical outcomes in terms of weight loss while other studies showed no significant contribution: this may reflect different surgery types included in the various studies [146-148]. In addition, Sarzynski *et al*. found evidence of one single nucleotide polymorphism (SNP) in the *FTO* gene as having an association with weight loss after surgery, while there were no associations in 10 other obesity candidate SNPs tested [149]. These contradictory results may reflect the limited power of small scale studies.

On other hand, GWAS studies looking at bariatrics surgery outcomes have been limited. One of the major studies conducted by Kuipers *et al*. identified 17 SNPs associated with post-surgery weight loss after 2 years, specifically, RYGB surgery [150]. Another GWAS analysis conducted by another group identified an association between a single SNP located near *ST8SIA2* and *SLCO3A1* and weight loss after RYGB surgery [151]. Nevertheless, the full role of genetic variants on bariatric surgery outcomes, specifically weight loss and diabetes remission, remain to be discovered. The contribution of such findings will be significant for healthcare providers and patients in determining who will benefit most from surgery.

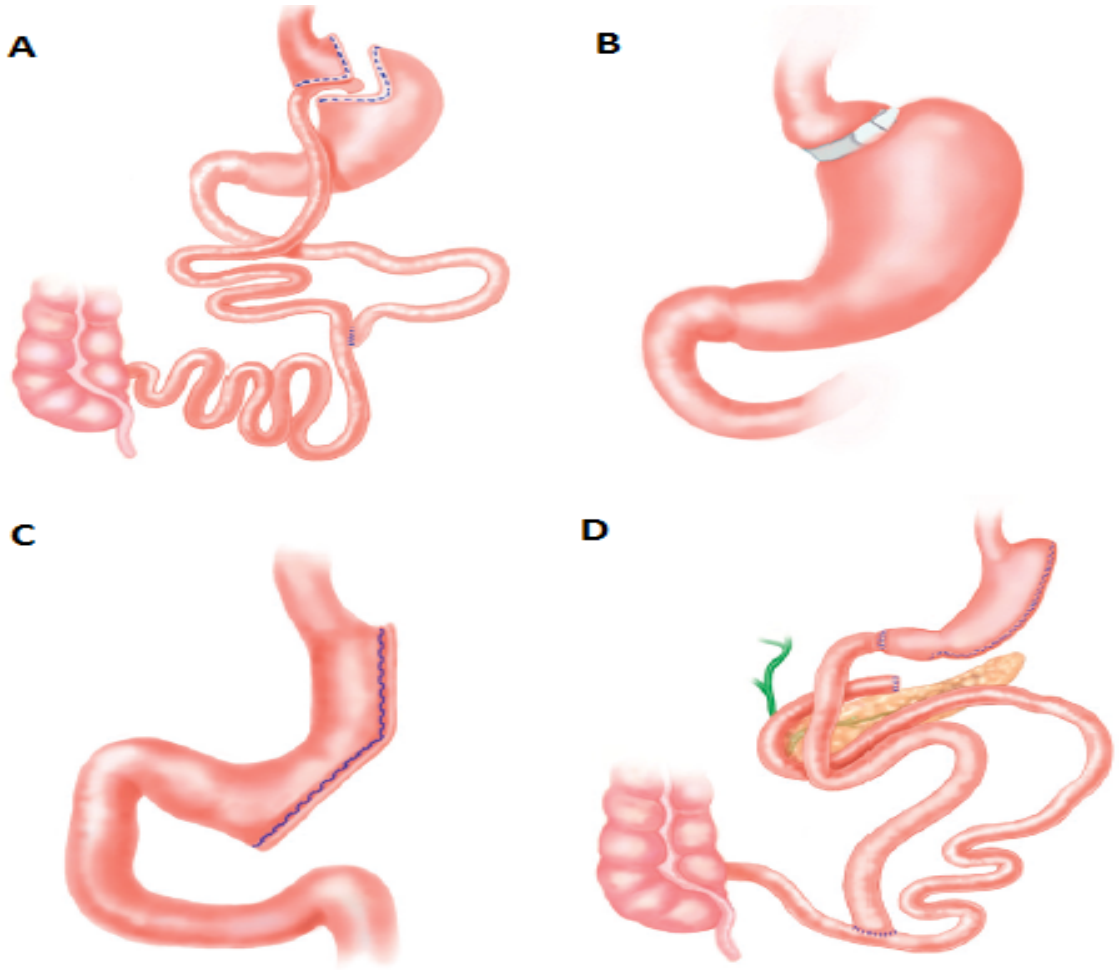


Figure 1.5: Diagram illustrating the different types of bariatric surgery: (A) Roux-en-Y gastric bypass (B) Adjustable gastric band (C) Vertical sleeve gastrectomy (D) Biliopancreatic diversion with duodenal switch. **(**Source: Dixon *et al.,* 2012 [152]) .

1.9 Genetics of diabetes

As with obesity, there is considerable evidence about the genetic causes and the pathologies of the monogenic forms of diabetes, which includes maturity-onset diabetes of the young (MODY), neonatal diabetes mellitus (NDM), and mitochondrial diabetes mellitus. These forms of diabetes follow a Mendelian pattern of inheritance, where the effect size of the variants is large and there is less environmental contribution compared to common forms of the disease [153-155]. The identification of genes causing monogenic forms of diabetes has provided great insights into the disease's mechanisms and pathways, including the role of the KATP channel pathway in causing hyperinsulinemia and hypoglycaemia, and the involvement of different cell types besides the Beta cell in hypoglycaemia [153-157]. This has enhanced patient care by enabling diagnosis and provision of the most appropriate treatment. For example introducing screening programs to diagnose diabetes in children caused by mutations in *HNF1A* or *HNF4A*, enables them to switch from insulin to sulfonylureas as therapy. So far 70 genes causing monogenic diabetes have been identified (with varying degrees of confidence) and these are shown in table 2. On the other hand, type 2 diabetes mellitus (T2DM) is the most common form of diabetes. Genome-wide association study (GWAS) has provided by far the biggest insight into the genetics of this form of the disease through identifying around 100 common genomic loci [154, 155]. However, the majority are mapped to noncoding regions of the genome and have modest effect risks (highest odds ratios (modified relative risks) around 1.1 to 1.2) . Although, this represents only a small portion of the polygenic heritability, there is an overlap between the regions of many those common variants genomic of diabetes mellitus and the monogenic diabetes genes, indicating

an overlap in aetiology It is currently unknown how rare and common genetic factors interact in predisposing to diabetes or treatment response [6, 155].

1.10 Overcoming the challenges of obesity and diabetes

With the dramatic global increase of obesity incidence and the growth of evidence of obesity and diabetes as a disorder, understanding and overcoming the challenges underlying the genetic susceptibility of obesity is a major research topic. Most analysis has been carried out using either population studies containing participants with a range of BMIs, or small cohorts or families of severely affected people. In this thesis, I describe a series of investigations of people with extreme obesity, both in individual families and as a large group.

There are several challenges that need to be considered. One major challenge is explaining the observed missing heritability– we do not yet know the contribution of rare variants to non-syndromic obesity in “normal populations”.

Other challenges include exploring incomplete penetrance of rare variants. Additionally, there has been little consideration of potential joint effects of rare and common variants and CNV interactions where each effect is studied independently. Finally, the implications of these rare variants for treatment outcomes are unknown. One obstacle in tackling these challenges has been the high cost of analysing a large number of participants, which limits the expansion of these studies and thus their statistical power.

With the advent of new technologies and large databases as well as a large number of study participants and carefully-phenotyped severely-affected study participants, it is now possible expand our knowledge of the architecture of human obesity by detecting and analysing larger numbers of rare deleterious variants in focused sample collections with extreme phenotypes, increasing statistical power.

1.11 Overall aims

The work described in this thesis aims to address the challenges described above in order to improve our understanding of the genetic architecture of obesity by identifying rare forms of obesity in families and participants of extreme phenotypes using advanced technologies at relatively low cost. We also aim to investigate the implications of rare variants for postsurgical outcomes. The individual objectives in furtherance of these aims are below.

1. **Familial analysis of monogenic obesity.**

A follow-up investigation that includes mutation confirmation and family segregation analysis in families with a proband identified previously as having a deleterious variant in monogenic obesity genes.

2. **Identification of new Mendelian forms of obesity through reanalysis of previous exome sequencing data from extreme obesity cases.**

Re-analysis of an existing whole exome sequencing (WES) data of 91 extreme obesity cases, from the PMMO cohort and another collaborative group. This was performed with improved filtering steps and a larger list of candidate genes.

3. **Design and application of a custom array to detect rare variants in known obesity genes and other candidate genes as well as common variants and CNVs.**

Creation of a customised array for obesity and diabetes that includes different forms of genetic variants: rare variants, common SNPs and CNVs. The array was applied to a total of 2,094 samples from patients with extreme obesity.

4. **Investigation of the implications of rare variants on postsurgical outcomes (diabetes improvement and weight loss) in patients undergoing bariatric surgery.**

A selection of participants who have undergone bariatric surgery have been chosen to further investigate the role of rare variants on bariatric surgery outcomes, which include weight loss and diabetes remission.

CHAPTER 2: COHORTS

2.1 Introduction

The work in this doctoral thesis is based on subjects and data from different cohorts and populations. Thus, this chapter is designed to give an overview of the different cohorts that have been used and that have contributed to the different studies described here. These cohorts include the Personalised Medicine of Morbid Obesity (PMMO) cohort, ENdobarrier, Geronimo, GLP-1 Receptor Agonist intervention for poor responders after bariatric Surgery (GRAVITAS), Obesity plus family, and the Imperial College Diabetic Center Abu Dhabi. These samples were included in the analysis as a part of collaboration via my Supervisor Alex Blakemore.

A summary of each cohort and type of analysis and number of samples that was used in each study is provided in the table 2.1 below. The PMMO cohort (led by Professor Alexandra Blakemore) represents the major cohort that contributes to most of the studies in this thesis.

Table 2.1: Overview of the cohorts used in the different chapters.

COHORT	FAMILY SEGREGATION ANALYSIS	RE-ANALYSIS OF WES	GENOTYPING BY THE CUSTOMISED ARRAY
PMMO		✓	✓
ENDOBARRIER			✓
GERONIMO			✓
GRAVITAS			✓
OBESITY PLUS FAMILY	✓	✓	✓
ICLDC	✓		

2.2 Cohorts

2.2.1 PMMO cohort

The PMMO study started in 2012 and was initiated by Professor Alex Blakemore. The study was designed with various aims, including to investigate the genetic architecture of obesity and diabetes; identify genetic factors influencing weight loss and diabetes remission after surgery; and mechanisms underlying this diabetes remission. The participants recruited under this cohort are classified into three groups: the first main group is pre-surgery participants who were preparing to undergo bariatric surgery, the second is post-surgery participants who had previously had bariatric surgery and the third group is morbidly obese individuals who are not intending to undergo surgery. The participants who were pursuing or had undergone bariatric surgery were recruited mainly at the Imperial Weight Centre, London, or at other NHS Trusts sites including Derby Royal Hospital (RDH), Chelsea Westminster Hospital (CWH), Croydon Health Service NHS Trust (CHS), Charing Cross Hospital (CXH) and Homerton University Hospital NHS Foundation Trust (HUH). The morbidly obese participants not seeking surgery were recruited from local GP practices or national support groups which are registered under Imperial College Healthcare NHS Trust or other NHS trusts in the UK. Table 2.2 shows the number of participants recruited at each site up to October 2019.

The study was approved by NRES Committee London (study reference: 11\LO\0935) and was performed according to the principles of the Declaration of Helsinki. To date, more than 2,000

patients have been recruited in the PMMO research study, where 1,078 were recruited before their surgery was performed while 191 individuals were recruited by their GP practices.

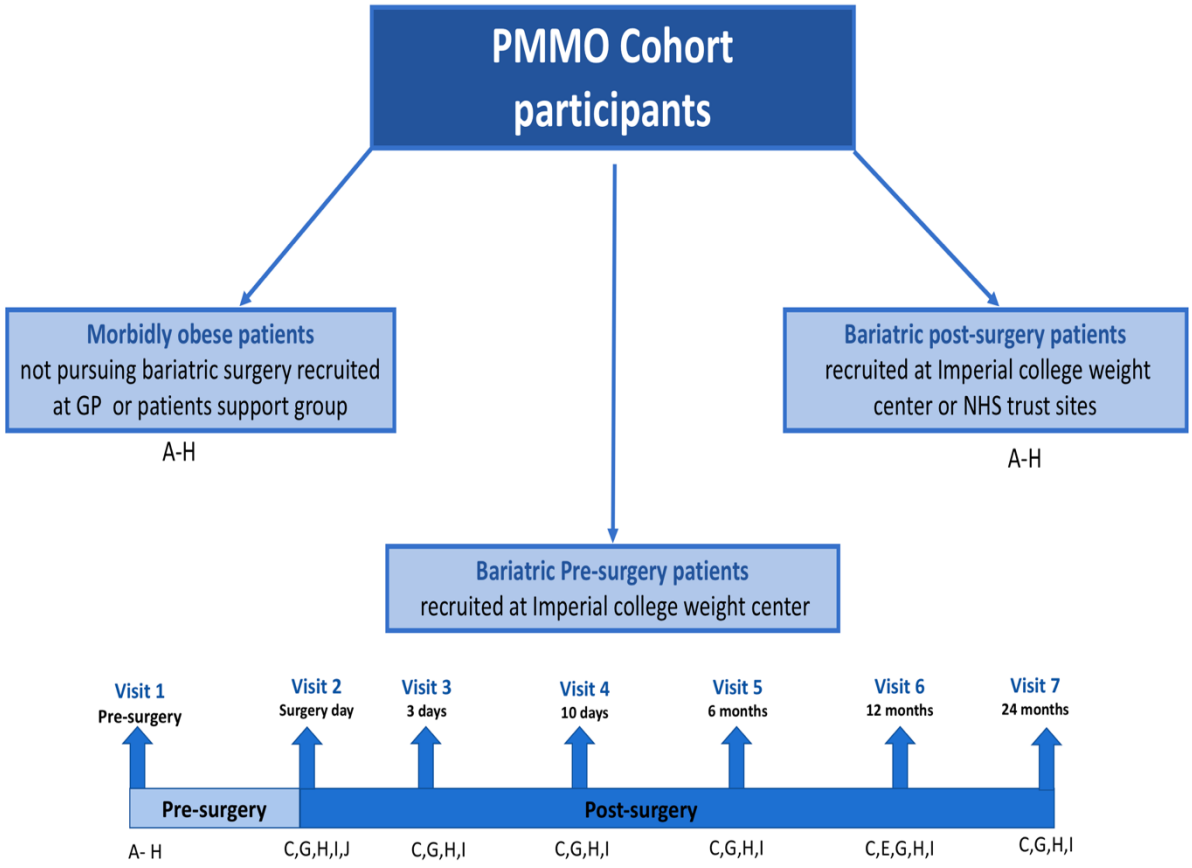
A schematic overview of the characteristics of the PMMO cohort is shown in Figure 2.1. Detailed clinical and phenotypic data, psychological questionnaires, and multiple sample types were collected at various time points from patients pursuing bariatric surgery, while patients who previously had bariatric surgery and were recruited at other health centres were required for only one research visit, illustrated in Figure 2.1. The collected blood samples are stored at -80° in a freezer. Written informed consent form was collected from all participants prior to their participation in the study.

PMMO Inclusion criteria

- Aged between 18 to 65 years.
- BMI > 35kg/m22/ pursuing/previously underwent bariatric surgery (for GP patients BMI> 35 kg/m^2 and not pursuing bariatric surgery)

PMMO Exclusion criteria

- Participants should not be receiving or planning to receive medications or treatments not approved by the European Medicines Agency
- Donation of blood in the last three months



Legend of the data collected and sample collection
A) Consent form/explanation of study
B) Demographics information: ethnicity, age, gender
C) Anthropometric (height, weight, BMI, blood pressure, pulse)
D) Family history questionnaires (ethnicity, obesity, T2D, learning difficulties, miscarriages, born defects)
F) General health, binge eating disorder, depression, other psychiatric disorders, heart disease, diabetes, bariatric surgery (type of surgery/related complication)
E) Psychological questionnaires
G) Values from routine blood tests (HbA1c, Insulin, Fasting Glucose, LDL, HDL, Cholesterol, Triglycerides, Thyroid and Kidney function, Vitamin D.
H) Blood or saliva for DNA/ RNA analysis (GP only saliva sample)
I) Urine and faeces
J) Muscle, liver, subcutaneous and visceral fat tissue

Figure 2.1: Schematic overview of the characteristics of PMMO participants

Table 2.2: Number of participants recruited at each site

SITE OF RECRUITMENT	NUMBER OF PARTICIPANTS
IMPERIAL WEIGHT CENTRE	1968
CWH: CHELSEA AND WESTMINSTER HOSPITAL	103
RDH: ROYAL DERBY HOSPITAL	50
CXH: CHARING CROSS HOSPITAL	107
HUH: HOMERTON UNIVERSITY HOSPITAL NHS FOUNDATION TRUST	100
CHS: CROYDON HEALTH SERVICE NHS TRUST	56
GP: GP PRACTICES THROUGHOUT THE UK	191

Clinical data collection of PMMO

The data collection for PMMO includes the following:

- Demographic information (gender, age, ethnic group)
- Anthropometric (height, weight, BMI, blood pressure, pulse)
- Family history (ethnicity, obesity, T2D, learning difficulties, miscarriages, birth defects)
- General health, binge eating disorder, depression, other psychiatric disorders, heart disease, diabetes, bariatric surgery (type of surgery/related complications)
- Values from routine blood tests: HbA1c, Insulin, Fasting Glucose, LDL, HDL, Cholesterol, Triglycerides, Thyroid and Kidney function, Vitamin D.
- Psychological questionnaires.
    - Health-related quality of life questionnaires; Short Form 36 health survey, and impact of weight on quality of life.
    - Eating behaviour questionnaires: Dutch eating behaviour questionnaire, three-factor eating questionnaire, eating disorder examination questionnaire
    - Mood questionnaires: Positive and Negative affect scale and Hospital Anxiety and Depression score

- The biological sample permissions for the PMMO cohort include:
- Blood or saliva sample (using Oragene DNA) for DNA and RNA analysis at the 7 visits for pre-surgery recruited patients and one sample for post-surgery recruited patients. GP services only collected saliva.
- Early morning urine and faces sample at visits 2-7.
- Muscle, liver, subcutaneous and visceral fat tissue during actual surgery.

2.2.2 EndoBarrier

Groups of participants were selected from the EndoBarrier cohort study which represents the largest randomised control trial (RCT) that compares participants with the EndoBarrier gastrointestinal liner device/duodenal-jejunal bypass sleeves (DJBS) against participants who receive best practice medical therapy, namely exercise and diet (control). The research study was approved by the Fulham Research Ethics committee, London (reference: 14/ LO/0871). The aims of the study were divided into two categories: primary objective: evaluating the effectiveness of the EndoBarrier DJBS compared to conventional medical therapy on glycemic control; and secondary objectives: investigating the effectiveness, acceptability and cost-effectiveness of each treatment. For the assessment of the objectives under both categories the following anthropometric measurements and medical records were collected during different visits for a period of 24 months: reduction in HbA1c, percentage of weight loss and other health changes including medication, cardiovascular factors and quality of life. In addition to the data collection, biological samples were collected across a number of visits which included plasma, urine, stool and blood samples as well as questionnaires. The blood sample from one of the visits was stored at -80° for our custom array genotyping [158].

The inclusion criteria to select the participants were: age 18-65, BMI 30-50 kh/m2, T2D for at least a year, oral hypoglycaemic medications and HbA1c 7.7%–11.0 while the exclusion criteria are listed in [158].There is a total of 80 participants for each of the EndoBarrier device treatment group and the conventional medical therapy treatment group. Further details into the recruitment and data collection of the study are described [158].

2.2.3 GERONIMO study

Other group of participants were included from the Geronimo cohort study which is based in Galway University Hospitals, Ireland, and was initiated in 2017 [159]. The participants in this cohort were recruited from the Milk-based Intensive Lifestyle Intervention Study (MILIS) cohort, which is a milk-based meal replacement programme and encompasses three phases (eight weeks each): initial weight loss phase, weight stabilisation phase and weight maintenance phase. The Geronimo study is designed to investigate the influence of genetic factors, specifically SNPs, in response to an intensive liquid diet and milk-based meal replacement in extremely obese individuals. The study approval was provided by the Galway University Hospitals' Central Research Ethics Committee and participants gave written informed consent for genotyping and study. Information and measurement data were collected at baseline including: weight; waist and hip circumference; blood pressure; glucose; insulin; HbA1c; cholesterol; HDL-cholesterol; triglycerides; C-peptide; liver function and lipids. Blood was stored at -80°C for DNA extraction and analysis [159].

GERONIMO Inclusion criteria

- Aged ≥ 18 year (Female and Male)
- Severe obesity with BMI ≥40 kg m^2
- Obesity with co-morbidities, such as type 2 diabetes or obstructive sleep apnoea syndrome with a BMI ≥35 kg m-2
- Commitment to attend all of the scheduled study visits.

GERONIMO Exclusion criteria

- BMI <35 kg m-2
- Inability to attend full program or clinical non-attendance after recruitment.

- Inability to walk more than 10 meters
- Female who are pregnant, or intending to become pregnant
- Females who are breast-feeding or using inadequate contraceptive methods
- Individuals with any of the follow: recent myocardial infarction (within period of 6 months), recent cholelithiasis (within period of 12 months), eating disorder, hepatic or renal dysfunction, major psychiatric disorders , type 1 diabetes, cancer, untreated arrhythmia, untreated left ventricular failure, major psychiatric disorders and previous bariatric surgery

2.2.4 GRAVITAS

Another cohort that was included in the study was the GLP-1 Receptor Agonist interention for poor responders After Bariatric Surgery (GRAVITS). This study represents a randomised controlled trial of the effect of a glucagon-like peptide-1 analogue/ GLP-1 hormone (Liraglutide) on level of weight and blood glucose level of the patients with T2D who undergo metabolic surgery/bariatric surgery. The study approval was obtained by the West London Research Ethics Committee (reference 15/LO/0780).

The inclusion criteria includes three main factors: Age 18-70; T2D; and metabolic surgery within $\geqslant$ 12 months. Written consent form was taken from all participants prior to their participation in the study. The participants were randomised into one of the two groups: (1) one time daily subcutaneous liraglutide or (2) a placebo along with increasing physical activity and a low-calorie diet. The primary outcomes include change in the overall HbA1c and weight over 26 weeks.

2.2.5 Obesity-Plus Family Cohort

Another cohort that is included in this thesis is the Obesity-Plus Family. This cohort was created by Dr Tony Goldstone at Imperial College Healthcare NHS Trust, and consists of obese probands and some members of their families. Clinical data, whole blood samples for DNA extraction and consent forms were collected for each participant. Ethical approval for the study was obtained from two National Research Ethics Service Committees (study reference 12/LO/0396), London West London and Fulham (Reference 07/Q0411/19).

2.2.6 ICLDC-Abu Dhabi Samples

From the Imperial College London Diabetic Centre (ICLDC) Abu Dhabi, a sample of a proband and his parents was used in chapter 4 for the family segregation analysis. Ethical approval for the study was obtained from the ICLDC Research Ethics Committee (IREC029).

CHAPTER 3: METHODS

3.1 Introduction

With advances genetic technologies, there are now various ways to investigate the genetics of human diseases. In this thesis, several procedures were used to investigate the genetics architecture of obesity using participants from the cohorts described in Chapter 2.

This chapter describes the genetic analyses used in this thesis. This includes the reanalysis of WES of obese individuals (Chapter 5), genotyping analysis using the customised array (Chapter 6,7,8) and Sanger sequencing analysis (Chapter 4).

3.2 List of candidate obesity and diabetes genes

3.2.1 Creation of the gene list

(A) Selection of Human Genes of obesity and diabetes

Prior to starting the different analyses in this thesis, it was vital to generate a comprehensive list of candidates genes for obesity and monogenic diabetes mellitus. Recently, many genes have been reported to be causative (or potentially causative) of Mendelian forms of obesity and diabetes [5, 6, 160, 161].

To determine genes to be investigated for the work described in this thesis, I first assembled a list of genes of interest. These candidate genes were those where there was existing evidence in the literature of being potentially causative of Mendelian obesity and diabetes. Subsequently, to expand the list further, and ensure coverage of all related genes, other available databases were used, including the Human Genetic Mutation Database (HGMD) and Online Mendelian Inheritance of Man (OMIM, http://www.omim.org/), which are further described below. Each database is summarised and described in Table 3.1.

The HGMD provides a comprehensive collection of all known germline variants in genes that are related to, or associated with, a disease. The collections of variants in HGMD are selected from literature reviews and classified into six different classes as presented Table 3.2. Variants that are reported and demonstrated to cause a clinical condition, according to the authors, are assigned in HGMD as a disease-causing mutation (DM), and variants that have some degree of uncertainty pertaining to the pathogenicity, in other words possibly disease-

causing, or another disease, are assigned as DM?. Thus, the database provides unified resource for almost all human inherited disease and functional variants. The HGDM database is available in two versions: public and professional. HGMD Professional was used in this study.

Using a phenotype-based search in the HGMD tool, the list of obesity and diabetes candidate genes was expanded to include more genes. Subsequently, using the classification features of variants in the HGMD database shown in Table 3.2, the list of obesity and diabetes genes was classified into two groups: Group 1: genes which had at least one variant causing the disease and Group 2: genes that had at least one pathogenic variant possibly causing the disease.

Online Mendelian Inheritance of Man (OMIM, http://www.omim.org/) is another valuable database that catalogues human genes and genetic disorders. This tool was used as a second source for searching for more human obesity and diabetes genes that were not in the initial list of genes to ensure the inclusion of all potentially causative genes. This was achieved by using the keywords obesity and diabetes to search and download the list of genes and then subsequently cross-checking it with the initial obesity genes list.

Table 3.1: Description of the databases used in generation of the list of genes of interest, and in the custom array design (details of which are given in Chapter 6 of this thesis).

	DATABASE	DESCRIPTION	REFERENCE
HGMD	Human Gene Mutation Database	Catalogue of known (published) gene lesions responsible for human inherited disease	www.hgmd.cf.ac.uk
NHGRI-EBI (GWAS catalog)	the National Human Genome Research Institute (NHGRI) and the European Bioinformatics Institute (EMBL-EBI)	GWAS Catalogue of published genome-wide association studies is provided within the NHGR-EBI	www.ebi.ac.uk
gnomAD	Genome Aggregation Database	Aggregating and harmonizing both exome and genome sequencing data from a wide variety of large-scale sequencing projects, and making summary data available for the wider scientific community.	https://gnomad.broadinstitute.org/
OMIM	Online Mendelian Inheritance in Man	Containing information on all known Mendelian disorders and over 15,000 genes. Focusing on the relationship between phenotype and genotype.	www.omim.org
MGI	Mouse Genome Informatics	Resource for the laboratory mouse, providing integrated genetic, genomic, and biological data to facilitate the study of human health and disease.	www.informatics.jax.org

Table 3.2: Description of HGMD variant classifications.

Variant		Description
DM	Disease-causing mutations	Reported to be pathogenic by the author corresponding report
DM?	Possible pathological mutation	Some degree of uncertainty
DP	Disease-associated polymorphisms	Evidence for a significant association with a disease/clinical phenotype along with additional evidence that the polymorphism is itself likely to be of functional relevance
FP	Functional polymorphisms	Reported that polymorphism in question exerts a direct functional effect but with no disease association reported as yet
DEP	Disease-associated polymorphisms with supporting functional evidence	Evidence for an association with a disease (but not significant)/clinical phenotype along with evidence of being of direct functional relevance

(B) Selection of rodent model genes of obesity and diabetes.

The role of mouse models in understanding the physiology of obesity and diabetes has been significant, especially in identifying obesity related genes such as *LEP, LEPR, POMC* and *MC4R*, as well as the more recent discovery of *CPE* mutation as a cause of human obesity and diabetes [162]. I, therefore, included genes identified in mice as associated with obesity and diabetes in the array. Mouse phenome databases were also searched to identify genes causing obesity or diabetes in mice when mutated. The mouse genome database (MGD, http://www.informatics.jax.org/), a widely-recognised resource for genetic, genomic and functional data on mouse models, was used to search for genes mutated in mouse models with diabetes and obesity. An overall flowchart of the process of selecting mouse model genes for obesity and diabetes is shown in Figure 3.1.

The mouse genes search began by using the search option “Obesity and Diabetes in the Human - Mouse: Disease Connection” to list the mouse model genes of the two diseases. This search option gives three further search categories, namely: 1) gene homologs x phenotypes diseases, 2) genes and 3) diseases, demonstrated in Figure 3.2. I initially started with the diseases option to select associated genes from the mouse models specifically with obesity (67 genes) and diabetes (186 genes). To expand the list further, the gene list from the genes categories option was downloaded for both obesity and diabetes. Then two types of genes were selected: genes reported to have T2D/obesity as associated human disease or genes reported to have growth/size/increased body weight as an abnormal mouse phenotype. Genes that overlapped with our initial gene list, or that do not have human homologs, and those linked with irrelevant diseases were excluded.

Finally, as the gene list for the list of the genes was finalised, a study by Hendricks et al., 2017 published a genetic analysis based on human and rodent obesity genes. Hence, a cross check between their list and our final was performed to expand the candidate genes list [163].

3.2.2 Final content of the gene list

For the obesity related genes, a total of 83 genes were selected (where 42 of those genes cause syndromic forms of obesity, and the rest caused monogenic obesity in humans). This gene list was created to be used in analysis described in Chapter 4 and 5, as well as in the creation of array content in Chapter 6.

For the diabetes-related gene content, a total of 70 genes have been found to cause or possibly cause rare types of diabetes or T2D. This list was created to be used in the array content creation in Chapter 6 only. An overall summary of obesity and diabetes genes list is given in supplementary file/Appendix 1.

Chapter 6 of this thesis describes the design of the custom array to include rare variants from the list of genes generated as described above.

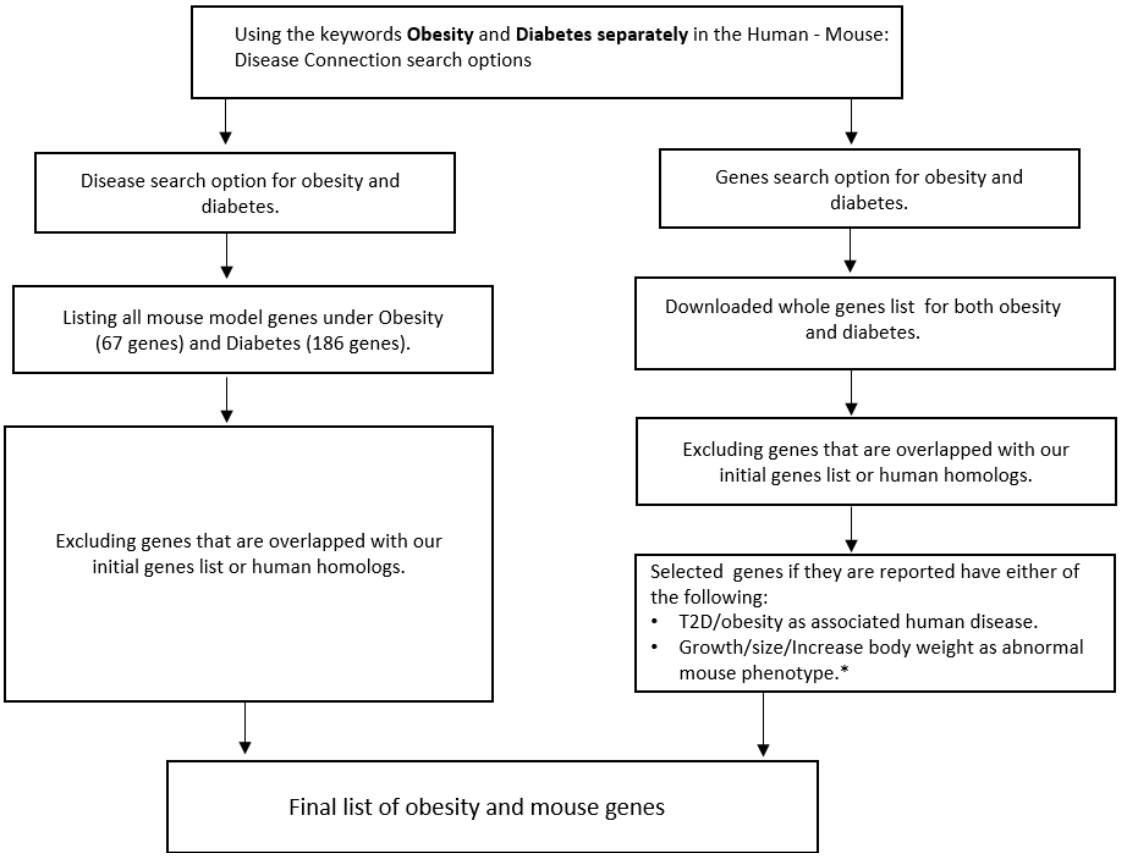


Figure 3.1 Flowchart of the process of selecting mouse model genes of obesity and diabetes.

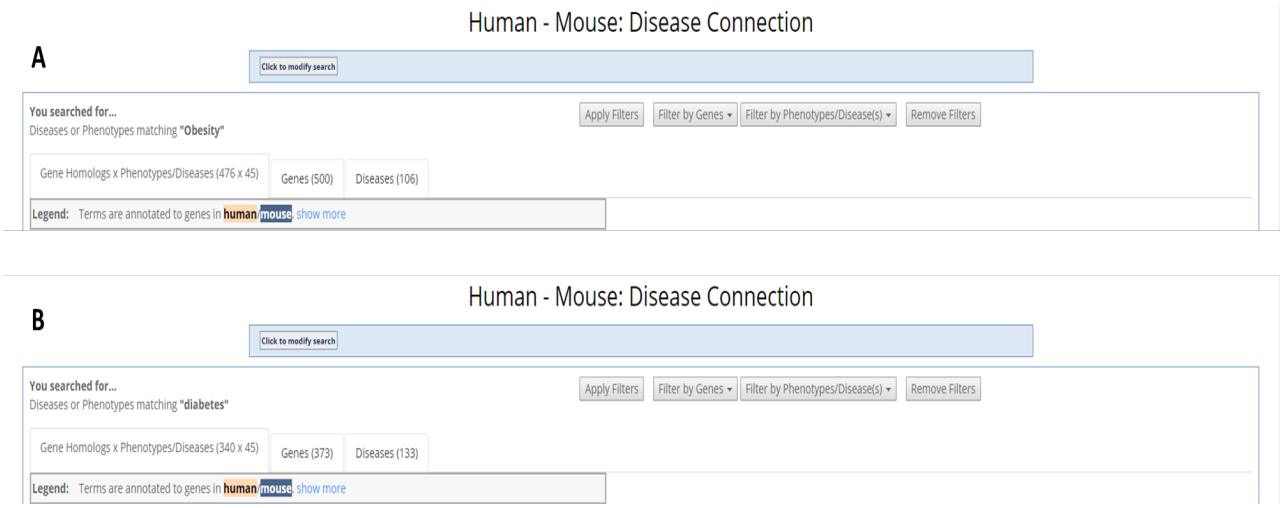


Figure 3.2: Screenshot of the three extended search categories which are given by search option Human-Mouse: Disease Connection option in Mouse Genomic Informatic database.

3.3 Whole exome sequencing

As described in the introduction Chapter, whole exome sequencing (WES) is a sequencing approach that covers only the coding and some regulatory regions of genes, representing 5% of the whole genome. Overall, an individual is estimated to differ from the reference sequence by around 30,000 variants in WES analysis, with >10,000 nonsynonymous SNV [164].

Existing WES data from 91 extreme obesity individuals from the PMMO cohort (originally analysed by Dr Sanne Alsters as part of her PhD, supervised by Professor Blakemore) were re-analysed to identify new Mendelian forms of diabetes and obesity. The WES was carried out on genomic DNA by the Genomics Laboratory of the MRC Clinical Sciences Centre, Imperial College London, UK. An enriched library was prepared using SureSelectXT Human All Exon V4+UTRs and sequencing was run on a HiSeq25000 platform, generating 100bp paired-end reads. The quality of the sequencing reads was checked with Fast QC and aligned to the reference genome build GRGh37 (hg19), using Burrows-Wheeler Aligner (BWA). For recalibration, realignments and variant calling the Genome Analysis Toolkit (GATK) were used. The annotated file was created using ANNOVAR.

Following annotation, variant interpretation was performed. In the last few years with the increased use of sequencing methodology, many commercial and open-source applications, with flexible interactive features, have been developed to facilitate variant prioritisation. These include Exomiser, The RD-Connect Genome-Phenome Analysis Platform, PhenIX (Phenotypic Interpretation of Exomes) and seqr (https://seqr.broadinstitute.org/). The preference of variant prioritisation pipeline differs among researchers and this relies on the

size of project, type of NGS project and type of questions asked. The variant analysis of the WES was carried out manually for the work described in this thesis.

For variant interpretation and analysis, a pipeline was created as shown in Figure 3.3. The first essential step is the variant filtering which includes the removal of variants with low-quality coverage variant (minimum read depth ≥10x), followed by filtering the variants according to the minor allele frequency (MAF) <1% , using public databases including gnomAD (http://gnomad.broadinstitute.org/), Exome Variant Server and the 1000 Genome Project (http://browser.1000genomes.org) or novel variants absent in those databases. Then from the retained variants, synonymous and intronic (except near the exonic region ≤ 5 bp and reported in HGMD or Clinivar) variants were excluded. The remaining variants were then screened against our list of obesity and diabetes genes as described above. Briefly, the list of genes included monogenic obesity genes, syndromic obesity genes as well as mouse model genes. Variants were then evaluated based on matching the inheritance pattern and pathogenicity prediction using different risk prediction tools summarised in Table 3.3, including CADD (Combined annotation-dependent depletion), FATHMM (Functional analysis through hidden markov models), SIFT and Polyphen, evolutionary conservation of the mutated variant and location in the gene/protein based on uniport and Protein structure effect (upon availability) [165-168].

Following this, the American College of Medical Genetic (ACMG) guidelines for variant pathogenicity classification and interpretation were applied to the candidate variants identified in this thesis. These guidelines represent the most usable approach for interpretation and evaluation of variants, relevant to Mendelian disease, in a clinical diagnostic laboratory setting. The evaluation and interpretation of the variants is based on proposed criteria that are described by Richards, et al. [36]. However, the classification of

each variant is based on all available data on this specific variant up to date, and therefore the classification may be changed as more evidence is obtained about the variant.

There are some ACMG criteria that are not applicable to monogenic obesity in comparison to other Mendelian rare diseases. Based on BS1 criterion, when the allele frequency is <1% in the general population but higher than expected for a rare disease, then it is considered benign. Despite the importance of variant frequency in assessing the potential pathogenicity of variants using public databases, these databases however consider obese individual as healthy adults. Therefore, the frequency of a candidate rare variant in such databases should be applied carefully in assessing the pathogenicity of obesity, as it can be predicted to be higher in general population than the other severe rare diseases.

Table 3.3: Risk prediction models description.

MODEL	DESCRIPTION
SIFT	Predicts whether a single amino acid substitution affects the protein function or not considering the degree of conservation for that site
POLYPHEN	Predicts the possible impact of a mutation on the structure and function of the encoded protein
CADD	Derived from 63 distinct annotations including conservation metrics such as GERP and phastCons, regions of DNAse hypersensitivity, transcription factor binding, and protein-level scores such as SIFT and PolyPhen.
FATHMM	Machine learning approach integrating 46 sequence conservation, histone modification, transcription factor binding site, and open chromatin annotations to assess the functional consequences of non-coding and coding variants

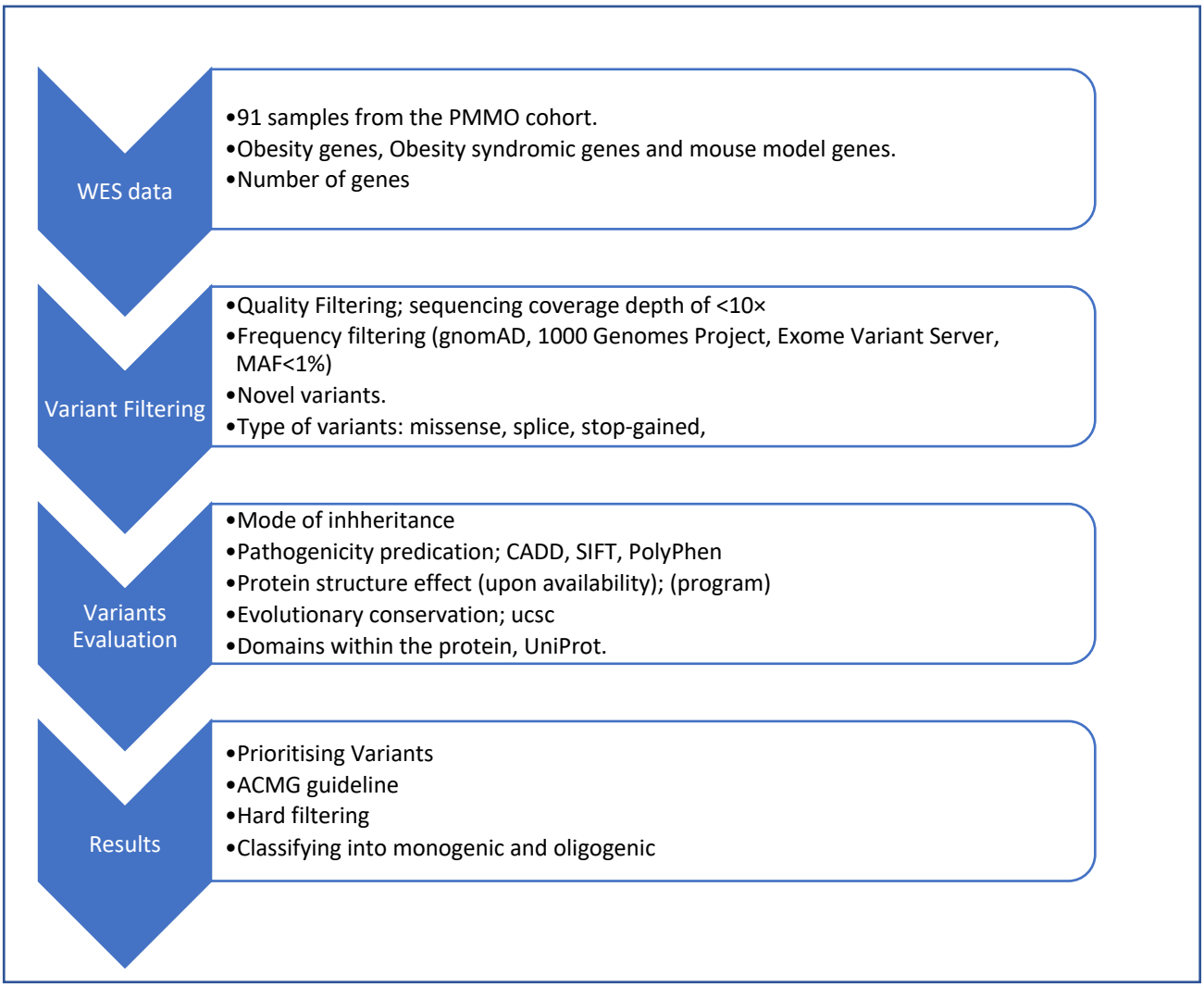


Figure 3.3 Flow chart of the whole exome sequencing analysis steps.

3.4 Application of Custom-designed Genotyping Array

This section of the methodology describes the genotyping analysis which includes study sample selection, sample processing, genotyping procedure, and the different types of QC applied (Figure 3.4). A custom Axiom array was produced by Thermofisher (originally known as Affymetrix), using variant content chosen as part of my research, as described in Chapter 6 (array design).

Figure 3.4 Flow chart of the steps involved in the in genotyping study.

3.4.1 Selection of participants for the genotyping array analysis.

For the analysis of the customised genotyping array, a total of 2068 participants were selected from all the cohorts described earlier in Chapter 2, except from the ICLDC cohort, which was not included in the genotyping analysis. The inclusion criteria for the genotyping study were: Obesity with BMI ≥35, availability of at least one biological sample (blood sample or DNA extracted from blood), and baseline information. All subjects involved in this genetic analysis have provided a signed consent form that is part of the original cohort consent form. The number of participants selected from each cohort is summarised in Table 3.4.

Below I describe the selection of subjects from each cohort. Eventually, I describe the overall characteristics of the selected samples from the different cohorts.

Table 3.4 Number of participants from each cohort included in the genotyping analysis.

Cohort	Number of participants
PMMO	1765
Endobarrier	146
GERONIMO	101
GRAVITAS	55
Obesity Plus Family	1

The PMMO cohort represents the main cohort for the genotyping analysis. Individuals lacking baseline information or who provided only saliva samples, especially the ones which were recruited at GP practice, were not included. A total of 1765 samples from the PMMO cohort were selected. As shown in Figure 3.5 (C), the selected PMMO subjects originated from different self-reported ethnic backgrounds, the majority of which were white ethnic. Those included British, Irish and any other white ethnic background, whereas the rest of the subjects belonged to Asian, Asian-British, and African populations. The high representation of female subjects (76.8%) selected for genotyping was due to the availability of the samples at that time point. The majority of the participants selected (n=879) were recruited before surgery, and 59.7% (n=525) of them have since completed the surgery. Volunteers who had undergone gastric bypass comprised the largest group of post-surgery participants. In addition, other subjects who underwent revision to other types of surgery have been included in the analysis. Figure 3.5 (A) demonstrates the number of participants in each surgery group.

The majority of the selected subjects from the PMMO cohort had a previously extracted DNA sample (n=1087), of which 304 were in plates and 783 in tubes. Around 479 samples were extracted by Dr Sanne Alsters, Dr Andrianos Yiorkas and Dr Hanis N. Ramzi between the year of 2014-2015 while the other 608 were extracted commercially at LGC group (https://www.lgcstandards.com/) . The remaining 674 subjects had blood samples collected at different time points.

The PMMO is a highly collaborative study. My contribution included cataloguing the EDTA and Paxgene tube samples from different visits and creating a sample log for each group in the cohort, that represents the location of each sample and the visit number. Previously-extracted DNA samples were re-sorted and a sample log was created to represent the location of each sample. This has improved sample tracking and makes it simple to determine whether a patient has an available extracted sample or needs one in future.

For the Endobarrier cohort, a total of 150 samples were received to be included for the genotyping analysis, However, 3 were excluded due to clotting of the collected blood. Of the 147 subjects, 72 were female and the majority of the subjects were of white ethnicity (78%), while the remaining were Asian (n=18), black (n=11), and Mixed (n=2).

For the Geronimo cohort, a total of 101 blood samples were received for the analysis in EDTA tubes, originating from 54 females and 47 males. All subjects belonged to the white ethnic group, specifically Irish white.

For the GRAVITAS cohort, a total of 55 samples were received for the analysis in EDTA tubes; 60% belonged to the white ethnic group, 32 were female and 23 were male.

A summary of all the samples selected for genotyping from the different cohorts is provided in Table 3.5. The majority of the samples are from white female volunteers.

Table 3.5 General characteristics of the samples selected for genotyping

BASELINE CHARACTERISTICS OF THE SAMPLE SELECTED FOR GENOTYPING	NUMBER	PERCENTAGE
TOTAL	2068	73
FEMALE	1466	27
MALE	547	
ETHNICITY		
BRITISH	857	42.5
IRISH	146	7.2
ANY OTHER WHITE BACKGROUND	312	15.4
WHITE AND BLACK CARIBBEAN	29	1.4
WHITE AND BLACK AFRICAN	9	0.4
WHITE AND ASIAN	15	0.7
ANY OTHER MIXED BACKGROUND	23	1.1
CARIBBEAN	106	5.2
AFRICAN	68	3.3
ANY OTHER BLACK BACKGROUND	24	1.1
INDIAN	103	5.1
PAKISTANI	32	1.5
BANGLADESHI	5	0.2
ANY OTHER ASIAN BACKGROUND	64	3.1
ANY OTHER	220	11.8
BASED ON AVAILABILITY OF DATA		

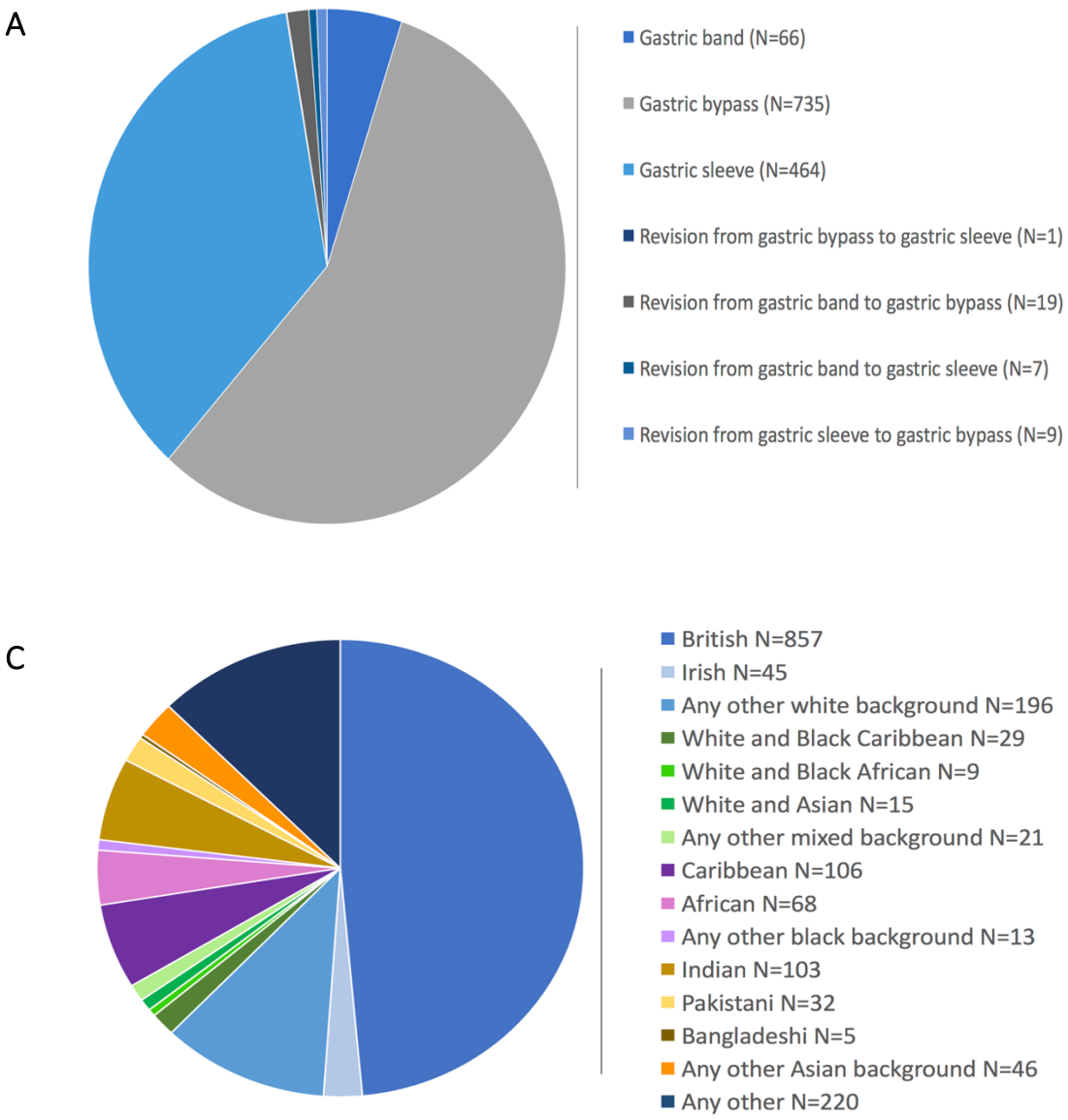


Figure 3.5 Summary of the characteristics the participants selected from the PMMO cohort.
(A) Pie chart showing the number of participants who underwent each type of surgery, majority of them belongs to the sleeve gastrectomy
(B) Bar chart showing the number of Females and Males selected from the PMMO cohort, the highest group is female.
(C) Pie chart showing the number of participants in each ethnicity group, the highest number of participants belong to the white ethnic group; specifically British.

3.4.2 Sample processing for the genotyping array

The sample processing for the genotyping array consisted of several steps including: preparation and extraction of samples, randomisation and plate layout preparation, and sample normalisation. An overview of the sample processing workflow is shown in Figure 3.6. To process the samples at high efficiency and to minimise mix-up, loss, contamination or degradation, most of the steps described here were automated at LGC based on an instruction sheet I created to demonstrate the steps.

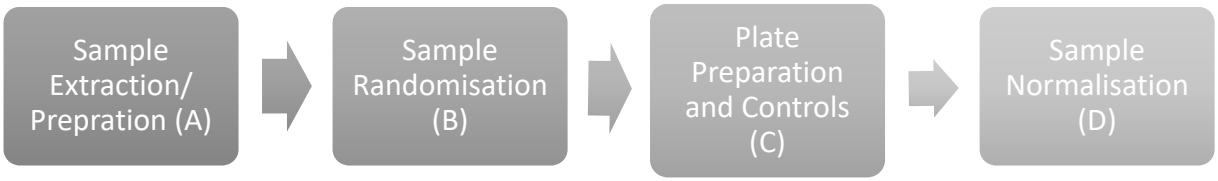


Figure 3.6 The workflow of the sample processing workflow.

(A) Sample preparation and DNA extraction

The Axiom genotyping application can be used on gDNA derived from a variety of human tissues including saliva and blood, and from cell lines. For studies described in this thesis, blood was the only source of gDNA used. According to the sample procedure collection in each cohort, the biological samples were collected and stored in EDTA (3.6 mg) blood collection tubes, except Endobrarrier samples, which were stored in cryo tubes at -80°C.

Overall, the biological samples of the selected participants for genotyping were divided into three categories: non-extracted blood samples, DNA extracted in tubes, and DNA extracted in plates. The previously extracted DNA samples were from the PMMO group; of these, a total of 799 samples were in tubes and 304 samples in different plates.

To assess the quality of the previously or newly extracted samples, DNA quantification was performed by LGC. Samples containing gDNA at concentrations lower than 50 ng/ul and volumes lower than 20 ul were replaced with another sample from the same individual, or a sample from a different individual. Samples were then transferred into Matrix 2D tubes barcoded by LGC to enhance the procedure of the next steps.

(B) Sample randomisation

Randomisation of samples is considered a vital experimental design step to obtain accurate results and to avoid experimental artifacts such as batch effects. In general, the samples selected for genotyping in this study differ in terms of study cohort, collection sites, processing time points and sites, and storage vessel (such plate or tube). Thus, to ensure the randomisation of as many processing and characteristic variables as possible across all the 22 plates, randomisation was introduced at four different levels: cohort group, ethnicity, sex, and sample extraction format. The randomisation was performed on Excel based on using the formula, =RAND().

(C) Plate layout

After the completion of sample randomisation, plate layouts were created to assign positions of samples and controls. The genotyping was carried out in 22 96-well size plates. Each plate contained the following: 94 samples (at least one sample was previously whole exome sequenced), 1 internal control representing a duplicate sample from the plate, and 1 Thermofisher control. The controls were located in the same positions in each plate. The layouts were used to transfer the samples from their original location in the 2D matrix tube to the Thermo-Fast 96-well skirted plate (Catalogue #AB-0800) in columns (A1, B1, etc) , using barcode by robots. Figure 3.7 demonstrates the overall plate layouts of the randomised samples on the 22 plates.

(D) Sample normalisation for the final plates

Normalisation was carried out after the completion of samples transfer according to their position on the final plate. All samples had a volume of 30 µl and were normalised to the minimum concentration of 10ng/ µl.

Once normalisation was completed, plates were sealed by adhesive seals and sent for genotyping at Oxford Genomics Centre. An excel file containing the plate layouts, samples and plate names were provided to Oxford Genomics Centre. Prior to genotyping, additional tests carried out on the samplesincluding quantification (Picogreen) and spectrophotometer readings, to ensure that they met the minimum requirements and quality control.


Column 1
Column 2
Column 3
Column 4
Plate 1
Plate 2
Plate 3
Plate 4
Plate 5
Plate 6
Plate 7
Plate 8
Plate 9
Plate 10
Plate 11

Plate 12

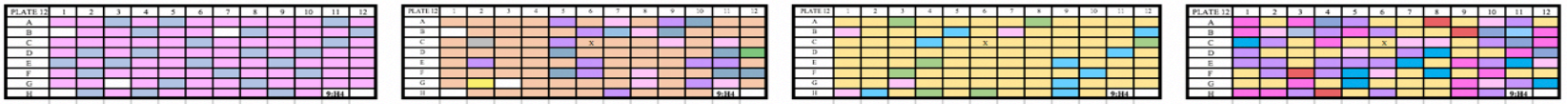


Plate 13

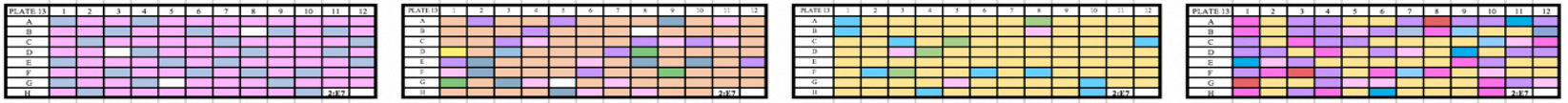


Plate 14

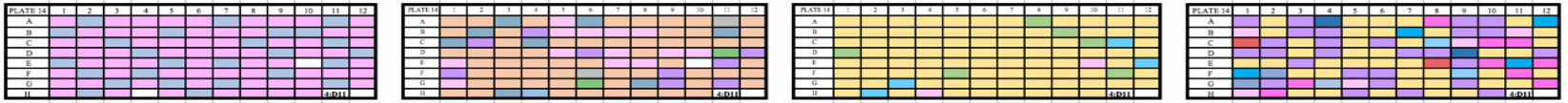


Plate 15

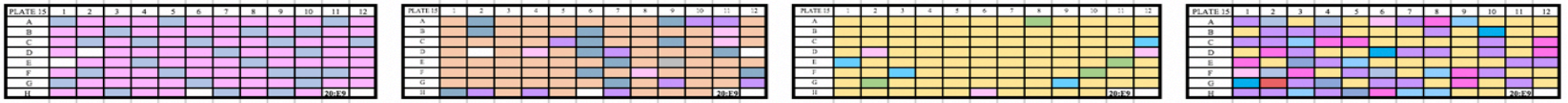


Plate 16

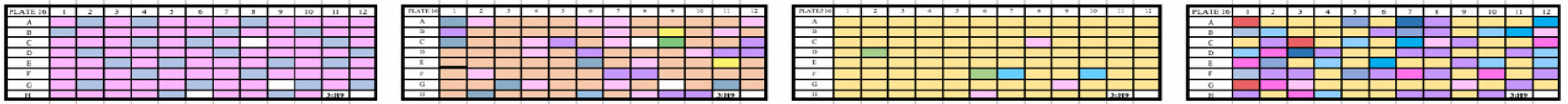


Plate 17

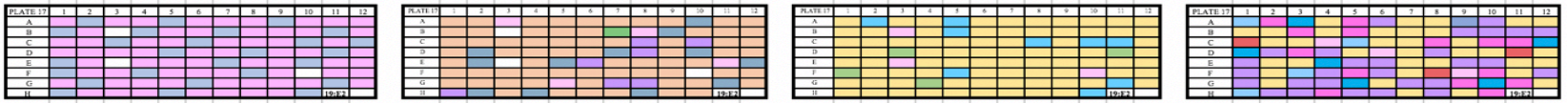


Plate 18

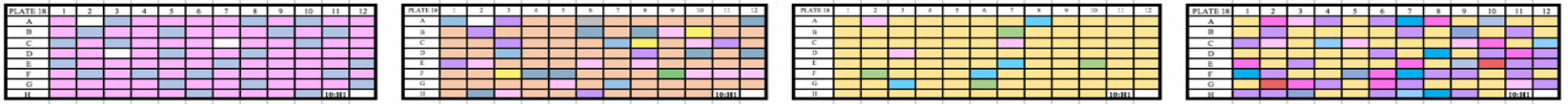


Plate 19

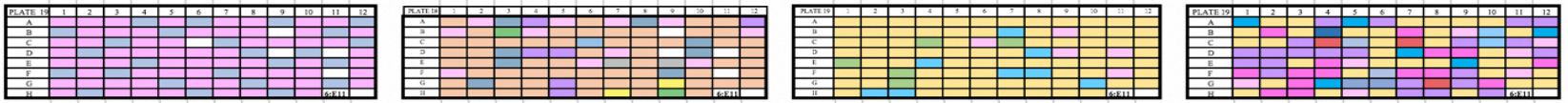


Plate 20

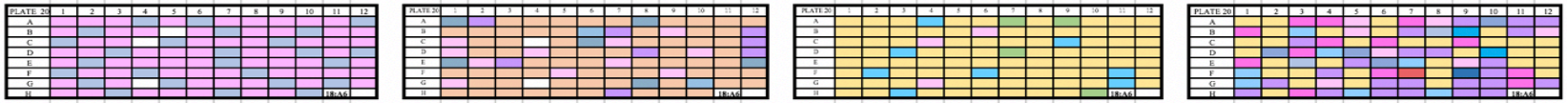
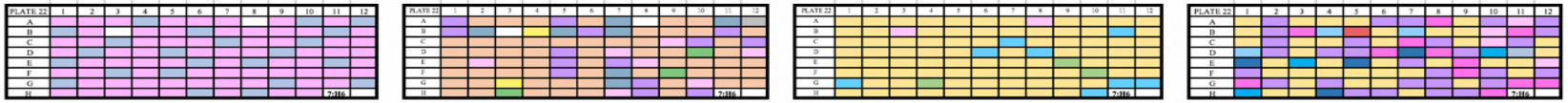


Plate 21

Plate 22



COLOUR APPENDIX	
First Column	Female
	Male
Second Column	White
	Mixed: white and Asian
	Mixed: White and Black Caribbean
	Any other Mixed
	Mixed: White and Black African
	Any other
	Asian/Asian British
	Black
Third Column	PMMO Cohort
	Geronimo Cohort
	Endobarrier Cohort
	Tricia Group
Fourth Column	Extraction Batch 1
	Extraction Batch 2
	Extraction Batch 3
	Extraction Batch 4
	Previously Extracted in tubes
	Previously Extracted in plate 1
	Previously Extracted in plate 2
	Previously Extracted in plate 3
	Previously Extracted in plate 4

Figure 3.7 Description of the samples randomisation for the final plates. The twenty two plates that were used in genotyping were randomized at four different level: sex (Column 1), ethnicities (Column 2), cohorts/research groups (Column 3), reference of samples processing (Column 4). The colors in the colors box corresponds to the colors in each column.

3.4.3 Genotyping at Oxford Genomics.

The genotyping procedure was performed according to the Thermofisher guidelines (https://tools.thermofisher.com/content/sfs/manuals/703434.pdf) at the Oxford Genomics Centre as a commercial service (https://www.well.ox.ac.uk/). The Axiom genotyping assay was processed by the GeneTitan Multi-Channel (MC) Instrument. An overview of the axiom genotyping procedure is shown in Figure 3.8 Initially, target preparation was carried out, where the DNA is amplified and fragmented into different sizes (25 bp – 235bp). The fragments underwent purification, resuspension and hybridization to the Axiom array plate. Subsequently, the bound target underwent a wash phase to ensure the removal of non-specific binding or background as well as to reduce any noise in the background due to random ligation events. Eventually, the arrays were stained ready for processing on the GeneTitan MC Instrument.

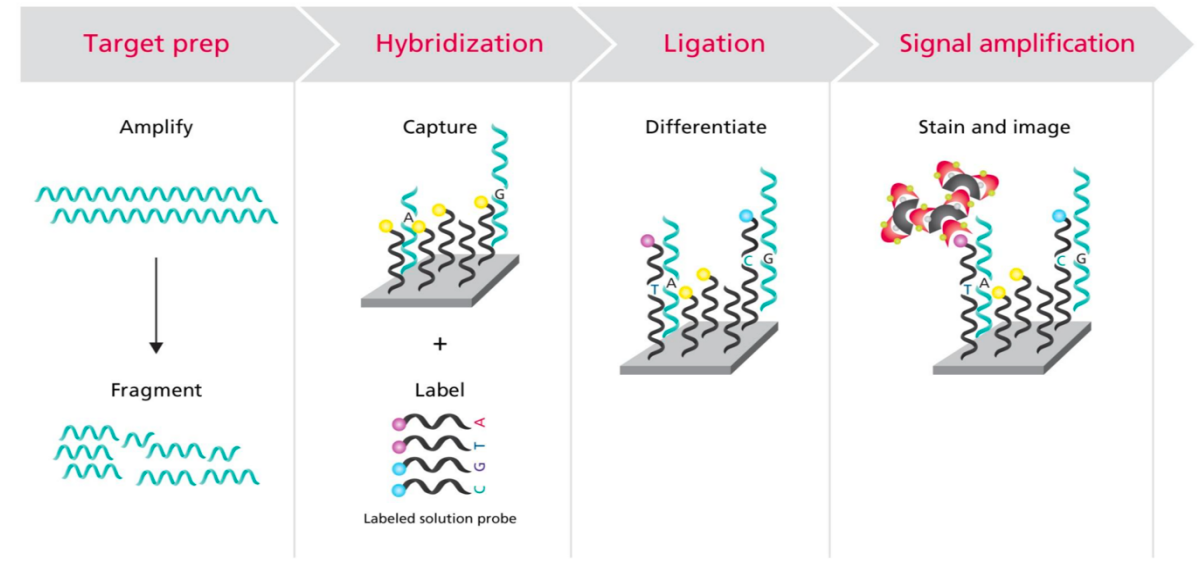


Figure 3.8: Overview of the steps involves in the Axiom Genotyping Assay (source: www.affymetrix.com)

3.4.4 Genotyping and Quality Control (QC)

3.4.4.1 Axiom GT1 genotypes and SNP cluster.

Upon the completion of the genotyping reaction in Oxford Genomics, cell intensity files (.CEL) were generated by the GeneTitan Multi-Channel (MC) Instrument. The cell intensity files of the genotyping of A and B alleles were converted into coordinate space X and Y by the Axiom GT1 algorithm to automatically produce the read count of the allele (genotyping call) and asses the quality of the called genotype.

The AxiomGT1 algorithm is a clustering algorithm based on pre-defined locations of the genotyping cluster. There are two types of priors in the AxiomGT1 algorithm: generic priors and SNP-specific priors. Generic priors describes where pre-positioned genotype cluster locations for all SNP are the same. It is recommended for large batches of >96 samples and diploid genome studies such as GWAS. The SNP-specific priors are pre-defined positions specific for each SNP. It is more appropriate for small batches or even large batches addressing monomorphic SNPs, genomic selection and allopolyploid genomes to provide accurate genotyping calls in the absence of intensities of the minor allele. As our array is based on rare variants, specific priors were used for the genotyping array to enhance the accurate detection of rare variants.

A SNP cluster plot, as shown in Figure 3.9, represents one probe set which interrogates a specific SNP. Each dot/point in the plot represents one sample after the genotype intensities of A and B allele have been converted into coordinate space X and Y by the Axiom GT1 algorithm. The X dimension corresponds to the main information for distinguishing genotype

clusters – the contrast - and is calculated as Log2(A_signal/ B_signal). The Y dimension is defined as a size that is [Log2(A_signal) + Log2(B_signal)]/2 (Size/Strength).

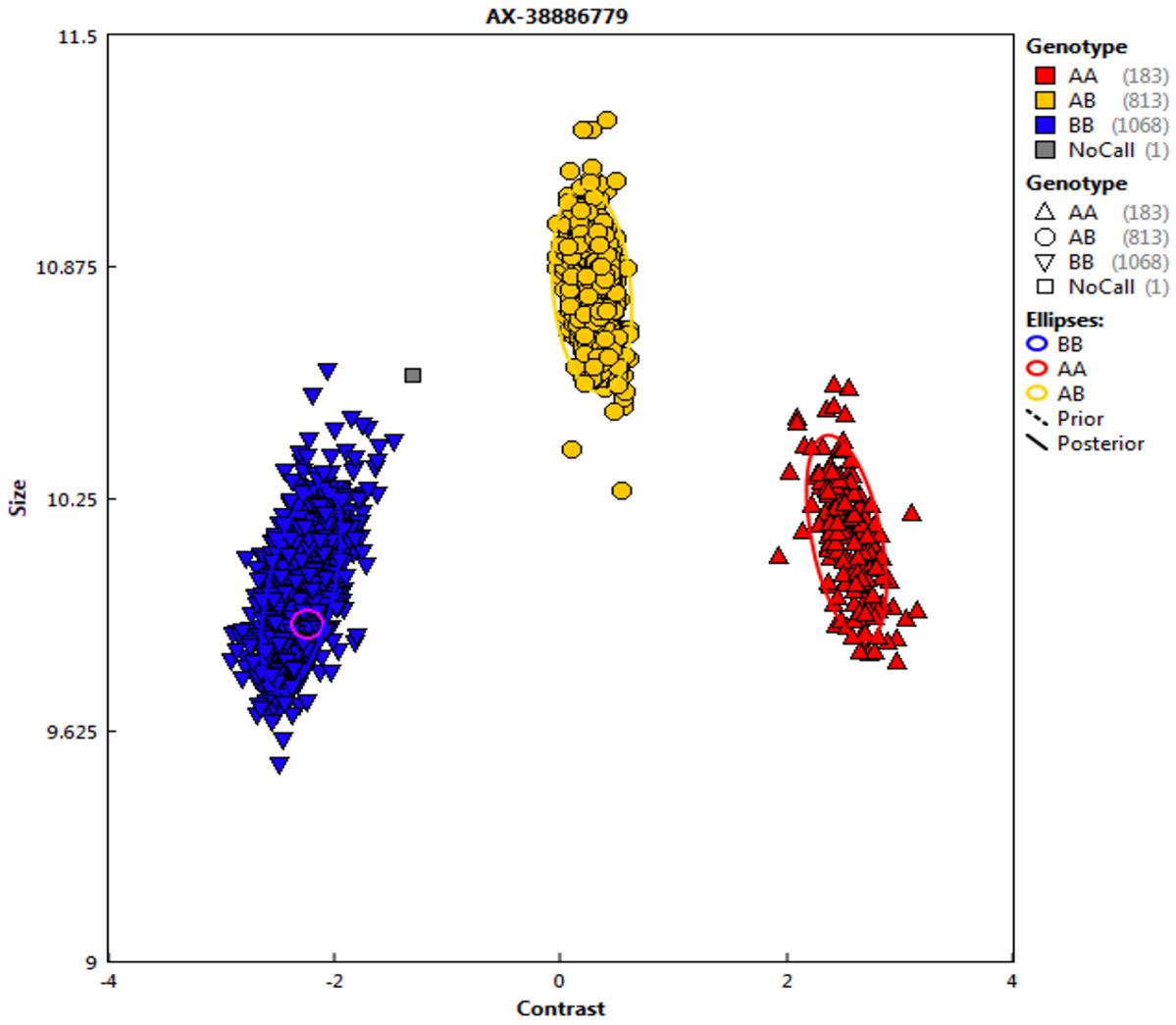


Figure 3.9 Example of a SNP Cluster Plot generated by the Axiom Analysis Suite. The SNP cluster plot represents one probe set that is interrogated a single SNP. The clustering is carried out in two dimensions: X and Y. The X dimension correspond to the main information for distinguishing genotype clusters and is calculated as Log2(A_signal/ B_signal). The Y dimension is defined as a size that is [Log2(A_signal) + Log2(B_signal)]/2 (Size/Strength). Samples are colored and shaped circles and triangles. AA calls (red triangles), BB calls (blue triangles upside down), AB calls (gold circles), AA calls as red triangles.

3.4.4.2 Quality Control for the Axiom array.

The raw genetic data have may contain errors due to technical and genotyping issues, such as poor quality of DNA samples, mix-ups of samples, contamination, poor variant calling/poor performance of genotyping probes and poor DNA hybridization in the array, that can lead to false results and major errors in analysing the data or affecting the cluster quality. Thus, one of the fundamental steps in dealing with the genetic data is the QC. Applying executive steps to check quality is important to ensure the generation of reliable data for the further analysis steps.

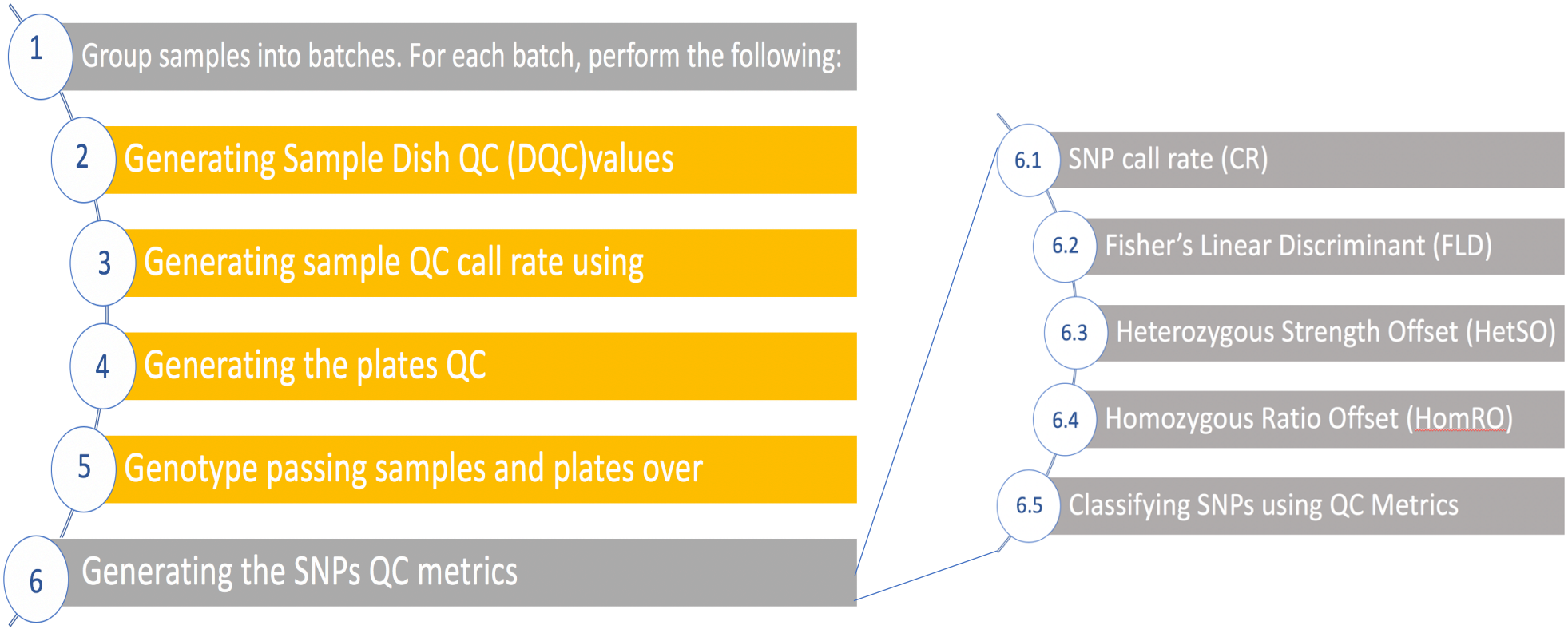
3.4.4.3 Pre-genotyping QC

The QC for prior to the genotyping of the array includes assessing the quality of the DNA to check that it meets the Genomic DNA input requirements according to Thermofisher Axiom: gDNA concentration 11.5 ng/ul , volume 8.7 ul and input mass per well 150 ng. The samples which do not meet the requirements should be either replaced with another samples or excluded from genotyping.

3.4.4.4 Post-Genotyping QC

The post genotyping QC was classified into two levels; Sample level QC (step1-5) which determined the quality of each genotyped sample and SNP level QC which detected problematic SNPs, and in total consisted of 5 steps, summarised in Figure 3.10 which will be discussed in detail below.

Figure 3.10: Summary of the Quality Control steps for the Axiom array.



(1) Group samples

The initial step prior to the QC is grouping samples into batches where the minimum batch consists of 96 samples and the maximum 5000 samples. It is recommended to group the plates in order based on the processing time. There are some exceptions to these batches, which include samples extracted from different tissues and samples processed/genotyped at different times, which should be grouped into separate batches.

For our study, all the DNA samples were extracted from blood and were treated as one batch by combining plates 1 to 22 in a single run (QC steps 2 to 6).

(2) Samples DQC Sample QC call rate

The quality of the individual samples was determined by performing Dish QC (DQC), call rate test (QC call rate), and average call rate of passing samples. The DQC is one of the most important and mandatory quality tests carried out by the Axiom platform. It is based on measuring the intensities of probe sequence at genomic sites (non-polymorphic genomic sites) that do not vary between individuals. The measurement is based on evaluating the overlap between the homozygous peaks AT and GC signals, and is calculated using contrast values from both probes AT and GC. The samples need to pass a threshold of the DQC score, which is ≥ 97%, in order to be genotyped. It is worth noting that the DQC is essential for assessing sample quality, but insufficient to detect sample contamination.

(3) Sample QC call rate.

The second QC to detect problematic samples is the sample QC call rate, defined as the proportion of called SNPs for each sample over all SNPs. The sample QC call rate is based on a subset of probes (around 20,000 probes) that have consistently showed good clustering. Thus this step is considered as a fundamental step in all genotyping data and the minimum standard quality threshold of the sample call rate value is 97%.

Generally, samples which do not match the default threshold of either DQC or sample QC call rate should be either excluded from the genotyping analysis or reprocessed at the Oxford Genomics Centre.

(4) Sample plate QC

Usually there is a clear correlation between high DQC values and high sample call rates. Therefore, the third step represents the plate QC which is designed to determine the plates with high number of samples failing DQC and QC call rate. The passing rate of plate is 95% for samples from blood and tissues. Affected plates should be removed or reprocessed separately.

$$\text{Plate pass rate} = \frac{\textit{Samples passing DQC and 97\% QC call rate}}{\textit{Total samples on the plate}} \text{ X } 100$$

(5) Sample passing samples QC

In this step, once the samples pass sample QC and plate QC, they are co-clustered and genotype calls are generated. The AxiomGT1 algorithm produces calls for all probe sets on the array.

(6) Generating the SNP QC

This step is applied to determine which probe sets performed well. The evaluation of the probe set for each SNP is based on two important features: well-clustered intensities and genotype calling meeting the recommendations for further statistical tests.

For some markers of the Axiom array, more than one probe set is used to interrogate one SNP, this is specially for SNP that have not been assessed before or are very important (e.g. rare). In such case one of the probe sets for each SNP is identified as the best probe set. This is achieved through steps of SNPs QC metrics which are described below and summarised in Figure 3.10 (6).

(A) Call Rate (CR): represents the proportion of the individuals that have a successful genotyping call for any of AA, BB, or AB for particular SNP over the total number of samples which attempt to call a genotype for the SNP. This SNP QC is important in measuring the quality of the genotyping cluster and data completeness. The recommended threshold is >95%.

SNP Call Rate = #Samples Called/N

(B) Fisher's Linear Discriminant (FLD), is defined as a direct measure of the cluster quality of a particular SNP that is based on how well separated the center of the cluster is and how low the variance is in and around the center. A high quality SNP will be narrowly clustered and well separated. The SNP (CR) and FLD are usually correlated but FLD can better detect other issues that are not detected with poor call rate. The threshold of the FLD is >3.6. HomFLD. SNPs with High FLD tend to have well separated cluster in the X-dimensional as shown in Figure 3.11.A, while SNP with low FLD has clusters that tend to be very close together to the center as in the X-dimensional shown in Figure 3.11.B.

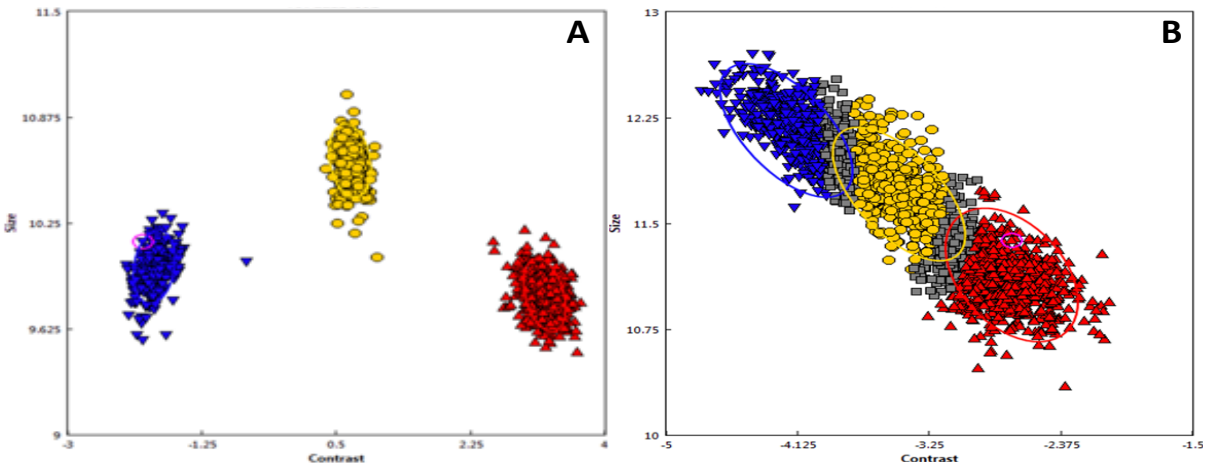


Figure 3.11. Example of variants with different values of FLD. (A) SNP with high value of FLD (10.167); the three clusters are well separate. (B) SNP with low value of FLD (3.404); the three clusters are close together.

(C) Strength Offset (HetSO), is a measurement in the Y-dimension of how far above or below the Heterozygous cluster is from the homozygous clusters locations. The recommended value of HetSO is 0.35. A low value for the HestSO indicates misclustering or variation between the sample and the reference genome. SNPs with low HestSO values should be removed or re-analysed further. Examples of low HesTSO values are shown in Figure 3.12.

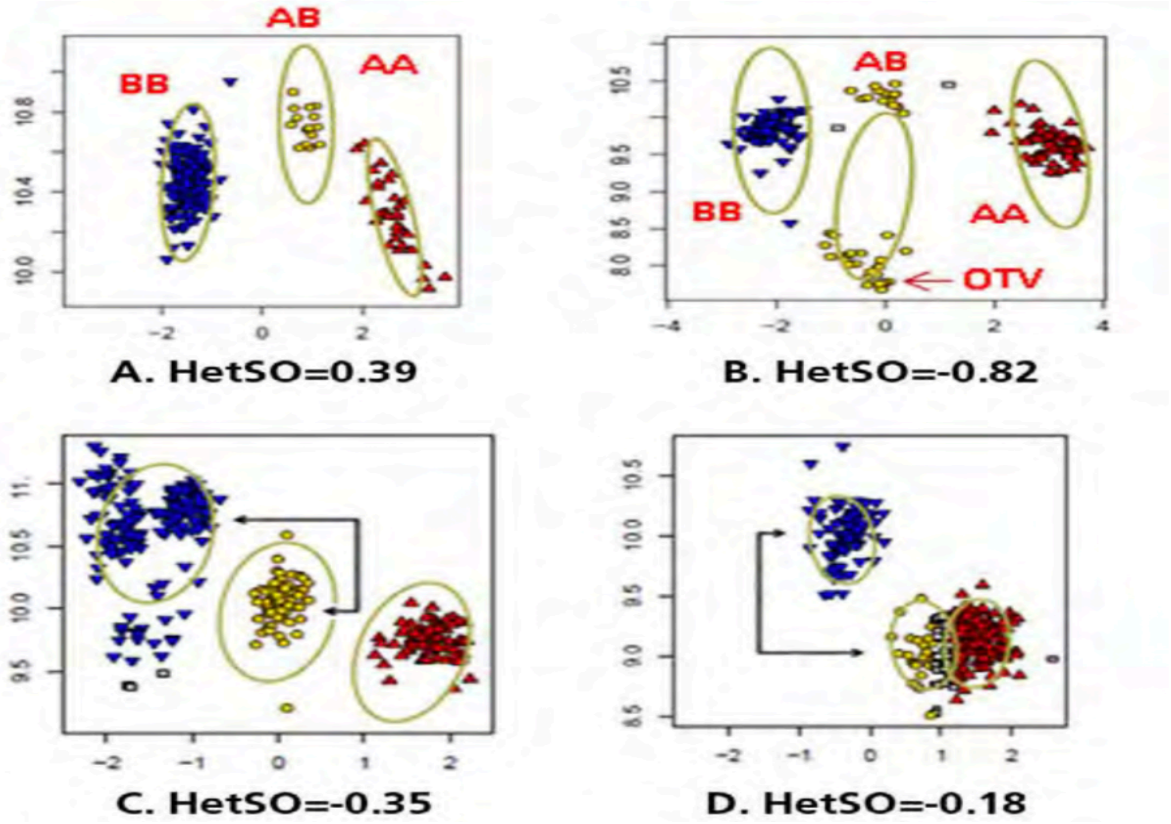


Figure 3.12. Examples of different low values of HetSO. (A) SNP with the recommended HetSO value. (B-D) SNPs with low values of HetSO. (Source: AxiomTM Genotyping Solution)

D) Homozygous Ratio Offset (HomRO), is defined as the distance to zero from the two homozygous clusters to heterozygous cluster location (which is at 0 on the X-axis) to identify dislocated clusters or unexpected positions. A negative HomRO value indicates that one of the homozygous cluster is misclustered and located on the wrong side. The recommended HomRO threshold is greater than 1. This is demonstrated in Figure 3.13.

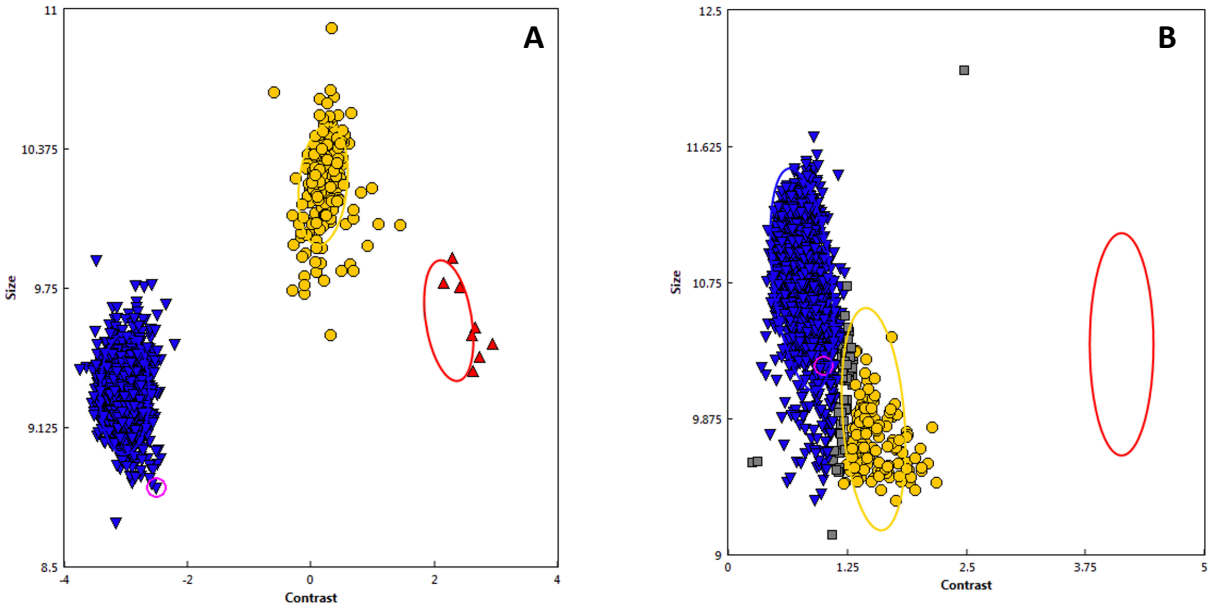


Figure 3.13. Example of variants with different values of HomRO. (A) SNP with high value of HomRO (1.002); the three clusters are well separated, with the two homozygous clusters on the correct side of the Heterozygous cluster (B) SNP with low value of HomRO (-0.775), dislocated clusters or unexpected positions.

(E) Classifying SNP: The selected SNP probe sets , which are based on the above default SNP QC criteria, are classified into one of the seven classification categories namely: MonoHighResolution, PolyHighResolution, NoMinorHom, off-target variants (OTV) and, CallRateBelowThreshold. These are demonstrated in figure 3.14 and described in the Table 3.6 below. Once the probe sets are classified, a list of "best and recommended" probe sets is generated for further analysis.

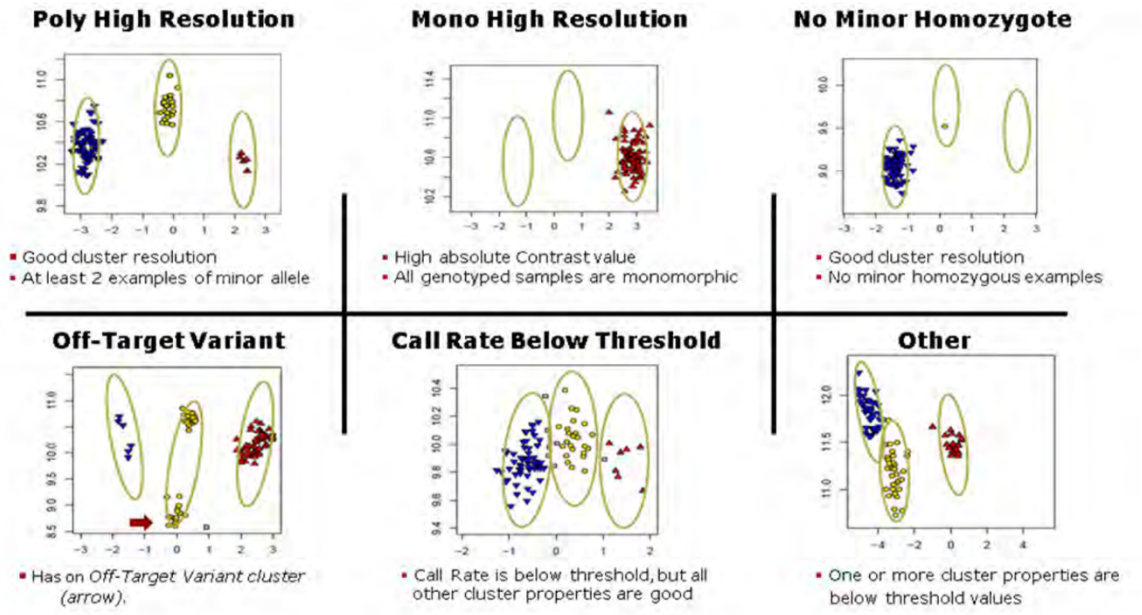


Figure 3.14 Examples of cluster plots of the different SNPs classifications categories. (Source: AxiomTM Genotyping Solution)

Table 3.6 Description of the SNP classifications categories clusters

SNP CLASSIFICATION CATEGORIES	DESCRIPTIONS
POLY HIGH RESOLUTION	When Three clusters with good separation are found; (polymorphic SNPs)
NOMINORHOM	When only Two clusters are found; one for a homozygous genotype and the Heterozygous genotype (No cluster for the other Homozygous)
OFF-TARGET VARIANTS (OTV)	When an additional (off-target) cluster is formed beside the three or two regular clusters is formed
MONO-HIGH-RESOLUTION	A single homozygous cluster is found (Monomorphic SNPs)
CALL RATE BELOW THRESHOLD	The SNPs fail the call rate (below the threshold 0.97)
OTHERS	SNPs can not be classified into any of the categories (unconverted SNPs)

3.4.4.5 WES previously detected variants

Another way of assessing the efficiency and validating the genotyping performance of the axiom is through confirming the presence of variants that have been identified in individuals by WES. This is in house QC which is not provided by Thermofisher is described in Chapter 7 of this thesis. A total of 77 samples who were previously whole exome sequenced were included in this genotyping analysis. One CNV, chr16:29675061-30215702, and 10 variants, have been selected for this QC stage. Table 3.7 shows the details of those variants.

Table 3.7 Previously identified variants from WES used for array QC.

	Variant Array ID	Gene	RS ID	Type of variant	MAF	Subject	Sanger Seq confirmed
1	AX-212366779	*IGSF1*	rs749977306	stopgain	N/A	S34	Yes
2	AX-314666608	*MYT1L*	rs201765281	Missense	0.0001365	S4	N/A
3	AX-212386946	*NTRK2*	rs753056075	Missense	0.000003977	S36	N/A
4	AX-212386946	*RAI1*	rs147844401	Missense	0.00009573	S28	N/A
5	AX-83359409	*ENPP1*	rs190947144	Missense	0.0004078	S37	N/A
6	AX-83586350	*MCHR1*	rs45439291	Missense	0.003407	S1	N/A
7	AX-82915434	*AFF4*	rs139490054	Missense	0.002310	S12	N/A
8	AX-82966004	*PEG3*	rs56237501	Missense	0.004053	S26	N/A
9	AX-83513405	*RAI1*	rs113208290	Missense	0.003726	S28	N/A
10	AX-88752465	*PTEN*	rs121909238	Missense	N/A	S35	Yes

MAF(%) gnomAD MAF.

3.4.4.6 Duplicate samples

Another in house QC is the inclusion of a total of 22 duplicate samples that were distributed across the 22 plates. The genotyping calls were assessed for each of the duplicates to determine the overall efficiency and performance of the Axiom array.

3.4.4.7 Sex discordance

Sex discordance is an important QC step for checking if samples have incorrect sex compared to the recorded/annotated sex in the database. This is a vital step to detect sample mix-up or other sample processing errors. Generally, the check for sex discrepancy is performed by assessing the SNP intensities on the sex chromosome (X and Y) for all samples. Males normally have a single copy of the X-chromosome and females have two copies. Identified SNPs on this chromosome are considered hemizygous (although they are called as homozygous by the genotyping software). Thus, the default threshold have been considered in assessing the sex discrepancies where males have higher (>0.8) “homozygosity” on the X chromosome while Females have a value of homozygosity of <0.2.

3.5 Copy Number (CN) Variation QC of the custom array genotyping

In addition to genotyping of SNP and rare variants, Axiom arrays are designed to detect copy number (CN) changes. The principle behind CN estimation is the use of log2 ratio and beta allele frequency (BAF). The CNV analysis was performed through Axiom Analysis Suite. A summary of the CNV analysis work flow is Fig 3.15

Figure 3.15 Summary of the CNVs analysis steps.

(A) Reference Creation

Generation of a reference is an important step in running the analyses which can affect the generated results. There are two ways of creating the reference, namely each plate's own reference, or universal reference. The plate's own reference is generated based on all samples in that particular plate. Universal reference is where a single reference is used for all of the plates - generated based on a selected number of samples or on samples in all the plates.

(B) Sample QC

The Axiom Analysis provide two types of QC metrics to assess the quality of the produced CNV data namely Median Absolute Pair-wise Difference (MAPD) and waviness-standard deviation (waviness-SD) metrics. The MAPD represents the median of the absolute value of all pair-wise differences between adjacent probesets, on other words, it is a measure of the variation of all probeset across the genome. The recommended threshold of the MAPD metrics is 0.35. The higher the value of MAPD the greater the genomic noise. Waviness-SD metric on the other hands, is calculated based on the variation of probesets that focuses on variation of long-range and insensitive to variation of short-range. The waviness-SD metric an indication of noise in the data which is due to processing or sample effect. The recommended threshold of waviness-SD is 0.1 and it is important in obtaining high accuracy of CN calls.

(C) Process intensity data to generate log 2 ratio

After creating the reference, Axiom Analysis Suite software calculates the CNV calling based on the log2 ratio and B allele frequency (BAF); which measures of the Heterozygosity of any location and is generated for each probset in the data. These two options are used either for fixed-region CN analysis, when the breakpoint of each CNV is defined, or for CN discovery analysis, where the breakpoints are unknown and need to be determined from the analysis. For the work in this thesis, I focused on fixed region analysis.

(D) Visualisation and identification of CNVs

The Axiom Analysis Suite application used Integrative Genomic Viewer for visualisation of the CNVs (www.thermofishehr.com).

3.6 Sanger sequencing

Sanger sequencing was used to confirm some of the identified variants from the whole exome sequences. The Sanger sequencing uses an amplified product of the polymerase chain reaction (PCR) to target specific regions of the genome.

PCR is a fundamental method for amplification of a specific region in the genome. It uses a DNA template, two sets of primers (matched to the sequence at both ends of the region of interest), a PCR reaction master mixture (buffer, MgCl2, dNTPs, distilled H2O and the thermostable enzyme Taq polymerase) and is carried out under optimal reaction conditions for each primer pair. In our study, to amplify the region covering the variants, the following PCR steps were used for all variants:

1. Two sets of primers were designed for each variant using Primer 3 Web version 4.0.0 (https://primer3.ut.ee/). The following criteria were considered in selecting the primers: (1) Primer Length: minimum 18bp long, (2) Base composition: G-C content between 40-60 to ensure optimal annealing, (3) Melting temperature: optimal annealing temperature for the primers was determined either by a temperature gradient PCR reaction or alternatively calculated by subtracting 5°C of the melting temperature (MT) from the pair of primers with the lowest melting temperature given by the manufacturer. The specificity of the primers was confirmed using tools available at https://genome.ucsc.edu/.

2. A PCR reaction master mixture was prepared through combining the following components in a sterile, nuclease-free microcentrifuge tube: (buffer, MgCl2, dNTPs, distilled H2O and the thermostable enzyme Taq polymerase) (Table 3.8). The master mix was from the GoTaq Flexi DNA polymerase kit, Promega, and dNTPs from Thermo Fisher Scientific.

Table 3.8 Component of each polymerase chain reaction (PCR).

Master Mix Component	Final Volume per reaction (50 µl)
5X Colorless GoTaq® Flexi Buffer1	10µl
MgCl2 Solution (25mM)	5 µl
dNTPs (10 mM)	1 µl
upstream primer (10µM)	1 µl
downstream primer (10µM)	1 µl
DNA template (100 ng/ µL)	1 µl
Taq DNA Polymerase (5 units/µL)	0.25 µl
Nuclease-Free Water to	30.75 µl

3. The standard PCR thermocycler protocol was carried out on a G-STORM GS4 thermal cycler (Somerton Biotechnology Centre, UK) for all the samples, with optimal reaction conditions for each primer pair as shown in Table 3.9.

4. Agarose gel electrophoresis was carried out to determine the quality and size of DNA fragments/PCR product, using a 1% agarose gel in TAE buffer at 60 to 80 V as follow:

    a. 0.5 of agarose powder was mixed with 50 mL 1xTAE in a Duran bottle

b. The mixture was heated up to boiling point (between 1-3 min) where the agarose is completely dissolved and then cooled down for around 2 mins

c. 2.5 µ of SafeView dye was added to the mixture and mixed by gentle swirling. Then it was poured into a gel tray. Once solidified, the agarose gel is placed into the electrophoresis unit.

d. 2 µ of loading buffer was added to 10 µ of each DNA sample from the PCR product. Followed by loading them and the molecular weight ladder into the lanes of the agarose gel.

e. Running the gel in TAE buffer at 60 to 80 for around about 1 to 1.5 hours then visualize the fragments of DNA using UV light.

5. Subsequently the PCR product was sent for purification and Sanger sequencing. This was carried out by the Genewiz company (https://www.genewiz.com). Sequencing was performed in both directions using the same primers that were used for the PCR amplification.

Table 3.9 PCR Thermal cycling conditions.

Steps	Temperature	time	Number of cycles
Initial denaturation	95°C	2 min	1
Denaturation	95°C	1 min	30 (Repeated)
Annealing	44°C - 68°C **	1 min per 1 kB	
Extension	72°C	1 min	
Final extension	72°C	5 min	1

** Based on the annealing temperature of the primers.

CHAPTER 4: ANALYSIS OF TWO UNRELATED PROBANDS IDENTIFIED AS CARRYING *SH2B1* MUTATIONS

4.1 Introduction

This Chapter describes the confirmation and subsequent family segregation analysis of participants who were identified to have a deleterious variant in a known obesity gene *(SH2B1),* through whole exome sequencing (WES) analysis, and for which biological samples were available from their parents.

Obesity is a heterogeneous disorder. Many gene variants have been shown to be associated with or to cause obesity. Recently, many approaches and techniques have been developed to expand our understanding of the genetic architecture of the diseases by facilitating the detection of highly penetrant variants and genes that are associated or linked with common or rare forms of the disease. Whole exome sequencing (WES) is one of the most efficient tools for examination of extreme forms of common diseases such as morbid obesity (which may represent rare Mendelian diseases hidden among the more common types of obesity). The development of this next-generation sequencing (NGS) technology has accelerated the process of sequencing, and can now be carried out at relatively low cost.

Nevertheless, the older Sanger sequencing method remains important and useful, especially for validating NGS results in probands and their family members, and thus avoiding reporting false positive results, and confirming the variant status in other family members.

Here we describe a follow-up analysis of two probands who were identified with novel variants in known and candidate obesity genes through WES carried out by a previous PhD student (Dr Sanne Alsters). The two cases were recruited from different sites. Case 1 was recruited with his

parents at the Imperial College London Diabetes Centre in Abu Dhabi, while case 2 was recruited with his mother at the Imperial College Healthcare NHS Trust (Hammersmith Hospital) under the Obesity Plus family cohort led by Dr Tony Goldstone (Imperial). Sanger sequencing analysis was performed both to confirm the presence of the novel variants in the two probands with extreme obesity and to determine the status of these variants in some of the family members with available biological samples.

4.2 Aim of the study

- To confirm the presence of the novel putatively deleterious *SH2B1* variants in each of two unrelated probands who previously underwent WES, using Sanger sequencing
- To re-examine the pre-existing WES data using a wider candidate genes list.
- To perform a family segregation analysis in each of the two families.4.3 Results

4.3 Results

4.3.1 Case 1

4.3.1.1 Case 1 participants' characteristics

Case 1 was a 22 year old male proband of south Asian ethnicity, recruited from Imperial College London Diabetes Centre, Abu Dhabi. He had a history of progressive weight gain starting in his mid-teens. He had problems with his school performance at around 15 years of age. There was no family history of note and the parents were not consanguineous. On examination at age 10 years, he was noted to have generalised obesity (BMI 38.6 kg/m2) and prominent acanthosis nigricans. He had normal blood pressure and no Cushingoid features. Biochemical and hormonal investigations were normal, apart from profound hyperinsulinaemia. The patient was commenced on metformin, without response in terms of his weight. He was also seen by psychiatrists and was treated with various drugs, including antidepressants.

4.3.1.2 WES and segregation analysis of case 1

Blood samples were collected from the patient and his parents. WES was performed for the proband, followed by Sanger sequencing to confirm the presence of the *SH2B1* variant and to check the segregation of the identified variant within the family. The overall quality of the WES, which was assessed by FastQC version 0.10.0, was high.

The proband had previously been predicted to carry a deleterious mutation in the sarcoma (Src) homology 2 B adaptor protein 1 (*SH2B1*) gene (C539T: p. S180F), with a CADD score of 23.6 (Table 4.1). The identified variant is located close to a variant (A175N) previously reported in an individual with severe early-onset obesity [3] which was shown to disrupt the function of NGF-induced neuronal differentiation. The presence of the variant (in heterozygous state) was confirmed in the proband. Mutations in this gene act in a dominant manner, so this was initially presumed to be the cause of the phenotype observed.

However, segregation analysis revealed that the same missense mutation was also present in heterozygous status in his father, who was not obese and had no relevant medical problems. The mother did not carry the deleterious variant (Figure 4.1).

The presence of the mutation in the unaffected father raised questions of penetrance. This led to a re-evaluation of the WES data of the proband with an updated list of obesity and obesity/developmental disorder genes (methodology chapter (3)/section 3.2/ Appendix 1). As described in the previous Chapter, this is an expanded list of genes compared to the list used in the initial analysis. The revaluation was performed to identify further candidate variants that might contribute to the phenotype, or affect penetrance.

This analysis revealed two further potentially deleterious variants in the proband, which had previously been overlooked: A3539T/p. D1180V in *MBD5* and C3022T/ p. L1008F in *POGZ*, which had CADD scores of 22.9 and 24.2 respectively (CADD score of 20 indicates the variant as 1% most deleterious substitutions in the human genome [165]), Figure 1.B-C. All these mutations segregating in this family were rare in the gnomAD database as shown in Table 4.1. Mutations in *POGZ* gene cause White-Sutton syndrome (which is an autosomal dominant condition). This has

a highly variable phenotype including developmental delay, language and speech delay, motor delay, microcephaly, non-specific vision problems and obesity [169-171]. *POGZ* encodes a zinc finger protein believed to have an important role in mitotic progression and possibly in neuronal proliferation [172, 173]. *POGZ* is particularly highly expressed in the pituitary gland, and the cerebellum, compared to lower constitutive expression in most tissues [169].

The third predicted-deleterious variant of interest in this proband was in *MBD5*, which encodes a member of the MBD family (which includes *MECP2* gene, a causal gene for Rett syndrome) [174]. The encoded protein is highly expressed in brain, oocytes and testis, and is thought to have a role in heterochromatin and epigenetic reprogramming [174]. The *MBD5* gene was defined as the causative locus for the 2q23.1 microdeletion syndrome, *MBD5*-associated neurodevelopmental disorder (MAND) (which is an autosomal dominant condition). [175, 176]. The clinical features of this syndrome include language impairment, microcephaly, development delay, motor delay, short stature, hyperphagia, behavioural problems, craniofacial abnormalities, sleep disturbance, epilepsy, and seizures [175-186]. However, the severity and complexity of the phenotypic spectrum of *MBD5* gene disruption differs from that of the high-penetrance 2q23.1 microdeletion [175, 177], and incomplete penetrance for *MBD5* mutations has been identified in several previous reports [182-185]. Missense variants in *MBD5* (which may be more likely to show reduced penetrance) are associated with risk of autism spectrum disorders [175] and may also contribute to schizophrenia and depression [186]. It is reported that 79% of cases with *MBD5* disruption showed hyperphagia and 26% had obesity [177, 179]. In addition, in a recent study of individuals with early-onset obesity, three copy number variants affecting this gene were

identified [187]. To further explore the network interaction among the three genes, an online tool called STRING has been used (Figure 4.2) (https://string-db.org/).

Table 4.1: Details of the putatively deleterious variants identified in proband 1

GENE	MODE OF INHERITANCE	VARIANT TYPE	VARIANT DETAILS	RS	MAF (gnomAD)	CADD SCORE
SH2B1	AD	Missense	NM_001145795:exon2:c.C539T:p.S180F	rs144107554	0.00002131	23.6
POGZ	AD	Missense	NM_001194937:exon19:c.C3280T:p.L1094F	rs768284272	0.0001591	24.2
MBD5	AD	Missense	NM_018328:exon12:c.A3539T:p.D1180V	rs752035001	0.00007163	22.9

Abbreviations are as follows: **MOI**. Disease Mode of Inheritance, **MAF(%)** gnomAD MAF, **RS**; Reference SNP, **CADD** (Combined Annotation Dependent Depletion); a tool for scoring and predicting the deleteriousness of variants in human genome, score of 20 indicates the variant as 1% most deleterious substitutions in the human genome [165].

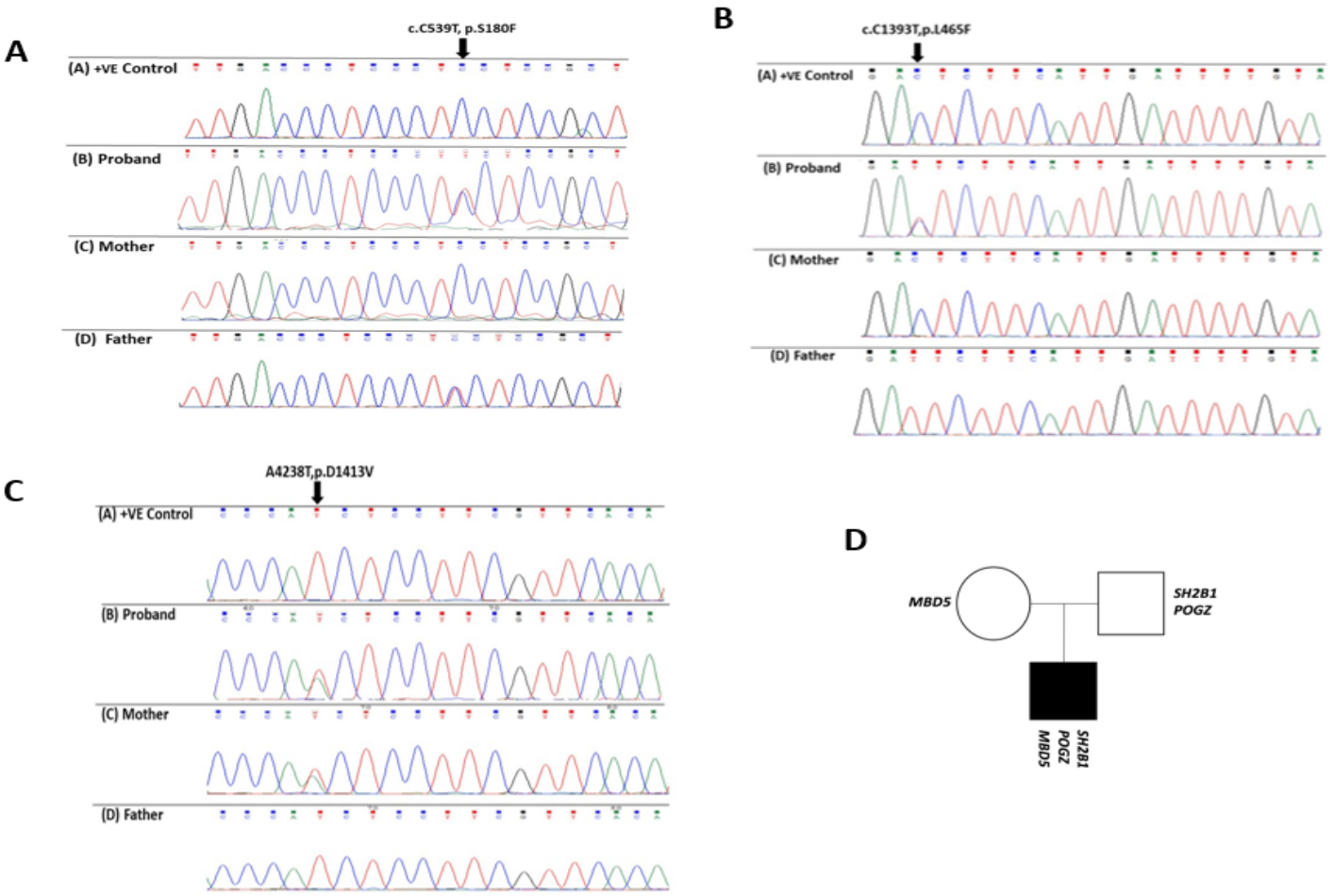


Figure 4.1 Sanger sequence electropherograms of case 1 and his parents. (A) Chromatogram of Sanger sequencing results demonstrating heterozygosity for SH2B1:c.C539T in the proband and his healthy father (B) Chromatogram of Sanger sequencing results, demonstrating heterozygosity for POGZ:c.C1393T in the proband and his healthy father. (C) Chromatogram of Sanger sequencing demonstrating heterozygosity for MBD5:c.A4238T in the proband and his healthy mother. (D) Pedigree of the family and the putatively deleterious variants in each individual.

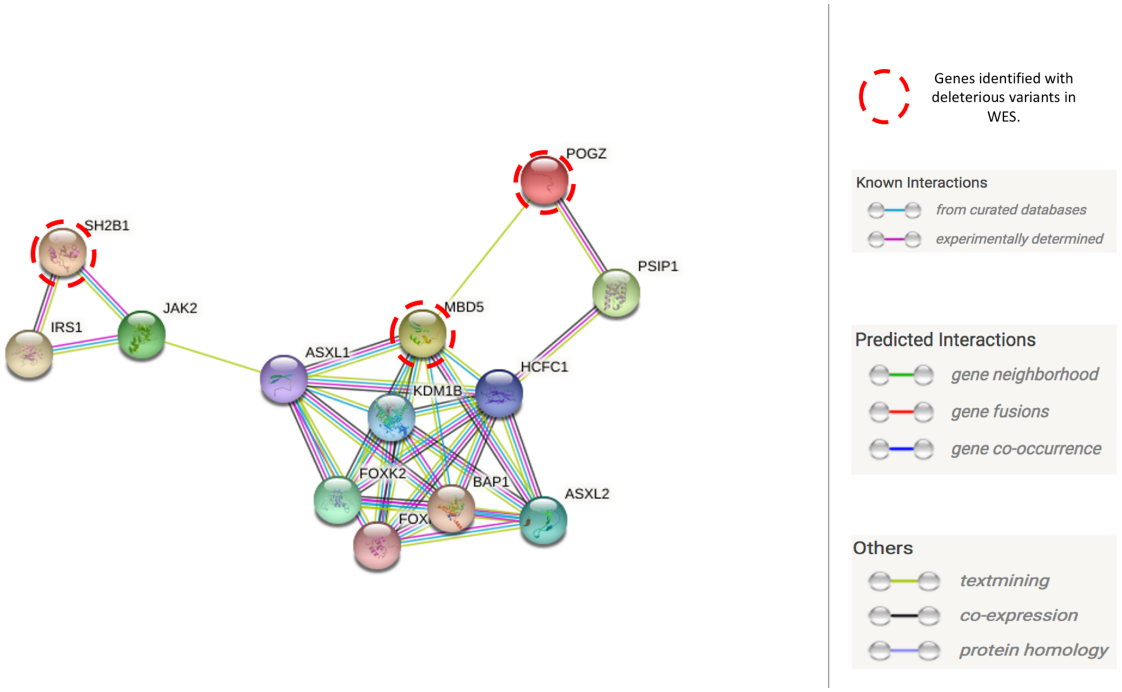


Figure 4.2 Interaction network among genes identified with potential rare deleterious variants in the first proband (predicted by STRING). Coloured nodes represent the different proteins/genes, dashed red circle represents the identified genes with potential rare deleterious variants. Different colour line represents the direct and indirect interactions between proteins where each colour indicates the type of evidence available for that interaction. (www.string-db.otg)

4.3.2 Case 2

4.3.2.1 Case 2: participants' characteristics

This case is the proband recruited by Dr Tony Goldstone (Imperial). On examination the proband was a 19 year old Male. His BMI was 35.9 (with historical maximum BMI of 47.6). The proband was clinically characterised with development delay, learning disability, delayed puberty, neonatal hypotonia, compulsive hoarding, behavioural rigidity, childhood-onset obesity with hyperphagia and mild insulin resistance.

4.3.2.2. WES and Familial Segregation Analysis of case 2

Previous WES analysis (by Dr Sanne Alsters) had predicted the presence of a deleterious variant, c.A1307G/Q436R, located in exon 1 of SH2B1 gene. The part of the protein encoded by this part of the sequence is involved in the JAK2 signalling pathway. The variant is a missense mutation that results in a change from the polar amino acid Glutamine to a positively charged Arginine. It has a CADD score of 25, so is predicted to be deleterious, and was presumed to be responsible for the observed phenotype in the proband. This variant was found in the gnomAD data set in a total of 13/281354 individuals, mostly European (non-Finnish), with minor allele frequency lower than 0.0007 in all ethnic populations.

Sanger sequencing confirmed the presence of the mutation (in heterozygous state) in the proband, and non-carrier status in the mother, whose sequence matches the wild type Figure

4.3. Since the father's DNA is not yet available, it is not possible to confirm whether or not the mutation is *de novo*. Samples from further family members are awaited to complete a segregation analysis to determine whether the mutation co-segregates with obesity in this family.

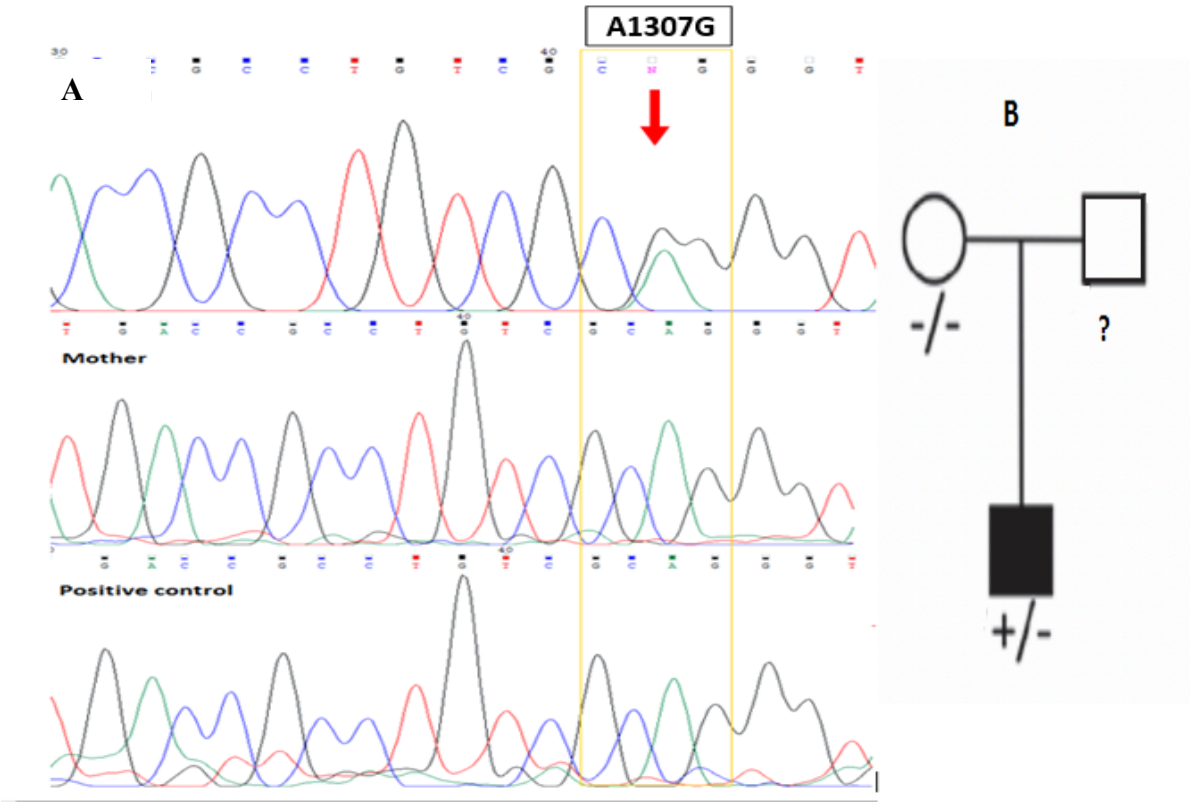


Figure 4.3 Sanger sequence electropherograms of case 2 and his mother.. Sanger sequence chromatograms show the heterozygous status of A1307G:p.Q436R in the proband, and non-carrier status in the mother, whose sequence matches the wild type. The arrow shows the mutation. [B] In the pedigree,+/- represents the heterozygous mutation in the proband, -/- represents the wild type status of the mother and (?) represents the undiagnosed status of the father. 'Males' and 'females' are indicated by squares and circles, respectively.

4.4 Discussion

Whole exome sequencing is a powerful tool for identifying inherited and *de novo* variants. In this Chapter I have described a follow-up study that uses Sanger sequencing analysis to confirm the presence of the novel rare deleterious variants in probands identified previously and assumed to be responsible for the observed phenotype acting in a monogenic fashion. The Sanger sequencing analysis did confirmed the identified variants in the probands and the parental status. However, the re-analysis of WES data led us to the detection of another possible mode of inheritance, called oligogenic, where additional mutations in different genes could also contribute to the phenotype/influencing penetrance.

The two probands described here each have a deleterious variant in the *SH2B1* gene, Q436R and S180F, respectively. The *SH2B1* protein product of this gene is a member of the Src homology 2B (SH2B) family of adaptor proteins, which contain a common domain structure N-terminal dimerization (DD), central pleckstrin homology (PH) and an Src homology (SH2) domain at the carboxy (C)-terminal. The gene encodes, via mRNA alternative splicing at the 3'end, four isoforms (α, β, γ, and δ). The four isoforms are identical except they differ in their C-termini. Although *SH2B1* is expressed in numerous tissues including brain, adipose, muscle and liver issues, each isoform has a different pattern of expression; SH2B1α and SH2B1δ are restricted to the brain and SH2B1β and SH2B1γ are widely expressed.

SH2B1 is involved in regulating several different signalling pathways mediated by cytoplasmic tyrosine kinase Janus Kinase (JAK), and tyrosine kinase receptors such as receptors for insulin (INS) and insulin-like growth factor-1 (IGF1), platelet-derived growth factor (PDGF), fibroblast growth factors (FGF) , nerve growth factor, brain-derived neurotrophic factor, and glial cell-derived neurotrophic factor. Evidence from several studies has implicated SH2B1 in glucose homeostasis, energy regulation, cell proliferation and motility, as well as neuronal differentiation. Animal studies have revealed that genetic deletion of the SH2B1 gene in mice results in leptin and insulin resistance, which leads to the development of severe obesity hyperphagia, type 2 diabetes and infertility [188-190]. These observations demonstrate the important roles of *SH2B1* in regulation of glucose metabolism, body weight and reproduction. Despite the key roles of *SH2B1* in the signalling pathways medicated by cytoplasmic tyrosine kinase JAK and tyrosine kinase receptors, the contribution and molecular mechanism of SH2B1 in energy and glucose homeostasis remain challenging.

Several studies have been conducted to identify mutations in SH2B1. Point mutations within the *SH2B1* gene [191, 192], and chromosomal deletions [193] including the *SH2B1* gene, are known to cause human obesity (with an autosomal dominant mode of inheritance). Doche, *et al.* 2012 screened a large cohort of patients with severe early-onset obesity and identified four rare variants in SH2B1 [192]. Subsequently, the same group reported the identification of an additional three variants and one previously reported variant in *SH2B1* through sequencing 500 individuals. Evidence from the functional studies showed that these identified variants impaired the cellular function of the *SH2B1*. This includes disrupting the ability of *SH2B1* to enhance nerve

growth factor (NGF) induced neuronal differentiation and impairing nuclear accumulation. Furthermore variants that reside in the 1–631 region of SH2B1 (SH2B1 β isoform) also alter the cellular function by reducing GH-induced macrophage motility [191, 192].

GWAS studies in populations of European ancestry have identified single nucleotide polymorphisms (SNPs) that are located near the SH2B1 gene, such as rs7359397 [191]. Interestingly, this SNP has a high linkage disequilibrium (LD) with another coding SNP in SH2B1 that is associated with obesity, identified as rs7498665. The association of those two common SH2B1 SNPs with increased BMI has been replicated in many studies, including of Swedish, German and Belgian populations [194-196]

In further analysis, Proband 1 was identified to have predicted deleterious variants in two other obesity-relevant genes: *POGZ* and *MBD5,* besides that in *SH2B1* which led to the work described in this Chapter. One of these variants is present in the healthy mother and two in the healthy father, which suggests that an oligogenic inheritance model might be responsible for the phenotype of the proband.

This has been instructive since the proband was initially diagnosed as having an *SH2B1* mutation causing monogenic obesity, with the other mutations only revealed when pedigree analysis gave an unexpected inheritance pattern resulting in reanalysis of the WES data to include genes causing syndromic obesity. Since individuals with variants in *SH2B1* exhibit a wide range of phenotypes, including obesity, insulin resistance and neurodevelopmental problems, while

variants and genomic alterations of *POGZ* and *MBD5* genes result in a diverse phenotypic spectrum of neurodevelopmental disorders, intellectual disability and obesity, we suggest that these three mutations may be acting in concert to produce the observed phenotype in our proband. Consistent with this hypothesis, Pearce, *et al.* also report that *SH2B1* mutations show incomplete penetrance and may require other genetic and environmental factors to manifest the disease [192].

We suggest that a proportion of severely-affected individuals may, in fact, have an oligogenic cause of their obesity. As a result of this experience, we urge caution in interpretation of sequencing results from individual candidate genes which, in this case, could have given rise to inaccurate genetic counselling. WES or whole genome sequencing (WGS) allows exploration of a wider range of genes, and allows reanalysis as new causative genes are discovered, and where resources allow, should be the approach of choice.

CHAPTER 5: RE-ANALYSIS OF WHOLE EXOME SEQUENCING DATA FROM PMMO PARTICIPANTS

5.1 Introduction

This Chapter describes the re-analysis of whole exome sequencing (WES) data from a total of 91 participants in the PMMO cohort. The aim was to further explore undiagnosed cases using an expanded list genes relevant to monogenic obesity and syndromic obesity and mouse models of obesity.

The development of the next generation of sequence technologies, such as whole exome sequencing, which covers around 95% of the exons and regulatory regions of known genes (around 5% of the whole genome), has revolutionised our understanding of, and ability to, diagnose Mendelian diseases in a timely manner.

The usefulness of this technology has been demonstrated by the accelerated number of reports of novel disease genes and variants. The continuous identification and characterisation of rare genetic disorders is a vital component in enhancing patient care. It can provide an accurate diagnosis to ensure the correct therapeutic approach and to undertake genetic counselling to the patient's family if required.

Despite the high utility of WES, there are several challenges in providing clinically meaningful sequencing results. One major issue is the interpretation of the functional implications of novel variants, which is particularly difficult when the identified variant does not have sufficient supportive evidence. The second challenge is the incomplete coverage of some regions by WES, which can be overcome by performing whole genome sequencing (WGS), which tends to give

more even coverage across the genome, or using long-read technologies that provide better resolution of repetitive regions.

As our understanding of pathogenic processes expands, it has become important to re-analyse previously generated WES data. In this context, a re-evaluation of previously generated WES data from our PMMO cohort was performed. The initial pilot study sequencing was performed in 2015 on a total of 40 white European individuals (out of 91 individuals) from the PMMO cohort selected for adult BMI>50, and family history of obesity. This analysis was based on 36 genes known to be causative of monogenic obesity at that time. A total of eight plausibly causative variants were identified at that time, in seven individuals (17.5% of the 40 samples).

The re-analysis described here differs in several ways from the initial analysis:

1) Examination of a larger number of genes relevant to monogenic obesity and syndromic obesity/ genes (methodology chapter (3)/section 3.2/ Appendix 1)

2) Application of an improved filtering pipeline for variant prioritisation, as described in of the genes (methodology chapter (3)/section 3.3

3) Assessment of variant quality by applying generic hard-filtering recommendations [197].

5.2 Aim of the study

1) To re-analyse the WES data of a selected subset of PMMO participants, including an updated list of genes that includes monogenic obesity genes, syndromic obesity genes and mouse model genes.

2) Investigate the prevalence and distribution of putatively causative mutations in this sub-set of patients.

5.3 Results:

5.3.1 Characteristics the participants

Full information about the PMMO cohort is given in Chapter 2. The overall characteristics of the 91 participants in the pilot WES study are shown in Table 5.1

The average age at recruitment was 56 years old (ranging between age of 24-69.), and 55% (n=50) of the participants are female. Nearly 87% of the participants are of white ethnicity and 31% had onset of obesity before the age of 10.

From the 91 individuals, initially two types of participants were selected for the analysis: 1) those in whom no putatively causative variant(s) was identified in the previous analysis; and (2) those not included in the previous analysis due to non-white ethnicity. In the light of our finding of potential oligogenic inheritance in Chapter four, all participants were then investigated to determine the distribution of variants in the PMMO cohort.

Table 5.1: General characteristics of the PMMO participants who were included in the re-analysis of WES data.

	PMMO SUBJECT INCLUDED FOR WES RE-ANALYSIS
N	91
AGE (YEARS) (MEAN ± SD)	57 ± 12
FEMALE	50/91
BMI KG/M2 (MEAN ± SD)	55 ± 11.3
WEIGHT KG (MEAN ± SD)	M=158.2 SD=36.1
HEIGHT (MEAN ± SD)	1.7 ± 0.10
ONSET OF OBESITY BEFORE THE AGE OF 10	24/91
TYPE 2 DIABETES (T2D)	23/91

5.3.2 Exome data re-analysis

The strategy for identification of putatively causative variants is described in the methodology Chapter; **Chapter 3/Section 3.3**. Briefly, the analysis was performed manually, focusing on variants in the coding and splicing regions of the selected genes. Variants within those genes were filtered and prioritised based on standard filtration and evaluation steps for WES data analysis. These included absence or minor allele frequency in public database (MAF<1), risk predication, and matching the relevant (known) mode of inheritance for each gene based on literatures and previous studies [198, 199].

5.3.3 Overall view of the findings from the re-analysis

Overall from the re-analysis, a total of 87 rare deleterious variants were identified in a total of 39 obesity and syndromic obesity genes. Thirty of those identified deleterious variants matched the expected mode of inheritance, as shown in **Table 5.2**.

Of these, 21 variants were detected from the current re-analysis and, 9 variants were detected in the previous analysis along with two copy number variants: chr9:87570170-87570529 and chr16:29675061-30215702. Thus, this has increased the number of individuals in the cohort identified with potentially causative obesity variants from 11 to 32 subjects.

Furthermore, from the mouse model genes, a total of 10 rare deleterious variants matching the mouse model inheritance pattern were identified in 11 individuals. Below, the Monogenic obesity human gene variants will be presented first, followed by the mouse model genes.

Table 5.2 Overall summary of the identified rare deleterious variants in monogenic obesity and syndromic obesity genes from WES.

Gene	*Mode of inheritance*	*Total Rare deleterious variant*	*Considered Variants*
AFF4	AD	3	3
ALMS1*	AR	12	0
BBS1*	AR	3	0
BBS10*	AR	2	0
BBS12*	AR	1	0
BBS2*	AR	1	0
BBS4*	AR	5	0
BBS5*	AR	2	0
BBS9*	AR	3	0
CARTPT	AD	1	0
CEP290*	AR	2	0
COA3	Compound Heterozygous	1	0
CPE*	AR	1	0
IGSF1‡	AR	2	2
KSR2*	AD	1	0
LEPR*	AR	2	2
LRP2	Compound Heterozygous	5	0
MAGEL2 ‡	AD	3	3
MANF	AR	1	0
MC4R*	AD	1	0
MCHR1	AD	2	1
MKKS*	AR	1	0
MRAP2	AD	1	0
MYT1L	AD	1	1
NTRK2*	AD	1	1
POGZ	AD	4	3
PTEN*	AD	1	1
RAI1	AR/AD	5	5
RBMX	XLR	2	2
SETD2	AD	5	1
SH2B1*	AD	3	3
SIM1	AD	1	1
SLC35D3	AD	1	0
TTC8*	AR	2	0
TUB*	AR	2	0
UCP1	AD/AR	1	0
UCP3	AD/AR	2	1
WDPCP*	AR	1	0
WNT10B	AD	1	0

Abbreviations are as follows: **, Variants within those genes identified in the initial analysis of WES data. **‡,** Variants within those genes identified in the initial analysis of WES data with an additional variant identified in the re-analysis. **Considered variants:** Rare deleterious variants match the expected mode of inheritance.

5.3.4 Classification of monogenic and oligogenic forms of obesity

The individuals with considered rare deleterious variants in obesity monogenic genes were categorised into two classes: monogenic and oligogenic. The monogenic class as clarified in the earlier Chapter represents the typical Mendelian mode of inheritance (dominant or recessive) that includes mutations in just one of the relevant genes. The oligogenic class represents cases with more than one deleterious variant in more than one gene. A total of 17 variants under the monogenic form were detected in 18 individuals, summarized in Table 5.3. Each of those variants and their related genes will be described further below. A total of three cases have been identified with oligogenic mode of inheritance, summarised in Table 5.4.

5.3.4.1 Monogenic forms of obesity.

Of the newly identified deleterious variants (Table 5.2), four of them are from obesity related genes: Melanin Concentrating Hormone Receptor 1 (MCHR1), Mitochondrial uncoupling proteins 3 (UCP3) and Immunoglobulin Superfamily Member 1 (IGSF1), while the other 15 variants were found in four of the syndromic obesity genes; pogo transposable element derived with ZNF domain (POGZ), AF4/FMR2 Family Member 4 (AFF4), RNA Binding Motif Protein X-Linked (RBMX) ,RAI1 (Retinoic Acid Induced 1) and MAGE Family Member L2 (MAGLE2). All identified variants are heterozygous and match the inheritance patterns, except for IGFS1 which follows the X-linked recessive mode of inheritance. Variants in each gene group are described in further detail below.

Table 5.3 Summary of cases in which mutations were detected in only one of the relevant genes.

ID	GENE	MOI	RS ID/ POSITION	VARIANT TYPE	HGVS	MAF (%)	CADD	GENOTYPE	PUBLICATIONS
S1	MCHR1	AD	rs45439291	Missense	p.R317Q	0.003407	24.5	Het	N/A
S2	UCP3	AD	rs142952570	Missense	p.R95H	0.0007876	26.9	Het	N/A
S3	IGSF1	XLR	rs146462069	Missense	p.N599T	0.009383	22.3	Hem	N/A
S4	MYT1L	AD	rs201765281	Missense	p.V3M	0.0001	22.4	Het	N/A
S5	POGZ	AD	N/A	Missense	p.G1334V	N/A	24.9	Het	N/A
S6	POGZ	AD	N/A	Missense	p.M167I	N/A	25	Het	N/A
S7	RAI1	AD	rs374187267	Missense	p.A1230E	1.676E-05	23	Het	N/A
S8	RAI1	AD	rs149716029	nonframeshift deletion	p.1259_1260del	0.002619	N/A	Het	N/A
S9	RAI1	AD	rs149716029	nonframeshift deletion	p.1259_1260del	0.002619	N/A	Het	N/A
S10	RAI1	AD	rs149716029	nonframeshift deletion	p.1259_1260del	0.002619	N/A	Het	N/A
S11	RAI1	AD	rs142981643	Missense	p.A1679V	0.0005855	23.5	Het	N/A
S12	AFF4	AD	rs139490054	Missense	p.Thr1107Ala	0.0022	22	Het	N/A
S13	AFF4	AD	N/A	Missense	p.S1079N	N/A	28.1	Het	N/A
S14	AFF4	AD	rs770827508	Missense	p.R209H	1.626E-05	22.6	Het	N/A
S15	RBMX	XLR		Fameshift		0.0042	N/A	Het	N/A
S16	RBMX	XLR	rs767553768	Non-nframeshift deletion	p.187_187del	0.0052	N/A	Het	N/A
S17	AFF4	AD	rs34527550	Missense	p.T136P	0.0000318543	17.60	Het	N/A
S18	MAGEL2*	AD	rs111759069	Missense	p.Ala360Glu	0.006048	19.04	Het	N/A
S18	MAGEL2*	AD		nonframeshift insertion	N/A	N/A	N/A	Het	N/A

Abbreviations are as follows: ID. participants research ID, **MOI**. Disease Mode of Inheritance, **HGVS**. Protein position HGVS, **MAF(%)** gnomAD MAF, **Genotype**. Subject Genotype (**Het**: Heterozygous, **Hom**: Homozygous, **Hemi**; hemi *; Further analysis is required to determine mutated allele is inherited from the mother or the father as the MAGEL2 gene is maternally imprinted, paternally expressed.

5.3.4.2 Oligogenic form of obesity

In addition to the monogenic causes of obesity, a total of three cases have been identified to carry more than one deleterious variant in different candidate genes of monogenic obesity and syndromic obesity. These cases are believed to follow the oligogenic inheritance model as more than one variant might be involved in the development of the phonotype.

The identified oligogenic events consist of the following combinations: *SH2B1*/*RAI1*, *MAGEL2*/*RAI1* and, SETD2/POGZ as summarised in Table 5.4. All variants are missense variants in heterozygous status that match the mode of inheritance and have a CADD score greater than 18. These variants affect evolutionarily conserved residues and none of the variants have been reported previously. Nevertheless, the pathogenic impact remains unknown.

Table 5.4 Summary of cases with oligogenic mode of inheritance.

Case	*Gender*	*Gene*	*MOI*	*Rs ID/ Pos*	*Type of variant*	*HGVS*	*CADD*	*GnomAD*	*Zygosity*	*Publications*
S28	F	SH2B1	AD	rs772678200	Missense	c.G1633A:p.G545S,	29	0.0000244	Het	*N/A*
		RAI1	AD	rs113208290	Missense	c.C1142T:p.A381V	24	0.003643	Het	*N/A*
S30	***F***	MAGEL2	AD	rs111759069	Missense	c.C1079T:p.A360V	23.1	0.0052	Het	*N?A*
		RAI1	AD	rs147844401	Missense	c.G5653A:p.D1885N	27.5	0.00008319	Het	*PubMed 21857958*
S31	***F***	POGZ	AD	1:151381292	Missense	c.G6703C:p.V2235L	20.9	-	Het	*N/A*
		STED2	AD	rs780019200	Missense	c.T1654C:p.Y552H	24.6	0.00001769	Het	*N/A*

Abbreviations are as follows: **ID**. participants research ID, **MOI.** Disease Mode of Inheritance, **HGVS**. Protein position HGVS, **MAF(**%) gnomAD MAF, **Zygosity**. Subject Genotype (Het: Heterozygous, Hom: Homozygous, Hemi; hemizygous

5.3.5 Description of the new identified variants in human obesity genes.

MCHR1

One patient was identified with a mutation in the MCHR1 gene, NM_005297.3:c.950G>A, p.Arg317Gln. The change at the amino acid level is from Arginine (R) basic to Glutamine (Q)/polar. The variant has been reported previously, as described below, and has a frequency of 0.003407 in gnomAD (equivalent of allele count of 963/282680 gnomAD controls). The amino acid position of the variant is highly conserved across species and located at the transmembrane domain receptor (rhodopsin family).

The *MCHR1* gene encodes an integral plasma membrane protein, which is a member of one of the largest protein families, the G protein-couple receptor (GPCR) family 1 [200]. The *MCHR1* consists of 353 amino acids and is highly homologous (more than 95% identity) as well as having similar tissues distribution across all vertebrates studied [201]. The MCHR1 has a high affinity for the natural ligand, around ∼1 nM, and upon activation couples to the Gi, Go, and Gq G proteins. The activation of the MCHR1 leads to an increase in the intracellular calcium accumulation. However, the MCHR1 is distinguishable from the other GPCR proteins, which are regulated by their own ligand, as the deficiency of the MCH does not change the expression of the MCHR-1.

MCHR1 knockout mice have been studied by different groups [3-5]. It was observed that Mchr1 knock results in reduced susceptibility to diet-induced obesity, leanness, low fat mass, hyperphagia and hyperactivity, the hyperphagia and hyperactivity manifesting on both HFD or regular chow. Thus, it was suggested that the observed lean phenotype is attributed to the

elevated metabolic rate and hyperactivity [202]. This phenotype is similar to that observed in mice lacking the MCHR-1 ligand MCH. A phenotypic difference due to gender was also observed in the mchr1 knockout mice, with male mice being significantly more hyperphagic than female mice. The differences observed in weight gain and body weight in response to MCHR-1 knockout differs with the background strain used, , related to specific metabolic and behavioural features. In addition, Although the phenotype of overexpression of MCHR1 has not been reported, the overexpression of the ligand MCH results in obesity. Based on the identified role of the MCH-MCHR1 system in the regulation of feeding , energy balance, and emotional processes in the mouse, the system has been highlighted as a potential target for treating obesity. However, the lean phenotype has not been always replicated in the MCH and MCHR1-deficient mice [203, 204].

On the other hand, two up-to-date human genetic studies have been performed to identify genetic variants of *MCHR1* and possible links to human obesity [205, 206]. The initial study by Gibson et al. (2004) was based on 106 obese subjects with early onset obesity and 192 normal-weight individuals as controls, and led to the identification of two variants, Y181H and R248Q. Both variants are predicated to be deleterious and were only found in obese subjects. The R248Q variant was found to co-segregate with obesity while family samples for the other variant were not available. Functional studies were done only for R248Q and it showed no functional difference [207]. The same study identified two common SNPs in linkage disequilibrium, but no association was found with obesity-related phenotypes. Another genetics study by Wermter *et al*. in 2005 based on two cohorts of German children and adolescents with extreme obesity

identified 11 rare variants and two SNPs in the *MCHR1* coding region. However, these findings weren't replicated in five independent cohorts [206].

On the other hands, a functional analysis of 11 previously reported *MCHR1* variants found effects on receptor function ranging from failing to respond to the ligand to decreased and increased activity, depending on variant [207].

The identified variant in our re-analysis is one of the 11 reported rare variants in Wermter et al. 2005 study which was identified in a 15-year-old male with a BMI of 43.24 kg/ m^2 who inherited the variants from his extremely obese father (age 42.83 years; BMI 44.66 kg/m^2) [206].

UCP3

A heterozygous missense variant, c.G284A p.R95H, was identified in the *UCP3* gene, in a 60-year-old male individual. . The variant p.R95H consists of a substitution from an arginine, a basic large-sized amino acid, to a Histidine, a polar medium-sized amino acid, in exon 3 of the *UCP3* gene, which possibly might alter the structure and function. The arginine in this position is evolutionarily conserved across species. The variant has a CADD score of 26.9 and has a frequency of 0.0007876 in gnomAD (equivalent of an allele count of 222/281880 gnomAD controls). The predicted effect on the protein structure is visualised in Figure 5.1.

The *UCP3* gene is mapped to a chromosomal region that is related to obesity and hyperinsulinemia, 11q13 [208]. The gene encodes an inner mitochondrial membrane protein, a member of the mitochondrial anion carrier proteins family (MACP). The *UCP3* gene consists of seven exons, where six encode the transmembrane regions. In humans and rodents, this protein is mostly expressed in skeletal muscle and is also found at a lower levels in brown adipose tissue and heart tissue. Its activity is regulated by many cellular metabolic and hormonal signals [209-211]. The protein has been shown to be involved in several roles as summarised in Figure 5.2, primarily fatty acid metabolism, through protecting the mitochondrial and the cell against the oxidative damage that is induced by reactive oxygen species (ROS) through reducing the production of ROS by mitochondria or storing the energy as fat [212, 213], exporting the LOOH from the mitochondria matrix, also involved in mild/partial mitochondrial uncoupling in the presence of FFA or reactive oxygen species (ROS) [214, 215].

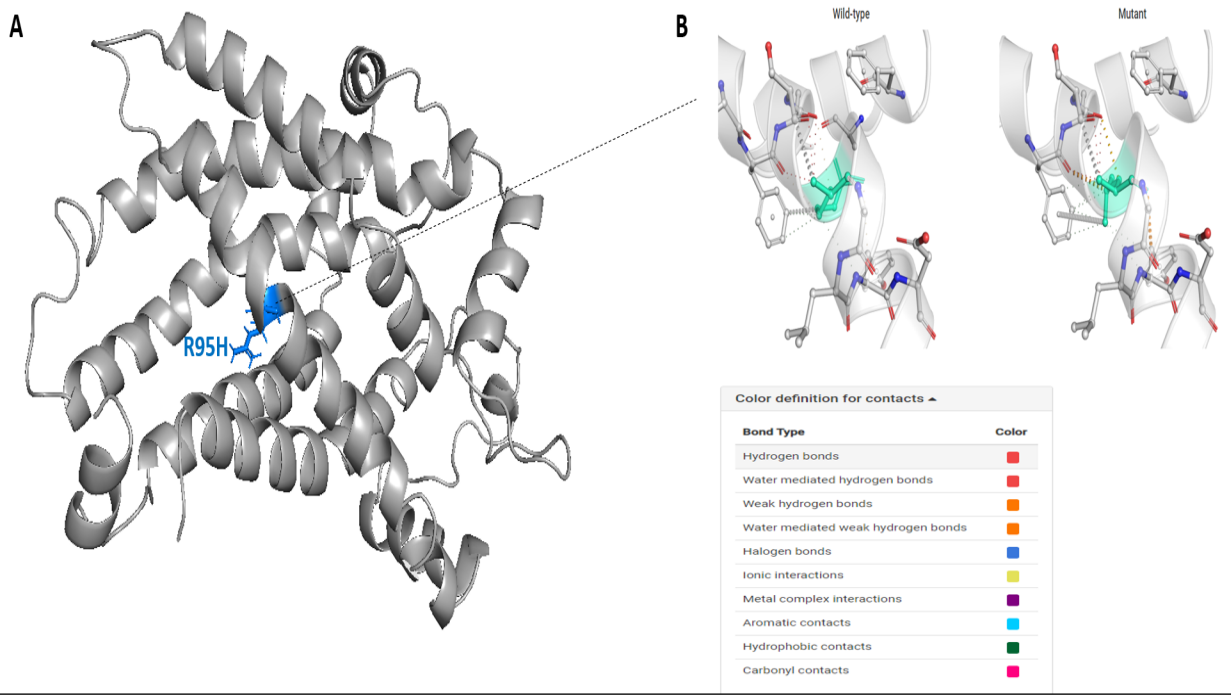


Figure 5.1: Protein structure of *UCP3*. A) A predicted three-dimensional structure of UCP3 (BD Bank file). The position corresponding to the variant is shown in blue. (B) Mutation effect prediction of the interatomic interaction of the wild type and variant residue are presented as sticks and highlighted in light green, generated by DynaMut. The colour definitions of the different types of interaction are shown in the Table.

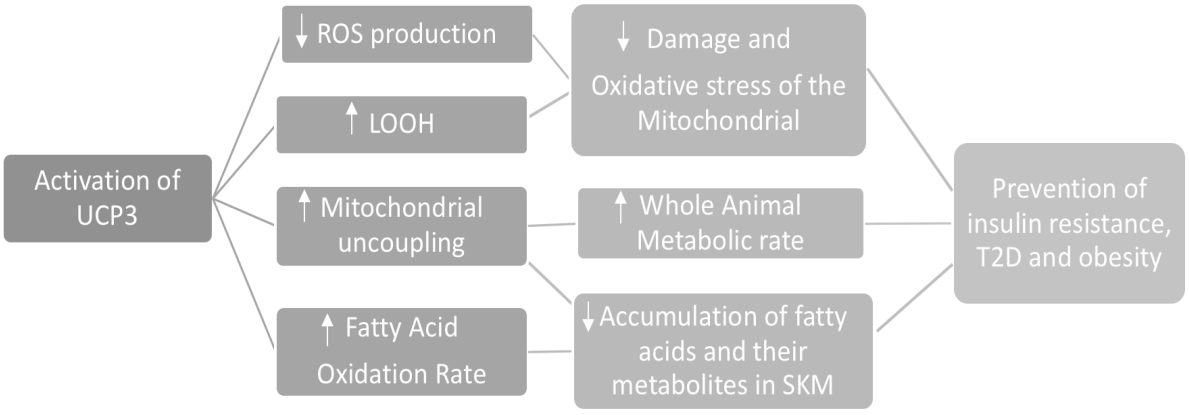


Figure 5.2. Schematic representation of the relation between the reactions influenced by *UCP3* and prevention and treatment of insulin resistance, and type 2 diabetes (T2DM) or obesity. The image is based on the illustration shown in Busiello *et al.* 2015.

Nevertheless, although the complete physiological role of the *UCP3* is not yet known, the evidence from the functional effects of *UCP3* (mentioned above) suggests its protective role against obesity, diabetes and cardiovascular dysfunction [214, 216, 217].

In addition, three different studies on morbidly obese individuals have identified a total of six rare variants, summarised in Table 5.5. Also, polymorphisms of the *UCP3* gene have been identified based on population studies and a European meta-analysis for associations with diabetes, obesity and fat metabolism [218-221].

Table 5.5: Previously reported variants in *UCP3.*

DPSNP ID	VARIANT TYPE	HGMD CLASSIFICATION	GNOM MAF	CADD	PROTEIN POS	STUDY
RS145163696	Missense/nonsense	DM	0.0002172	1.6	56	Musa (2011) *Int J Obes (Lond)* epub:epub
RS17848368	Missense/nonsense	DM	0.0001011	25.1	70	Brown (1999) *Hum Mutat* 13:506
RS74907838	Missense/nonsense	DM	0.001156	14.45	111	Musa (2011) *Int J Obes (Lond)* epub:epub
RS104894319	stop gained	DM	0.0007602	38	143	Argyropoulos (1998) *J Clin Invest* 102:1345
RS373468564	Missense/nonsense	DM	3.184E-05	17.76	192	Musa (2011) *Int J Obes (Lond)* epub:epub
RS765633988	stop gained	DM	7.073E-06	38	252	Musa (2011) *Int J Obes (Lond)* epub:epub
RS142952570	Missense/nonsense		0.0007876	26.9	95	PMMO cohort (our study)

IGSF1

A missense deleterious homozygous variant, c.A1811C:p.N604T (rs146462069), in the *IGSF1* gene (NM_001555.5) was identified in an obese male. The variant is located in exon 12 and consists of a substitution of the evolutionarily conserved residue asparagine (Asn) with the smaller size amino acid threonine. The variant has a CADD score of 23.1 and gnomAD allele frequency of <0.01.

The *IGSF1* gene is located at the Xq25 of the X chromosome. The encoded protein is transmembrane glycoprotein that consists of a transmembrane domain, cytoplasmic domain and 12 immune loops and is expressed in several tissues including the brain and liver, and particularly highly in the testes and pituitary gland [222-224]. However, the precise molecular functions and mechanism is still unknown [224]. Defects in this gene can causean X-linked disorder that is characterised by congenital central hypothyroidism (C-CH) which is due to deficiency in the normal thyroid gland, macroorchidism and central, obesity and testicular enlargement [225, 226]. To date, a total of 31 cases with a pathogenic variant in the *IGSF1* gene have been previously described;, the variant reported here had not been identified before [225, 226]

MYT1L

An individual was identified with a deleterious missense variant in one recent candidate obesity gene: *MYT1L* (NM_015025, OMIM 613084). The variant, c.G7A:p.V3M, causes a change from the non-polar valine amino acid to non-polar methionine. This valine in exon 1 of the MYT1L gene is highly evolutionarily conserved across species from Zebrafish to X-tropicalis. The variant has a CADD score of 23.2, and a frequency of 0.0001365 in gnomAD, which is the equivalent of an allele count of 34/249080 gnomAD controls. The substitution with methionine possibly introduces an alternative start site that might alter the structure of the protein.

This gene is a member of the Myt1 family which is characterised by the highly conserved (C2HC-) zinc finger domain that has an important role in neuronal development [227]. The encoded protein is highly expressed in neuronal tissues and especially during the early stages of foetal brain development [228]. It is considered to be a pro-neuronal transcription factor and in conjunction with other transcription factors such as Brn2, Ascl1 and NeuroD1 can directly reprogramme human fibroblasts into functional neurons [229]. It has also been shown to act as a transcriptional repressor by repressing Notch signalling and subsequently promoting the differentiation of the neuronal cells [230].

The *MYT1L* is one of the genes implicated in a syndrome that is caused by a deletion involving the chromosome 2p25.3 [231-234]. The syndrome is characterised by intellectual disability and obesity [231-234]. The region of deletions varies in size, with the smallest consisting of two genes: *PXDN* and *MYT1L* [231]. Furthermore, genetic studies have reported cases of individuals with SNV a*t MYT1L* showing the same phenotype as that observed in individuals with 2p25.2

deletion syndrome, while SNVs and CNVS identified in the *PXDN* gene are reported to cause congenital cataracts but not the intellectual disability or obesity associated with 2p25.3 deletions [228, 231]. This suggests that the *MYT1L* is the key gene for the phenotype observed in the 2p25.2 deletion syndrome. The clinical features of the subjects reported with a *MYT1L* variant are shown in Figure 5.3. In addition, functional studies in an experimental zebrafish showed that loss of function of the *MYT1L* gene led to dysregulation of gene expression, reduction of the brain neuropeptide and hormone oxytocin, and disruption of hypothalamic development, similar in effect to that observed with loss of function of the *SIM1* and *PO3F2* genes [228]. This finding demonstrates the relation between *MYTL1* loss of function and development of obesity and intellectual disability. Figure 5.4 shows the structure of the *MYTL1* with all reported SNVs along with the variant identified in our study.

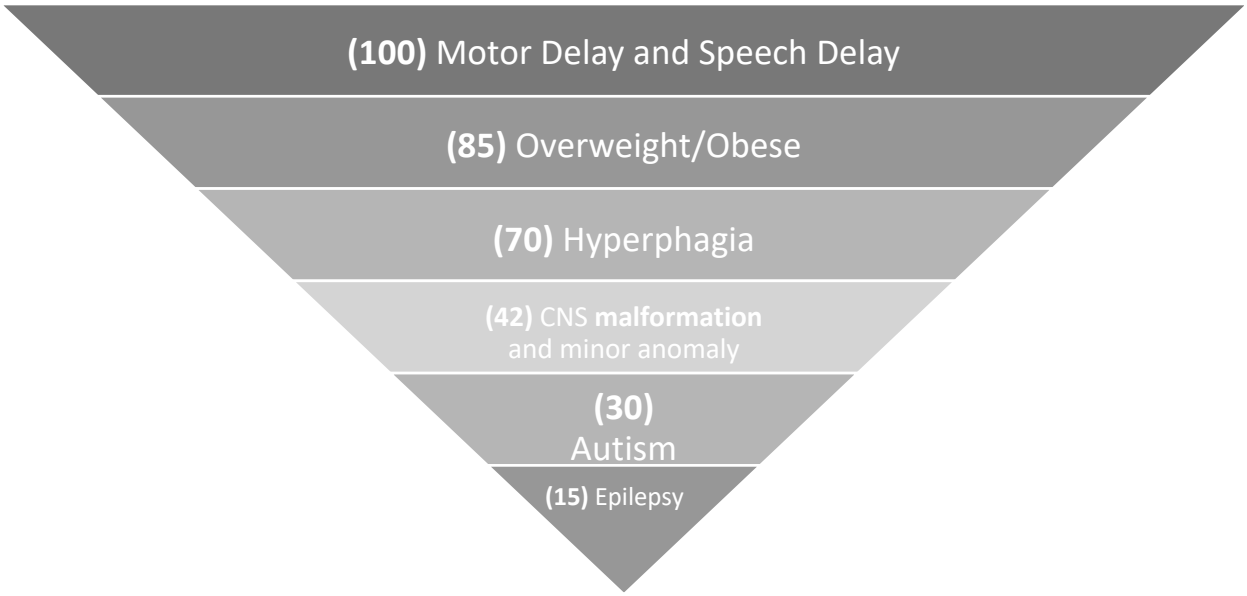


Figure 5.3 Clinical features of subjects reported with *MYT1L* variants.

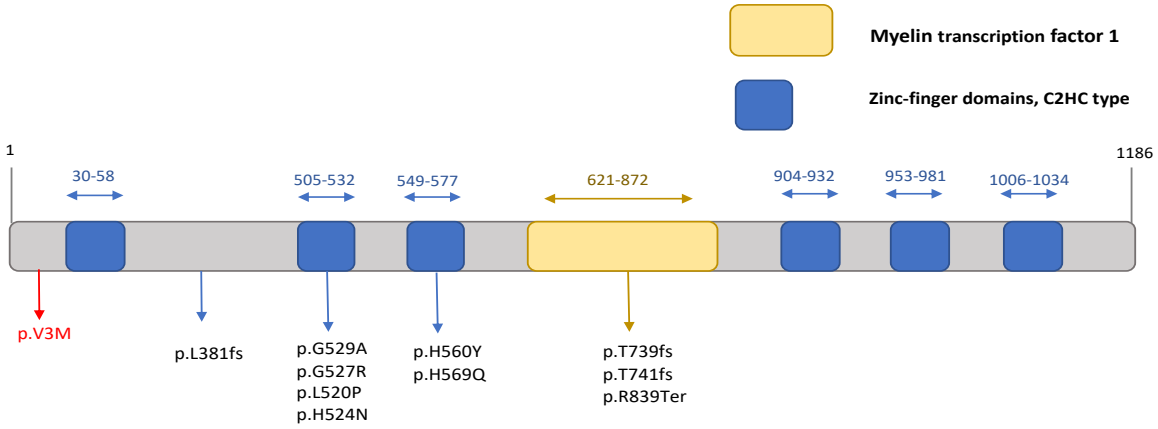


Figure 5.4: Schematic representation of the *MYT1L* gene and all previously reported variants. The ***MYT1L*** gene consists of six domains of the Zinc-finger domains and C2HC type and the Myelin transcription factor 1 domain.All previously reported pathogenic variants and the identified variant from the re-analysis ares shown in red.

POGZ

A total of three missense variants were identified in the *POGZ* gene (NM_015100, OMIM: 614787), in three individuals. These variants were c.G501C:p.M167I, c.T1939C:p.Y647H, exon19, c.G4001T:p.G1334V. The three variants were identified in the Heterozygous state; two are novel (not found in the public database and not reported previously) while p.Y647H has an allele count of 1 in gnomAD.

The first variant, c.G501C:p.M167I, is located in exon 5 and results in change from a methionine (Met) highly conserved across species to an isoleucine (IIe), an amino acid with a hydrophobic side chain. The second variant, c.T1939C:p.Y647H, is located in exon 13 and causes a change from the hydrophobic amino acid tyrosine (Tyr) to the electrically charged amino acid histidine (His), which might result in functional and structural changes. The residue Tyr is evolutionarily conserved across species to X-tropicalis. The third variant, c.G4001T:p.G1334V, causes a change from the simple amino acid glycine (Gly) to the hydrophobic valine (Val). The Gly residue is highly conserved across species to Zebrafish. The subject with this variant has another deleterious variant in another gene (*STED2*, rs780019200) which is described further in the oligogenic section.

The crystallographic structure of the *POGZ* protein is not yet available to demonstrate the predicted effect on the structure. Figure 5.5, adapted from Batzir, *et* al. [33] , shows all previously reported variants to date along with the variants identified in our study.

The *POGZ* gene is mapped to 1q21.3 and encodes a Hetrochromatin protein 1 a-binding protein with multiple domains including a C2H2-type zinc-fingers cluster, a centromere protein-B-like DNA-binding domain (CENP), HP1-binding motif and DDE domain. It is widely expressed in many

human tissues including the brain and has been reported to function as a transcriptional regulator with a significant role in neuronal proliferation and function, gene transcription and mitosis through regulating mitotic progression for chromatin remodelling and chromosome segregation [15, 16]. A mutation in *POGZ* has been described in individuals with an intellectual disability, autism, White-Sutton syndrome and schizophrenia. The phenotypic features are diverse among individuals. A recent study investigated the phenotypic features of patients with variants of *POGZ* through analysing the clinical and molecular data of 25 cases with deleterious mutations in the *POGZ* gene. They identified common features including severe delay in language and speech, motor and coordination and variable levels of intellectual disability, obesity, vision problems, microcephaly and hyperactivity. Stessman et al. (2016) identified 13 out of 23 cases to be obese and another group identified obesity in older children, suggesting obesity becomes an issue in later stages [169]

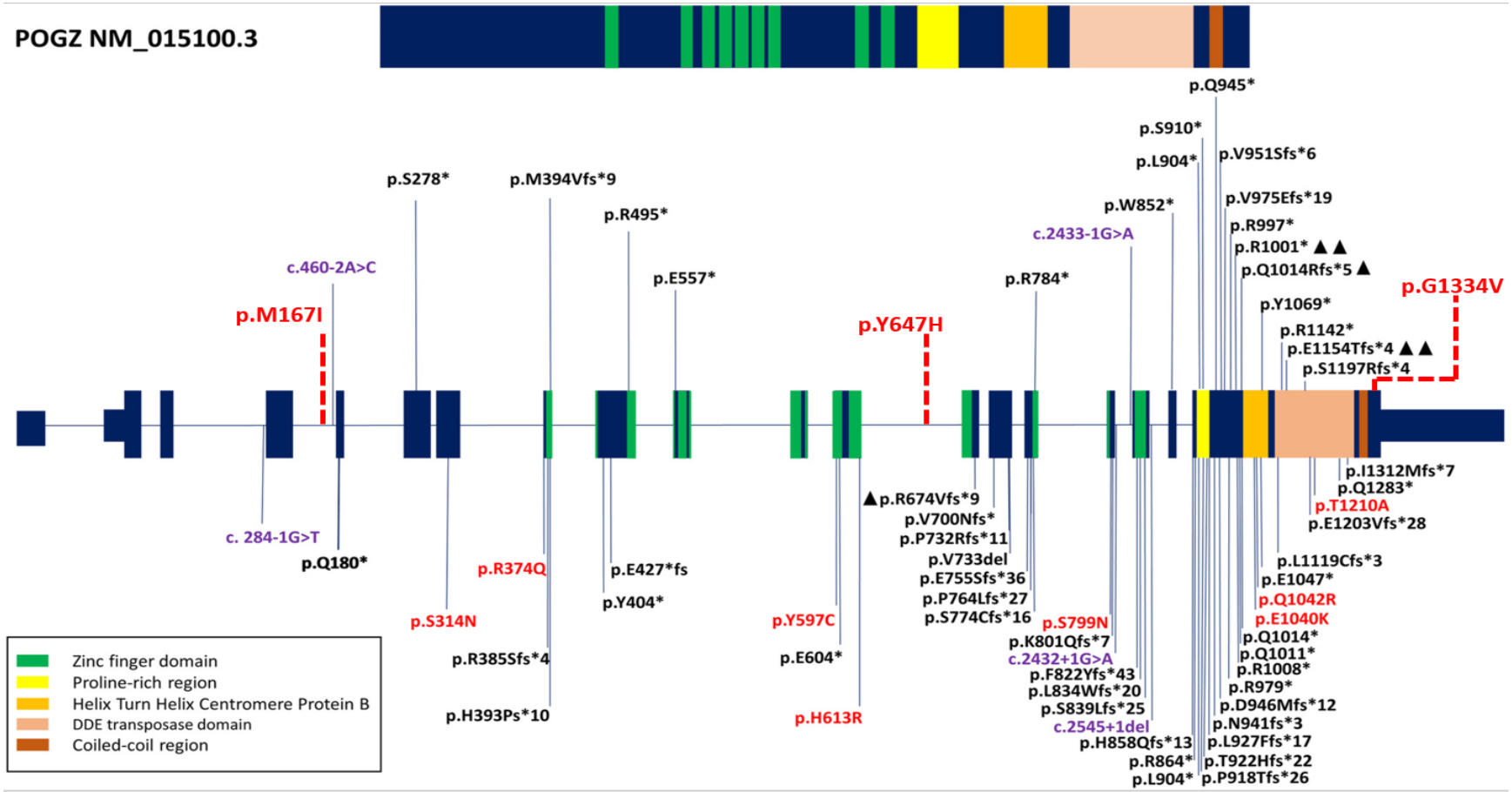


Figure 5.5: Schematic representation of the *POGZ* gene and all reported variants. The *POGZ* gene has a range of domains demonstrated in the top key box. Previously reported variants in the literature are shown in the below exon structure and variant reported in the [33] are shown above. The missense variants identified in our re-analysis are shown in the dashed red line. Variants' colours: splice variants (Purple), nonsense/ frameshift variants (Black), missense variants (Red). (Original source before the edit: Assia Batzir *et al.* (2020)[33])

RBMX

Two hemizygote variants were found in the *RBMX*, (NM_002139, OMIM: 300199), in two male individuals. The first variant is a frameshift deletion that results in deleting two base pairs at a highly conserved region. The variant has a frequency of 0.003019; with only two cases in gnomAD as hemizygote. The second variant, c.559_561del:p.187_187del, is a non-frameshift deletion that resulted in the deletion of three base pairs that encodes the polar uncharged amino acid threonine which is conserved across species. The minor allele frequency of the variant in gnomAD is 0.004495, where 234 cases are hemizygote.

The *RBMX* gene is an RNA binding protein which is homologous (homology) to the candidate Y chromosome gene *RMBY*. *RMBX* is conserved in all mammalian X chromosomes and highly expressed in the body [235]. Zebrafish studies have implicated the *RBMX* gene in the regulation of genes and pathways important in the brain development through controlling transcription repression and activation, chromatid cohesion, splicing and expression [236]. In addition, further evidence has shown the involvement of *RBMX* in cancer development and other diseases such as Shashi type [235]. Shashi type is an X-linked recessive inheritance that is characterised by moderate intellectual disability, coarse facial features, large ears and obesity [237]. *RBMX* is believed to regulate brain development and intellectual ability through interactions with proteins that have roles in brain development, splicing controls and DNA damage response [235].

AFF4

Three deleterious variants were identified in another gene that is involved in syndromic obesity, AF4/FMR2 Family Member 4 (NM_014423, OMIM:604417), *AFF4*. These are missense variants in the Heterozygous state, and include: c.G626A:p.R209H, c.G3236A:p.S1079N and, c.3319A>G, p.Thr1107Ala.

The first missense variant consists of a substitution of the evolutionarily conserved residue arginine (Arg) with histidine (His), both positively charged amino acids. The allele has a CADD score of 22 and frequency of the variant in gnomAD is 0.00003535, which is equivalent to 10/282866. The second variant, c.G3236A:p.S1079N, results in a change from the amino acid serine (Ser) to asparagine (Asn), both amino acids with polar-uncharged side chains. The residue (Ser) is highly conserved across species. This variant is located in the c-terminal homology domain (CHD) and has a CADD score of 28.1. The variant is not found in the public database and has not been reported before. The third variant, c.3319A>G, p.Thr1107Ala, causes a substitution of the polar uncharged amino acid threonine (Thr) with the hydrophobic amino acid alanine (Ala). The Thr residue is highly conserved across species to Zebrafish. The variant has not been reported previously and has a frequency of 0.00231 in gnomAD which is equivalent to an allele count of 653/282676 gnomAD controls. The variant has a CADD of 22 and is mapped to the CHD domain of the *AFF4* gene.

The gene encodes the AFF4 scaffold protein that belongs to the AF4/FMR2 family proteins. Other members include AFF1, AFF2 and AFF3 (6) [238]. The AFF4 protein is a fundamental competent of the super elongation complex (SEC), which plays an important role in the transcription

regulation through governing the transcriptional elongation checkpoint control (TECC). The SEC regulates this process by mobilisation of the paused RNA polymerase II (RNAP2) [238]. The AFF4 protein by itself or together with AFF1 acts as a central scaffold to assemble the complete SEC through direct interactions with other subunits, including the positive transcription elongation factor (P-TEFb), ENL family protein, either ENL or AF9 and an elongation factor RNA polymerase (ELL2).

Missense variants in this gene are known to cause the CHOPS syndrome [239]. The CHOPS abbreviation refers to multiple features of the disease which include cognitive impairment, coarse facial features, heart defects, obesity, lung (pulmonary) involvement, short stature, and skeletal abnormalities [238, 239]. The disease follows the autosomal dominantmode of inheritance, and to date five *de novo* variants have been reported in the literature, and a total of six variants have been reported in the *AFF4* gene, in a total of 11 individuals. All of the six variants lie within the evolutionarily highly conserved ALF (AF4/LAF4/FMR2) homology domain that involves recruiting ELL. Such mutations in *AFF4* are predicted to lead the accumulation of RNA polymerase II (RNAP2), causing transcriptional abnormalities.

From the re-analysis approach, two of the variants were mapped to C-terminal homology domain (CHD) which is preserved among the other members of the AF4/FMR2 protein family. The CHD domain is found to be involved in regulating the formation of both the AFA (homodimer) and AFF1-AFF4 (Hetrodimer). Additionally, the surface loop region in AFFF4-CHD has been shown to act as substrate for the P-TEFb. An ***in vitro*** study also identified interactions of the AFF4-CHD with

DNA and RNA [240]. Figure 5.6 shows a schematic representation of the *AFF4* gene and all reported.

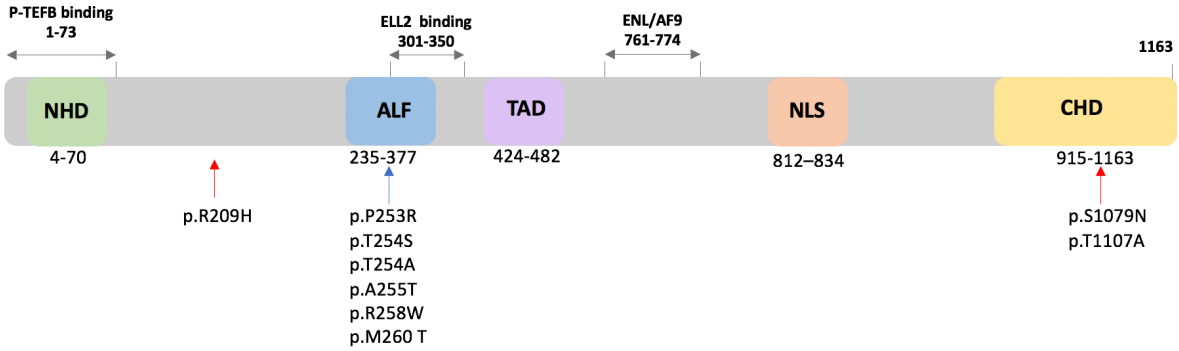


Figure 5.6: Schematic representation of the *AFF4* gene and all reported variants. The *AFF4* gene's four domains are the N-terminal homology domain (NHD), the AF4/LAF4/FMR2 homology domain (ALF), a serine-rich transactivation domain (TAD), a bipartite nuclear/nucleolar localization sequence (NLS) and a C-terminal homology domain (CHD). Previously reported variants in the literature are shown in blue arrow. The pathogenic identified in our re-analysis are shown in dashed red arrow.

RAI1

A total of five variants have been identified in the *RAI1* gene that is known to cause another form of syndromic obesity, Smith-Magenis syndrome (SMS). Three of the variants, c.C3689A:p.A1230E, c.3781_3783del p.Glu1261del and, c.C5036T:p.A1679V, were identified in five individuals as shown in Table 5.3. The remaining two variants (c.C1142T:p.A381V , c.G5653A:p.D1885N) exist with other rare deleterious variants in other genes in the oligogenic form, as shown in Table 5.4.

The first variant is a missense rs374187267, c.C3689A:p.A1230E, which is located in the C-terminal region between the NLS2 and PolyS. The variant has a CADD score of 23 and results in a change from the negatively charged glumatic acid (Glu) to the hydrophobic side chain amino acid alanine (Ala). The residue, Glu, is evolutionarily conserved across species including mice, dogs and elephants, and has a frequency of 0.00001069 in gnomAD. The variant has not been reported in the literature before. Evidence from the functional studies based on previously reported variants in the C-terminal region, p.Arg114Gln, p.Arg1217Gln, p.Gln1389Arg, p.Gln1562Arg and p.Ser1808Asn, has shown that these variants lead to similar effects, reducing and interfering with the activity regulation of the transcription factors protein, effecting the DNA direct or indirect binding to the *RAI1* gene [241].

The second variant, rs149716029 NM_030665.3:c.3781_3783del p.Glu1261del, is a Heterozygous non-frameshift deletion (inframe deletion) that was found in three subjects. The variant results in deleting three bp that encode the highly conserved negatively charged amino acid glutamic acid (E). The same variant has been reported previously in a five-year-old individual

and his unaffected father [242]. The same group showed that expression of *RAI1* mRNA was severely decreased compared to the control. The variant has a frequency of 0.002816 in gnomAD.

The third variant, c.C5036T:p.A1679V, is a Heterozygous missense variant that is located in exon 3 in the C-terminal between the Poly-S domain and the PHD domain. The altered amino acid is highly conserved across species including human, mouse, dog, elephant, dog and tropicalis. The variant has a CADD score of 23.5 and allele frequency of 0.0005835 in gnomAD and was not reported previously. The function effect could be similar to the one described above of the previously reported variants found in the same region: p.Arg114Gln, p.Arg1217Gln, p.Gln1389Arg, p.Gln1562Arg and p.Ser1808Asn.

The forth variant, c.C1142T:p.A381V, is a Heterozygous missense located in exon 3 in the region between PolyS and PolyQ. The allele has a CADD score 24, and frequency in gnomAD is 0.003643. As described earlier this variant exists with another rare deleterious variant in the form of Oligogenic mode of inheritance, *SH2B1*;c.G1633A:p.G545S which was found in the initial analysis of the WES data.

The last variant, c.G5653A:p.D1885N, is Heterozygous missense located in exon 4 in the PHD domain. The allele has a CADD score 27.5, and its frequency in gnomAD is 0.00008319. This variant along with another silent variant in *RAI1* were reported previously in a proband diagnosed with SIM1 syndrome, but was also found in his unaffected father [242]. In our case, this variant exist with another rare deleterious variant in the *MAGEL2*, c.C1079T:p.A360V as an oligogenic mode of inheritance.

The Smith-Magenis syndrome is a complex neurobehavioral disorder with intellectual disability that is caused by either deletion of the interstitial 17p11.2 region (representing 90% of SMS patients) or mutations in the *RAI1* gene. The encoded protein is a nuclear transcription factor and is expressed throughout the body tissues (around 84 human tissues), with high expression in neurons, brain and heart tissue [86, 87]. The *RAI1* protein encompasses conserved functional domains (Figure 5.7). The first one is at the N-terminal which includes a polyglutamine tracts domain (PolyQ) , a polyserine tract domain (PolyS), and the second functional domain group located at the C terminal includes bipartite nuclear localization signals (NLS1 and NLS2) and an extended plant homeo-domain/zinc finger domain (PHD) at the C-terminus, which is commonly found in chromatin-associated proteins [86]. The first functional domain at the N-terminal is responsible for the transcription activity while the second group at the C terminal is involved in multiple roles including nuclear localisation signals, putative repressor or regulator of transactivation activity and as a DNA-binding domain [243, 244].

Several mouse models of rai1 have been studied [86, 245, 246]. The Heterozygous Rai1 mice displayed differences in gene expression with a minor phenotype of SMS including craniofacial abnormalities, obesity, and circadian abnormalities, while the homozygous null mice exhibited lethality d [86, 245, 246]. To date, little is known about the function of *RAI1*. It is believed to be involved in the regulation of cell growth and cell cycle, lipid and glucose metabolism, and skeletal, bone and neuronal development.

Analysis of genotype-phenotype correlation in SMS showed that *RAI1* haploinsufficiency is the primary gene contributing to most of the SMS features and the severity and variability of the

phenotype varies as the deletion increases, encompassing more genes [247, 248]. Patients with mutations in *RAI1* tend to have less severe or absence of features of cognitive impairment, heart and rental defect, short stature, hearing loss and motor delay compared with the patients with the 17p11.2 deletion [249-251]. However, patients that harbour mutations in *RAI1* exhibit the other core features of SMS such as craniofacial abnormalities, sleep disturbance, intellectual disability and, more frequently, overeating and obesity [249-251].

To date, more than 30 variants have been reported to be linked to SMS [86]. The majority of those variants are located in exon 3 which represents around 95% of the coding region [86]. The locations of known point mutations in *RAI1* are shown in Figure 5.7. From the re-analysis of the WES, we have identified five variants that are found in a total of seven individuals and described individuals below: three missense variants and one in frameshift deletion.

Overall, the identified *RAI1* cases described here require further clinical analysis including detailed clinical descriptions to investigate the phenotypic feature of the cases in comparison to the previously reported RAI1 cases.

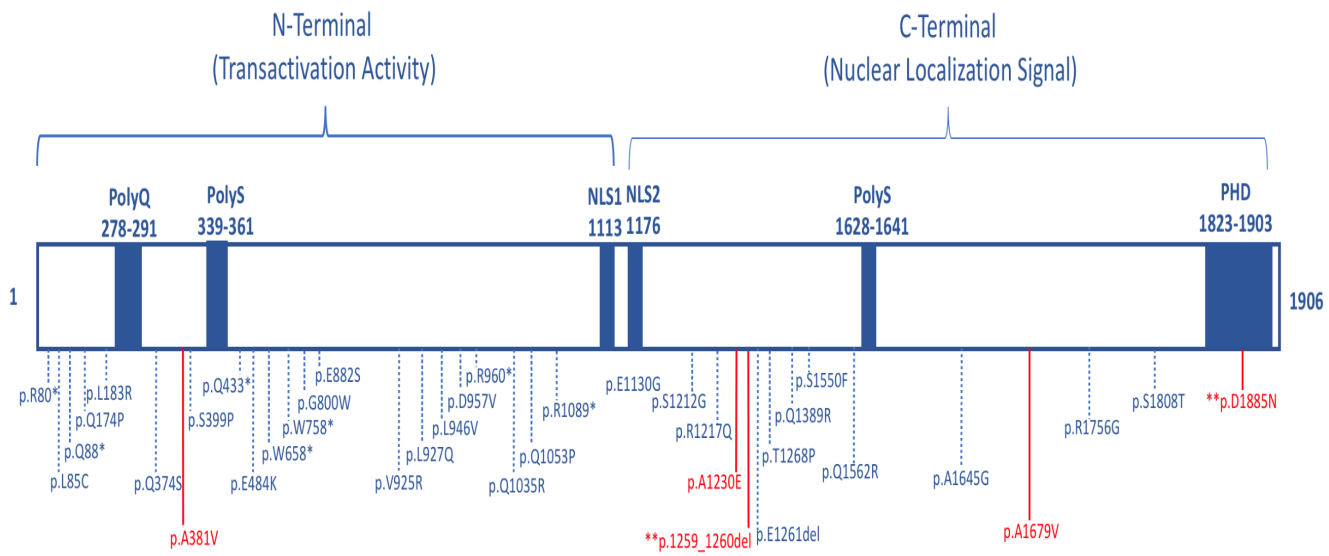


Figure 5.7: Schematic representation of the *RAI1* gene and all reported variants. The RAI1 gene four domains domain (PolyQ), a polyserine tract domain (PolyS), nuclear localisation signals (NLS1 and NLS2) and the plant homeo-domain/zinc finger (PHD domain). All reported point mutation identified in SMS patients in literature is shown in blue. In addition, the identified variant from the re-analysis is shown in red.

* indicates the variants identified in our study and have been reported before.

SETD2

From the 3 cases that follow the oligogenic mode of inheritance, one of them has a variant in the SET Domain Containing 2, Histone Lysine Methyltransferase (*SETD2).* The variant is a missense in heterozygous status c.T1654C:p.Y552H (rs780019200) which exists with another rare deleterious variant described earlier in the *POGZ* gene, c.G6703C:p.V2235L.

The variant rs780019200 is located at exon 15/21 and consists of a substitution of the evolutionary conserved Tyrosine (Y) with Leucine (L), both of which are amino acids with hydrophobic side chains.. The allele has a CADD score 24.6 , and a frequency in gnomAD is 0.00001769 which is equivalent to 1/251308.

The *SETD2* gene encodes a histone-modifying enzyme, methyltransferase, that catalyses H3K36 trimethylation (H3K36me3), which is an important histone in the active transcription including transcription activation, DNA repair, transcription fidelity, RNA splicing and chromatin organisation. The encoded enzyme is considered the primary methyltransferase for H3K36me3. In studies in mice and flies, ablation of *SETD2* (deficiency in *SETD2*) leads to lethality. Mutations in *SETD2* have been reported to cause Luscan-Lumish syndrome, which is characterised by various features including macrocephaly, postnatal overgrowth, obesity, intellectual disability and advanced carpal ossification [252]. To date, 13 patients have been reported to have mutation in *SETD2*, 10 classified as a Luscan-Lumish syndrome, while 3 others were reported with Autism. Current phenotype comparison analysis by Mariz et al.,2019 on these 13 patients showed an important variability in the clinical phenotype and also emphasised other features such as high susceptibility to obesity and high prevalence of autism and behaviour difficulties [253].

5.3.6 Findings from the mouse model obesity genes

Mice models play an important role in understanding obesity. Through including 67 obesity mouse genes in the gene panel, variants in seven genes were identified in 11 individuals. These genes include: Androgen Receptor (*AR*), Ataxin-2 (*ATXN2*), Histamine Receptor H3(*HRH3*), Corin Serine Peptidase (*CORIN*) and Phosphatidylinositol-4-Phosphate 3-Kinase Catalytic Subunit Type 2 Gamma (*PIK3C2G*). A summary of the variant details are provided in Table 5.6 and each gene will be discussed separately below.

Table 5.6 Summary of the identified variants in the mouse model obesity genes from the re-analysis of the WES.

Patient ID	Gender	Gene	Mouse Mode of inheritance	RS ID	Variant type	HGVS	gnomAD	CADD	Het/ Hom
S20	**F**	***HRH3***	AR	rs202179023	Missense	p.Arg420His	2.45E-05	19.23	Hom
S21	**F**	***CORIN***	AR	rs138837512	Splice Acceptor	c.618-2A>T	0.00009112	27.6	Hom
S22	**M**	***PIK3C2G***	AR	rs61754413	Missense	p.Leu488Phe	0.003408	26.5	Hom
S23	**F**	***AR***	AR	rs200185441	Missense	p.Gln58Leu	0.0001415	21.4	Hom
S24	**F**	***AR***	AR	rs200185441	Missense	p.Gln58Leu	0.0001415	21.4	Hom
S25	**M**	***ATXN2***	AD	11:111990210	Missense	p.H44Y	N/A	27.5	Het
S26	**F**	***PEG3***	AD	rs56237501	Missense	p.Ala1298Gly	0.0041	24.4	Het
S27	**M**	***PEG3***	AD	rs35087473	Missense	p.Ser543Ala	0.001944	10.48	Het
S29	**M**	***PEG3***	AD	19: 57329159	Missense	p.H273Y	N/A	26.1	Het
S31	**M**	***PEG3***	AD	rs765326136	Missense	p.Lys297Thr	0.0000176	22.7	Het
S33	**M**	***PEG3***	AD	rs149044578	Missense	p.His863Arg	0.0004554	20.9	Het

5.3.6.1 Description of the new identified variants in mouse obesity genes

AR

In the PMMO, a missense variant, c.A173T:p.Q58L, in the homozygous state was identified in the two females. The variant has a CADD score of 21.4 and allele frequency of 0.0001415 in gnomAD. The variant is located in exon 1 at the beginning of polymorphic trinucleotide repeat in transactivation domain which results in a change from Leucine (Leu), amino acid hydrophobic side chain to Glutamine (Gln) , uncharged side chain.

The *AR* (Androgen Receptor) is mapped to Xq11-q12 and encodes a protein that is expressed widely in several types of cells and tissues [254, 255]. The encoded *AR* protein is a ligand-dependent transcription factor and is a member of the nuclear receptor family, NR3C4. One of the main roles of the androgen receptor is regulating the actions of androgens of both men and women, which have wide biological functions. It has three homology functional domains, namely: the N-terminal domain, DNA-binding domain (DBD) and the ligand-binding domain. The N-terminal domain contains a variable size of a trinucleotide repeat known as polymorphic repeat, CAG, which encodes polyglutamine regions. The normal length of the CAG repeat is 9 - 38 residues and variations of this repeat have been found to effect transactivation potential, reduce the androgen receptor, and possibly correlate with many clinical factors such as androgen insensitivity, prostate cancer, male infertility (risk of defective spermatogenesis) and spinal and bulbar muscular atrophy (SBMA). Beside the very important role of *AR* in the reproductive system through mediating the function of the male sex hormones, androgens, the receptor also has important roles in the cardiovascular system, musculoskeletal system, haemopoietic system, immune system and neural system [254, 255]. The specific biological and physiological effects and roles of the *AR* have been identified

through studying animal models, especially male AR knockout mice, as summarised in Davey et al.'s review [255]. In comparison, only a few studies have studied the phenotype of female AR knockout mice.

Nevertheless, a recent study by Fagman et al. 2015 has demonstrated that the depletion of the AR receptor promotes atherosclerosis, obesity on a high fat diet, and dyslipidaemia [256]. These findings correlate with the male AR knockout mice studies showing that exposure to a HFD led to the development of late-onset obesity and metabolic dysfunction in male but not female ARKO mice [257-261].

The identified variant is located in exon 1 at the beginning of polymorphic trinucleotide repeat in transactivation domain which results in a change from Leucine, an amino acid with a hydrophobic side chain, to Glutamine, which has an uncharged side chain, causing shortening of the Leucine tract and lengthening of glutamine tract The same variant was reported previously in two males with infertility. It is predicted that the change changes the domain confirmation and influences the interaction with other repressors and activators [262-264]. Another two studies identified the same variant in two children: a 19 months old boy characterised with isolated hypospadias and the a 7 month new-born of a patient diagnosed with partial androgen insensitivity [262-264].

In the *AR* gene, a missense variant in a homozygous state was identified in two female individuals. The variant, c.A173T: p.Q58L, is located in exon 1 in the polymorphic trinucleotide repeat.

ATXN2

A missense variant, *ATXN2*:NM_002973:exon5:c.C925T:p.H309Y, with a CADD score of 27.5, was identified in the ATXN2 gene.

The *ATXN2* gene is evolutionarily conserved from yeast to mammals and is mapped to 12q24.1. It consists of 25 exons and encodes the protein *ATXN2*, which contains 1,312 amino acid residues [265-267]. The protein is expressed widely in various adult and embryonic cells such as the brain, liver, nervous system, heart and muscle and has a nuclear/cytoplasmic localisation [265, 266, 268, 269]. It plays an important role in several biological process including RNA processing, translation regulation, R-Loop regulation, cytoskeleton organisation and calcium homeostasis, as reviewed by Ostrowski et al. 2017 [269].

The protein contains a polyglutamine tract (polyQ) which is formed from a tandemly repeated unit CAG/CAA within the first exon of N terminal the gene [265-267]. Expansion of this repeat can result in different diseases such as spinocerebellar ataxia type 2 (SCA2), amyotrophic lateral sclerosis (ALS) and neurodegenerative disease [270, 271]. Within the terminal region, the protein contains two globular domains that are conserved across species, LSm and LSmAD, and which are important in mediating the RNA processing [272]. Additionally, Ataxin-2 contains another two conserved domains, Proline and Pam2, which regulate the interaction with PABP2 [273, 274].

Several lines of evidence suggest a possible role of ataxin-2 in body metabolism including fat distribution, lipid metabolism, body weight, obesity, and insulin sensitivity/resistance [275-278]. Ataxin-2 deficiency in different animal models caused obesity, hyperphagia, and altered lipids [275-278]. In addition, Scoles et al. identified hyperphagia in both knockout and

Heterozygous mice [279]. The development of obesity is predicted to be related to dysfunction of this protein in the hypothalamus [279-281].

In genetic linkage analysis studies of obesity-related traits, a correlation was identified between the chromosome 12q24 locus, which includes the ATXIN-2 gene, and obesity [reviewed in [282]. On the other hand, another study reported the occurrence of obesity and polyphagia in a family's members segregating with SCA2, which was the first evidence that supported the relation between ATXIN-2 and obesity [282].

From the re-analysis a total of five variants in the *ATXN2,* located in exon 1, were identified. Three non-frameshift insertions and one Heterozygous frameshift deletion which are located at the polQ tract and within the standard length; <34 in length. The last variant is missense, *ATXN2*:NM_002973:exon5:c.C925T:p.H309Y, which is predicted to be deleterious with a CADD score of 27.5, and is located at one of the RNA-binding domains known as Like-Sm (LSm) which has an important role in RNA metabolism and RNA processing [283-285]. This variant is novel as it is not found in gnomAD and has not been reported before.

HRH3

The *HRH3* encodes a G-protein couple receptor widely expressed by neurons in the Central Nervous System (CNS)[286, 287]. The H3HR protein mediates regulatory functions through acting as a presynaptic inhibitory autoreceptor, regulating histamine turnover and also acting as Hetroreceptor to modulate the release and synthesis of histamine and various neurotransmitters including dopamine, noradrenaline, acetylcholine and serotonin [288]. HRH homozygous null mice have increased body weight, food intake, fat mass, and leptin and insulin levels, and reduced energy expenditure [288-290]. This phenotype is thought to be due to dysregulation of histaminergic neurons and indicate that the histamine receptor H3 is important in the regulation of body weight, food intake and energy expenditure.

In the PMMO cohort, a homozygous missense variant, rs202179023 c.1259G>A (NM_007232.2) p.Arg420His (NP_009163.2), was identified in one obese individual. The variant is predicted to be deleterious with a CADD score of 26.8 and has a frequency of 0.00001236 in gnomAD.

CORIN

CORIN is a transmembrane serine protease that expressed in the cardiac atrium [291, 292]. It is involved in the conversion of the peptide hormone, atrial natriuretic peptide (ANP) inactive precursor into the active ANP in response to various pathophysiological and physiologic conditions such as pressure overload [291, 292]. ANP is regulates cardiorenal homeostasis and blood pressure. In mice, expression of *CORIN* was identified in several cell lines from tissues such as kidney, bone and testes. *CORIN* deficiency in mice abolishes the conversion of ANP, leading to the development of hypertension, cardiac hypertrophy and alterations of sodium homeostasis [292, 293]. In addition the knockout mice had increased body weight at the age of 4 months, for which the physiological mechanism is unknown [293].

One splice acceptor variant in the homozygous state was identified in a female with a BMI of 83 pre-surgery. The variant is located straight before exon 3 and has a CADD score of 27.6. The variant has a frequency of 0.0009112 in gnomAD (2/ 21950). Consistent with the mouse model, the patient had hypertension and early onset obesity as well as hypothyroidism diagnosed at the age of nine.

PEG3

Paternally expressed gene 3 (*PEG3*) is an imprinted gene that is localised at a well-conserved genomic region, 500-kb domain, on the human 19q13.4 chromosome that contains six other imprinted genes. Imprinted genes refer to the genes that have parent-specific allelic expression, where one allele is expressed and the other allele is silenced dependent on their parent of origin. These types of genes are classified mainly into two groups: paternally expressed genes and maternally expressed genes. They have a range of important roles in development and metabolism.

The *PEG3* gene consists of 9 exons and encodes C2H2 type zinc finger protein that is expressed during embryogenesis and also in adult skeletal muscle, brain, testes and ovaries [294-297]. The protein is involved in several cellular roles such as apoptotic pathways and cell proliferation [298, 299].

The mouse model of *PEG3*, Peg3+/ −, has been shown to develop a high level of body fat that includes subcutaneous, abdominal and intra-scapular fat in both males and females. Although the mouse model was hypophagic, the increase in the different adiposity tissues is predicted to occur due to lower metabolic activity and reduced energy expenditure, which could result from the hypothalamic dysfunction [300].

From the analysis, we have identified a total of four missense variants in Heterozygous state in four individuals. All of those variants have a CADD score of greater than 20 and allele frequency of <0.01 in gnomAD, and none of them have been reported previously. Nevertheless, although these variants match the mode of inheritance of the mouse, further

confirmation from analysis of parental samples is required to determine the parental origin of the variants.

PIK3C2G

Phosphatidylinositol-4-Phosphate 3-Kinase Catalytic Subunit Type 2 Gamma (*PIK3C2G*) encodes an enzyme that belongs to a family of lipid kinases; the phosphoinositide 3-kinases (PI3K) [301]. The PI3K enzyme family is divided into three classes: class I, class II (which contains three members: PIK3C2A , PIK3C2B, PIK3C2G) and class III. The protein is expressed in the liver, pancreas, breast, prostate and small intestine, and has been shown to contribute to several cellular signalling pathways and functions [302-304]. Unlike the other members of class II, the PIK3C2G does not possess the clathrin-binding domain; rather it has an important role in vesicular trafficking and metabolic signalling through the generation of the lipid substrate PI and P2. The formation of P1 and P2 is important for continuing Akt2 activation which subsequently sustains the effects of insulin on a cell. Results from the mouse model revealed that the loss of this enzyme led to reduction of the endosomal pool of PI(3,4)P2 that subsequently reduced the prolonged insulin-dependent Akt2 phosphorylation and glycogen synthase (GS) activation. Thus Pik3c2g-deficient mice exhibited the development of insulin resistance, adiposity and hyperlipidemia, due to the reduction in the GS (activity that is required for the insulin receptor stimulation and metabolic responses in the liver [302, 304]. In addition, results from several GWAS studies have shown association of PIK3C2G SNPS and several other metabolic phenotypes including diabetic nephropathy, development of T2D, hyperlipidemia, myocardial infarction and BMI [305-308]. These results support the evidence from the mouse model regarding the role of PIK3C2G in the regulation of cell metabolism.

In the PMMO cohort, a total of six variants were found, as shown in Table 5.7, but only one carrier matched the mode of inheritance of the mouse model. This variant, rs61754413 NM_001288774:exon16:c.C1462T:p.L488F is a missense variant that is located at a highly evolutionarily conserved amino acid Leucine and has a CADD score of 26.5. The allele frequency in gnomAD is 0.003408 where there were only three individuals with a homozygous state.

Table 5.7: Summary of the identified variants in the *PIK3C2G* gene in the PMMO cohort.

Variant type	Variant Details	dbSNP	gnomAD	CADD	Subject	MOD
frameshift deletion	NM_001288774:exon19:c.1961delT:p.I654fs	N/A	N/A	N/A	AB303	Het
nonsynonymous	NM_001288774:exon32:c.C3536T:p.A1179V	rs77070108	0.003358	11.27	AB48	Het
nonsynonymous	NM_001288774:exon16:c.C1462T:p.L488F	rs61754413	0.003408	26.5	AB166	Hom
					AB170	Het
					AB64	Het
					EF100	Het
nonsynonymous	NM_001288774:exon29:c.A3254C:p.H1085P	N/A	N/A	24.1	AB257	Het
nonsynonymous	NM_001288774:exon29:c.C3241T:p.P1081S	rs146312199	0.005018	24.3	CD50	Het
stoploss	NM_001288774:exon33:c.G3794C:p.X1265S	rs61757718	0.01083	13.8	EF70	Het

5.4 Discussion

In this Chapter, a re-analysis of pre-existing WES data was performed on a total of 91 cases, to further explore undiagnosed cases, using an expanded list of genes relevant to monogenic obesity, syndromic obesity and/or mouse models of obesity. The re-analysis revealed possible causative variants in monogenic/syndromic obesity genes in 21 cases, where previous analysis had explained only 11 cases. In addition, 11 candidate variants in genes from the rodent models were identified. In total the potential diagnostic yield was 31%. A detailed description of each variant and the gene involved is given in the results section of this Chapter.

From the obesity candidate genes, variants were identified specifically in new obesity genes that were not investigated in the initial analysis , namely *UCP3* and *MCHR1*. The variant identified in *MCHR1* has been reported previously in an obese proband and his father.

Of note, the identification of several variants in genes implicated in syndromic obesity, in individuals not showing obvious signs of the wider syndromic phenotype (as we seen in people with *RAI1* mutations, which causes Smith-Magenis Syndrome, but with a highly variable phenotype), is a vital indication that deficiency in such genes could be found in obese individuals that lack a cognitive phenotype.

Another key outcome was confirmation of further cases of apparent oligogenic inheritance. The occurrence of multiple rare deleterious variants in obesity and obesity syndromic genes highlights oligogenic inheritance, in which multiple variants could be contributing to the disease, as a novel mechanism in obesity. In parallel with the findings from Chapter 4, where

I discussed a proband with an oligogenic form of obesity, including variants in *SH2B1*, *POGZ* and *MBD5*, we have found another case of oligogenic obesity that involve *SH2B1*, together with variants in other obesity genes.

Recently, several diseases have been reported to show oligogenic inheritance, such as cardiovascular disease, autism, BBS and, ciliopathies [309-312]. Thus, taking the evidence presented here and in Chapter 4 along with literature from other conditions, we suggest that obesity can exhibit a different mode of inheritance than those already recognised. Furthermore, the identified variants in the mouse model genes represent a novel group of variants and candidate genes which haven't been reported to be implicated in human disease before.

Overall, the majority of the described variants are not found in ClinVar and require further supporting evidence in order to confirm their impact through functional studies, careful and detailed clinical evaluation (especially for the syndromic obesity gene variants) and screening in a larger cohort. Despite these limitations, these findings suggest the importance of performing re-analysis of exome sequence data in a timely manner, e.g. 6–42 months after the initial analysis, informed by the continual advances in bioinformatics and the expansion of the genomic databases and literature base [313-315]. The effectiveness of the re-analysis observed here is consistent with other re-analysis studies which have increased the diagnostic yield in a similar way [313-315].

In summary, the identification of additional putatively pathogenic variants implicated in the development of obesity and diabetes, especially in this extreme cohort, provides further

information on the genetic architecture of severe obesity and reveals new insights into the disease. This is demonstrated by the detection of potential oligogenic inheritance, and identification of new deleterious alleles in previously under-studied obesity candidate genes. The work described here highlights the utility of genetic investigation of patients with severe obesity, which may have implications for their management and potential genetic counselling.

These preliminary findings required replication in a larger cohort. In this context, a customised genotyping array focusing on obesity and diabetes mellitus (T2D) was designed and applied to a large cohort of patients with extreme obesity. The design of this array is described in Chapter 6, the results of the genotyping are described in Chapter 7, and implications for responses to bariatric surgery are explored in Chapter 8.

CHAPTER 6: DESIGN OF THE CUSTOMISED GENOTYPING ARRAY

6.1 Introduction

This Chapter describes the design of a custom genotyping array for rapid and cost-effective screening of large numbers of obese individuals for genetic variants potentially causative of monogenic obesity and /or diabetes. In order to facilitate exploration of the effects of "genetic background", common SNPs identified in GWAS of obesity, type-2 diabetes and related traits are also included in the design.

As described earlier, obesity is a complex disease with monogenic, syndromic and common forms. Although Mendelian (monogenic and syndromic) forms of obesity and diabetes are considered to be rare in the general population, their prevalence in people with severe obesity (eg. Bariatric surgery patients) is currently underexplored and may be under appreciated.

The majority of the genetics studies in obesity and diabetes mellitus have studied the genetics factors of the two forms of the diseases separately. Nevertheless, the potential importance of rare deleterious variants contributing to common disease has been supported by the observed overlap between genes implicated in monogenic forms of the disease and the common variants detected by GWAS. For example, nearly 10% of the SNPs associated with T2D are located within 250KB of a gene linked to diabetes, while some of those SNPs are located within the same region of regulatory networks that are affected by monogenic genes, such as transcription sites of MODY (Figure 6.1) [6]. In obesity, many SNPs have been identified in or near genes that are linked to the monogenic obesity, including *MC4R, SH2B1, BDNF, SIM1* and *SIM1*.

Recent genetics studies have been carried out using a wide range of research strategies, including next-generation sequencing (whole genome sequencing, whole exome sequencing and panel genes sequencing) and high-throughput genotyping. The choice of sequencing approach varies according to the question investigated and type of analysis used.

There are some advantages and disadvantages of each approach. Undoubtedly, the sequencing approach has shown remarkable success in studying monogenic forms of various diseases. However, the cost of sequencing is still high for large cohorts, it requires very high storage of datasets and intensive analysis and many variants identified by WGS cannot be interpreted.

Therefore, one way of reducing the cost of studying the genetics of obesity in a large cohort is to use a genotyping array. In particular, customised genotyping arrays allow an efficient and precise approach to address specific traits or diseases through avoiding or focusing on particular regions of the genome or variants. In addition to the cost effectiveness of the customised array, there are several other advantages such as fully customisable design in terms of variants content, generation of rapid results by providing automated allele calling and high precision, which allows the detection of very rare variants and easier quality assessment.

The technology can be used to address several biological questions such as investigating specific regions or medical conditions of interest, validating and replicating previously identified markers from GWAS or sequencing, investigating the contribution of low-frequency and rare variants which are not sufficiently detected by most current genotyping arrays to address the missing heritability. Therefore, these advances allow the detailed examination of candidate loci, including both rare and common variants (including CNVs), that are related to

obesity and diabetes, on a large number of samples at cost-effective prices and on short timescales.

Hence, in this Chapter, a new approach is proposed to investigate the genetic architecture of obesity and diabetes, through use of a customised genotyping array. Separate strategies were used for the array design to detect (i) rare variants that might cause Mendelian disease; (ii) CNVs; and (iii) common variants identified by GWAS analyses.

Figure 6.1: Overlap between regions of common variants associated with diabetes mellitus, T2DM, or levels of insulin and glucose, and genes and variants relevant to monogenic forms of the disease. (source: Flannick *et al*. (2016) [6]). The first three outer circles indicate the common variants associated withdifferent diabates mellitus traits and the three inner circles show the variants with higher effect that are associated with the different diabetes mellitus traits.. The pink lines indicate the monogenic genes which reported to have association with common and low frequency variants.

6.2 Aims of the study

This Chapter covers the design process of the customised array for obesity and diabetes, through the use of literature review and publically-available databases. The custom array is targeted to include detection of the following:

- Rare predicted-deleterious variants in genes putatively causative of Mendelian (monogenic or syndromic) forms of obesity and diabetes.
- Candidate copy number variations (CNVs) affecting the same genes
- Common SNPs that have been associated with obesity, type-2 diabetes and related traits in GWAS analyses, at genome-wide significance

6.3 Design process

6.3.1 Type of the Customised genotyping array.

6.3.1.1 Choice of the customised genotyping array provider.

Customisation of genotyping array is offered by two companies: Illumina and Thermo Fisher (previously known as Affymetrix). Each company uses a different scientific approach to genotyping, though both are characterised by high performance and efficiency. Nevertheless, the flexibility in the design contents and minimum number of samples per customised array vary between the two companies as Illumina requested a minimum number of samples of approximately 100,000 per customised design array order. Thus, for our study we have used the technology offered by Thermo Fisher, known as the Axiom myDesign genotyping array, which allows the creation of fully flexible, customisable panels, including 1,500 to 2.6 million markers per sample, and requires as few as 480 samples per order,. This is the most economically and efficiently appropriate option for small study designs.

6.3.1.2 Axiom myDesign™ TG Array Plate.

There are several available options/formats of the Axiom myDesign Arrays depending on the desired number of markers. The initial plan of our study was to use the Axiom myDesign 384HT custom array (catalogue number: 000870), which has capacity for up to 50,000 markers. However, as one of the main aims of the study is to include, alongside common variants, rare variants that were not previously included in a previous array, nor in the

available Thermofisher database which includes 11 million validated assays for genetic variants (and for which the assays are, therefore, not pre-optimised – requiring extra probes for confident genotyping) , we were later offered a large-sized array, the Axiom myDesign TG Array (catalogue number: 000842). This array has a capacity of up to 90,000 markers per array, in 96 well array format.

The Axiom array design technology of the variants is based on probe sets, which refers to the combined intensities of one or more sequences of probes to confirm the marker site. Once the two probe sets are applied, the best probe will be added to the recommended probe set list where it can always be used for that particular marker. Thus, the validated markers in the Axiom Genomic Database, which have been previously tested, genotype accurately and produce good clustering and therefore usually require one or two probe sets for the forward and reverse genomic strand, where each probe set is genotyped separately. However, for our type of study, since most of the variants are not validated, the Thermo Fisher design team recommend tiling both strands and having at least two replicates (with up to eight probes for multi-allelic loci).

6.3.2 Specific methods used to design the array

6.3.2.1 Overview of the design for genes and variants.

The overall flowchart of the obesity-diabetes custom array design is shown in Figure 6.2.

To include the most candidate variants related or associated with obesity and diabetes, the array design process was divided into three different variant categories: rare variants, common variants, and CNVs. Each category followed a different methodology as described in details below.

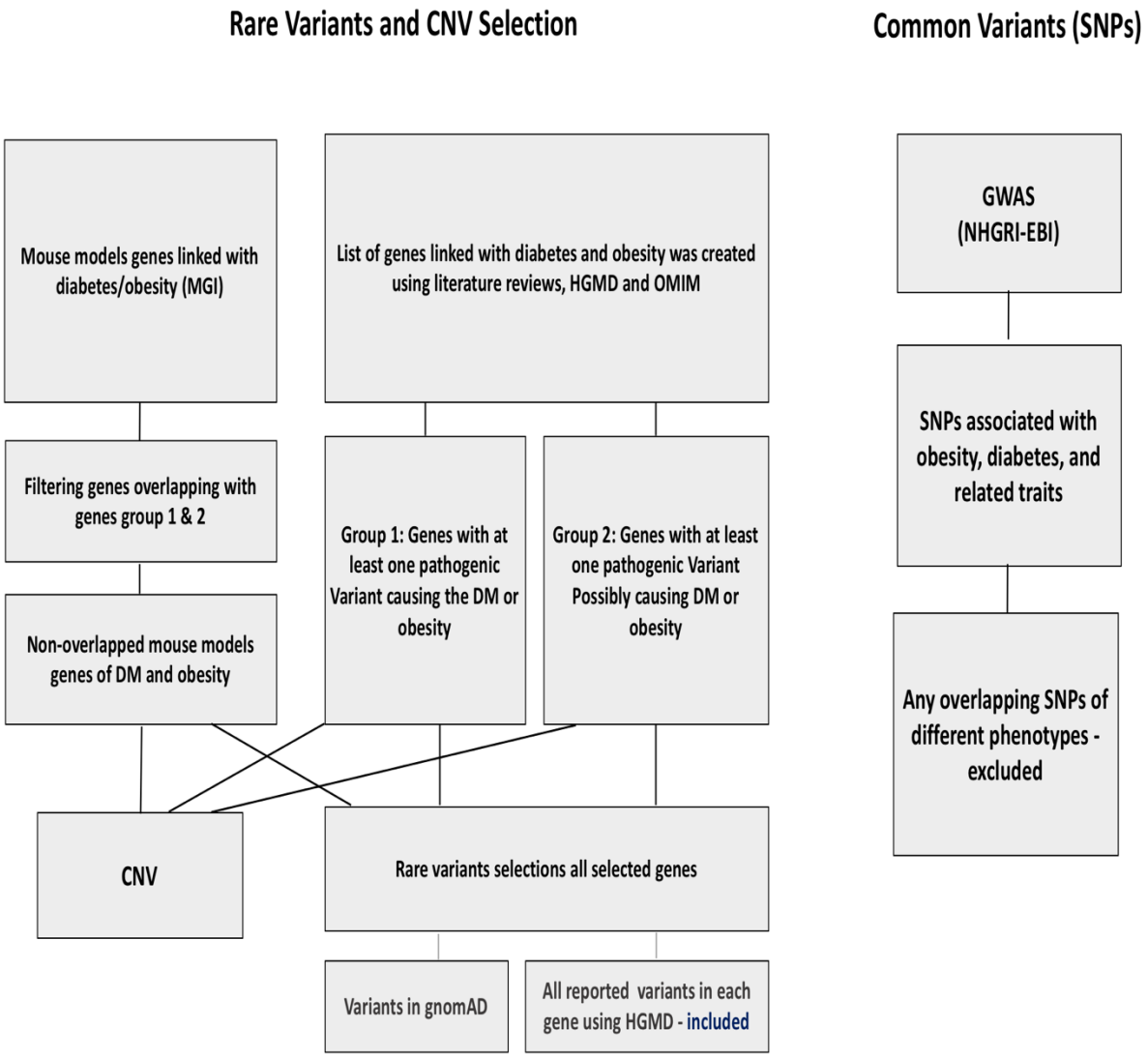


Figure 6.2: The overall flowchart of the obesity-diabetes custom array design. Three categories of variants were included in the customised genotyping arrays: rare variants, common variants and copy number variations. The steps involved in selecting each group of mutation are shown in the grey squares.

6.3.2.2 Selection of candidate genes and rare variants for inclusion on the array

6.3.2.3 Selection of candidate genes.

Selection of candidate gene of obesity and diabetes in Human and mouse.

Prior to the rare variants selection, it was important to generate a list of the candidate genes implicated in obesity and diabetes. The selections of obesity and diabetes genes was as the described list in the methodology chapter (3)/section 3.2. Subsequently, the genes were classified into priority one (group 1), genes that are known to cause obesity or diabetes, and priority two (group 2), genes that possibly cause obesity or diabetes.

6.3.2.4 Selection and prioritisation of sequence variants in the candidate genes.

After completing the list of candidate human and mouse genes of obesity and diabetes, the next step was to identify deleterious variants within these genes. A list of specific types of candidate variants was selected using HGMD and gnomAD databases, a database of variants detected by NGS of over 140,000 individuals.

HGMD was used to select all reported/published variants in each gene on our list of human and mouse genes linked to obesity and diabetes. Due to the high cost of download access to HGMD, the list of reported variants was created manually by going into each variant within the gene and listing the relevant information including chromosome, position, ref and alt allele etc.

The Genome Aggregation Database (gnomAD) was used as another source for selecting

deleterious rare variants in our candidate genes. This database includes data from a wide variety of large-scale sequencing projects of around 140,000 independent individuals in total from several population genetic studies; 125,748 WES and 15,708 whole genome sequences. It has been widely used as a reference data set for assessing the frequency of rare variants (as a part of a protocol of pathogenicity prediction in clinical work) and also in predicating the enrichment of rare variants within a gene. Thus, this database is considered an important source of studying rare variants. To our knowledge it has never been used to design a customised genotyping array as presented here.

The data set of gnomAD is very large, and it was necessary to prioritise variants for inclusion on the array : only specific variants were selected in each of the human and mouse model genes, including missense, frameshifts and stop-gain variants. For the non-synonymous/missense variants, further filtering was applied using three different risk prediction tools: SIFT, PolyPhen, and CADD, to select the most deleterious variants [165-167]. The tools were applied differently based on the gene group priority, as exhibited in Table 6.1. Genes in priority one (group one) are genes known to cause obesity or diabetes and, genes in priority two (group two) are genes that possibly cause obesity or diabetes.

Table 6.1: Types of missense filtering applied for each genes groups of the array.

Array gene group	gnomAD Missense variants selection
Obesity and diabetes genes group 1/priority one	CADD score> 15, predicted to be pathogenic in Silico by either SIFT or Polyphen
Obesity and diabetes genes group 2/priority two	CADD score> 20, predicted to be pathogenic in Silico by either SIFT or Polyphen
Mouse Model genes of obesity and diabetes	CADD score> 20, predicted to be pathogenic in Silico by either SIFT or Polyphen

Missense variants obtained from gnomAD database were selected and assessed based on the risk prediction tools including SIFT, Polyphen and, CADD. Different selection criteria was applied to each group of gene: group 1, group 2 and mouse model genes.

6.3.2.5 Selection of CNV

Another very distinct phase in the array design was the addition of known copy number variants (CNVs) of obesity and diabetes genes. CNVs represent an important category of genetic variation that contributes to various types of disease. Several databases have been established to catalogue and characterise CNVs, such as the database of genomic variants, Decipher, and the Database of Genomic Variants (DGV), and CNVs are also included in HGMD. For the array design, important CNVs in the candidate genes of obesity and diabetes were included using the HGMD database. Only those reported to have functional significance or to contribute to disease have been included.

Additionally, we have included other potential CNV regions that that were identified by Dr Nikman Adli Bin Nor Hashim, a former PhD student of Professor Alex Blakemore, as potentially causative of obesity in his work on the Northern Finnish Birth Cohorts of 1966 and 1986 (unpublished data)..

After the list of genes was created, several steps were applied to create the final list of CNVs, demonstrated in Figure 6.3. This involved going through each of the published articles to identify the start and end position of the CNV in each gene, converting them into the genome build GRGh38 and excluding the ones with unclear locations. Since the array space is limited in terms of number of probes to be used for CNVs, and we did not want to eliminate any of the CNVs from the array, we introduced a strategy to minimise the number of probes used for the CNV detection while maintaining the overall number of CNVs. It was performed through integrating CNVs which have overlapping positions into a single region, using the farthest start position of the integrated CNVs as a start point, and the furthest position of

integrated CNVs as an end point. From this, we identified the major overlap region which represents the overlap among all the CNVs (and includes key genes of interest). The major overlap region arranged to have sufficient resolution/probes compared to the rest of the integrated region. As an example shown in Figure 6.4, for the CNVs in the *SIM1* gene, there were 14 CNV regions integrated into a single region start, referred to as the major overlap region.

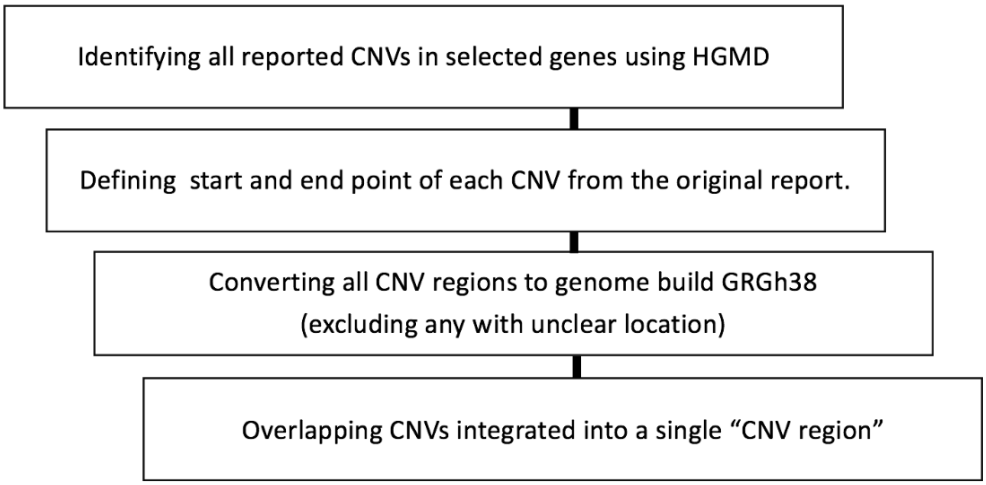


Figure 6.3: A representation of the steps in the CNV regions selection. The Figure shows the steps taken in creating the list of CNVs in the obesity and diabetes related genes

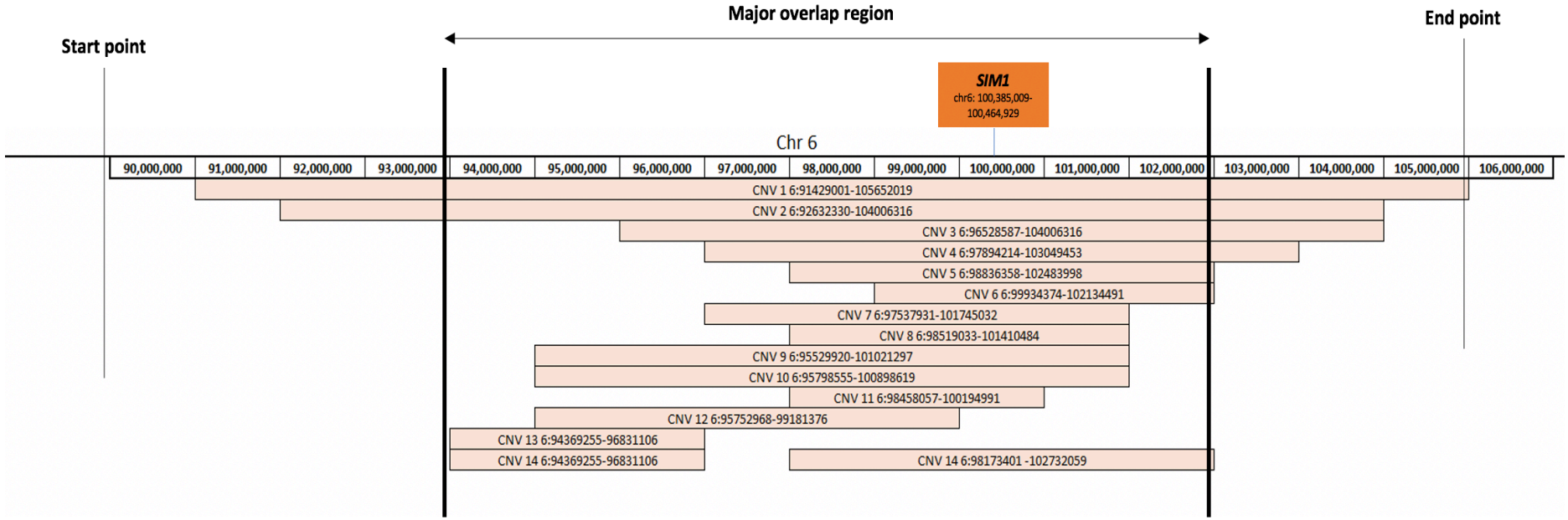


Figure 6.4: Integration of 14 CNV regions in the *SIM1* gene into a single region named: the major overlap region. Each orange color bar represents a CNV, the start point is defined by the farthest start position of the integrated CNVs, the end point is defined by furthest position of integrated CNVs and the major overlap region represents the of overlap among all the CNV

6.3.2.6 Strategy for the selection of common variants for inclusion on the array

To examine the contribution of common variants and to perform other follow-up studies of them, SNPs associated with diabetes, obesity and related phenotypes with a genome-wide significant p-value of $\leq 5X10^{-8}$ were selected. For the common variant selection, the GWAS EMBL-EBI catalogue was used, employing a phenotype-based search of the words 'obesity' and 'diabetes'. Any duplicate SNPs and unrelated SNP associations were removed. In addition, the reported SNPs by Pedersen et al. (2016), which are associated with diabetes remission after bariatric surgery, were added [316].

Furthermore, SNPs from the following studies were also included in the array: anti-psychotic weight gain [317-321] diabetes remission after bariatric surger, percent weight loss after RYGB [322] and, Type 2 diabetes Diagram Consortium (Diabetes Genetics Replication And Meta-analysis) [323]

6.4 Results

6.4.1 Arrays genes list.

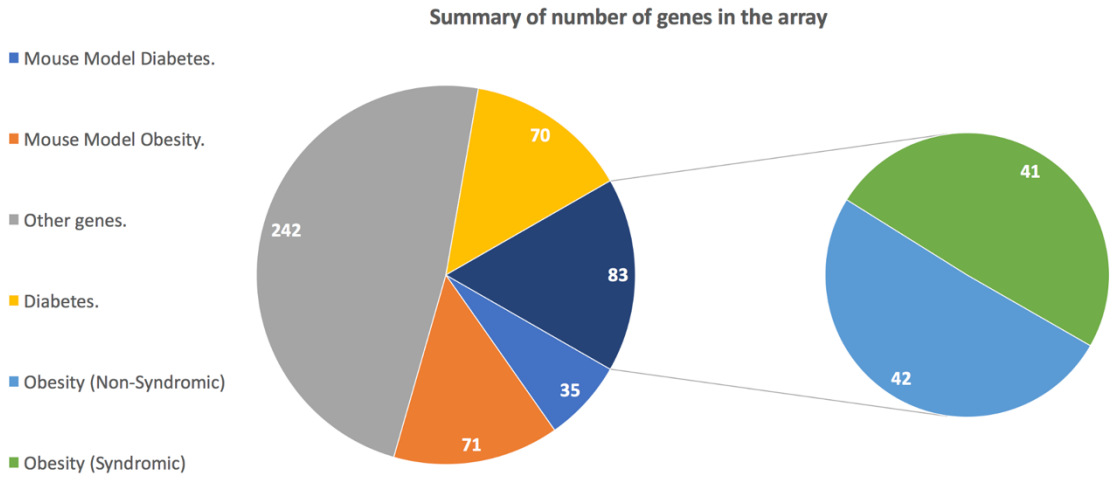
Using several resources to identify genes that have at least one pathogenic or possibly pathogenic variant causing obesity and diabetes, a total of 153 genes were identified. For the purpose of the array, the genes were classified into two priorities: priority one: genes known to cause obesity or diabetes, and priority two, genes that possibly cause obesity or diabetes.

A total of 83 human genes linked to obesity were identified; 42 linked to non-syndromic obesity and 41 to syndromic obesity. Sixty-eight of them had obesity-causing variants (priority 1 genes) and 15 genes were considered to possibly cause obesity (priority 2 genes).

For the diabetes related content of human genes, a total of seventy genes have been found, 58 of them having pathogenic variants thought to cause rare types of diabetes or T2D (priority 1 genes), and the other 12 genes possibly causing diabetes (priority 2 genes). Note additional genes linked to metabolic syndrome or other research purposes or collaboration, were annotated under priority 2 genes. A summary of the identified human obesity and diabetes genes are listed in Table 6.2-6.3

A list of genes from the mouse models with obesity and diabetes was created using the MGI database and literature. A total 71 of genes had been found to be related to rodent obesity, and another 35 were related to diabetes in mouse models . The mouse model genes were grouped under priority 2, demonstrated in Table 6.4 and 6.5

An overall summary of the custom array genes and variants groups is shown in **Figure 6.5**



	Diabetes		Obesity		Mouse Model Obesity
	Group 1	Group 2	Group 1	Group 2	
Genes	59	12	61	15	132
HGMD variants	4931		2587		1611
gnomAD Stop gained variants	587	120	455	125	803
gnomAD Framshift variants	828	140	509	99	1545
gnomAD missense	10496	1404	10732	855	10052

Figure 6.5. An overall summary of the custom array genes and variants groups. (A) Pie chart represents the different genes group and number of genes under each group included in the array. **(B)** Overview of the total number of variants in obesity and diabetes related genes identified in human and rodent models. Priority one refers to the genes with variants which cause the diseases and priority two refers to the genes that have variants possibly causing the diseases.

Table 6.2: Overview of the identified human obesity genes and their selected variants for the array.

GENE	HGMD VARIANTS	MISSENSE GNOMAD	FRAMESHIFT GNOMAD	STOP GAINED GNOMAD	GENE	HGMD VARIANTS	MISSENSE GNOMAD	FRAMESHIFT GNOMAD	STOP GAINED GNOMAD	GENE	HGMD VARIANTS	MISSENSE GNOMAD	FRAMESHIFT GNOMAD	STOP GAINED GNOMAD
ADIPOQ	21	67	10	10	*FAM3C*	0	46	4	3	*PCSK1*	36	213	0	13
AFF4	3	286	7	2	*FOXA3*	7	95	8	0	*POGZ*	40	351	5	2
ALMS1	262	1072	125	103	*GHSR*	17	123	7	6	*POMC*	36	98	12	5
ARL14EP	2	60	2	0	*GNB3*	8	124	15	4	*POU3F2*	0	60	0	0
ARL6	17	54	6	3	*HDAC8*	34	9	2	1	*PRKD1*	4	175	9	17
BBIP1	1	8	8	4	*IFT27*	1	50	6	5	*PTEN*	510	30	5	4
BBS1	88	222	17	12	*IGSF1*	3	131	12	7	*PYY*	7	36	3	2
BBS10	98	135	39	16	*INPP5E*	0	201	13	8	*RAB23*	0	99	6	4
BBS11	0	206	16	10	*KSR2*	28	222	6	0	*RAI1*	40	505	8	6
BBS12	59	154	29	17	*LEP*	19	38	4	0	*RBMX*	0	6	0	2
BBS13	0	185	21	18	*LEPR*	32	229	21	9	*RETN*	3	20	5	2
BBS2	87	252	24	18	*LRP2*	28	0	0	0	*RORA*	0	93	9	11
BBS4	37	182	27	20	*LZTFL1*	3	46	10	9	*SDCCAG8*	17	245	35	18
BBS5	24	144	18	6	*MAGEL2*	13	193	3	3	*SETD2*	9	614	11	13
BBS7	33	209	28	12	*MANF*	1	21	6	1	*SH2B1*	11	232	10	7
BBS9	35	284	35	27	*MBD5*	14	502	3	1	*SIM1*	26	295	5	4
BDNF	8	74	11	4	*MC3R*	28	145	12	9	*SLC35D3*	2	67	8	6
CADM2	1	46	2	1	*MC4R*	158	162	17	7	*TAOK2*	0	386	16	5
CARTPT	2	36	3	1	*MCHR1*	10	198	18	7	*TMEM18*	1	24	12	7
CEP19	1	57	6	5	*MIR137*	2	0	0	0	*TTC8*	4	174	10	9
CEP290	250	572	97	59	*MKKS*	57	158	36	32	*TUB*	1	226	17	5
CLMP	9	69	7	10	*MRAP2*	5	77	4	3	*UCP1*	7	78	13	13
COA3	2	41	4	5	*MYT1L*	2	157	1	1	*UCP3*	9	115	11	5
CPE	3	132	4	4	*NMB*	4	23	4	3	*VPS13B*	0	0	101	73
CXORF36	1	25	3	4	*NPY2R*	9	62	7	4	*WDPCP*	7	198	28	12
DNAAF1	23	147	37	18	*NPY4R*	2	141	18	13					
DYRK1B	2	130	8	4	*NTRK2*	0	145	1	0					
FAAH	4	158	13	8	*NUCB2*	2	69	17	9					

Table 6.3: Overview of the identified human diabetes genes and their selected variants for the array.

GENE	HGMD VARIANTS	MISSENSE GNOMAD	FRAMESHIFT GNOMAD	STOP GAINED GNOMAD	GENE	HGMD VARIANTS	MISSENSE GNOMAD	FRAMESHIFT GNOMAD	STOP GAINED GNOMAD	GENE	HGMD VARIANTS	MISSENSE GNOMAD	FRAMESHIFT GNOMAD	STOP GAINED GNOMAD
ABCC8	591	434	23	30	*HMGA1*	1	74	63	12	*PTF1A*	**10**	**68**	**6**	**1**
ACACB	6	644	60	50	*HNF1A*	496	203	13	19	*PTPRD*	**6**	**937**	**8**	**5**
AKR1B1	4	165	30	4	*HNF1B*	158	177	2	1	*RFX6*	**21**	**245**	**12**	**10**
AKT2	2	107	10	13	*HNF4A*	143	111	4	16	*SLC19A2*	**46**	**167**	**13**	**6**
APOE	57	98	10	7	*IAPP*	4	18	6	3	*SLC2A2*	**77**	**131**	**20**	**9**
APPL1	2	224	15	6	*IER3IP1*	3	46	2	4	*SLC2A4*	**3**	**126**	**9**	**14**
BLK	15	260	24	14	*INS*	70	22	2	2	*SREBF1*	**0**	**291**	**24**	**7**
CASR	389	249	7	6	*INSR*	161	291	12	17	*TBC1D4*	**2**	**473**	**26**	**30**
CAT	18	246	12	9	*IRS1*	23	453	6	8	*TRMT10A*	**5**	**92**	**16**	**13**
CDKAL1	0	81	9	12	*IRS2*	12	112	4	1	*UCP2*	**7**	**131**	**15**	**13**
CEL	10	235	58	16	*KCNJ11*	171	129	6	3	*WFS1*	**343**	**499**	**32**	**34**
CISD2	3	24	2	0	*KLF11*	5	190	21	13	*ZFP57*	**13**	**76**	**20**	**9**
DNAJC3	0	117	7	7	*KRT17*	18	187	15	12					
EIF2AK3	69	247	14	14	*LIPE*	6	351	43	23					
FN3K	0	124	13	12	*MAPK8IP1*	2	205	5	2					
FOXA2	3	139	3	3	*MNX1*	57	71	10	6					
FOXP3	81	16	0	2	*MTNR1B*	30	136	9	10					
GATA6	65	141	2	1	*MTTP***	64	153	23	12					
GCGR	0	108	17	20	*NEUROD1*	15	114	2	3					
GCK	770	79	5	3	*Neurog3*	9	84	10	5					
GH1	85	73	5	2	*NKX2.2*	0	77	3	2					
GIPR	0	134	26	20	*PASK*	2	382	66	28					
GLIS3TV1	0	0	0	0	*PAX4*	12	130	8	9					
GLP1R	10	107	6	6	*PCBD1*	11	34	3	4					
GLUD1	37	103	7	5	*PDX1*	28	87	11	3					
GPD2	4	208	28	12	*PLAGL1*	1	87	5	3					
HADH	24	104	20	6	*PPARA*	7	82	1	6					
HFE	56	72	8	10	*PPARG*	49	357	42	19					
HK2	6	348	13	5	*PPP1R3A*	4	236	37	29					

Table 6.4: Overview of the identified mouse model obesity related genes and their selected variants for the array.

GENE	HGMD VARIANTS	MISSENSE GNOMAD	FRAMESHIFT GNOMAD	STOP GAINED GNOMAD	GENE	HGMD VARIANTS	MISSENSE GNOMAD	FRAMESHIFT GNOMAD	STOP GAINED GNOMAD	GENE	HGMD VARIANTS	MISSENSE GNOMAD	FRAMESHIFT GNOMAD	STOP GAINED GNOMAD
ACOT11	0	146	52	17	***GUCY2C***	0	324	34	30	***PROX1***	*1*	*80*	*3*	*1*
ADCY3	1	179	14	13	***HCRT***	2	16	8	1	***ROCK1***	*0*	*275*	*2*	*4*
ADIPOR2	0	28	10	1	***HDC***	0	209	3	7	***ROCK2***	*2*	*0*	*0*	*0*
ADRB2	8	58	6	1	***HMGA2***	3	9	4	3	***RSC1A1***	*0*	*79*	*26*	*15*
ADRB3	4	79	21	5	***HRH1***	0	122	6	11	***SCD***	*1*	*46*	*2*	*0*
AGRP	5	28	8	4	***HRH3***	0	102	1	5	***SDC3***	*3*	*102*	*5*	*1*
AHSG	4	85	4	12	***ICAM1***	0	80	8	5	***SOCS3***	*4*	*16*	*0*	*0*
ANGPTL6	0	106	14	11	***IL18***	5	3	2	2	***SST***	*0*	*18*	*0*	*0*
ANKRD26	20	199	64	44	***IL6***	14	28	12	4	***STAT3***	*130*	*34*	*0*	*0*
APOC3	17	15	9	0	***IRX3***	1	59	34	5	***TLR5***	*0*	*182*	*38*	*27*
AR	552	49	4	2	***KDM3A***	2	184	12	7	***TYK2***	*14*	*224*	*18*	*20*
ASIP	1	24	1	2	***MYD88***	7	54	18	4	***UBB***	*0*	*0*	*12*	*4*
ATXN2	0	435	36	8	***NCOA1***	1	195	5	4					
BRS3	0	31	6	1	***NEIL1***	5	99	28	7					
CCL2	3	7	4	1	***NMU***	2	29	6	2					
CLOCK	4	117	16	4	***NMUR2***	0	105	12	8					
CNR1	5	55	6	4	***NPY1R***	2	45	4	4					
CORIN	0	417	43	0	***NPY5R***	0	57	14	5					
CRTC1	0	199	1	2	***NTSR1***	0	215	13	8					
ESR1	0	176	0	2	***OTP***	0	128	1	1					
ESRRA	1	42	10	0	***PAM***	0	385	21	9					
FABP2	2	17	10	3	***PEG3***	0	387	26	20					
FABP4	0	43	4	2	***PIK3C2G***	2	281	61	36					
FABP5	0	24	4	1	***PPARGC1A***	4	9	4	5					
FEN1	0	154	6	8	***PPARGC1B***	3	239	36	10					
FOXO1	0	68	0	0	***PRKAA2***	2	86	24	12					
FTO	12	110	28	16	***PRKAB2***	0	66	3	2					
GPR12	0	47	4	3	***PRKAG1***	2	49	13	14					
GPR45	0	126	4	3	***PRLH***	0	34	3	2					

Table 6.5: Overview of the identified mouse model diabetes related genes and their selected variants for the array.

GENE	HGMD VARIANTS	MISSENSE GNOMAD	FRAMESHIFT GNOMAD	STOP GAINED GNOMAD	GENE	HGMD VARIANTS	MISSENSE GNOMAD	FRAMESHIFT GNOMAD	STOP GAINED GNOMAD
ARHGEF11	3	341	26	5	***SLC5A2***	84	223	40	13
ARNTL	0	61	12	3	***SNAP25***	3	10	0	1
CAPN10	5	141	38	22	***TBC1D1***	5	271	39	24
CTF1	2	18	8	2	***TCF7L2***	4	70	16	4
CYB5R4	2	75	10	9	***TGM2***	9	166	24	18
ENPP1	71	151	22	17	***TP53INP1***	0	41	7	5
FEM1B	0	52	2	3	***VDR***	63	77	8	5
FOXM1	1	181	17	5					
GADD45GIP1	0	43	10	7					
GHR	0	143	16	9					
HMOX1	4	85	12	4					
HTR2C	6	20	1	5					
IFNGR2	0	12	0	1					
IGF2	7	42	2	3					
IGF2BP2	0	72	4	6					
IL1R1	0	85	4	6					
LIPC	19	156	19	11					
MADD	3	432	33	28					
MAFA	0	34	1	3					
NFE2L1	0	123	18	2					
OAS1A	0	97	19	18					
PHOX2A	5	40	1	0					
PPARD	2	73	6	1					
PPP1R3C	2	75	10	8					
PRCP	1	105	15	17					
PRKCI	0	57	11	5					
PTPN1	1	28	14	2					
PTPN22	0	127	0	17					
SIRT1	19	98	9	7					

6.4.2 Rare variants of the selected genes.

The HGMD was the main initial source for identifying rare variants. All variants in HGMD of each of the selected genes were selected irrespective of their phenotype or degree of pathogenicity. The total number of variants in diabetes genes was 4,931, with 2,706 being diabetes-causing/possibly causing variants. On the other hand, 2,587 variants were identified in the obesity genes and 1,237 of those variants are annotated as either obesity-causing/possibly causing variants. For the mouse model genes, 1611 variants have been identified.

Overall the majority of the HGMD variants, as demonstrated in Figure 6.6, are missense/nonsense variants representing single base-pair substitutions in the coding region. Most of the variants belong to the DM variant category of HGMD - a representation of the selected HGMD variants subdivided into the different variant classes is shown in Figure 6.7.

To expand the list of rare variants, gnomAD was used to select specific variants in each group of genes. A summary of the number of variants selected in each group is shown in Table 6.2-6.5 .

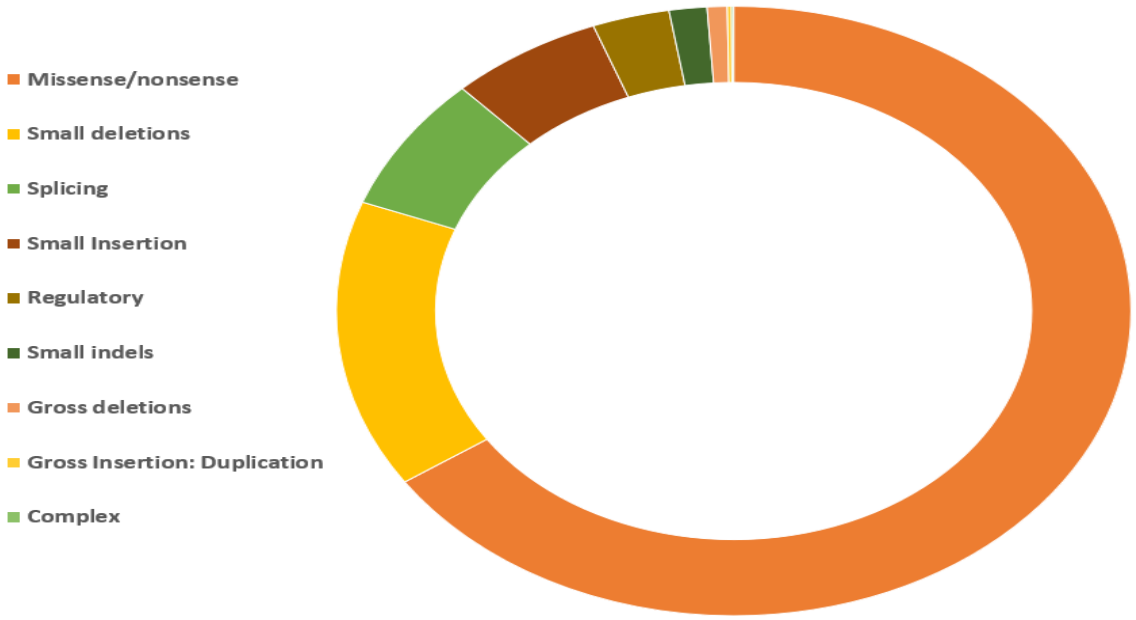


Figure 6.6: The types of variants in diabetes and obesity genes shown in percentage. Missense/nonsense: Single base-pair substitutions in coding regions. Complex: complex arrangement variant as extremely variable quality of the original data reported. Regulatory: Substitutions causing regulatory abnormalities. Gross insertion/deletion: 20 bp or more insertion/deletion. Splicing: Mutations with consequences for mRNA.

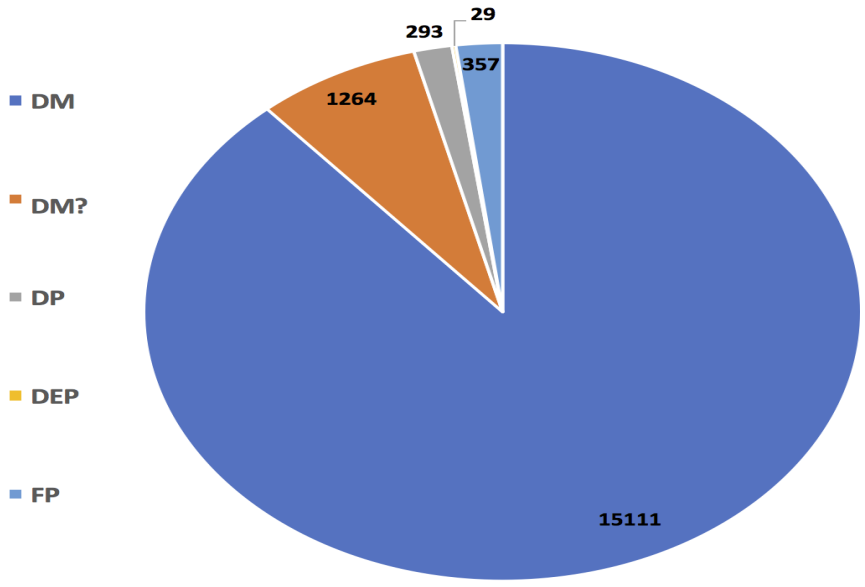


Figure 6.7: An overview of the total selected variants in HGMD subdivided into the HGMD variants classes. Definition of each classes is represented in Table 3.2.

6.4.3. Common variants.

A total of 874 SNPs were associated with diabetes, obesity and related phenotypes with a genome-wide significance level of ≤ 5X10-8. Figure 6.8 represents an overview of the total SNPs associated with diabetes, obesity and related traits. Some of the SNPs show association with multiple phenotypic/diseases, which indicates the importance of not only including SNPs associated with the diseases but also phenotypes/traits related to the disease of interest to highlight significant associations and loci.

6.4.4 CNVs

Another new feature of the arrays involved the inclusion of CNVs. The original list of 458 CNV regions was eventually shortened to 265 CNV regions. However, for the integrated CNV regions whose major overlap region required more than 1,000 probes, we had to shorten the region further to include only the exons of the genes. This strategy maintains the coverage of the key region of the major overlap region. The full list of the identified CNV major overlap regions is shown in Table 6.6.

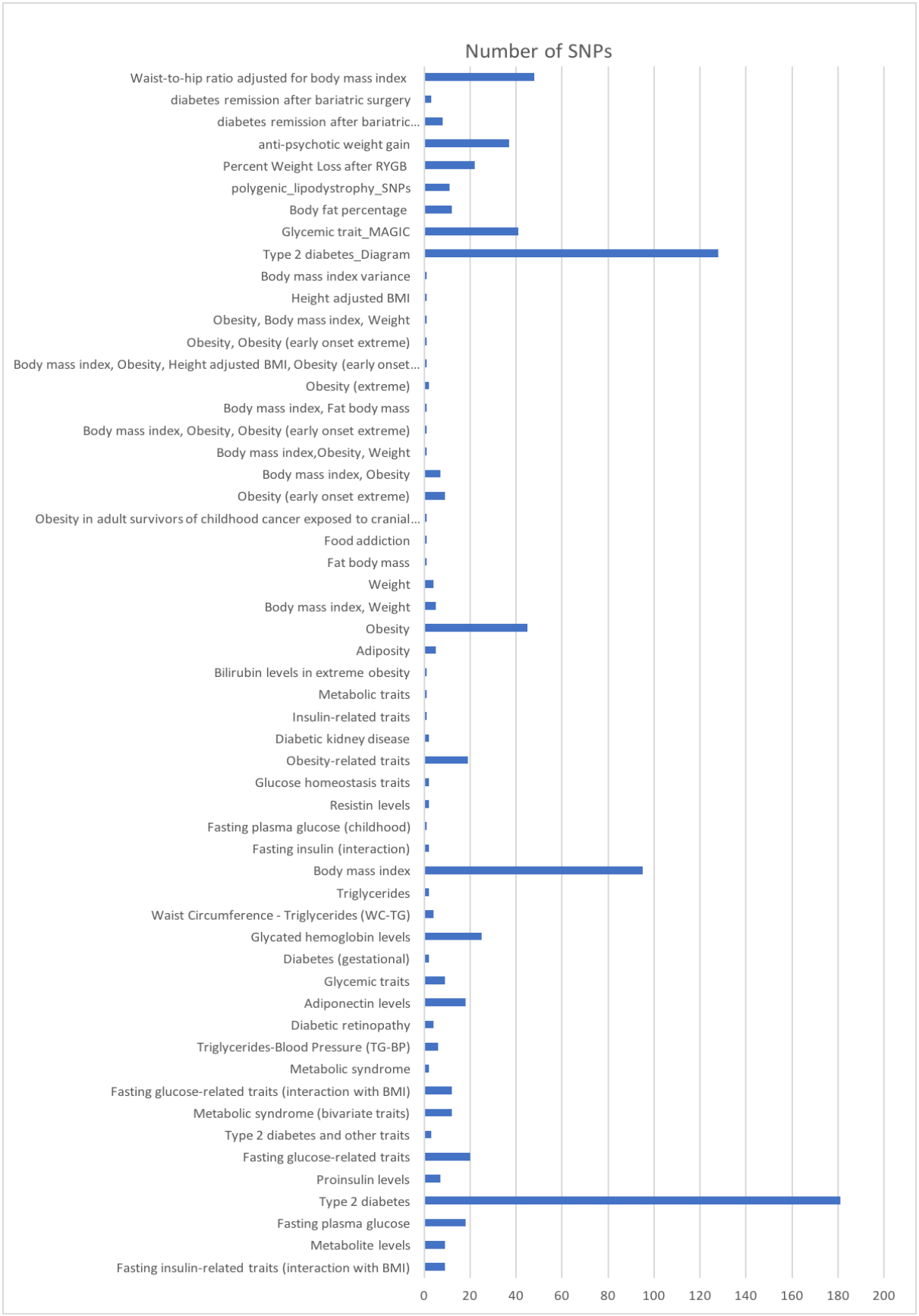


Figure 6.8: A bar chart representing the number of SNPs in obesity and diabetes and related traits.

Table 6.6: A summary of the identified CNVs in each gene and the created critical region which they belong to.

Main gene	Chr	Critical Region	Name	CNVs regions within Integrated large CNV	Reference
BDNF (22880391-31780895;8.9Mb)	**chr11**	**11:27,654,893-27,722,058**	**BDNF.chr11.Obesity**	chr11:27654893-27722058	PubMed 18753648
				chr11:31784792-31817961	PubMed 18753648
				chr11:24650080-31284456	PubMed 18753648
				chr11:27279397-29612514	PubMed 18753648
				chr11:26133522-31780895	PubMed 18753648
				chr11:24650080-31284456	PubMed 18753648
				chr11:27072499-29571990	PubMed 18753648
				chr11:23506076-27879805	PubMed 18753648
				chr11:23024064-27978597	PubMed 18753648
				chr11:27279397-29612514	PubMed 21567907
				chr11:26133522-31780895	PubMed 21567907
				chr11:24650080-31284456	PubMed 21567907
				chr11:27072499-29571990	PubMed 23044507
				chr11:23506076-27879805	PubMed 23044507
				chr11:23024064-27978597	PubMed 23044507
				chr11:25670994-31588476	PubMed 23044507
				chr11:22880391 --29088197	PubMed 23044507
				chr11:27589482-27598519	PubMed 24643514
				chr11:27201781-28859543	PubMed 23044507
SIM1 (95552124-101552124; 6Mb)	**chr6**	**6:100,385,009-100,464,929**	**SIM1.chr6.Obesity**		PubMed 18925680
				chr6:97894214 -103049453	PubMed 25351778
				chr6:95798555 -100898619	PubMed 25351778
				chr6:97537931 -101745032	PubMed 25351778
				chr6:96528587 -104006316	PubMed 25351778
				chr6:98519033 -101410484	PubMed 25351778
				chr6:99934374 -102134491	PubMed 25351778
				chr6:92632330 -104006316	PubMed 25351778
				chr6:95752968 -99181376	PubMed 25351778
				chr6:91429001 -105652019	PubMed 25351778
				chr6:98836358 -102483998	PubMed 25351778
				chr6:94369255-96831106 + chr6:98173401 -102732059	PubMed 25351778
				chr6:94369255-96831106	PubMed 25351778
				chr6:95529920-101021297	PubMed 25351778
				chr6:98458057- 100194991	PubMed 25351778
HDAC8 (72329516-73214492; 884977)	**chrX**	**X:72,329,516-72,573,103**	**HDAC8.chrX.Obesity**	chrX:72527131-73038566	PubMed 24403048
				chrX:72329454-72331280	PubMed 24403048
				chrX:72350240-72535599	PubMed 24403048
				chrX:72462003-73214492	PubMed 24403048
				chrX:72573022-72667386	PubMed 24403048
				chrX:72512183-72731334	PubMed 24403048
				chrX:72371425-72492425	PubMed 24403048

IGSF1 (131176931-131864792;328449)	**chrX**	**X:131,273,506-131,578,899**	**IGSF1 .chrX.Obesity**	chrX:131252293-131378028	PubMed 23143598
				chrX:131203600 -131864792	PubMed 25052774
LEP	**chr7**	**7:128,241,202-128,257,629**	**LEP.chr7.Obesity**	chr7:7:128213154 -128248822	PubMed 23275889
				chr7:7:128213154 -128251076	PubMed 23275889
LEPR	**chr1**		**LEPR.chr1.Obesity**	chr1:65592657-65598804	PubMed 25751111
				chr1:64472675 -67646613	PubMed 21416589
PCSK1 (96279156-96935983; 656828)	**chr5**	**5:96,390,336-96,434,143**	**PCSK1.chr5.Obesity**	chr5:96279156-96935983	PubMed 23800642
				chr5:96279156 -96790605	PubMed 22665139
BBS1 (66510609-66523882;13274)	**chr1**	**11:66,510,606-66,533,627**	**BBS1.chr1.Obesity**	chr11:66510609 -66523882	PubMed 24746959
BBS2 (56502808-56501353; 1456)	**chr16**	**16:56,466,836-56,520,283**	**BBS2.chr16.Obesity**	chr16:56502673-56,501,353	PubMed 20177705
ARL6 (97764521-97798023; 33503)	**chr3**	**3:97,762,581 to 97,812,585**	**ARL6.chr3.Obesity**	chr3:97788120-97798023	PubMed 19858128
				chr3:97769347 -97769348	PubMed 21642631
				chr3:97764521-97801107	PubMed 22773737
BBS4 (72709700-72727994;18294)	**chr15**	**15:72,686,179-72,738,476**	**BBS4.chr15.Obesity**	chr15:72709700-72715402	PubMed 11381270
				chr15:72727994-72716850	PubMed 15666242
				chr15:72727994-72716850	PubMed 20177705
				chr15:72722794-72727994	PubMed 20177705
BBS5 (169492874-169506655; 13782)	**chr2**	**2:169,479,178-169,506,655**	**BBS5.chr2.Obesity**	chr2:169492874-169493836	PubMed 21344540
				chr2:169499486-169506655	PubMed 16380913
BBS7 (121833395-121870278;36885)	**chr4**	**4:121,824,329-121,870,497**	**BBS7.chr4.Obesity**	chr4:121870278-121861679	PubMed 16380913
				chr4:121833395-121848844	PubMed 16877420
BBS9 (33177478-33606068; 428591)	**chr7**	**7:33,129,244-33,635,769**	**BBS9.chr7.Obesity**	chr7:33533954-33606068	PubMed 20120035
				chr7:33177478-33534176	PubMed 21344540
					PubMed 24400638
				chr7:33257236-33264374	PubMed 22353939
				chr7:33273012-33273956	PubMed 20177705
					PubMed 20177705
				chr7:33287572-	PubMed 22912587
TRIM32 (116685531-117021623; 336093)	**chr9**	**9:116,687,302-116,701,300**	**TRIM32.chr9.Obesity**	chr9:116685587 -116809984	PubMed 25351777
				chr9:116685531 -117021623	PubMed 25351777
				chr9:116687302-116701300	PubMed 23541687
					PubMed 19492423
CEP290 (86595095-88432959;1837865)	**chr12**	**12:88,049,013-88,142,216**	**CEP290.chr12.Obesity**	chr12:86595095-88432959	PubMed 23954617
SDCCAG8 (243286272-243304777;18506)	**chr1**	**1:243,255,419-243,500,09**	**SDCCAG8.chr1.Obesity**	chr1:243286272-243304777	PubMed 20835237
ALMS1 (73455162-73599521;144360)	**chr2**	**2:73,385,758-73,610,793**	**ALMS1.chr2.Obesity**	chr2:73572262-73573424	PubMed: 21877133 P
PTEN (86755786-89575528; 1828074)	**chr10**	**10:87,863,113-87,971,930**	**PTEN.chr10.Obesity**	chr1010:87518427 -89575528	PubMed 25288137
				chr10:87925513- 87933251	PubMed 20848651
				chr10:87863438 -87971930	PubMed 12844284
				chr10:86755786-87863438)	PubMed 22673385
				chr10:87863113- 87864548	PubMed 12844284
				chr10:87863113- 87933251	PubMed 12844284
				chr10:87863113- 87971930	PubMed 9286463
				25 bp nt. 520 cd. 174*	PubMed 9467011
				75 bp cd. 121-145*	PubMed 9699651

chr10:87863438 -87971930	PubMed 22382802
chr10:87863438 -87971930	PubMed 22382802
chr10:87863438 -87971930	PubMed 22382802
chr10:87952118-87971930	PubMed 22382802
chr10:87925513-87971930	PubMed 22382802
chr10:10:87503263 -88041263	PubMed 21926107
chr10:87863438 -87971930	PubMed 18456716
chr10:87518427 -89575528	PubMed 23132533
chr87894025-87958019	PubMed 23512313
chr87863438-87971930	PubMed 18456716
chr10:87659719- 87971930	PubMed 18456716
chr10:87659719-88583860	PubMed 18456716
chr10:86755786-87863438	PubMed 20600018
chr10:87894025-87958019	PubMed 20600018
>107447bp incl entire gene	PubMed 21194675
chr10:87863113- 87952259	PubMed 21194675
chr10:87863113- 87933251	PubMed 21194675
chr10:87894025-87958019	PubMed 21194675
chr10:87952118-87958019	PubMed 21194675
>107447bp incl entire gene; chr10:87863438-87971930	PubMed 25669429
chr10:87863113- 87958019	PubMed 25669429
chr10:87863113- 87971930	PubMed 9491322
chr10:87894025- 87971930	PubMed 25669429
chr10:87925513- 87933251	PubMed 25669429
chr10:87952118- 87952259	PubMed 25669429
chr10:87952118- 87971930	PubMed 24609522
chr10:87957853 87958019	PubMed 25669429
chr10:87960894 87961118	PubMed 25669429
chr10:87859161-87971930	PubMed 21671387
chr10:87894025- 87971930	PubMed 23399955
chr10:87863113- 87952259	PubMed 23399955
chr10:87952118- 87952259	PubMed 23399955
32bp c.955_986*	PubMed 23335809
chr10:87863438-87971930	PubMed 23335809
chr10:87863438-87971930	PubMed 23335809
chr10:87894025- 87971930	PubMed 22266152
chr10:87952118- 87971930	PubMed 22266152
chr10:87863113- 87952259	PubMed 23335809
chr10:87894025-87958019	PubMed 24778394
chr10:87863438- 87952259	PubMed 16287957
18595bp incl ex. 6	PubMed 24375884
36 bp within in. 7	PubMed 9688299
chr10:87863113- 87933251	PubMed 22382802
chr10:87863113- 87952259	PubMed 25669429
chr10:87952118- 87952259	PubMed 25669429
chr10:87863438- 87952259	PubMed 22595938

				c.153T>C and c.163A>C, p.Asp51Asp and p.Arg55Arg.*	PubMed 22320991
				chr87863113- 87971930	PubMed 23335809
PYY (43990147-43997159;7013)	**chr17**	**17:43,952,733-44,024,260**	**PYY.chr17.Obesity**	chr17: 43952733-44024260	
GCK (44143213-44189439;46227)	**chr7**	**7:44,143,213-44,198,170**	**GCK.chr7.Diabetes**	chr7:44143213- 44189439	PubMed 23074679
				chr7:44158073-44189423	PubMed 19790256
				chr7:44158073-44145280	PubMed 22060211
				chr7:44144271-44145730	PubMed 19790256
				chr7:44149969-44150064	PubMed 19790256
ABCC8 (17,392,498-17,476,879)	**chr11**	**11:17,392,498-17,476,879**	**ABCC8.chr11.Diabetes**	chr11:17476629-17453283	PubMed 21978130
				chr11:17428670- 17428565	PubMed 21378087
				chr11:17476629-17410653	PubMed 23345197
				chr11:17410653-17410516	PubMed 20685672
				chr11:17448671 -17,48516	PubMed 20685672
				chr11:17415304-17412746	PubMed 23275527
				chr11:17443178-17432244	PubMed 23771172
HNF4A (44355700-44434596;77044)	**chr20**	**20:44,355,700-44,434,596**	**HNF4A.chr20.Diabetes**	chr20:120978543-120997665	PubMed 23348805
				chr20:120,978,515 to 121,002,512	PubMed 21105491
HNF1B (37686431-37745078;58648)	**chr17**	**17:37,686,431-37,745,247**	**HNF1B.chr20.Diabetes**	chr17:36476549-37883408 (according to UCSC)	PubMed 24835530
				Entire gene (37,686,431-37,745,078)	PubMed 18411231
				Entire gene (37,686,431-37,745,078)	PubMed 16249435
				Entire gene (37,686,431-37,745,078)	PubMed 24905847
				31,474,577 (not deleted) and chr17:31,929,967 (hg18)	PubMed 20175044
				Entire gene (37,686,431-37,745,078)	PubMed 21380624
				Entire gene (37,686,431-37,745,078)	PubMed 19417042
				Entire gene (37,686,431-37,745,078)	PubMed: No PubMed ID
				chr17:33562645 -34996659	PubMed 25270717
					PubMed 16912708
				chr17:36463828-37844128	PubMed 21055719
				chr17:36460073-37886264	PubMed 25893603
				Entire gene (37,686,431-37,745,078)	PubMed 22583611
				Entire gene (37,686,431-37,745,078)	PubMed 19417042
				Entire gene (37,686,431-37,745,078)	PubMed 19417042
				Entire gene (37,686,431-37,745,078)	PubMed 19417042
				chr17:36365470 -36441333, chr17:37890226 -37937509	PubMed 25256560
				chr17:37739440-37739440	PubMed 17924346
				Entire gene	PubMed 22706971
				Entire gene (37,686,431-37,745,078)	PubMed 22706971
				Entire gene (37,686,431-37,745,078)	PubMed 16971658
				chr17:37710503-37691928	PubMed 24382792
				chr17:37744541-37736354	PubMed 22060211
				chr17:37744541-37691928	PubMed 20206420
				chr17:37733557-37736354	PubMed 22060211
				chr17:37731595-37736354	PubMed 16371430
				chr17:37710503-37691928	PubMed 23539225

				Entire gene (37,686,431-37,745,078)	PubMed 26340261
				chr17:37744541-37703733	PubMed 22233861
				chr17:37710503-37715204	PubMed 26123568
				chr17:37731595-37715204	PubMed 21540130
HNF1B (36123078-37937509;1814432)	**chr17**	**7:37,686,431-37,745,247**	**HNF1B.chr20.Diabetes**	chr17::39219904 -41015172	
HNF1A (120978515-121002512;23998)	**chr12**	**12:120,977,787-121,002,512**	**HNF1A.chr12.Diabetes**	chr12:120988833-120993706	PubMed 23707370
				chr12:120978543-121002512	PubMed 23348805
				chr12:120978515-121002512	PubMed 17828387
				chr12:120988833-121002512	PubMed 17828387
				chr12:120978543-120979094	PubMed 17828387
				chr12:120993520-120994405	PubMed 23348805
				chr12:120997474-120999627	PubMed 22060211
				chr12:120979095-120996540	PubMed 24905847
WFS1 (6300657-6303265;2609)	**chr4**	**4:6,269,849-6,303,265**	**WFS1.chr4.Diabetes**	chr4:6300657-6303265	PubMed 22694282
INS (2159779-2161209;1431)	**chr11**	**11:2,159,779-2,161,341**	**INS.chr11.Diabetes**	chr11:2160862-2160839	PubMed 20133622
				chr11:2159779- 2161209	PubMed 21823539
				chr11:2159779-2160988	PubMed 24411943
				chr11:2159779- 2160988	PubMed 20133622
EIF2AK3 (88558917-88570873;11957)	**chr2**	**2:88,556,741-88,627,576**	**EIF2AK3.chr2.Diabetes**	chr2:88558917-88570873	PubMed 25893603
FOXP3 (49250436-49266505;16070)	**chrX**	**X:49,250,436-49,269,727**	**FOXP3.chrX.Diabetes**	chrX:49250436-49266505	
GATA6 (22087101-26821912;4734912)	**chr18**	**18:22,169,437-22,202,528**	**GATA6.chr18.Diabetes**	chr18:22087101 -26821912	PubMed 23696469
				chr18:22169437-22202528	PubMed 22318994
				chr18:22168345 -22171492	PubMed 22219654
PTF1A (23173481-23192990;19510)	**chr10**	**10:23,192,327-23,194,252**	**PTF1A.chr10.Diabetes**	chr10:23173481 -23181096	PubMed 24212882
TRMT10A (999546707-99564057;17351)	**chr4**	**4:99,546,707-99,564,057**	**TRMT10A.chr4.Diabetes**	chr4:99546707-99564057	PubMed 26297882
INSR (7112255-7294034;181780)	**chr19**	**19:7,112,255-7,294,034**	**INSR.chr.Diabetes**	chr19:7143507-7206857	PubMed 8175972
				chr19:7267438 -8003803	PubMed 9249867
				chr19:7112255-7294034	PubMed 23824322
				chr19:7122614-7122989	PubMed: No PubMed ID
				chr19:7267345-7267896	PubMed 7693131
				chr19:7184316-7184637	PubMed 23637016
				chr19:7163032-7152927	PubMed 25358339
				chr19:7128852-7132317	PubMed 23969187
DNAJC3 (95760040-95794989;34950)	**chr13**	**13:95,677,139-95,794,989**	**DNAJC3 .chr13.Diabetes**	chr13:95760040-95794989	PubMed 25466870
glis3tv1 (4152067-3828408;323659)	**chr9**	**9:3,824,127-4,348,392**	**glis3tv1.chr9.Diabetes**	Exons 1– 4 del	PubMed 26259131
				chr9:4152067-3828408	PubMed 26259131
				chr9:3824127-3828408	PubMed 26259131
				chr9:4117768-4118881	PubMed 26259131
				chr9:3932360-3937189	PubMed 26259131
				chr9:3829310-3828408	PubMed 26259131
				chr9:4125734-3828408	PubMed 23771172
SLC19A2 (169468836-169463909;4928)	**chr1**	**1:169,463,909-169,486,079**	**SLC19A2.chr1.Diabetes**	chr1:169468836-1169,463,931	

HMGA1 (34236800-34246231;9432)	**chr6**	**6:34,236,800-34,246,231**	**HMGA1.chr6.Diabetes**	chr6:34,236,800-34,246,231	
GPD2 (156324432-156586403;261972)	**chr2**	**2:156,435,290-156,613,735**	**GPD2.chr2.Diabetes**	Chr2: 156324432-156586403	
AR (67544032-68433841;889810)	**chrX**	**X:67,544,032-67,730,619**	**AR.chrX.Mouse Model**	chrX:67544036-67686126	PubMed 21710452
				complete deletion of coding region	PubMed 2050265
				chrX:67544032- 67730619	PubMed 9039995
				chrX:67544036-67686126	PubMed 24321103
				chrX:67544036-67730619	PubMed 25613104
				chrX:67544036-67686126	PubMed 25674389
				chrX:67544036-67686126	Source: LSDB
				chrX:67544036-67730619	PubMed 8990010
				chrX:67686010-67696075	PubMed 1508223
				chrX:67686010-67696075	PubMed 9007482
				chrX:67686010-67730619	PubMed 8339746
				chrX:67694673-67694673	PubMed 1750490
				chrX:67694673-67730619	PubMed 3186717
				partial	Source: Meeting abstract
				chrX:67544032- 67730619	PubMed 9007482
				chrX68042344-68433841	PubMed 27301361
				chrX:67544036-67686126	PubMed 10690872
ENPP1 (131886562-131895155;8594)	**chr6**	**6:131,808,016-131,895,155**	**ENPP1.chr6.Mouse Model**	chr6:131886562-131895,155	PubMed 20137773
IFNGR2 (33402896-33421685;18790)	**chr21**	**21:33,402,896-33,479,348**	**IFNGR2.chr21.Mouse Model**	chr21:33402896-33415020	PubMed 23459074
				chr21:33410836-33421685	PubMed 25592983
GPR12 (26758955-26761685;2731)	**chr13**	**13:26,755,200-26,760,785**	**GPR12.chr13.Mouse Model**	chr13:26231092–26233822	
E2F1 (33575580-33713849;138270)	**chr20**	**20:33,675,486-33,686,404**	**E2F1.chr20.Mouse Model**	chr20:33675486-33686404	PubMed 25439843
				chr20:33575580-33713849	PubMed 25439843
				chr20:33675486-33686404	PubMed 25439843
EIF2AK3 (87399366;90233829;2834463)	**2**	**87399366-90233829**	**EIF2AK3_Diabetes**	chr2:87399366-90233829	PubMed 25893603
IGSF1 (131176931;131505379;328448)	**X**	**131176931-131505379**	**IGSF1 _Obesity**	Chrx:131176931-131505379	PubMed 23143598

6.5 Discussion.

The success of GWAS and exome sequencing has provided fundamental insights into the genetics of diabetes and obesity. In this Chapter, we introduce a new unique strategy to address the genetic architecture of diabetes and obesity, which allows the analysis of both rare and common variants simultaneously. The custom array can be used to explore the genetics of disease cost-effectively in several ways. Applications of using this approach include to identify the frequency of particular rare variants in a cohort of people with extreme phenotypes (in this case obesity), identify rare variants in patients as a diagnostic panel testing, and to perform rare variant association studies. The common variants can be used to perform a follow-up analysis on previously-analysed common SNPs, generating a genetic risk score. Innovatively, this approach could be used to develop analysis that addresses the contribution of both rare and common variants to a disease, examining the possible mediating effects of common variants that may affect penetrance or deleterious rare variants.

The methodology does have limitations and challenges. Firstly, it is unable to include all the potential variants such as start loss, start gain and other types of frameshift, due to the size limitation of the array, which could have a significant impact in obtaining a complete picture. Secondly, as the largest ethnic group in the gnomAD database is European and most of the array variants come from gnomAD, it is not yet clear how effective the array will be on non-European populations. Lastly, the detection of rare variants using genotyping arrays is still relatively new and, we are therefore not yet certain of the efficiency of this methodology in detecting all of the rare variants. To examine this, we have included a group of 77 PMMO participants for whom exome sequence and SNP data is already available – the results of this

quality control analysis are presented in Chapter 7. This is a new methodological approach which is likely to require a series of validation analyses and re-design of array content as the identification of other new variants continues and the tested rare variants are validated.

CHAPTER 7: APPLICATION OF THE CUSTOM GENOTYPING ARRAY TO PMMO PARTICIPANTS

7.1 Introduction

This Chapter describes the quality control and initial results after application of the custom Axiom array, described in Chapter 6.

The custom Axiom array was designed to include as many as possible of the known or projected deleterious rare variants previously in genes causing monogenic/syndromic obesity and diabetes mellitus, in humans or rodent models. Additionally, probes to identify copy number variants within those genes have been included. To complement the rare variant analyses, a number of as well-studied common SNPs identified through GWAS as associated with obesity, diabetes and other related traits were also included, so that GRS (genetic risk scores) could be constructed.

In general, the use of genotyping arrays is very well established and they have been proved to perform excellently in genotyping common SNP variants associated with many diseases and traits. They are considered reliable and cheap for studying common variants. Nevertheless, due to their small effect-sizes, information on common variants is rarely used in clinical practice.

Rare variants are more likely to have profound effects on phenotype and disease risk, so have greater potential in clinical utility. Studying rare variants using custom genotyping arrays is still a more unusual approach and not very well-established. Many challenges in array-based genotyping have already been observed, especially with the problems of false positive genotyping results for rare variants using the previous Axiom UK Biobank and UK Bileve microarrays. These false positives arose from problems of clustering of sparse datapoints – for common variants, three genotype clusters are expected, but for rare variants, only two

clusters may be seen (common homozygotes and rare Hetrozygotes). The original algorithm, optimised for use with common variants, did not deal well with this.

Therefore, to improve rare variant genotyping analysis and achieve accurate results, very recently a new genotyping algorithm has been introduced by Thermofisher Scientific, namely Rare Heterozygous Adjusted Genotyping. The new algorithm now can cope with only two clusters. An additional refinement is based on intensities of probe sets; when a marker has multiple probes, the variability in intensities of these probes can be used to improve the reliability of calling of rare Heterozygous variants. This algorithm has been applied and tested on many different databases that showed it to eliminate false Heterozygous genotyping calls, maintain the het calls and achieves excellent positive predictive value (PPV), which represents a portion of correct het calls from all the het calls (not available publicly yet)

With our custom genotyping array for obesity and diabetes, both the original and improved genotype calling algorithms was tried. The analysis of genotyping results was performed based on the Axiom suite software. The analysis was performed twice: first with the previous algorithm and once again with the improved algorithm.

In this thesis, the results of the improved algorithm will be presented as they are considered the most stringent. Overall, the array was applied to more 2000 samples from people with extreme obesity, of whom around 35% have T2D. These genotyped participants belong to diverse ethnicities where white ethnicity and female are the highest. Full details of the sample selection and preparation are described in the methodology Chapter (Chapter 3/section 3.4). The inclusion, in the custom array design, of different variants, specifically concentrating 0n rare variants and CNVS is an innovative approach to investigating the genetic architecture of obesity and T2D in a cost-effective manner. To our knowledge, this is the first such attempt.

Thus, it is important to carry out effective quality control procedures, since such an approach has not been tried before.

A number of objectives and further analyses have been planned, using the data generated using this custom array. However, a relatively limited number of these analyses will be included in this thesis, due to time constraints. In this Chapter, a full description of the sample QC and marker QC results will be presented, as well as the outline results.

7.2 Aims of the study

- To carry out effective QC of data from the custom-designed genotyping array.
- To assess the prevalence of rare variants for monogenic obesity, and of cases with putative oligogenic inheritance.
- To analyse known obesity-related CNVs, through fixed region copy number analysis.

7.3 Results

7.3.1 Quality control of the Axiom array

Twenty-two 96-well plates were generated for a total of 2122 samples. The plates were processed and treated as one batch by the Oxford Genomic Centre. Genotyping calls were generated for all the samples using the Axiom Analysis Suite (1.1.0.616). To obtain high-quality genotyping results, the Axiom recommended best practice workflow was applied as described in the methodology Chapter (Chapter 3/Section 3.4). Quality control was performed at both sample and SNP levels. The overall details of the quality control steps are described in the methodology chapter (Chapter 3/Section 3.4).

7.3.1.1 Sample-level and plate QC

It is important to remove data from poor quality samples that have not genotyped well. The sample-level QC for the axiom array is assessed based on two parameters (described in the methodology chapter): the DQ, threshold > 0.82, and sample call rate, <97%. An overall summary of the sample-level QC is shown in Table 7.1.

Table 7.1 Summary of sample-level quality control.

Number of input samples	2112
Samples passing DQC	2103 out of 2112
Samples passing DQC and QC CR	2065 out of 2112
Samples passing DQC, QC CR and Plate QC	2065 out of 2112 (97.775%)
Number of failing samples	47
Number of input samples without QC information	0
Number of Samples Genotyped	2065
Average QC CR for the passing samples	99.767
Gender Calls Counts	female=1501 male=600 unknown=2

An overall summary of the samples passing the quality control thresholds. Three quality criteria were used to assess the samples as follow. The dish quality control (DQC value >0.82), call rate (CR value >0.97) and, plate QC (the average Call Rate value ≥50 and the percent of passing samples value ≥50 per plate). Beside the samples QC, the number of genders found in analysis.

Following the best practice guidelines by the manufacturer (Thermofisher), a total of 2103 samples passed DQ and a total 2065 samples passed both DQC and sample call rate. This indicates that a total of 47 samples out of the 2112 genotyped samples failed at least one of the sample QC standards: the majority failed at sample call rate.

In addition, all samples passed the plate-level QC, which was applied to detect poorly performed plates as samples with systematically different intensities on plates could impact negatively in genotyping analysis down. A summary of the plate QC Table is provided in Table 7.2.

Table 7.2 Plates Quality Control Summary

Plate Barcode	Result	Number of files in a batch	Number of files failing dish QC	Number of files failing QC Call rate	Number of samples that passed	Percent of passing samples	Average call rate for passing samples	Filtered Call Rate
5509734391235042121078	PASSED	96	1	0	95	98.958	99.702	99.886
5509734391235042121079	PASSED	96	0	1	95	98.958	99.721	99.819
5509734391235042121080	PASSED	96	0	1	95	98.958	99.831	99.883
5509734391235042121081	PASSED	96	0	2	94	97.917	99.653	99.804
5509734391235042121082	PASSED	96	0	0	96	100	99.77	99.828
5509734391235042121083	PASSED	96	0	1	95	98.958	99.833	99.889
5509734391235042121084	PASSED	96	1	2	93	96.875	99.763	99.887
5509734391235042121085	PASSED	96	0	3	93	96.875	99.799	99.924
5509734391235042121086	PASSED	96	0	0	96	100	99.762	99.857
5509734391235042121087	PASSED	96	1	2	93	96.875	99.84	99.923
5509734391236042121092	PASSED	96	0	0	96	100	99.817	99.916
5509734391236042121093	PASSED	96	1	1	94	97.917	99.827	99.941
5509734391236042121094	PASSED	96	0	1	95	98.958	99.854	99.945
5509734391236042121095	PASSED	96	0	1	95	98.958	99.69	99.854
5509734391236042121096	PASSED	96	0	3	93	96.875	99.789	99.91
5509734391236042121097	PASSED	96	0	2	94	97.917	99.83	99.926
5509734391236042121098	PASSED	96	0	1	95	98.958	99.834	99.924
5509734391236042121099	PASSED	96	2	1	93	96.875	99.75	99.914
5509734391236042121101	PASSED	96	1	2	93	96.875	99.834	99.931
5509734391236042121102	PASSED	96	0	2	94	97.917	99.819	99.912
5509734393343042221938	PASSED	96	1	4	91	94.792	99.736	99.88
5509734393344042221826	PASSED	96	1	8	87	90.625	99.407	99.6

An overall summary of each plate QC. This includes the number of samples failing the DQC, number of samples failing QC call rate, percentage of passing samples, an average call rate for passing samples and filter call Rate for your passing samples.
Plate Barcode: refers to the plate reference name (no=22)

7.3.1.2 Sex discordance

A standard “sense check” includes checking that the recorded sex of individual participants is in accord with the results of genotyping. This can detect errors such as sample or plate mix-ups. As described in the methodology Chapter, for samples’ sex confirmation, a set of markers located on the X chromosome were used to determine sex. The sex of two samples could not be unambiguously determined by the genotyping: one from the PMMO cohort and the one from the GERONIMO group. Another 13 samples showed different sex to that recorded in the database. These include eight samples from the Gravitas group, four samples from PMMO and one sample from the Endobarrier group. The Gravitas mismatched samples were found

to be due to a mistake in sample labelling by the provider of the samples. Therefore, all these sex-mismatched samples were excluded from the analysis.

A total of 2052 (97.1%) samples were considered for genotyping after passing the sample QC, plates QC and sex discordance QC.

7.3.1.3 Marker QC

It is also necessary to remove data from probes that did not perform consistently well. This was particularly important in our study, since the custom array design included a high proportion of probes that had never been included in any previous array design, and were therefore "untested". The SNP-level QC was performed to detect probe set that did not perform well, based on achieving well-clustered intensities and genotyping meeting the recommended statistical tests, so that they could be excluded from the analyses. For markers with more than one probe set, the best one of the two probe sets was selected.
This SNP QC was achieved based on default parameters that were recommended by Axiom best practice workflow, shown in Figure 3.10. (Chapter 3/Section 3.4)

A summary of the marker categories that were obtained from the genotyping is shown in Table 7.3. Overall, genotype calls were assigned for a total of 108894 probe sets, which is equal to 83157 markers. Around 76578 are considered best and recommended by Thermofisher, which is equivalent to 91.735%. This percentage of successful marker calling is considered high for rare variants according to Thermofisher, which they have advised us indicates high-quality genotyping.

Table 7.3: Summary of ProbeSet and markers from the genotyping data.

CONVERSIONTYPE	COUNT (PERCENTAGE %)	
	ProbeSets	Markers
MONOHIGHRESOLUTION	82802 (76.039%)	65445 (78.701%)
NOMINORHOM	13278 (12.194%)	6582 (7.915%)
OTHER	7205 (6.617%)	5995 (7.209%)
POLYHIGHRESOLUTION	4734 (4.347%)	4551 (5.473%)
OTV (OFF-TARGET VARIANTS)	698 (0.641%)	460 (0.553%)
CALL RATE BELOW THRESHOLD	177 (0.163%)	124 (0.149%)

The table summarises the number of probesets and markers from the genotyping according to the marker categories that were obtained from the genotyping. Description of each category is demonstrated in methodology Chapter.

7.3.1.4 QC based on WES previously detected variants

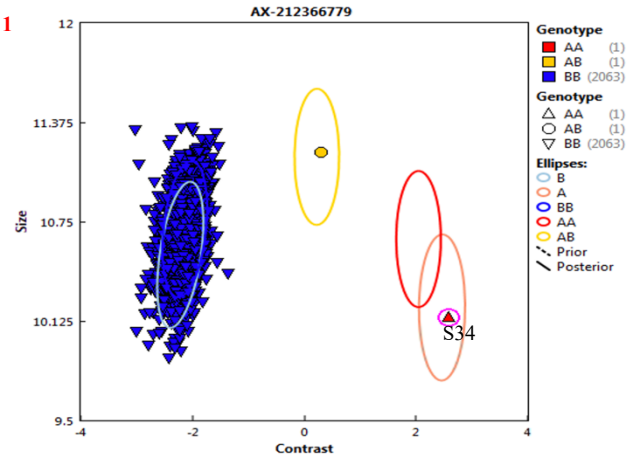
As an additional step, to assess the efficiency and validate the genotyping performance of the Axiom array, a number of samples with previous WES data were included in the genotyping. All the examined variants detected initially by WES were confirmed by the genotyping, except for one missense variant in PTEN, in the participant listed in Table 7.4 as S35. This variant was categorised as uncalled, but looking at the genotyping cluster figure, the cluster of the wild type AA allele carriers is well separated from the misclassified carrier (Figure 7.1). A summary Table of the tested variants is shown in Table 7.4 and genotype clusters for each are shown in Figure 7.1.

Table 7.4: Summary of the examined variants which were detected initially by WES.

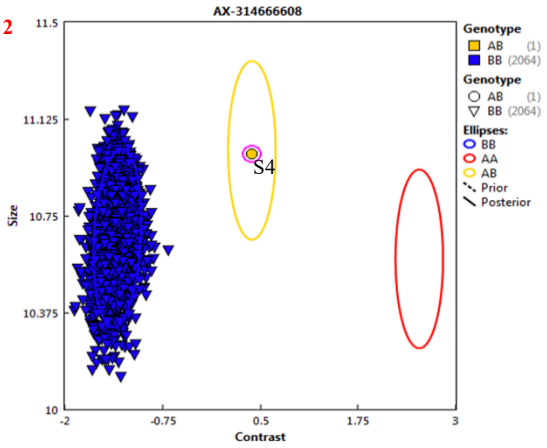
	Variant Array ID	Gene	RS ID	Type of variant	MAF	Subject	Genotyping confirmation
1	AX-212366779	*IGSF1*	rs749977306	stopgain	N/A	S34	Confirmed
2	AX-314666608	*MYT1L*	rs201765281	Missense	0.0001365	S4	Confirmed
3	AX-212386946	*NTRK2*	rs753056075	Missense	0.000003977	S36	Confirmed
4	AX-212386946	*RAI1*	rs147844401	Missense	0.00009573	S28	Confirmed
5	AX-83359409	*ENPP1*	rs190947144	Missense	0.0004078	S37	Confirmed
6	AX-83586350	*MCHR1*	rs45439291	Missense	0.003407	S1	Confirmed
7	AX-82915434	*AFF4*	rs139490054	Missense	0.002310	S12	Confirmed
8	AX-82966004	*PEG3*	rs56237501	Missense	0.004053	S26	Confirmed
9	AX-83513405	*RAI1*	rs113208290	Missense	0.003726	S28	Confirmed
10	AX-88752465	*PTEN*	rs121909238	Missense	N/A	S35	MISSCLASSIFIED

MAF(%) gnomAD MAF.
All the exmained variants detected initially by WES were confirmed by the genotyping, except for one missense variant in PTEN, as shown in Figure 7.1 (10).

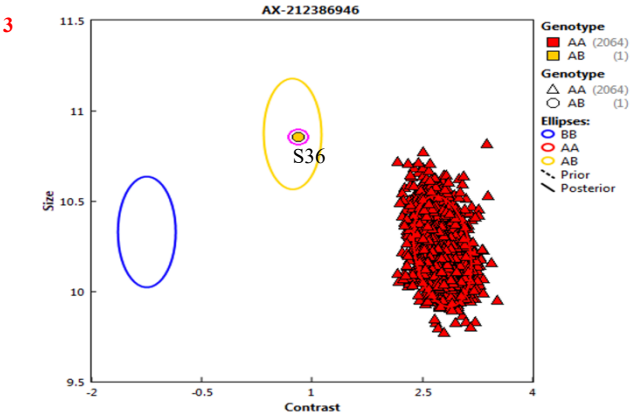
1


AX-212366779
Genotype
AA (1)
AB (1)
BB (2063)
Genotype
AA (1)
AB (1)
BB (2063)
Ellipses:
B
A
BB
AA
AB
Prior
Posterior
S34
Size
Contrast

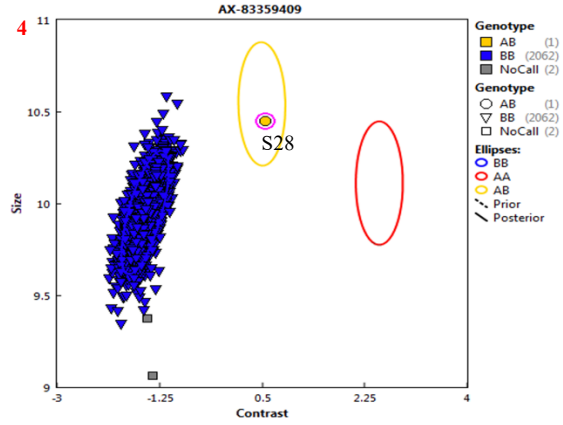
2


AX-314666608
Genotype
AB (1)
BB (2064)
Genotype
AB (1)
BB (2064)
Ellipses:
BB
AA
AB
Prior
Posterior
S4
Size
Contrast

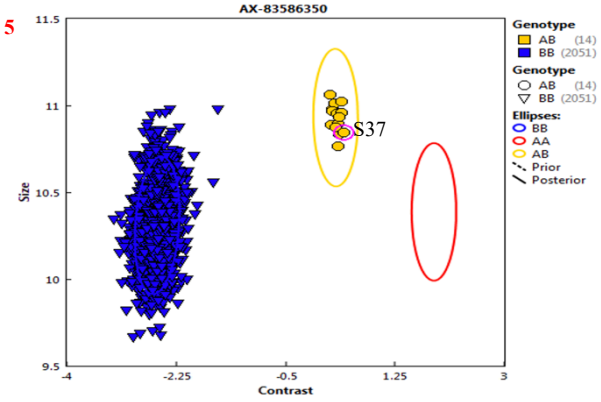
3


AX-212386946
Genotype
AA (2064)
AB (1)
Genotype
AA (2064)
AB (1)
Ellipses:
BB
AA
AB
Prior
Posterior
S36
Size
Contrast

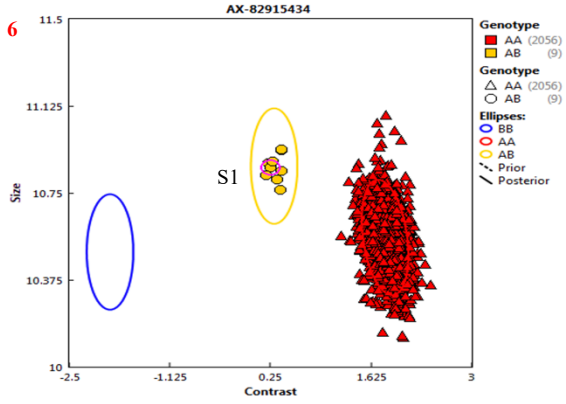
4


AX-83359409
Genotype
AB (1)
BB (2062)
NoCall (2)
Genotype
AB (1)
BB (2062)
NoCall (2)
Ellipses:
BB
AA
AB
Prior
Posterior
S28
Size
Contrast

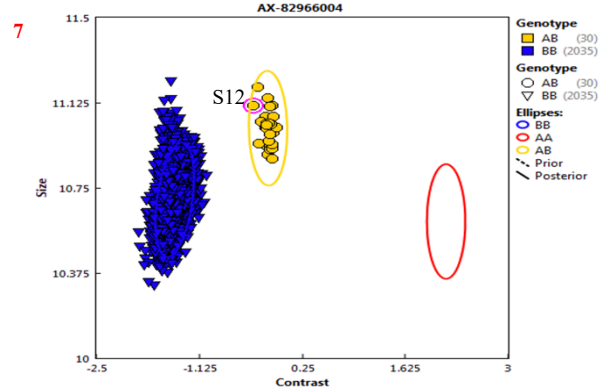
5


AX-83586350
Genotype
AB (14)
BB (2051)
Genotype
AB (14)
BB (2051)
Ellipses:
BB
AA
AB
Prior
Posterior
S37
Size
Contrast

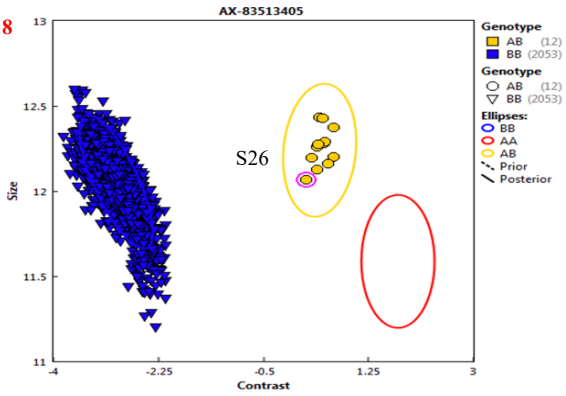
6


AX-82915434
Genotype
AA (2056)
AB (9)
Genotype
AA (2056)
AB (9)
Ellipses:
BB
AA
AB
Prior
Posterior
S1
Size
Contrast

7


AX-82966004
Genotype
AB (30)
BB (2035)
Genotype
AB (30)
BB (2035)
Ellipses:
BB
AA
AB
Prior
Posterior
S12
Size
Contrast

8


AX-83513405
Genotype
AB (12)
BB (2053)
Genotype
AB (12)
BB (2053)
Ellipses:
BB
AA
AB
Prior
Posterior
S26
Size
Contrast

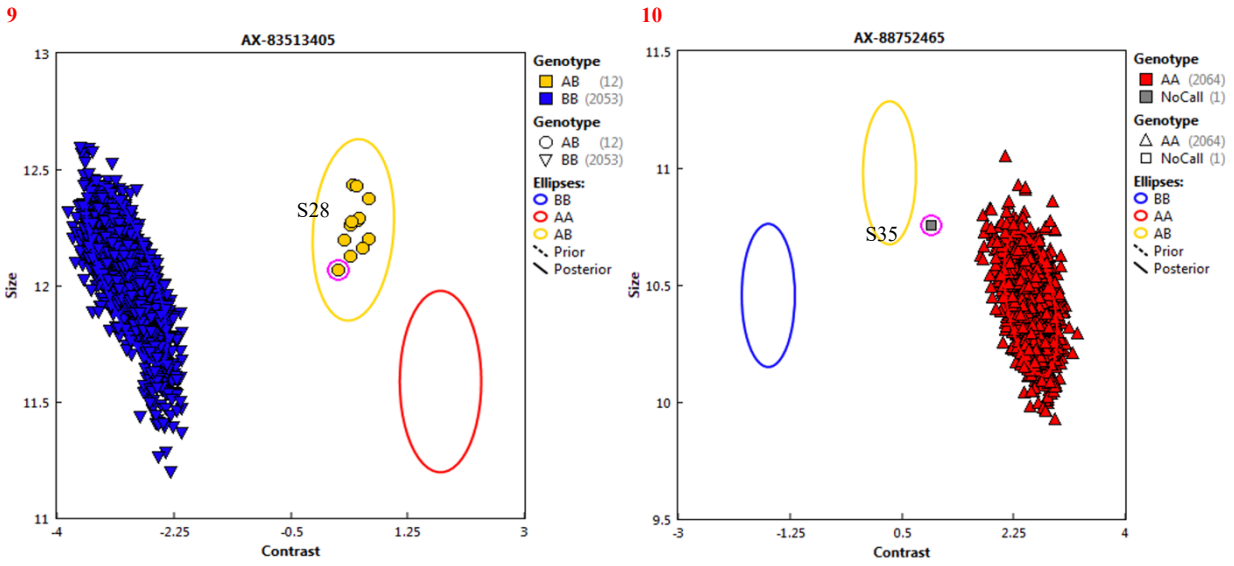


Figure 7.1 Figures of SNP clusters of the confirmed WES variants through genotyping.
The SNP cluster plots are numbered according to variants in Table 7.4.
Each SNP represents one probe set that is interrogated into a single SNP. The clustering is carried out in two dimensions: X and Y. The X dimension correspond to the main information for distinguishing genotype clusters and is calculated as Log2(A_signal/ B_signal). The Y dimension is defined as a size that is [Log2(A_signal) + Log2(B_signal)]/2 (Size/Strength). Samples are colored and shaped circles and triangles. AA calls (red triangles), BB calls (blue triangles upside down), AB calls (gold circles). The confirmed variant is shown in pink circle.

7.3.2 Overview of genotyping results for rare variants

7.3.2.1 Rare deleterious variants in monogenic and syndromic obesity genes

As described in Chapter 3, methodology, the markers are classified into different cluster categories (Figure 3.14). For the rare variants analysis, two types of SNP clusters were analysed, namely NoMinorHom (markers where no participants homozygous for the rare variant were observed – ie. Where only two clusters were expected), and PolyhighResolution with MAF<0.01, which represents the variants the homozygous and Heterozygous carriers of the mutation when three clusters with good separation are found (uncommon variants, but where minor homozygotes are detected – ie. Three clusters are expected). Due to the thesis scope and PhD timeline, only markers within monogenic and syndromic obesity genes were analysed. A total of 6582 (7.915%) markers were grouped under NoMinorHom, and a total of 4551 (5.473%) markers were grouped under PolyHighResolution. A flowchart of variant selection in obesity and syndromic obesity genes is shown in Figure 7.2. A summary of variants detected is shown in Figure 7.2 .

The next step was to consider whether these rare variants might affect phenotype. Since at this stage only genes causing monogenic or syndromic obesity were considered, the expected mode of inheritance was known. Of a total of 393 markers in the NoMinorHom category, 149 variants matched the mode of inheritance expected to result in a phenotype in Hetrozygotes or hemizygotes (autosomal dominance mode of inheritance, or XLR in a male). Of 29 variants were found in the PolyhighResolution class, 12 variants match the mode of inheritance expected to result in a phenotype in homozygotes for the rare allele (autosomal recessive mode of inheritance or XLR in a female). Thus, in total, 161 rare deleterious variants have

been identified as potentially pathogenic variants from both NoMinorHom and PolyhighResolution group in a total of 40 genes. A summary of the number of rare deleterious variants and considered variants in each are summarised in Table 7.5.

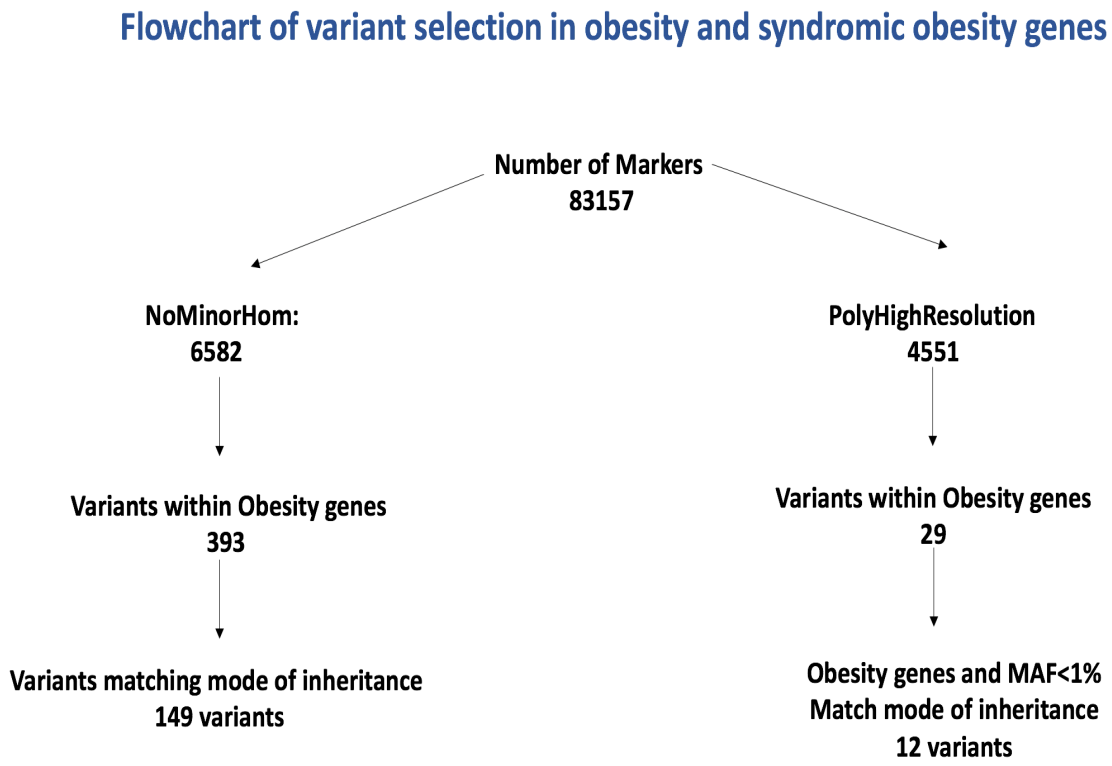


Figure 7.2: Flow chart of the variants selection in obesity and syndromic obesity genes.

For the rare variants analysis, two types of SNP clusters were analysed, namely NoMinorHom and PolyHighResolution. **NoMinorHom** represents the Heterozygous carriers of a mutation (when only two clusters are found, one for a homozygous genotype (normal) and the Heterozygous genotype (no cluster for the other homozygous). **PolyHighResolution** represents the variants the homozygous and Heterozygous carriers of the mutation when three clusters with good separation are found (polymorphic SNPs). A total of 162 rare deleterious variants have been identified from both groups.

Table 7.5: An overall summary of the rare deleterious variants and considered variants in each gene of obesity and syndromic obesity genes

GENE	CHR	MOI	NUMBER OF RARE DELETRIOUS VARIANTS IN EACH GENE	CONSIDERED RARE VARIANTS IN EACH GENE	TOTAL NUMBER OF CARRIERS OF THE CONSIDERED VARIANTS IN EACH GENE
ADCY3	2	AR	6	6	9
AFF4	5	AD	7	7	10
ALMS1	2	AR	1	1	1
ARL14EP	11	AR	1		
ARL6	3	AR	3		
ATXN2	12	AR	1	1	1
BBIP1	10	AR	1	1	1
BBS1	11	AR	14	1	1
BBS10	12	AR	7		
BBS11	9	AR	16		
BBS12	4	AR	8		
BBS13	17	AR	5		
BBS2	16	AR	10		
BBS4	15	AR	7	1	1
BBS5	2	AR	7		
BBS7	4	AR	5		
BBS9	7	AR	16		
BDNF	11	AD	2	1	1
CARTPT	5	AD	4	2	2
CEP290	12	AR	31	1	1
COA3	17	AR	2		
CPE	4	AR	3		
DNAAF1	16	AR	16		
DYRK1B	19	AD	7	5	6
FAAH	1	AD	3	2	3
FOXA3	19	AD	2	2	2
IFT27	22	AR	6		
IGSF1	X	XLR	1	4	5
KSR2	12	AD	5	5	9
LEPR	1	AR	14	1	1
LRP	2	AR	2		
LZTFL1	3	AR	1		
MAGEL2	15	AR	6	5	6
MBD5	2	AD	16	16	17
MC3R	20	AD	4	3	3
MC4R	18	AD	7	6	15
MCHR1	22	AD	9	6	11

MGLL	3	AD	1		
MMKS	20	AR	4		
MRAP2	6	AD	2	2	2
MYT1L	2	AD	2	2	4
NPY4R	10	AD	5	3	3
NR0B2	1	AD	2	2	3
NTRK2	9	AD	3	2	4
NUCB2	11	AD	2	2	2
PCSK1	5	AD	6	5	5
POGZ	1	AD	7	7	9
POMC	2	AR	7	4	3
POU3F2	6	AD	2	1	1
PPARG	3	AD	1	1	1
PTEN	10	AD	9	9	13
RAI1	17	AD	18	14	26
RBMX	X	XLR	1		
SDCCAG8	1	AR	10		
SETD2	3	AD	22		
SH2B1	16	AD	9	9	10
SIM1	6	AD	7	7	8
SLC35D3	6	AD	4	4	10
TTC8	14	AR	6		
TUB	11	AR	14		
UCP1	4	AD	5	4	5
UCP3	11	AD	3	3	6
WDPCP	2	AR	9		
WNT10B	12	AD	5	4	8
TOTAL			422	161	229

MOI: modes of inheritance
Considered Variants: variants that follow the mode of inheritance.
ǂ: Genes that follows the autosomal recessive patterns of inheritance and have been reported to have Heterozygous mutations possibly contributing to obesity.

Summary details of the considered variants, which refer to rare deleterious variants that follow the expected mode of inheritance, are shown in Table 7.6. From these variants, 145 are missense, 9 frameshift, 4 stop gained, 2 at the 5' untranslated region and 1 at splice region (Figure 7.3). A total of 7 of the considered deleterious rare variants are not found in gnomAD but they have been reported in the literature.

There is a small number of obesity genes that follow the autosomal recessive patterns of inheritance, but have been reported to have Heterozygous mutations possibly contributing to obesity. These genes, which include *POMC*, *ADCY3* and *PCSK1*.

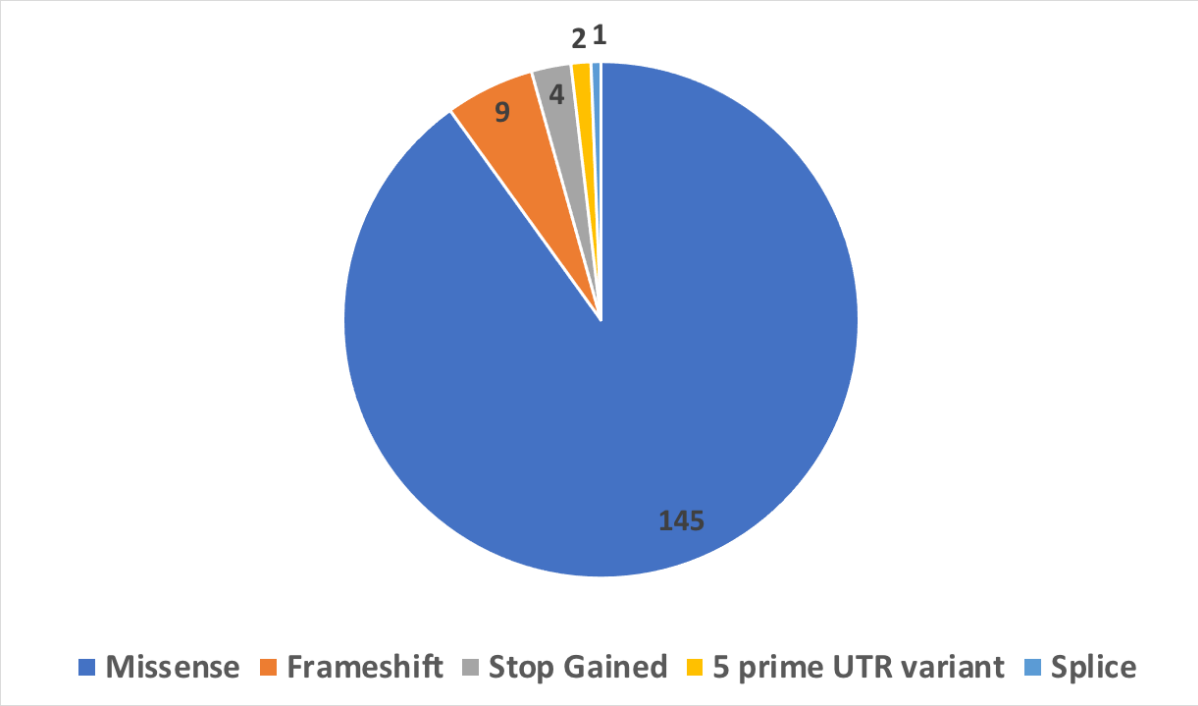


Figure 7.3: Summary of the different types of potential pathogenic variants identified in obesity and syndromic obesity genes.

7.3.2.2 Description of the rare deleterious variants identified in obesity genes

The details of identified candidate rare deleterious variants in obesity-related genes are shown in Table 7.6. Overall, a total of 101 variants were found in 27 obesity-related genes; ten of these genes are related to the leptin-melanocortin pathway, namely one homozygous variant in LEPR and Heterozygous variants in the following genes: 9 variants in *SH2B1*, 6 variants in *MC4R*, 3 variants in *MC3R*, 4 variants in *POMC*, 5 variants in *PCSK1*, 2 variants in *MRAP2*, 7 variants in *SIM1,* 1 variant in *BDNF* and 2 variants in *NTRK2* (6 variants). The other considered rare deleterious variants were found in *DYRK1B, DNAAF1, KSR2, ADCY3, IGSF1, NPY4R, UCP1, GHSR, WNT10B, FOXA3, MCHR1, UCP3, CARTPT, NUCB2 NR0B2, POU3F2, PPARG, SLC35D*3 and *FAAH*.

For genes that follow the autosomal recessive mode of inheritance, but are also reported to have an intermediate obesity phenotype in Heterozygous mutation carriers, including POMC, ADCY3 and PCKS1, a total of 15 variants have been identified. Details of the variants are summarised in Table 7.7.

Nine of these considered variants in obesity-related genes have been reported previously; this includes *SH2B1* (rs190981290), *MC4R* (rs121913563, rs13447325, rs138281308 and rs372794914) and *PTEN* (10:87933124, 10:87864534, 10:87894025 and rs202004587). Seven of the identified variants are not found in gnomAD.

To further explore the biological network interaction among those genes with considered rare deleterious variants, the online tool (STRING) was used and is shown in Figure 7.4 All genes seem to have some direct or indirect interaction to the leptin-melanocortin pathway except for three genes: *IGFS1*, *SLC35D3*, *DYRK1B* and, *FOXA3*.

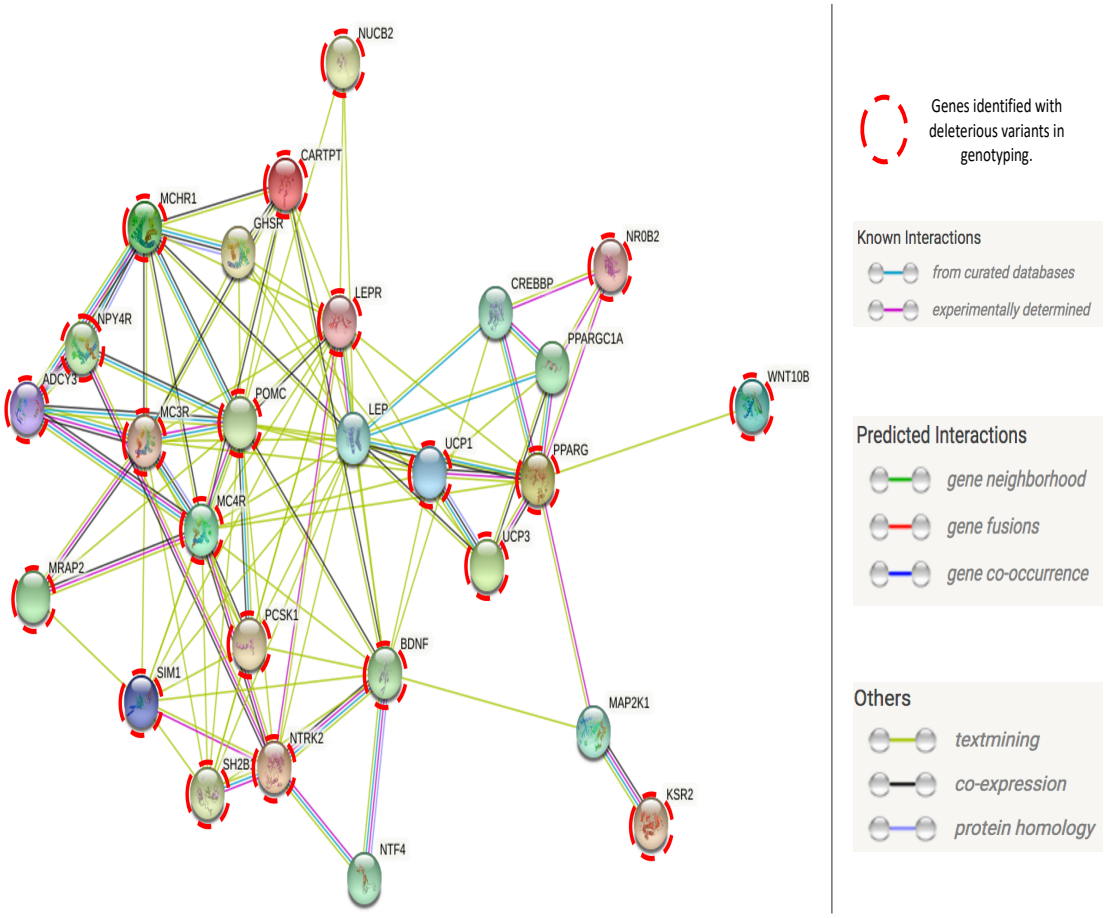


Figure 7.4: Interaction network among genes identified with potential rare deleterious variants from the genotyping analysis. Identification and prediction of interaction network among the identified obesity genes with rare deleterious variants. Coloured nodes represent the different proteins/genes, dashed red circle represents the identified genes with potential rare deleterious variants. Different colour line represents the direct and indirect interactions between proteins where each colour indicates the type of evidence available for that interaction. Image generated by STRING (www.string-db.otg)

7.3.2.3 Rare deleterious variant identified in syndromic obesity genes

Despite the identification of potential rare deleterious variants in the known obesity genes that are related to leptin-melanocortin pathway and known to cause obesity along with an intellectual disability phenotype such as NTRK2 and BDNF, further candidate variants were identified in other syndromic obesity genes. A total of 61 variants were found in 15 syndromic obesity genes, which include *BBS1, BBS4, CEP290, MKKS, ALMS1, ATXN2, AFF4, RAI1, MAGEL2, BDNF, MBD5, BBS1, MYT1L* and *POGZ*.

Four of those variants are related to Bardet-Biedl syndrome genes group, namely *BBS1*, *BBS4*, *MKKS* and *CEP290*.

Around 11 of these variants have been reported previously, which include *BBS4* (rs34620165), *CEP290* (rs183655276), MKKS (rs74315394), *ALMS1* (rs45501594), *ATXN2* (rs140262591), *RAI1* (rs142981643), *RAI1* (rs549244691), *RAI1* (rs376429075), *RAI1* (rs745320469), *MBD5* (rs536900412) and *BBS1* (rs35520756). Two of the identified variants are not found in gnomAD.

Table 7.6: Details of the considered variants of obesity and syndromic obesity genes.

GENE	AX ID	INFO SUMMARY	GENO TYPE	TYPE OF VARIANT	HGVS	MAF GNOMAD	CADD	CLINVAR	PUBLICATION	NO OF CARRIERS
AFF4	AX-175171397	rs200076214	Het	Missense	p.His194Asn	0.000007953	22.8	-	No	1
AFF4	AX-181564899	rs796481281	Het	Missense	p.Arg177Cys	0.00009146	28.7	-	No	1
AFF4	AX-212365184	rs535607337	Het	Missense	p.Pro708Ala	0.00003979	22.7	-	No	1
AFF4	AX-213430847	rs138207289	Het	Missense	p.Ser174Cys	0.000008791	29.7	-	No	2
AFF4	AX-213431280	rs1170657071	Het	Missense	p.Ser1082Leu	0.000007959	33	-	No	1
AFF4	AX-82915434	rs139490054	Het	Missense	p.Thr1107Ala	0.002388	22.1	Benign	No	1
AFF4	AX-83420563	rs141068875	Het	Missense	p.Ser2Thr	0.0008073	22.5	Likely benign/Chops syndrome	No	3
ALMS1	AX-83103610	rs45501594	Hom	Missense	p.Thr3543Ser	0.006639	20.3	Benign	1	1
ATXN2	AX-83572698	rs140262591	Hom	Missense	p.Val1000=	0.002779	14.81	Uncertain significance	1	1
BBIP1	AX-83228465	rs201753066	Hom	Missense	p.Arg31Gln	0.0002366	20.5	-	No	1
BBS1	AX-30203871	rs35520756	Hom	Missense	p.Glu234Lys	0.007676	24.2	Benign	3	1
BBS4	AX-39943719	rs34620165	Hom	Missense	p.Gly250Arg	0.001121	24.7	Benign	1	1
BDNF	AX-212394690	rs748154574	Het	frameshift	p.Gly36AlafsTer33	0.000004168	-		No	1
CARTPT	AX-212362302	rs77043527	Het	Missense	p.Val79Ile	0.000007953	22.8	-	#N/A	1
CARTPT	AX-213426360	rs766852629	Het	Frameshif	p.Asp30ThrfsTer24	0.00000797	23.6	-	No	1
CEP290	AX-83307392	rs183655276	Hom	Missense	p.Asp1413His	0.002316	23.8	Uncertain significance	4	1
DYRK1B	AX-113915693	rs757189339	Het	Missense	p.Arg433Leu	0.0000192	27	-	No	1
DYRK1B	AX-212361690	rs375929795	Het	Missense	p.Glu105Lys	0.00000795	27.5	-	No	1
DYRK1B	AX-213438849	rs374944306	Het	Missense	p.Asp95Asn	0.0000437	24.5	-	No	1
DYRK1B	AX-213456043	rs35858874	Het	Missense	p.Ser234Gly	0.00000399	23.3	-	No	1
DYRK1B	AX-314672684	rs562169098	Het	Stop Gained	p.Tyr271Ter	0.00001596	20.2	-	No	2
FAAH	AX-213455215	rs138437161	Het	Missense	p.Arg524Gly	0.0000398	14.97	-	#N/A	2
FAAH	AX-83560961	rs144534314	Het	Missense	p.Val346Met	0.0001273	15.41	-	#N/A	1
FOXA3	AX-175234417	rs145277260	Het	Missense	p.Arg110Trp	0.0000241	26.9	-	#N/A	1
FOXA3	AX-317020312	rs780181879	Het	Missense	p.Asp321Tyr	0.00000398	25.8	-	#N/A	1
IGSF1	AX-212393848	rs749977306	Hemi	Stop Gained	p.Arg1281Ter	-	38	-	No	1
IGSF1	AX-314675127	rs142822502	Hemi	Missense	p.Pro237Ser	0.0006018	18.24	-	#N/A	1
IGSF1	AX-317021587	rs1185245635	Hemi	Missense	p.Pro627Thr	0.00000546	23.3	-	No	1

IGSF1	AX-317024590	rs757080190	Hemi	Missense	p.Ala634Thr	0.0000164	24.7	-	No	1
KSR2	AX-114012359	rs376902515	Het	Missense	p.Ala344Ser	0.0001948	18.92	-	No	2
KSR2	AX-212372740	rs576454880	Het	Missense	p.Gln526Pro	0.0000127	24.6	-	No	3
KSR2	AX-213434282	rs751356379	Het	Missense	p.Pro553Leu	0.0000202	24.2	-	No	1
KSR2	AX-213434340	rs552480492	Het	Missense	p.Thr395Met	0.0001165	25	-	No	2
KSR2	AX-326369165	rs771733781	Het	Missense	p.Pro329Leu	0.0000283	25.2	-	No	1
LEPR	AX-314677925	rs34499590	Hom	Missense	p.Thr699Arg	0	23.6	Benign	2	1
MAGEL2	AX-212364645	rs778127117	Het	Missense	p.Gly127Asp	0.000004052	14.53	-	No	1
MAGEL2	AX-212385443	rs576711188	Het	Missense	p.Val444Ala	0.000556	25.5	Likely benign	No	2
MAGEL2	AX-212389810	rs371613799	Het	Missense	p.Arg440His	0.00001782	18.23	-	No	1
MAGEL2	AX-213425652	rs200958282	Het	Missense	p.Glu117Gln	0.00005499	23.6	-	No	1
MAGEL2	AX-213457006	rs1391195912	Het	Missense	p.Ala247Thr	0.00000713	16.32	-	No	1
MBD5	AX-147188889	rs752035001	Het	Missense	p.Asp1180Val	0.0000716	27.3	Uncertain significance/Intellectual Disability, Dominant	No	1
MBD5	AX-164001116	rs201668347	Het	Missense	p.Gly800Asp	-	24	Uncertain significance	No	1
MBD5	AX-169020776	rs201334086	Het	Missense	p.Gly1199Arg	0.0000677	28.8	Likely benign	No	1
MBD5	AX-175182497	rs536900412	Het	Missense	p.Gln371Glu	0.001561	24.7	Uncertain significance/Intellectual Disability, Dominant	1	1
MBD5	AX-212357956	rs1466233766	Het	Missense	p.Leu98Val	0.000003986	24.7	-	No	2
MBD5	AX-212367049	No dbSNP ID	Het	Missense	p.Asp654Glu	0.00000698	21	-	No	1
MBD5	AX-212381031	rs770801894	Het	Missense	p.Ala92Thr	0.0000638	28	Uncertain significance/History of neurodevelopmental disorder	No	1
MBD5	AX-212393905	rs761118931	Het	Missense	p.Val412Ile	0.0000438	23.3	Uncertain significance/Mental retardation, autosomal dominant 1	No	1
MBD5	AX-213428863	rs748142226	Het	Missense	p.Ile1130Asn	0.0000358	26.5	Uncertain significance/Mental retardation, autosomal dominant 1	No	1
MBD5	AX-213434182	rs757256547	Het	Missense	p.Ile19Met	0.00000796	17.38	-	No	1
MBD5	AX-213456379	rs142913108	Het	Missense	p.Pro1250Thr	0.00000796	19.86	-	No	1
MBD5	AX-82890333	rs138639760	Het	Missense	p.Pro721Leu	0.0001715	23.6	Likely benign	No	1
MBD5	AX-83069920	rs201695275	Het	Missense	p.Ser512Phe	0.0002799	20.3	Likely benign	No	1
MBD5	AX-83300134	rs143028540	Het	Missense	p.Gln1015Arg	0.0005098	26.1	-	No	1
MBD5	AX-83431587	rs145808884	Het	Missense	p.Thr352Ile	0.0000797	24.4	Likely benign	No	1
MBD5	AX-83566711	rs145475623	Het	Missense	p.Thr1048Ile	0.0004813	19.42	Uncertain significance	No	1
MC3R	AX-212369150	rs574634142	Het	Missense	p.Asn68Lys	0.0000119	22.9	-	No	1
MC3R	AX-212387413	rs373708098	Het	Missense	p.Gly249Ser	0.0000677	30	-	No	1
MC3R	AX-212391314	rs777078457	Het	Missense	p.Pro272Ser	0.0000199	23.8	-	No	1

MC4R	AX-147908172	rs372794914	Het	Missense	p.Thr11Ala	0.00006	20.1	-	1	2
MC4R	AX-212380155	rs773199277	Het	Missense	p.Ile185Val	0.00000796	24	-	No	1
MC4R	AX-213430957	rs1226923092	Het	Frameshift	p.Ser116PhefsTer6	0.000006987	-	-	No	1
MC4R	AX-317015454	rs138281308	Het	Missense	p.Phe202Leu	0.0006923	20.2	Uncertain significance/ obesity	7	9
MC4R	AX-82935491	rs13447325	Het	Missense	p.Asp37Val	0.0000676	21.2	Uncertain significance/ obesity	6	1
MC4R	AX-86564484	rs121913563	Het	Missense	p.Ala175Thr	0.0005411	19.5	Pathogenic/obesity	2	1
MCHR1	AX-212389578	22:40681355	Het	frameshift	p.Ile233ValfsTer40	0.00000406	N/A	-	No	1
MCHR1	AX-213433144	rs368183075	Het	Missense	p.Val53Met	0.000016	22.2	-	No	5
MCHR1	AX-213455206	rs765877994	Het	Missense	p.Ala133Val	0.0000278	20.9	-	No	1
MCHR1	AX-317025913	rs201404111	Het	Missense	p.Ile316Asn	0.00000398	19.26	-	No	1
MCHR1	AX-82936288	rs143942609	Het	Missense	p.Val135Met	0.0002545	17.12	-	No	1
MCHR1	AX-83276573	rs150436727	Het	Missense	p.Gly264Ser	0.0000798	23.5	-	No	3
MKKS	AX-83259327	rs74315394	Hom	Missense	p.Ala242Ser	0.00521	25.5	Uncertain significance	4	1
MRAP2	AX-181519031	rs749880950	Het	Missense	p.Val57Met	0.0000795	29.3	-	No	1
MRAP2	AX-212389609	rs749283207	Het	Missense	p.Pro167Ala	0.00004956	25.7	-	No	1
MYT1L	AX-164394519	rs1191064463	Het	Missense	c.1315G>A	0.000008023	24.6	-	No	1
MYT1L	AX-82918030	rs200235250	Het	Missense	p.Gly273Ser	0.0004952	20.7	-	No	3
NPY4R	AX-119041350	rs149136223	Het	Missense	p.Pro96Leu	0.0000716	26.5	-	No	1
NPY4R	AX-182725553	rs373849471	Het	Missense	p.Cys168Tyr	0.0000915	22.2	-	No	1
NPY4R	AX-212370506	rs575872443	Het	Missense	p.Asn187%3D	0.00000398	18.38	-	No	1
NR0B2	AX-212364698	rs367827644	Het	Missense	p.Arg213His	missense	26.5	-	No	1
NR0B2	AX-212388987	rs140901243	Het	Missense	p.Pro115Ser	missense	22.4	-	No	2
NTRK2	AX-212374429	rs200900730	Het	Missense	p.Asn825Asp	0.0000239	24	-	No	1
NTRK2	AX-314672219	rs76950094	Het	Missense	p.Ser167Tyr	0.0009584	19.81	-	1.	3
NUCB2	AX-212364970	rs200431249	Het	Missense	p.Lys370Thr	0.00001432	27.2	-	No	1
NUCB2	AX-86731621	rs756249891	Het	Frameshift	p.Leu14Ter	0.00002496	N/A	-	No	1
POGZ	AX-212366913	rs373783340	Het	Missense	p.Pro931Leu	0.0002	20.5	Association	No	2
POGZ	AX-212369012	rs776985962	Het	Missense	p.Gln133Arg	0.000007203	20.3	-	No	1
POGZ	AX-212377957	rs768284272	Het	Missense	p.Leu1103Phe	0.0001591	15.41	-	No	2
POGZ	AX-212380768	rs151243063	Het	Missense	p.Gln277His	0.00001194	15.17		No	1
POGZ	AX-212381481	rs1312798265	Het	Missense	p.Glu1347Gln	0.00003184	23.4	-	No	1
POGZ	AX-314675255	rs72996036	Het	Missense	p.Arg617Gln	0.0004	25.5	Uncertain significance;Hereditary disease	No	1
POGZ	AX-82938151	rs141132016	Het	Missense	p.Asn136Ser	0.0006	20.8	-	No	1

POU3F2	AX-212361621	rs1321043636	Het	Missense	p.Gln315His	0.000003977	28.6	-	No	1
PPARG	AX-213456800	rs758594107	Het	Missense	p.Ile331Val	0.00003188	24.7	-	No	1
PTEN	AX-119127349	rs765433422	Het	Missense	p.Gly209Arg	0.00000801	25.7	-	No	1
PTEN	AX-122990578	rs11202592	Het	5 prime UTR variant	c.511C>G	0.003765	22.2	Benign	6	2
PTEN	AX-147852202	rs587779994	Het	5 prime UTR variant	c.-315C>T	0.00003202	22.1	Uncertain significance/Cowden syndrome 1	No	1
PTEN	AX-164176837	10:87894025	Het	Missense	c.80A>G	N/A	26.2	-	1	1
PTEN	AX-164419336	10:87933124	Het	Missense	c.365T>G	N/A	28.9	-	1	1
PTEN	AX-317025423	rs773176120	Het	Stop Gained	p.Tyr68Ter	-	37	Pathogenic	No	1
PTEN	AX-83279280	rs202004587	Het	Missense	p.Ala79Thr	-	20.9	Uncertain significance/Cowden syndrome 1	9	1
PTEN	AX-86650465	10:87931034	Het	Splice	c.210-7_210-3del5	0.0002814	-	Conflicting-Interpretations-Of-Pathogenicity	0	1
PTEN	AX-90032807	10:87864534	Het	Missense	c.65A>G	N/A	25.7	-	1	1
RAI1	AX-113515809	rs745320469	Het	Missense	p.Pro890Leu	0.0000883	23	Likely benign	1	1
RAI1	AX-146352297	rs549244691	Het	Missense	p.Ala617Asp	0.0002684	24.8	Uncertain significance	1	1
RAI1	AX-168884593	rs376429075	Het	Missense	p.Pro1735Leu	0.0000441	16.1	Uncertain significance	1	1
RAI1	AX-175202319	rs772449638	Het	Missense	p.Lys1417Arg	0.0000159	24.2	Uncertain significance	No	1
RAI1	AX-175232258	rs141317462	Het	Missense	p.Arg1559Gln	0.0005058	25.1	Uncertain significance	1	4
RAI1	AX-181503748	rs201842299	Het	Missense	p.Arg7Gly	-	27	-	No	1
RAI1	AX-212367190	rs774984414	Het	Missense	p.Gly1410Ser	0.0000119	25.8	-	No	1
RAI1	AX-212385081	rs372896387	Het	Missense	p.Arg80Gln	0.0000573	26.5	-	No	1
RAI1	AX-212394732	rs1201201830	Het	Missense	p.His208Tyr	0.00000398	23.6	-	No	1
RAI1	AX-212397797	rs764594243	Het	Missense	p.Arg1576His	0.000012	25.4	-	No	1
RAI1	AX-317022819	rs1386459275	Het	Missense	p.Glu514Ala	0.00000408	24.3	-	No	1
RAI1	AX-83106472	rs142415050	Het	Missense	p.Arg1217Gln	0.0006023	23.6	Likely benign	1	4
RAI1	AX-83555249	rs142981643	Het	Missense	p.Ala1679Val	0.0005847	23.5	-	1	6
RAI1	AX-83565099	rs149701833	Het	Missense	p.Arg44Gln	0.0001131	24.6	-	No	2
SH2B1	AX-119163961	rs1198783948	Het	Missense	p.Arg156His	0.00000796	25.2	-	No	2
SH2B1	AX-212377854	rs369196090	Het	Missense	p.Arg330Trp	0.0000199	28.1	-	No	1
SH2B1	AX-212379912	rs372038271	Het	Missense	p.Gly205Arg	0.0000519	24	-	No	1
SH2B1	AX-212380441	rs779520493	Het	Missense	p.Ser202Phe	0.00000851	25.5	-	No	1

SH2B1	AX-212393643	rs772678200	Het	Missense	p.Gly545Ser	0.00002	28.4	-	No	1
SH2B1	AX-213432141	rs747007890	Het	Missense	p.Glu353Lys	0.00000398	24.6	-	No	1
SH2B1	AX-213432535	rs749598219	Het	Missense	p.Pro22Gln	0.0000655	21.9	-	No	1
SH2B1	AX-82968702	rs142515048	Het	Missense	p.Ser616Pro	0.0003142	23	-	No	2
SH2B1	AX-82974958	rs190981290	Hom	Missense	p.Ala663Val	0.007079	18.6	Benign	1	1
SIM1	AX-212371572	rs763116551	Het	Missense	p.Arg550Cys	0.00001194	34	-	No	1
SIM1	AX-212372142	rs770543844	Het	Missense	P.Ile564Thr	0.00006364	20.7	-	No	1
SIM1	AX-212378385	rs199543656	Het	Missense	p.Ser622Phe	0.0001309	18.54	Uncertain significance/Obesity due to SIM1 deficiency	No	1
SIM1	AX-213424575	rs776861600	Het	Missense	p.Ser440Leu	N/A	24.4	-	No	1
SIM1	AX-213429395	rs1435329451	Het	Missense	p.His520Arg	0.00003185	18.91	-	No	1
SIM1	AX-82979717	rs138546433	Het	Missense	p.Ile128Thr	0.001333	25.8	Conflicting interpretations of pathogenicity Benign(2);Uncertain significance(1)	No	2
SIM1	AX-83223382	rs201812554	Het	Missense	p.Thr464Ser	0.00007078	18.22	-	No	1
SLC35D3	AX-213427074	rs767002059	Het	Missense	p.Val277Ala	0.0000677	23.9	-	No	3
SLC35D3	AX-317017436	rs1258598258	Het	Frameshift	p.Glu362GlyfsTer16	0.00003188	-		No	1
SLC35D3	AX-83162938	rs144509540	Het	Missense	p.Tyr401Ser	0.0003263	24.3	-	No	3
SLC35D3	AX-83306630	rs200232532	Het	Missense	p.Ala194Thr	0.0001231	17.73	-	No	3
UCP1	AX-169123249	rs141520915	Het	Missense	p.Gly57Ser	0.0002346	26	N/A	No	2
UCP1	AX-212379114	rs375694859	Het	Missense	p.Glu69Lys	0.0003062	27.1	-	No	1
UCP1	AX-83324147	rs150886806	Het	Missense	p.Gln44Arg	0.0004295	28.2	-	No	1
UCP1	AX-83346312	rs201976256	Het	Missense	p.Arg153His	0.0000915	25.5	-	No	1
UCP3	AX-181511055	rs763845388	Het	Missense	p.Pro237Leu	0.0000159	31	-	No	1
UCP3	AX-213430494	rs1179184494	Het	Frameshift	p.Glu205ArgfsTer37	0.00000398	33	-	No	2
UCP3	AX-83576137	rs74907838	Het	Missense	p.Ala111Val	0.0008888	14.45	-	No	3
WNT10B	AX-175167975	rs183587423	Het	frameshift	p.Gly119ArgfsTer?	0.00003189	7.118	-	No	3
WNT10B	AX-212368800	rs760486607	Het	Missense	p.Arg302Pro	0.0000439	26.7	-	No	1
WNT10B	AX-212391893	rs144672721	Het	Missense	p.Arg302Cys	0.0001435	31	-	No	1
WNT10B	AX-83089829	rs146010731	Het	Missense	p.Ile285Thr	0.0006309	26.7	-	No	3

Abbreviations are as follows: **MOI**. Disease Mode of Inheritance, **HGVS**. Protein position HGVS, **MAF(%)** gnomAD MAF, Subject Genotype (Het: Heterozygous, Hom: Homozygous, Hemi; hemizygous

Table 7.7 Details of variants in autosomal recessive obesity genes that have been reported to have heterozygous mutations possibly contributing to obesity.

GENE	*AX ID*	*MOI*	*Info Summary*	*Genotype*	*Type of variant*	*HGVS*	*MAF GnomAD*	*CADD*	*Clinvar*	*Publication*	*No of carriers*
ADCY3	AX-213439512	AR	rs779014882	Het	Missense	p.Arg478Cys	0.00003535	24.7	-	-	1
ADCY3	AX-213444255	AR	rs768733760	Het	Missense	p.Gln275His	0.000003986	33	-	-	2
ADCY3	AX-213445133	AR	rs777689844	Het	Missense	p.Arg453His	0.0000177	27.2	-	-	1
ADCY3	AX-213446975	AR	rs762205806	Het	Missense	p.Arg1122Trp	0.00001428	23.3	-	-	1
ADCY3	AX-213450173	AR	rs752934161	Het	Missense	p.Ile1042Thr	0.00001592	34	-	-	3
ADCY3	AX-213452686	AR	rs529302905	Het	Missense	p.Thr646Met	0.00002392	24.7	-	-	1
POMC	AX-212363058	AR	rs760210424	Het	Missense	p.Thr38Ile	0.00001193	22.9	-	-	1
POMC	AX-212389097	AR	rs1476776978	Het	stop gained	-	-	-	-	-	1
POMC	AX-317019782	AR	rs1300055734	Het	frameshift	p.Gly101LysfsTer58	0.00003393	-	-	-	1
POMC	AX-317022348	AR	rs8192606	Het	Missense	p.Pro132Ala	0.000007843	9.9	-	-	
PCSK1	AX-212385699	AD	rs937916641	Het	Missense	p.Ala36Thr	0.00000398	27.3	-	No	1
PCSK1	AX-115097541	AD	rs368253923	Het	Missense	p.Arg740Trp	0.00000796	24.4	-	No	1
PCSK1	AX-83177353	AD	rs146545244	Het	Missense	p.Met125Ile	0.0001804	21.8	-	No	1
PCSK1	AX-212393935	AD	rs539542819	Het	Missense	p.Thr377Met	0.0000119	20.2	-	No	1
PCSK1	AX-181569008	AD	rs372128150	Het	Missense	p.Val261Met	0.0000318	23.2	-	No	1

Abbreviations are as follows: **AX ID**: SNP ID in Axiom Array. **MOI.** Disease Mode of Inheritance, **HGVS**. Protein position HGVS, MAF(%) gnomAD MAF, **Subject Genotype** (Het: Heterozygous, **Hom**: Homozygous, Hemi; hemizygous

7.3.3 Carriers of potential rare deleterious variants in obesity and syndromic obesity genes

The variants have been found in a total of 212 individuals, which represents a total of 11.1% of the overall samples that have successfully passed genotyping quality control. As shown in Figure 7.5, the highest number of individuals belong to the PMMO and white ethnic groups, which is expected as the PMMO and white ethnic groups are the dominant group of the genotyping samples. The percentages of affected participants in the different cohorts are 86.7%, 5.1%, 2.7% and, 5.1% for the PMMO, Endobarrier, Geronimo and, Gravitas cohorts respectively. The percentages of affected participants in each self-reported ethnic group are 104, 44, 34 and, 30 for the White, Black, Asian and Mixed ethnic groups.

Furthermore, the analysis showed that the black ethnic group has a higher percentage of carriers, 22%, compared to the white ethnic group, which is considered to be around 8%. These over-representations could be due to various factors. The first is that the African population has more genetic variations generally compared to the white ethnic group. Secondly gnomAD was used a reference to assess the allele frequency of rare variants, and variants that appear rare in gnomAD may be more common in less well-studied populations such as the black ethnic group. Because of this we may have inadvertently included variants that are not, in fact, rare in certain ancestry groups. Thirdly, it is possible that what we are seeing is a real difference, but it may reflect an ascertainment bias – for example, people of white ancestry may be referred to bariatric services with less severe or complex obesity phenotypes than people with black, Asian or other ethnicities (either by self-selection or the physician's unconscious bias).

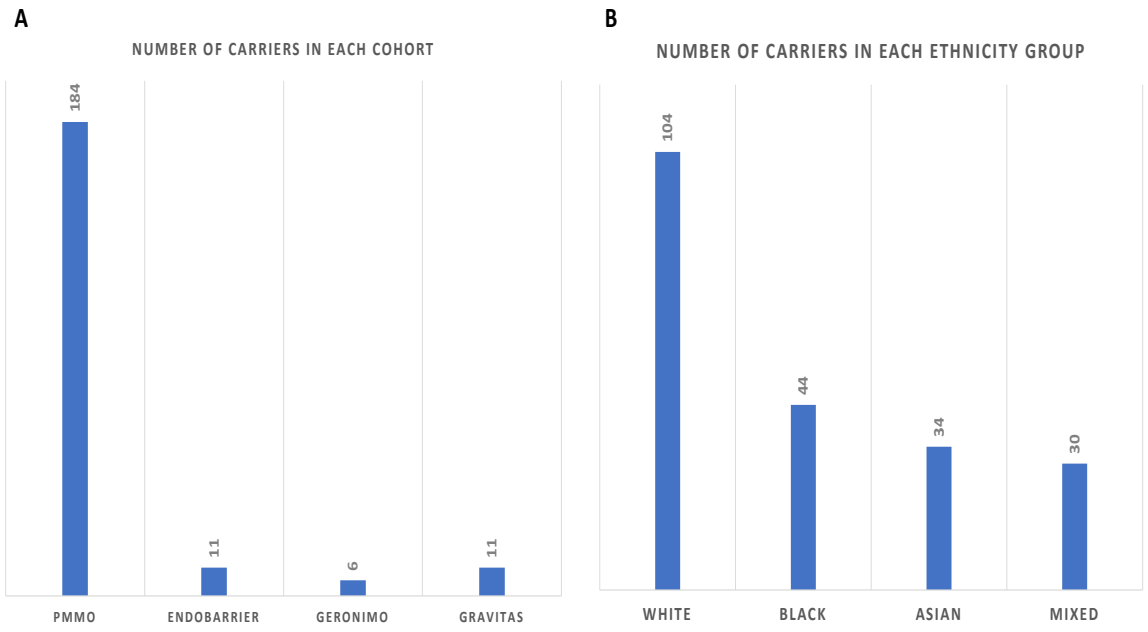


Figure 7.5: Number of the identified subjects with potential rare deleterious variants in each research group (A) and ethnicity (B).

As observed in the earlier analysis of the WES, the subjects identified here with rare deleterious variants also follow either the typical monogenic form of inheritance or a potentially oligogenic form of inheritance. In total, 198 individuals have a single deleterious rare variant as monogenic form of inheritance, and 14 follow a potentially oligogenic form of inheritance. The identified oligogenic events are summarised in Table 7.8. Some of the oligogenic combinations were observed in earlier analysis as *SH2B1-MBD5* in Chapter 4 and, *RAI1-MAGEL2* in Chapter 5.

Table 7.8 Summary of the identified cases with potentially oligogenic mode of inheritance.

CASE	GENE	PROBESET_ID	RS ID/POS	TYPE OF VARIANT	CADD	GNOMAD
1	*PTEN*	AX-164419336	10:87933124	Missense	28.9	-
	SLC35D3	AX-83306630	rs200232532	Missense	17.73	0.0001231
2	*DYRK1B*	AX-213456043	rs35858874	Missense	23.3	0.00000399
	MBD5	AX-83566711	rs145475623	Missense	19.42	0.0004813
3	*MBD5*	AX-83431587	rs145808884	Missense	24.4	0.0000797
	SH2B1	AX-119163961	rs1198783948	Missense	25.2	0.00000796
4	*MBD5*	AX-82890333	rs138639760	Missense	23.6	0.0001715
	MCHR1	AX-212389578	22:40681355	frameshift	N/A	0.00000406
5	*MBD5*	AX-213456379	rs142913108	Missense	19.86	0.00000796
	RAI1	AX-83106472	rs142415050	Missense	23.6	0.0006023
	RAI1	AX-83106472	rs142415050	Missense	23.6	0.0006023
	MAGEL2	AX-213457006	rs1391195912	Missense	16.32	0.00000713
6	*MBD5*	AX-147188889	rs752035001	Missense	27.3	0.0000716
	PTEN	AX-122990578	rs11202592	5 prime UTR variant	22.2	0.003765
7	*UCP3*	AX-213430494	rs1179184494	Frameshift	33	0.00000398
	POMC	AX-212389097	2:25161633	stop gained	-	-
8	*MCHR1*	AX-213433144	rs368183075	Missense	22.2	0.000016
	MAGEL2	AX-212385443	rs576711188	Missense	25.5	0.000556
9	*RAI1*	AX-168884593	rs376429075	Missense	16.1	0.0000441
	SLC35D3	AX-317017436	rs1258598258	Frameshift	-	0.00003188
10	*AFF4*	AX-213430847	rs138207289	Missense	29.7	0.000008791
	MCHR1	AX-213455206	rs765877994	Missense	20.9	0.0000278
11	*RAI1*	AX-146352297	rs549244691	Missense	24.8	0.0002684
	MAGEL2	AX-212364645	rs778127117	Missense	14.53	0.000004052
12	*PTEN*	AX-164419336	10:87933124	Missense	28.9	-
	SLC35D3	AX-83306630	rs200232532	Missense	17.73	0.0001231
	POU3F2	AX-212361621	rs1321043636	Missense	28.6	0.000003977
	ADCY3	AX-213452686	rs529302905	Missense	24.7	-
13	*MC4R*	AX-317015454	rs138281308	Missense	20.2	0.0006923
	KSR2	AX-212372740	rs576454880	Missense	24.6	0.0000127
14	*CEP290*	AX-83307392	rs183655276	Missense	23.8	0.002316
	AFF4	AX-83420563	rs141068875	Missense	22.5	0.0008073

Abbreviations are as follows: AX ID: SNP ID in Axiom Array. **MOI**. Disease Mode of Inheritance, **HGVS**. Protein position HGVS, **MAF(%)** gnomAD MAF, **Subject Genotype** (Het: Heterozygous, Hom: Homozygous, Hemi; hemizygous

7.3.4 CNV analysis of the fixed region

In the array design, probes were included both to detect known CNVs with defined breakpoints (fixed region analysis), and discovery of novel CNVs in regions containing genes of interest. For the scope of this thesis, only the fixed region analysis was performed. Fixed region analysis is applied when the breakpoints of the CNVs (228 pre-specific regions) are defined and probe sets covering that specific region of interest included on the custom array.

CNV calling was created based on the universal reference, as all plates were processed at the same time in the same lab under similar conditions. The quality overall looked fine (Appendix 2). Overall, all samples passed the Axiom QC metric for CNV MAPD and Waviness-SD except for 274 individuals (Figure 7.6). Plate reference instead of the universal reference was used for two plates which showed a high QC failure among the samples. The plate reference has greatly improved the QC of those two plates.

From the fixed region CNV analysis, three type of CNVs were identified that are hemizygous in the obese subject. This includes 10q23.31 in four subjects, 20q13.3 in one subject and 16p11.2 in three subjects (summarised in Table 7.9). One of the subjects was known to carry a deletion of 16p11.2 as it was detected initially by WES and was used as a control the genotyping to confirm efficiency of the array in detecting CNV.

Table 7.9 Summary of copy number variations detected from fixed region CNV analysis.

REGION_NAME	START (BP)	END (BP)	NO OF HEMIZYGOUS CARRIERS
OBESITY-10Q23.31	89623195	89728532	4
OBESITY-20Q13.3	62300001	62400000	1
OBESITY-16P11.2-V1	29500001	30100000	3

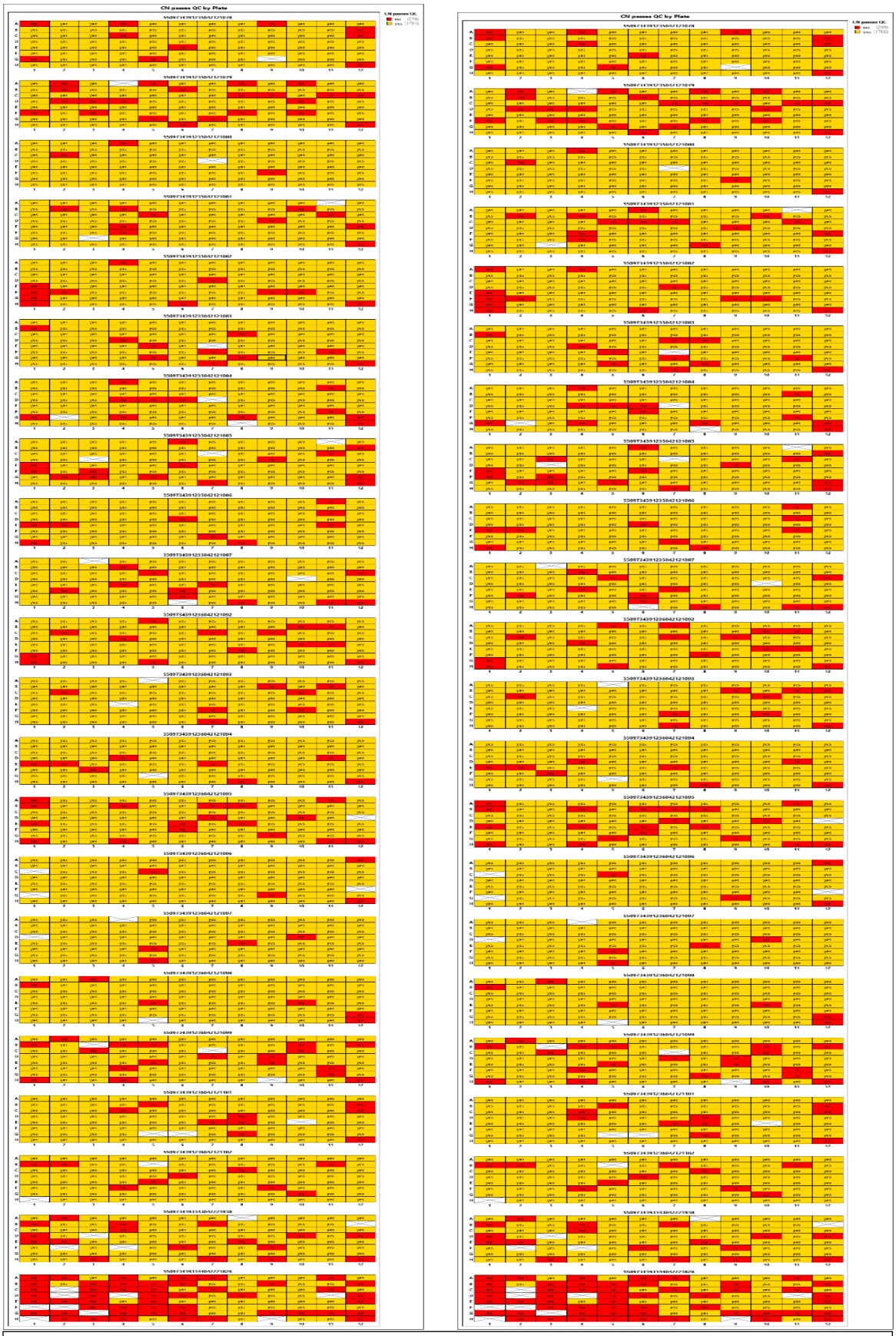


Figure 7.6: Layout of all plates indicating samples passing CNV Quality Control. The total number of plates included in the genotyping is 22. Yellow indicates the samples that have passed QC and red indicates the samples that failed QC in each plate.

16p11.2 deletion (593kb)

I identified three individuals with a heterozygous 16p11.2 deletion. The length of this CNV is 593 kilobases and covers around 30 genes. As described earlier in the introduction, this CNV 16p11.2 is highly associated with obesity and morbid obesity and has been found in several independent studies [88, 89]. Duplication of this region is known to be associated with underweight, schizophrenia and autism [89]. Previous reported this CNV with a frequency ranging from 1% to 7% [88, 324, 325].

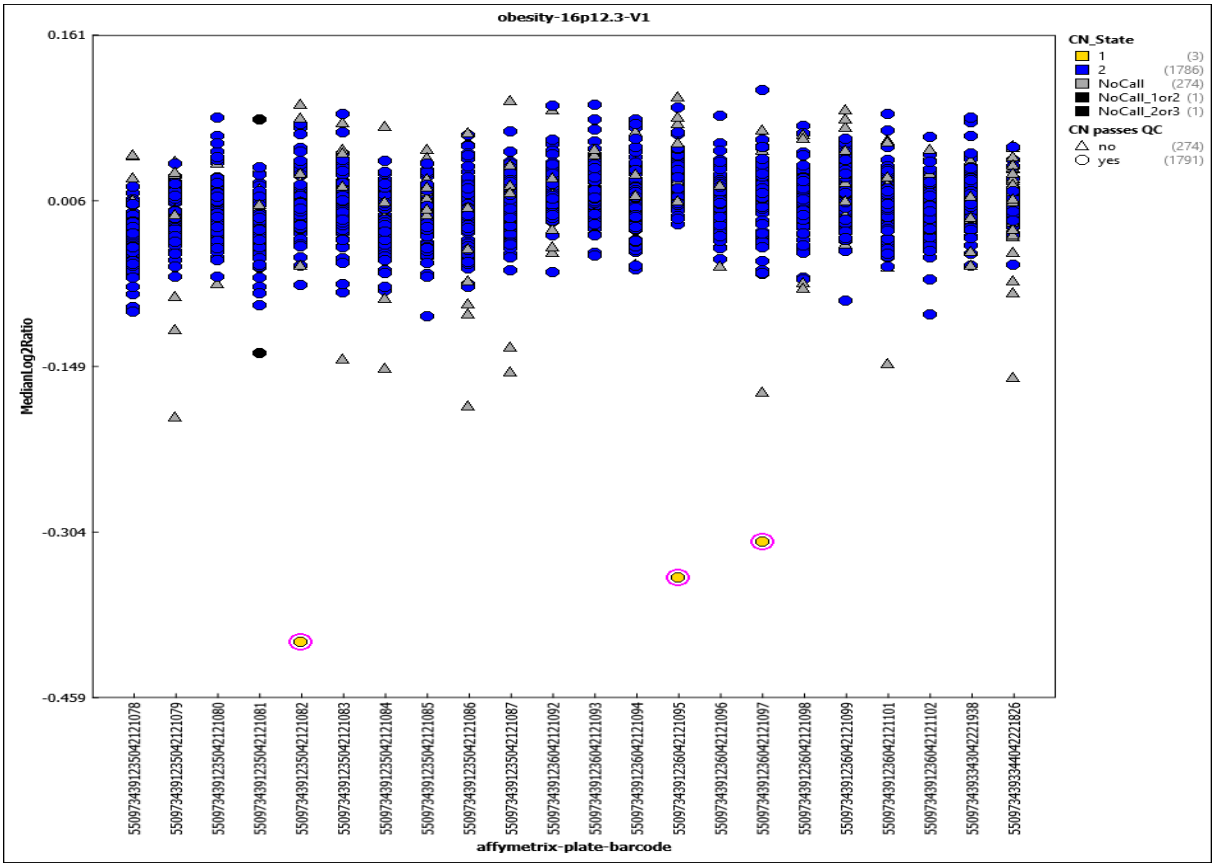


Figure 7.7 Fixed region CNV plots of the subjects with 16p11.2. The X axis is represents 22 included in the genotyping and Y Axis represents calculation of the log2 ratio which is comparison of the intensity of a probe set of a sample according to reference total intensity. Yellow circle indicates the subjects with hemizygous 16p11.2 CNV in each plate. Blue circle indicates the subjects with 2 copy/with no deletion.

20q13.3

One individual was identified with a deleterious deletion of 20q13.3 CNV. The obesity-20q13.3 CNV was initially reported as 77 KB length and includes seven genes where the strongest candidate gene that affects BMI is *ARFRP1* (ADP Ribosylation Factor Related Protein 1) [326]. The protein encoded by this gene is involved in the G protein couple receptor signaling pathway. As with 16p11.2, duplication of this region has been found to be associated with leanness [326]. This region is not found in public database.

10q23.31

Four individuals were identified with a heterozygous deletion of 10q23.31 CNV. This CNV encompasses the PTEN gene which is a tumor suppressor gene that is involved in regulation of cell proliferation [327]. It is also involved in wider cellular roles including cell process, survival proliferation, energy metabolism and, insulin pathway [327]. Loss of function or reduced activity of one copy through a point mutation or deletion can lead to the development of numerous types of cancers [328]. The most commonly known is Cowden syndrome which is characterised by cell hamartomas, macrocephaly, thyroid problems, mental retardation/developmental delay [329]. The features of Cowden syndrome vary amongst carriers of *PTEN* mutations, that is reviewed in [329].

The percentages of affected participants in the different cohorts based on the findings from rare variants and CNV analysis are 87.2%, 5.%, 2.7% and, 5.% for the PMMO, Endobarrier,

Geronimo and, Gravitas. None of those carriers have additional variants as found with the oligogenic mode of inheritance.

7.4 Discussion

In this Chapter, I have described the quality control results at sample and marker levels and results of a partial CNV analysis of rare and putatively pathogenic variants, using data from the custom-designed Axiom array described in Chapter 6.

The genotyping methodology for detecting rare variants described here is still a developing approach, and best practice is not yet well established, despite the fact that the array producer ThermoFisher characterises the approach as a highly accurate in detecting extremely rare variants and common variants. Thus, I aimed to test the ability and efficiency of this approach in detecting rare variants on a large number of obese individuals. One of the ways to test this approach was to compare the results with the previously exome sequenced samples and also through the inclusion of duplicate samples.

The genotyping analysis was performed on a total of 2112 samples, and the overall quality of both samples and markers was successful at a high quality rate. Additionally, the samples with previous WES data all confirmed the same detected variant through genotyping, except one, which was assigned as misclassified, although it was clearly visible in the cluster plot. This increases confidence that the approach is conservative. When the COVID situation allows, it is planned to confirm all predicted genotypes by a separate sequencing approach, but this was not possible within the timescale of this PhD.

From this first analysis of the genotyping data focusing on the monogenic and syndromic obesity genes, a total of 161 potential causative variants have been identified in a total of 40 genes. Of these, 149 follow the autosomal dominance mode of inheritance, and 9 the autosomal recessive

mode of inheritance and 4 follow the X-linked recessive mode of inheritance. In addition, a total of 3 types of rare CNV that are known to be associated with obesity, 10q123.31, 20q13.3O, 16p11.2-V1 , were identified in a total of 8 individuals.

Therefore, a potential diagnosis has been found in 220 individuals which represents a prevalence of 11% of the participants(combined from the PMMO, Endobarrier, Geronimo and, Gravitas cohorts.

In comparison to the frequency obtained from the re-analysis of WES described in Chapter 5 (which is around 31%), this is much lower. This is not unexpected since one of the limitations of the array-based genotyping technology used is inability to detect novel variants – it is entirely possible that participants carry variants not included on array that might have been detected by WES.

Another possible limitation is that the updated Thermofisher genotype calling algorithm could error on the side of being too stringent, resulting in false negatives, as with the PTEN mutation shown in Figure 7.1 (10) . In the initial genotype calling, based on the old algorithm, the number of detected rare deleterious variants in obesity and syndromic obesity was much higher (possibly due to false positives), but the call rate dropped dramatically with the newer algorithm. For example, for the *SH2B1* gene, a total of 27 variants were identified by the first algorithm, whereas only 9 variants were detected by the second algorithm. I have used the more conservative figures in my analyses for this thesis, but it is unknown how many false negatives (like the PTEN mutation) there might have been. Thus, this needs to be further investigated by sequencing to determine which samples carry variants and which do not.

Comparing the frequency with other studies, which differ in many respects such as study design and methodology, gives variable results. A study in Netherlands based on 1230 obese individuals has identified a frequency of 3.9% [330]. While another study based on obese children from Guadeloupean Afro-Caribbean found a frequency of >15% [331]. In addition, a recent paper study based on 92 subjects with severe early-onset childhood obesity identified a frequency of 8% [332].

In parallel with the findings from Chapter 4, where I discussed a proband with an oligogenic form of obesity, including variants in *SH2B1, POGZ and MBD5,* and from Chapter 5 (the re-analysis of WES) where we identified two cases of oligogenic obesity that involve *SH2B1* together with variants in other obesity genes, here from the genotyping analysis I have identified an additional 14 cases of apparent oligogenic inheritance. Two of the oligogenic gene combinations were observed in earlier analyses (*SH2B1-MBD5* in Chapter 4 and, *RAI1-MAGEL2* in Chapter 5). The data presented in this thesis strongly suggest that obesity could exhibit a different mode of inheritance than the ones already recognised clinically.

There is also an ambiguity about the status of heterozygotes. Consistent with several studies that have reported that some obesity genes (*POMC*, *ADCY3* and *PCKS1*) with autosomal mode of inheritance, also have intermediate obesity phenotypes in heterozygous individuals, a total of 15 heterozygous rare predicted deleterious variants were observed in 18 individuals. Further functional studies and family segregations to explore the natural history of the disease in these participants are needed [66, 332, 333].

Several of the identified cases here will require further confirmation and clinical diagnosis and evaluation. Individuals with rare deleterious variants in PTEN, which is known to increase the risk of developing Cowden syndrome and various type of cancers, autism and, macrocephaly, require earlier genetics counselling [328]. Patients with potential variants in syndromic obesity genes such as *ALMS1*, *RAI1*, *BBS* and, *NTRK2* will require detailed clinical examination to assess the other feature associated with each syndrome, such as behavioural problems, intellectual disability, cognitive impairment and dysmorphology. This is could be vital for future treatment options and decisions with respect to reproductive choices, both for the proband, and for family members. It is important to remember that whereas the phenotypes may be only partially penetrant, with absent or mild clinical features in the proband themself, their children and family members may experience more severe effects. Also, probands may require enhanced investigation and monitoring: for example. individuals carrying rare deleterious variants in IGFS1, which is known to cause congenital central hypothyroidism (C-CH), will require regular thyroid gland screening [225, 226].

The current analysis is not yet completed to provide the full picture as the discovery analysis of CNV and mouse model obesity genes will need to be addressed in the next stage of this study. Also, laboratory-based validation of all the mutations identified could not be done during my PhD due to the COVID-19 pandemic. Nevertheless, the identification of a number of individuals with putative Mendelian causes of obesity in a cost-effective way indicates that the approach could optimised to be useful as a first level diagnostic screening in severe obesity. The majority of the identified variants have not been reported before which will require a further molecular

investigation at cellular and functional level. The implications of the variants identified for baseline patient characteristics and for outcomes of bariatric surgery or other therapeutic approaches to weight loss and diabetes remission will be explored in the next Chapter.

CHAPTER 8: IMPLICATIONS OF RARE MENDELIAN FORMS OF OBESITY FOR TREATMENT OUTCOMES.

8.1 Introduction

This Chapter describes the implication of rare Mendelian forms of obesity for outcomes after two different therapeutic interventions: bariatric surgery and the “milk diet”.

Nowadays, there are a wide range of treatments available to treat obesity such as (1) lifestyle intervention, (2) pharmacotherapy and (3) surgery. The treatment option depends on various factors including the patient’s BMI, measurements of WHR and WC, the patient’s health and other treatments, as well as local access to care [128].

In most cases, the first therapeutic intervention tried in severe obesity is caloric restriction [334]. This can be achieved by a range of methods, with proponents for low-carbohydrate, low fat, ketogenic or other approaches [335, 336]. In clinical services, one very commonly used intervention is the “milk diet”, which is a total diet replacement approach, whereby patients consume approximately 1200 kcal per day for 8 weeks followed by 16-week period reintroduction of food [159, 337, 338]. The “milk diet” is most often used in the preparatory period before bariatric surgery, either to achieve pre-surgery weightless into to demonstrate compliance, and/or to shrink the liver in order to facilitate laparoscopic surgery.

Bariatric surgery, especially RYGB and VSG, has shown excellent outcomes after surgery including long-term weight loss and improvement or permanent remission of co-morbidities including T2D, hypertension, dyslipidaemia and other health factors [140].

Despite the successful outcomes, responses of patients to treatment and surgery vary in terms of weight loss and diabetes remission [339]. Around 10-40% of individuals who undergo bariatric surgery do not achieve a successful long-term weight loss (>20% weight loss) or experience significant weight regain. In addition, other studies have shown that some individuals do not achieve diabetes remission, or experience relapse of T2D five years post-surgery. Several studies have reported a contribution of genetics to the variability of weight loss after bariatric surgery [340, 341].

8.2 Aims of the study

- To examine the baseline characteristics of participants with putative Mendelian forms of obesity, compared to the rest of the cohort (for PMMO and Geronimo cohorts separately)
- To explore whether there are discernible effects on weight loss during dietary intervention in the Geronimo cohort, or after bariatric surgery in the PMMO cohort.
- To analyse whether there is preliminary evidence for differences in the risk of longer term weight regain (>2 years post-surgery) in the PMMO cohort
- To explore data on diabetes remission after bariatric surgery in the PMMO cohort

8.3 Results

To investigate the percentage of weight loss in subjects with rare deleterious variants in monogenic obesity genes, two groups of subjects were included: PMMO and Geronimo participants. These two groups have weight information pre- and post-treatment/surgery and were analysed separately.

8.3.1 Comparison analysis of the percentage of weight loss between carriers and non-carriers in the Geronimo group (after "milk diet")

In the Geronimo group, a total of 101 subjects were included for genotyping. Of these, seven individuals were identified as having putative Mendelian obesity. The average percentage weight loss of the non-carrier group (n=94) is 16.2±6.33 while the average percentage of weight loss of the carriers group (n=7) is 15.5 ± 4.23 (Figure 8.1). Unsurprisingly, given the limited power of this small cohort, there is no significant difference between the two groups (p-value of 0.89).

To explore the data further, the percentage weight loss trajectories of each carrier was plotted separately along with the whole group mean, as shown in Figure 8.2. This reveals that 5/7 of the carriers had a lower percentage weight loss than the mean for the non-carriers at 24 months.

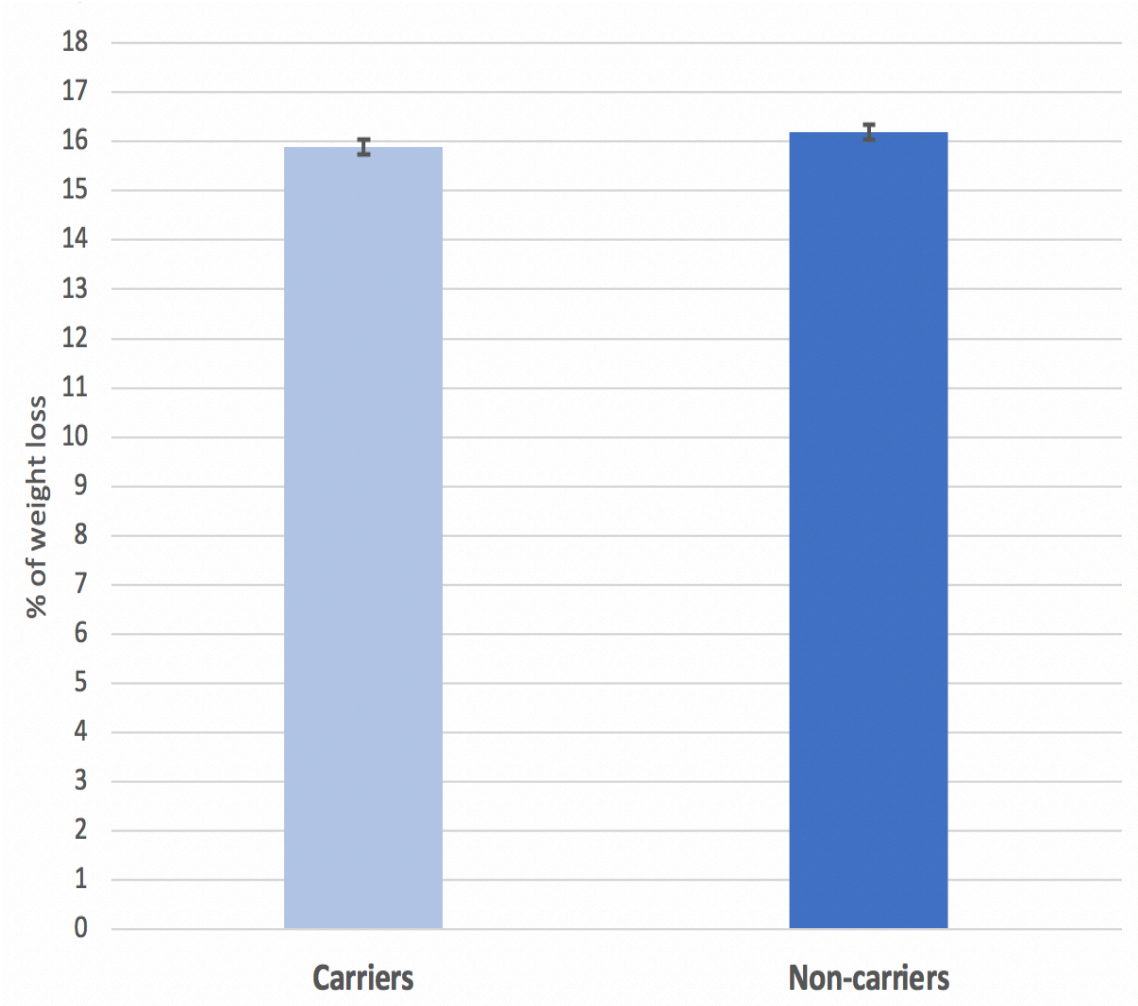


Figure 8.1 Percentage of weight loss in the Geronimo group . Carriers (no=7) is 15.5 ± 4.23 and non-carriers (no=94) is 16.2±6.33.

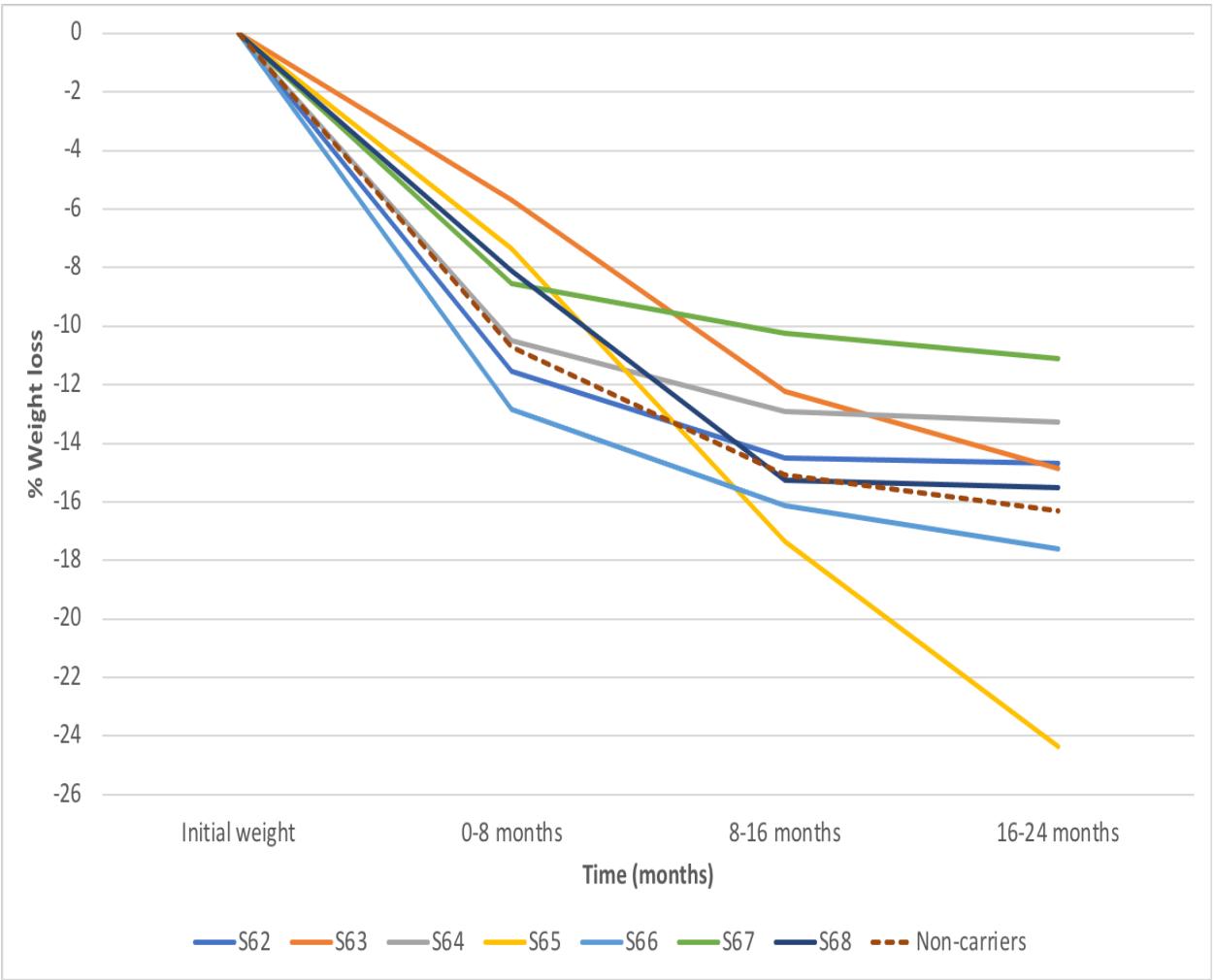


Figure 8.2 Percentage of weight loss trajectory between the carriers and non-carriers of the Geronimo group. The chart shows the percentage of weight at different time points from 0-24 months. Solid lines (carriers) Dashed line (non-carriers). Details of the identified mutation in each carrier is shown in Table 8.1 below.

Table 8.1 Details of the identified mutations in each carrier shown in Figure 8.2.

SUBJECT	GENE	RS ID	TYPE OF VARIANT	GNOMAD (MAF)	CADD
S65	*RAI1*	rs142981643	Missense	0.0005847	23.5
S66	*MBD5*	rs748142226	Missense	0.0000358	26.5
S67	*PCSK1*	rs146545244	Missense	0.0001804	21.8
S68	*IGSF1*	rs142822502	Missense	0.0006018	18.24
S64	*PTEN*	10:87931034	Splice	0.0002814	25.7
S63	*ADCY3*	rs768733760	Missense	3.986E-06	33
S62	obesity-16p11.2-V1				

8.3.2 Comparison analysis of the percentage of weight loss between carriers and non-carriers in the PMMO cohort

8.3.2.1 Initial weight loss after bariatric surgery (PMMO cohort)

As described in the cohort Chapter, the PMMO subjects have weight information pre-surgery and post-surgery at year 1 and year 2. To determine the implications of Mendelian forms of obesity on the initial weight loss after bariatric surgery, an overall comparison took place between subjects with potentially-causative rare deleterious variants (described here as carriers) and the rest of the PMMO group who were not identified as having relevant variants, from WES or genotyping analysis (non-carriers). The comparison was done for the BMI baseline and percentage of weight loss at year 2 post-surgery. Since PMMO is an ongoing study, some patient information was not available as some of the research participants had not had surgery or been followed-up at the time this analysis was done. The number of individuals varies in each analysis depending on the availability of weight information in each group, as summarised in Table 8.2. Overall t-test was performed using SPSS. Then the same analysis was performed for each surgery type separately.

Table 8.2. Summary of number of individuals with available weight information in each group (carriers and non-carriers).

		BMI BASELINE	YEAR 1	YEAR 2	YEAR 1 AND YEAR 2
CARRIERS	Whole group	200	121	75	68
	Gastric Band	6	4	3	3
	Gastric bypass	87	69	42	39
	Gastric sleeve	59	46	28	25
NON-CARRIERS	Whole group	1456	896	501	429
	Gastric Band	49	29	20	12
	Gastric bypass	648	469	318	283
	Gastric sleeve	397	257	145	128

Year 1: Individuals with weight information at baseline and year 1 post-surgery
Year 2: Individuals with weight information at baseline and year 2 post-surgery
Year 1 and Year 2: Individuals with weight information at baseline, years 1 and 2 post-surgery

To examine the baseline characteristics of participants with putative Mendelian forms of obesity, compared to the rest of the cohort, T-test was used. There was a significance difference between the two groups of carriers and non-carriers (p-value=0.04)

A total of 121 carriers and 896 non-carriers have weight information at year 1. The mean (± SD) percentage weight loss at year 1 for the carriers is 27.84 (10.23) while for non-carriers it is 26.77 (12.04). The distributions of percentage of weight loss at year 1 for the carriers and non-carriers group are shown in Figure 8.3. At year 1, 21 of the carriers (17.3%) had lost less than 20% of their initial weight, while 201 from the non-carrier group (27.4%) had lost less than 20%.

A total of 75 carriers and 501 non-carriers had weight information at year 2. The mean (± SD) percentage weight loss at year 2 of the carriers is 26.94 (11.75) while for non-carriers it is 28.06 (11.99). The distribution of percentage of weight loss at year 1 and year 2 of the carriers and non-carriers group is shown in Figures 8.3 and 8.4

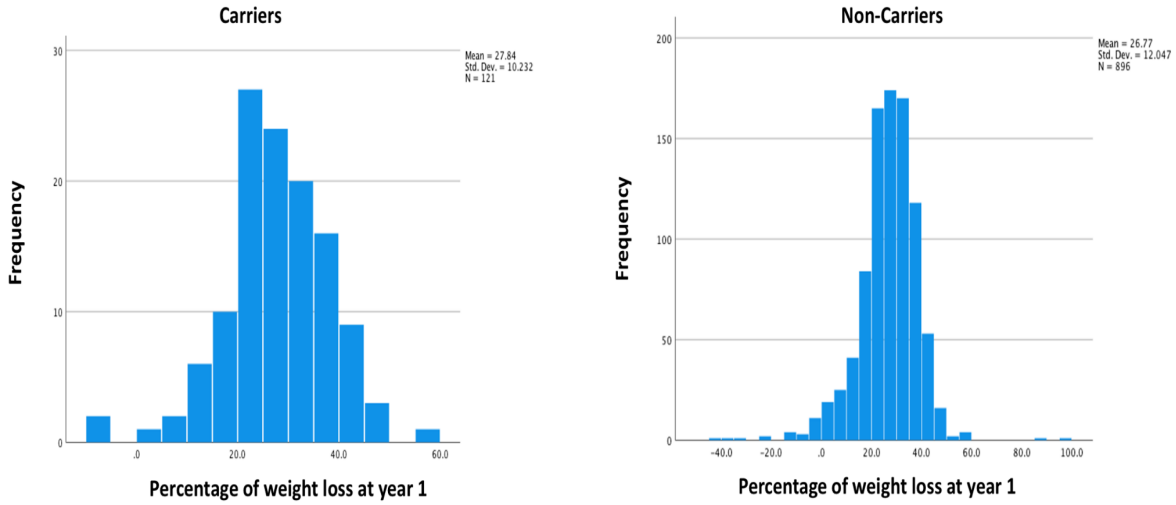


Figure 8.3 Distribution of percentage of weight loss at year 1 post-surgery of the carriers and non-carriers group. Histogram of the percentage of weight loss at year one post-surgery for carriers (n=121) and non-carriers (n=896) group in the PMMO.

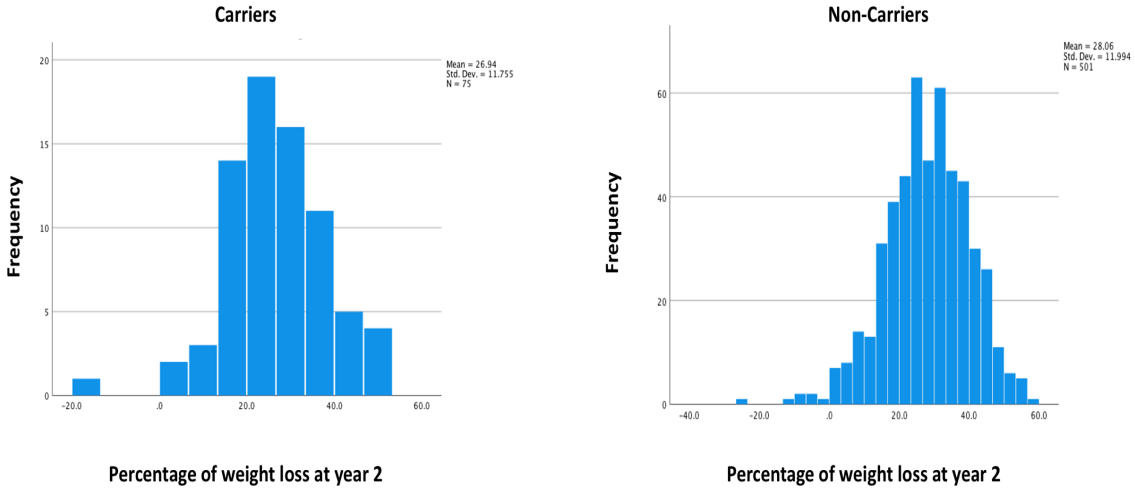


Figure 8.4 Distribution of percentage of weight loss at year 2 post-surgery of the carriers and non-carriers group. Histogram of the percentage of weight loss at year two post-surgery for carriers (n=75) and non-carriers (n=501) group in the PMMO.

A statistical analysis was carried out of weight loss 2 years after surgery of the whole two groups: carriers (N=75) and non-carriers (501) using an independent t test. As shown in Table 8.3 and Figure 8.5 (A), there is no statistical significance difference in percentage weight loss between the whole two group Tables.

8.3.2.2 Initial weight loss after different types of bariatric surgery

Following this, the relation between carriers and non-carriers in each surgery group was investigated. Figures 8.5, 8.6 and Table 8.3 show the comparison analysis between each of the surgery groups. Overall, in the sleeve gastrectomy group, the non-carriers group (25.3±12.6) had a higher percentage weight loss compared to the carrier group (20.4±11.8). However, this did not reach statistical significance.

On the other hand, there was almost no difference in percentage weight loss between the two groups: carriers (30.4±9.7) and non-carriers (30.3±10.8)in the gastric bypass. In the gastric band group, there is a huge difference in terms of the number of individuals in each group and that could be introducing some uncertainty and bias. Overall, there is no significant statistical difference between the two groups: carriers (34.0±15.8) and non-carriers (18.3±14.0).

Table 8.3 Summary statistic of the comparison analysis between carriers and non-carriers at 2 years time-point

Group		N	Mean	Std. Deviation	Std. Error Mean	Sig.
Whole group	Carriers	75	26.94	11.75	1.36	0.685
	Non-carriers	501	28.06	11.99	0.54	
Gastric Band	Carriers	3	34.61	15.83	9.14	0.847
	Non-carriers	20	18.34	14.02	3.13	
Gastric Bypass	Carriers	43	30.44	9.73	1.48	0.688
	Non-carriers	316	30.38	10.84	0.61	
Gastric Sleeve	Carriers	28	20.45	11.86	2.24	0.612
	Non-carriers	145	25.35	12.67	1.05	

The t-test analysis was performed on the whole group of carriers and non-carriers and separately for each surgery group. N= number of participants in each group. Sig= p-value

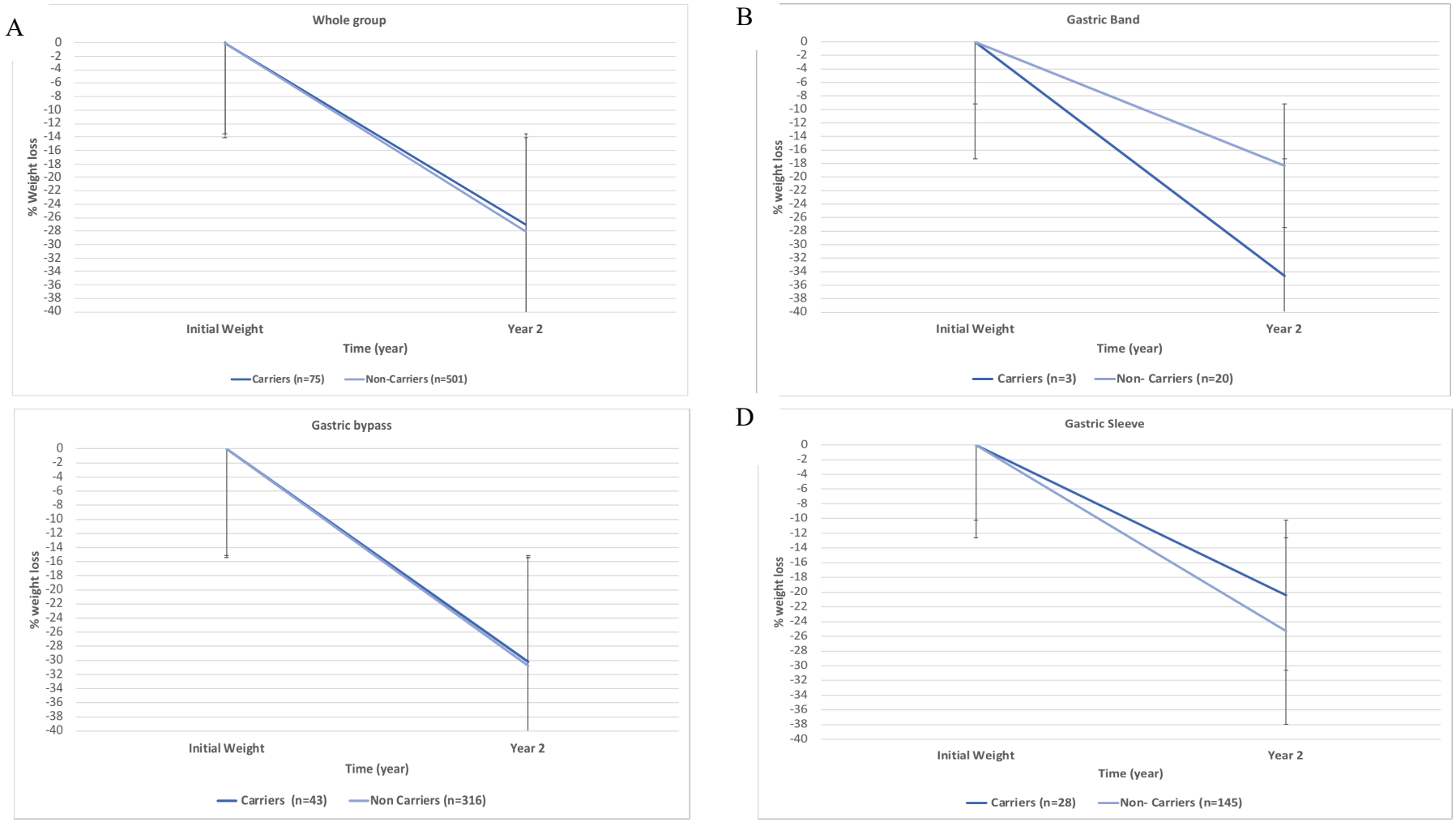


Figure 8.5 Comparison of the average weight loss percentage between carriers and non-carriers at 2 years timepoint. The whole group of carriers (26.9±11.7) and non-carriers (28.0±11.9). (B) The gastric band carriers (34.0±15.8) and non-carriers (18.3±14.0). (B) The gastric bypass carriers (30.4±9.7) and non-carriers (30.3±10.8). (B) The gastric sleeve carriers (20.4±11.8) and non-carriers (25.3±12.6). Data presented as mean±SEM

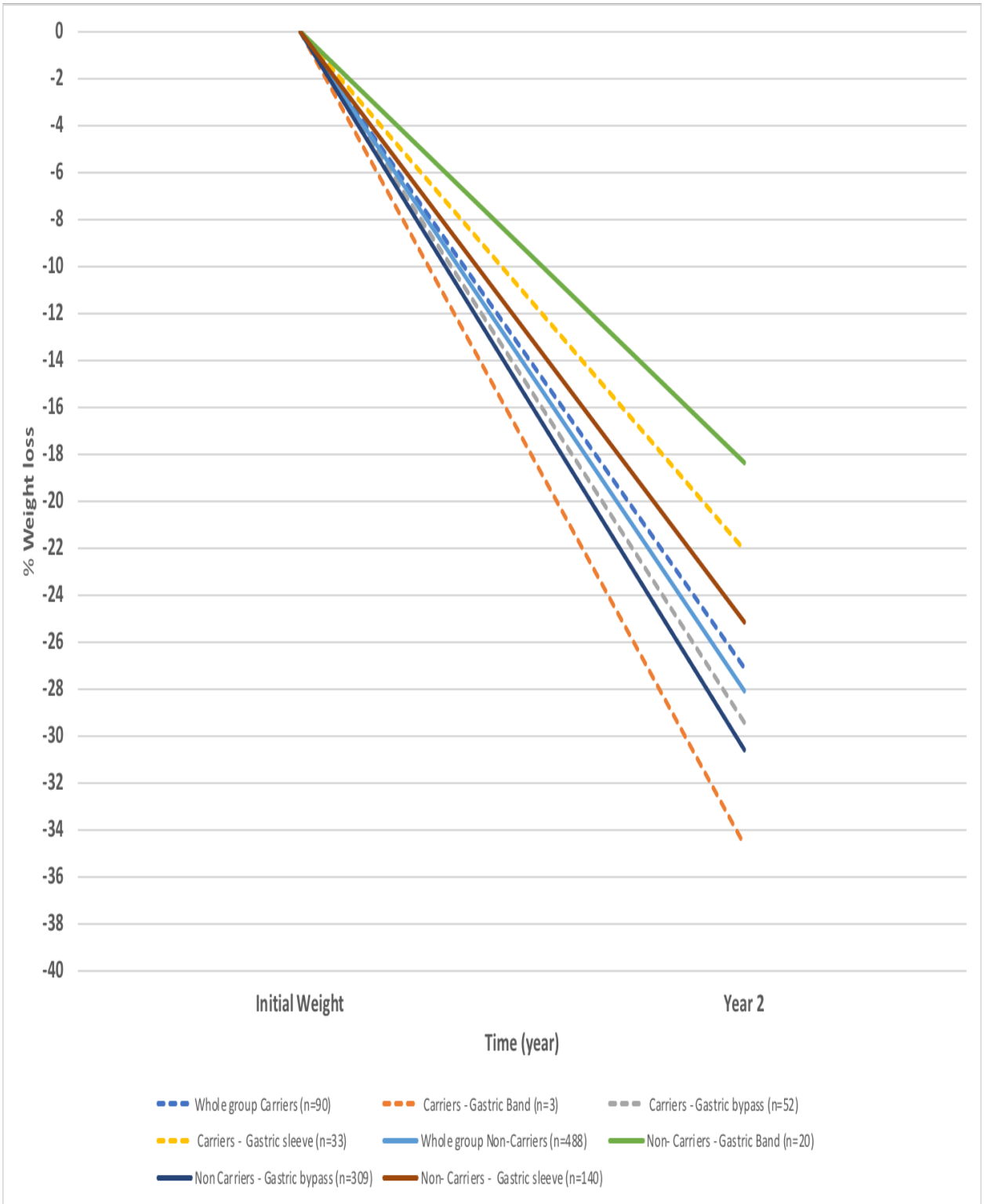


Figure 8.6 Percentage of weight loss between the carriers and non-carriers at 2 years timepoint. Dashed colored lines indicate the carriers group and solid colored lines indicate the non-carriers group.

8.3.2.3 Weight beyond two years post-surgery

There are around 20 carriers who have weight information beyond two years post-surgery. To explore the percentage weight loss two years post-surgery of individuals with monogenic obesity mutations, the carriers were categorised according to surgery type: bypass and sleeve (no participant had the gastric band surgery). The percentages of weight loss trajectories were mapped according to the year recorded post-surgery. As shown in the Figures below (Figure 8.7 and 8.8), there was a wide diversity in the percentage of weight loss amongst carriers, where some of them have almost regained the whole weight by six and seven years post-surgery.

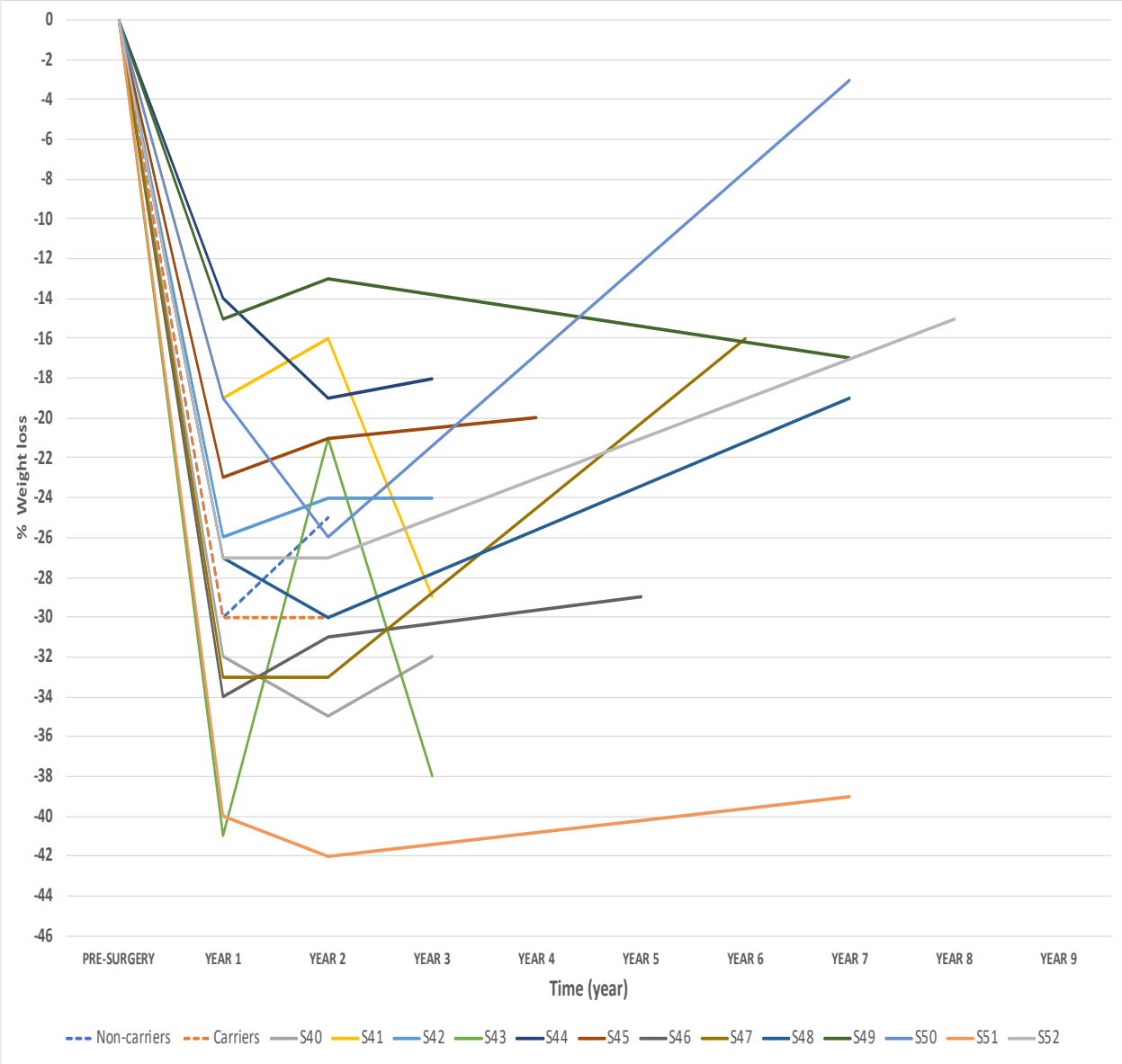


Figure 8.7 Percentage of weight loss trajectory of carriers who have weight information beyond two years after gastric bypass surgery. The chart shows the percentage weight at different time points from 0-9 years depending on weight information available for each subject Dashed blue line refers to the mean percentage of weight loss of the bypass non-carriers group. Dashed orange line refers the mean percentage of weight loss of the bypass carriers group. Each of the solid colored lines represents a carrier that is defined by a study number shown in the bottom legend. Details of the identified mutation in each bypass carrier is shown in Table 8.4 below.

Table 8.4 Details of the identified mutations in each carrier shown in Figure 8.7.

SUBJECT ID	GENE	TYPE OF VARIANT	RS ID	CADD	GNOMAD (MAF)
S40	*DYRK1B*	Missense	rs375929795	27.5	0.00000795
S41	*PCSK1*	Missense	rs937916641	27.3	0.00000398
S42	*RAI1*	Missense	rs1386459275	24.3	0.00000408
S43	*MBD5*	Missense	rs145808884	24.4	0.0000797
S45	*FOXA3*	Missense	rs780181879	25.8	0.00000398
S46	*WNT10B*	Missense	rs144672721	31	0.0001435
S47	*NTRK2*	Missense	rs76950094	19.81	0.0009584
S48	*GHSR*	Missense	rs1370841716	24.6	0.00000399
S49	*RAI1*	Missense	rs774984414	25.8	0.0000119
S50	*ALMS1*	Missense	rs45501594	20.3	0.006639
S51	*POU3F2*	Missense	rs1321043636	28.6	3.977E-06
S52	*RAI1*	Missense	rs142981643	23.5	0.0005847
S44			obesity-16p11.2-V1		

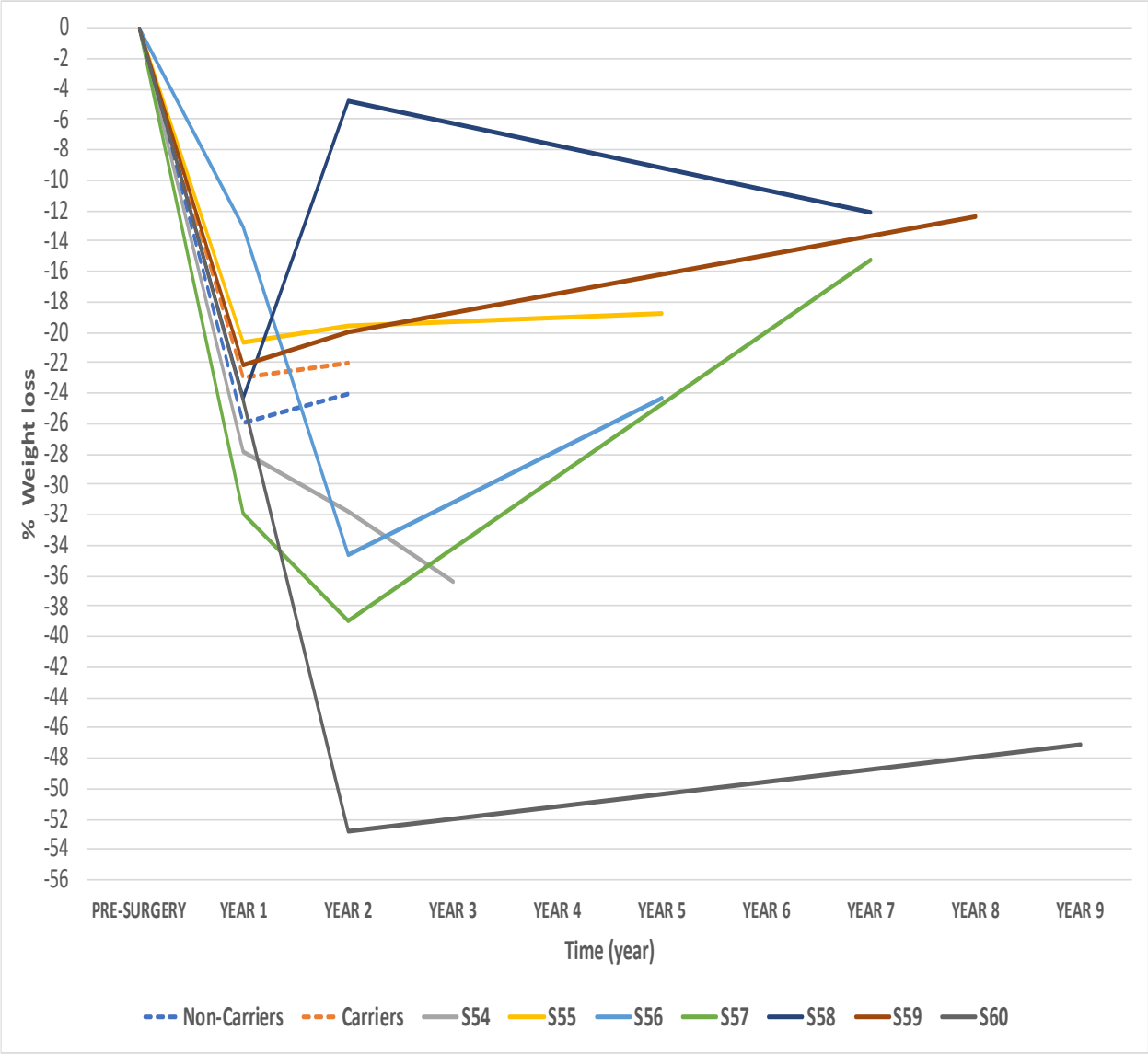


Figure 8.8 Percentage of weight loss trajectory of sleeve gastrectomy carriers who have weight information beyond two years. The chart shows the percentage of weight at different time points from 0-9 years depending on weight information available for each subject Dashed blue line refers to the mean percentage of weight loss of the sleeve gastrectomy non-carriers group. Dashed orange line refers the mean percentage of weight loss of the sleeve gastrectomy carriers group. Each of the solid colored lines represents a carrier that is defined by a study number shown in the bottom legend. Details of the identified mutation in each sleeve gastrectomy carrier is shown in Table 8.5 below.

Table 8.5 Details of the identified mutations in each carrier shown in Figure 8.8.

SUBJECT ID	GENE	TYPE OF VARIANT	RS ID	CADD	GNOMAD (MAF)
S54	*RAI1*	Missense	rs201842299	27	-
S55	*POGZ*	Missense	rs72996036	25.5	0.0004
S56	*WNT10B*	Missense	rs146010731	26.7	0.0006309
S57	*SH2B1*	Missense	rs115698674	24.3	0.0002033
S58	*RAI1*	frameshift deletion	rs149716029	-	0.002
S59	*POGZ*	Missense	-	25	-
S60	*MC3R*	Missense	rs373708098	30	0.0000677

8.3.3 Diabetes remission after bariatric surgery of subjects with potential deleterious variants in monogenic obesity genes in the PMMO cohort.

As it has been mention earlier, bariatric surgery has consistently been shown to enhance diabetes status of (type two diabetes) T2D through either decreasing the use of medication and insulin or achieving diabetes remission. Diabetes remission is defined as a reduction of HbA1c (<40 mmol/mol) without diabetes medication of at least 1 year [342, 343]

From the whole genotyped and sequenced subjects in the PMMO group, there are around 24% with T2D. To analyse the relation between T2D improvement and variants in the monogenic obesity gene, a comparison analysis was performed between subjects with T2D who are carriers of monogenic obesity, and non-carriers. Overall in the carrier group (n=252), there are around 47 individuals with T2D (38.8%). Of these, 33 individuals had available diabetic status post-surgery, where 11 of them had diabetes remission. This is equivalent to 33%.

On the other hand, in the non-carrier group (n=1568) there are around 370 individuals with T2D (%23.5). Of those, 167 had an available diabetic status post-surgery (45%): whereas for 63 of them their diabetes was in remission (normal). This is equivalent to a percentage of 37.7. To statistically evaluate the proportion of the observed diabetes remission in the two groups, the Chi-square test was used and show no significance difference (P-value=0.616). Overall difference in the diabetes remission rate between the two groups is not statistically significant.

.

8.4 Discussion.

In this Chapter, I explore the impact of Mendelian of obesity on treatment outcomes including weight loss and diabetes remission. The study performed here on weight loss was confined to the subjects with available data at least 0-24 months post-surgery. For a small number of PMMO participants, weight at later time-points was also available.

Aside from the effectiveness and success that bariatric surgery has reached in treating obesity, lifestyle intervention, which includes reduction in food intake, eliminating certain types of food or using meal replacements, is still considered a fundamental approach. Evidence from several reports has suggested that differences in the responses to lifestyle intervention may be partly due to genetic factors [344]. In our analysis investigating the weight loss difference between carriers and non-carriers of the Geronimo cohort who underwent the Milk-Based Meal Replacement Programme at 24 months, there was no statistical difference in percentage weight loss between the small number of participants with putative Mendelian forms of obesity and non-carriers.

As described earlier, bariatric surgery is considered the most effective long-term treatment not only for weight loss but also diabetes mellitus (T2DM) remission and improving comorbidities of metabolic syndrome. Patients' responses to surgery vary in terms of weight loss and diabetes remission. The diversity in surgery outcomes among individuals is believed to be likely to be related to heritability [340, 341]. Several recent studies investigated the influence of common or

rare variants on the postsurgical weight loss, but large scale studies of a range of Mendelian forms of obesity are urgently required: the majority of existing rare variant studies focus on subjects with *MC4R* mutations.

Overall our comparison of weight loss between carriers of Mendelian form of obesity and non-carriers showed no significant difference in weightless in response to bariatric surgery therapeutic intervention. These findings are in accordance with several previous studies that showed that gastric bypass and gastric band patients with MC4R mutations showed no difference in weight loss to patients with no molecular diagnosis and were able to lose as much weight as them [345-348].

The comparison based on the sleeve gastrectomy group done here showed a numerically lower percentage weight loss in carriers than non-carriers, but this did not reach statistical significance. Our data tend to offer some (weak) support to previous reports that carriers of MC4R and other obesity gene mutations had significantly lower percentage weight loss compared to non-carriers with no molecular diagnosis, particularly after sleeve gastrectomy [349, 350].

Some studies have suggested that the effect on weight loss may depend on the type of variant and its molecular impacts or a combination of other genetic / non-genetic factors [345, 346]. Human and mouse model data have shown that a modestly deleterious variant in one copy of MC4R gene (maintaining one active copy of MC4R, and one with residual activity) does not have a large impact on weight loss after surgery compared to homozygous or loss of function

mutations [351]. The general effects of bariatric surgery, which include reducing hunger, improving gut microbiome and mediating gut hormones, is believed to rely largely on roles of hypothalamic neurons. Thus, loss of function *mutations* in key genes in the energy balance centre of the hypothalamus have contributed to a reduction in weight loss post-surgery [349, 352]. This can be re-assess in the future when post-surgery weight for most carriers and non-carriers are available at least 2 years post surgery.

There is a relationship between weight loss and diabetes T2D remission or improvement, but in bariatric surgery, weight loss independent mechanisms are also important. Numerous studies have reported effective outcomes of bariatric surgery in terms of improving glycemic control and achieving diabetes remission [353, 354]. Nevertheless, the mechanisms underlying the metabolic improvement following the surgery are not yet fully clear [355]. Also, little is known about the relationship between the Mendelian forms of obesity and diabetes remission. In our analysis, diabetes remission two years post-surgery has been observed in a total of 11/33 (33.3%) carriers and 63/167 (37.3) non-carriers but this did not reach statistical significance. In contrast, Li *et al*. [349] investigated the influence of monogenic obesity of a variety of genes on weight loss and diabetes remission, have found the same rate of diabetes remission at two years post-surgery in the two groups: carriers and non-carriers, while another study that focused on MC4R mutations has reported a subject with unresolved diabetes [347].

There are various limitations to this analysis, including the effect of common variants on weight loss and diabetes remission, which is out of my PhD scope and is being addressed currently by

another PhD student, the long-term follow-up of weight information and missing weight information for some individuals due to missing some appointments, which I plan to address later, possibly through retrieving information from bariatric service and GP records. Thus, the current study has underperformed as there was no statistical significance despite the adjustment for multiple testing. This will require further re-analysis in the future as more information is obtained.

CHAPTER 9: CONCLUSIONS

9.1 Conclusions

Obesity and diabetes have gradually increased worldwide, creating a serious health issue. In recent decades, a considerable amount of work has been done to study the pathogenicity underlying those diseases, which has led to valuable insights into the genetic basis, treatment and prevention of obesity and diabetes. This includes the identification of several genes causing monogenic forms of obesity and diabetes, in addition to multiple variants statistically associated with common forms of the diseases.

Studies focused on rare forms (monogenic form) of the disease, which are mainly a result of highly penetrant genetic mutations and are much less affected by environmental factors, have provided substantial insight into the mechanisms of obesity and diabetes mellitus. In contrast, less is known about the pathophysiology and inheritance pattern of the common forms (polygenic form) of obesity and diabetes, which result from complex interactions between environmental and more subtle genetic factors of small effect size. Nevertheless, both the identified rare and common variants have so far been found to account for only a small percentage of the total estimated heritability [1-4].

With the ongoing advent of genetic technologies, such as next-generation sequencing and a variety of genotyping options, it has become more applicable and simpler to study the genetic factors underlying those diseases. However, despite the improvement in attempts to study the genetic factors related to the etiology of diabetes and obesity, it is not yet sufficient.

Thus, further characterisation of rare variants and study of the contribution of rare variants to common forms of obesity and diabetes in large cohorts and affected families are still desirable to enhance our understanding and reveal new insights into disease etiology. The initial focus of this thesis was to study the monogenic type of obesity and diabetes, but due to unexpected complications and PhD timeline being affected by the COVID-19 pandemic, obesity was fully addressed but diabetes was only included in the study design for future analysis after the PhD.

9.2 Summary of overall findings

In summary, **chapter 2** describes the cohorts included in the thesis's analysis. **Chapter 4** describes confirmation and subsequent family segregation analysis of participants who were identified to have a deleterious variant in a known obesity gene *(SH2B1).* This analysis led to the identification of an oligogenic mode of inheritance in obesity – the was the first evidence from our data, and provide the impetus to further explore the phenomenon of oligogenic inheritance in severe obesity in a larger clinical cohort (PMMO). **Chapter 5** describes the re-analysis of a pre-existing WES data of 91 individuals from the PMMO cohort at 24 months, utilising a larger list of genes relevant to monogenic obesity, syndromic obesity and/or mouse models of obesity. This re-analysis revealed an additional 21 possible causative variants in monogenic/syndromic obesity genes in 21 individuals, where 3 of those individuals follow the oligogenic mode of inheritance. In addition, 11 candidate variants were identified in the genes suggested by rodent models of obesity and/or diabetes. **Chapter 6** describes the design of the custom genotyping array focusing

on obesity and diabetes mellitus (T2D), and monogenic forms of obesity) to be applied to a larger cohort: allowing the exploration of the genetics underlying these diseases in a cost-effective and efficient manner. The array was designed to include rare and common variants in obesity and diabetes genes as well as CNVs in obesity and diabetes. In **Chapter 7**, I explored the overall quality control of the array and conducted a partial analysis focusing only on obesity/syndromic obesity genes and also rare CNVs related to obesity. The results demonstrated the success of the array in analysing rare variants, and that was validated by the confirmation of the variants that were seen by previous exome sequencing. The analysis led to the identification of a total of 161 potential causative variants in 40 monogenic obesity/syndromic obesity genes, affecting a total of 212 individuals, where a total of 14 cases showed an oligogenic mode of inheritance. Additionally, a total of 3 rare CNVS, known to be causative of obesity, were identified in 8 individuals. **Chapter 8** describes the implications of rare Mendelian forms of obesity on treatment outcomes, specifically weight loss and diabetes remission. Overall, a statistical difference was observed between carriers and non-carriers of genetic diseases at baseline BMI , but no statistically significant difference in the percentage of weight loss found at 2 years post-surgery. On the other hand, diabetes remission (HbA1c <40 mmol/mol)) was found to be higher in the non-carriers compared to the carriers but that did not reach statistical difference.

9.3 Contribution of the study to the field

The potential diagnostic yield of our array-based approach was 11%, which is likely an under-estimate of the prevalence of Mendelian forms of obesity among patients attending bariatric surgery services in the UK. This indicates a significant unmet need for WES/WGS and genetic counselling services in this patient group. There is also rising interest in identifying such cases, who may be eligible for a new range of drugs about to become available – including some specifically targeted to patients with Mendelian forms of obesity such as setmelanotide (https://www.rhythmtx.com/) which has recently received FDA approval for use in patients with *POMC, PCSK1* or *LEPR* deficiency [356].

We also confirm the findings of other authors [345-348], who report that bariatric surgery (especially RNY gastric bypass) is effective in achieving weight loss and diabetes remission in people with Mendelian forms of obesity, although data on long-term outcomes remains sparse.

Overall, the data reported in this thesis have contributed to new understanding in other ways. The first is the identification of oligogenic forms of obesity, which has not been reported previously and might be importance in explaining variable penetrance of known pathogenic mutations. Consistent with this, there is evidence on *SH2B1* mutations with incomplete or variable penetrance and may require other genetic and environmental factors to manifest the disease [192].

These oligogenic cases have been replicated in our larger cohort of genotyping over 2,000 obese subjects. Previous studies have identified oligogenic forms of other diseases including autism and cardiovascular diseases [309-311]. Triallelic inheritance has also long been suggested as an important mechanism in Bardet-Biedel syndrome [312, 357]

We also identified a number of people who were heterozygous carriers of mutations expected to cause monogenic obesity and diabetes. Additionally, the identification of heterozygous cases in obesity genes that follow the autosomal recessive mode of inheritance such as *ADCY3*, *PCSK1* and *POMC* confirms the findings of previous studies reporting obesity subjects with heterozygous deleterious mutation. This absolutely will require further functional studies to confirm the molecular impacts of those variants and their pathogenicity [66, 332, 333]. This is also notable in the family of genes causing Bardet-Beidl syndrome, raising the possibility that such mutations in heterozygous form may show a partial phenotype including obesity, especially where there are also other mutations in related pathways that might have the cumulative effect of compromising appetite regulation.

In addition, the re-analysis of the whole exome sequencing sheds light on the importance of re-analysing the WES data at different times as the findings underlying the genetics of obesity and diabetes expand, which might lead to diagnosis for other cases. The benefits of this approach of re-analysis have been proven in other reports [313-315]. Here we describe the identification of possibly pathogenic mutations in genes identified from rodent models of obesity and diabetes.

This confirms the utility of screening such genes, as previously suggest by analysis carried out by the group led by Sadaf Farooqi [163].

Furthermore, the use of the cost-effective genotyping array in analysing rare variants, as we have demonstrated in obesity, has validated the improvement, success and efficiency of this technology in studying and detecting rare variants. Further replicated studies using this approach can confirm its efficiency as the first step of a diagnostic screening for monogenic obesity.

Also the work in this thesis has shed light on new variants and inadequately reported obesity genes. With further clinical confirmation, some of the findings here can lead to personalised treatments ranging from surgery, lifestyle, and specific therapy, especially for subjects with causative mutations in *LEPR*, *LEP* and *POMC*.

9.4 Limitations

There were various limitations to the work performed in this thesis.

The majority of identified variants lack the functional studies that are important to confirm their molecular impact and genetic pathogenicity for clinical diagnoses. Therefore, this highlights a critically important problem in genetic counselling as many of the subjects that are identified here with possible causative variants might be missed if they are evaluated according the clinical genetics criteria (ACMG), where they would be categorised as variants of unknown significance (VUS).

Secondly, for many of the subjects identified with variants in syndromic obesity genes we do not have their full clinical features to draw a final diagnosis. This will require further assessments by a specialist clinical geneticist to confirm their diagnosis and potential order for further investigations. This would include assessments of potential dysmorphology, cognitive issues or related sub-clinical conditions. Participants with *PTEN* mutations require cancer screening follow-up. Assessment of methylation status/parent of origin analysis will also be required for imprinted genes, such as *MAGEL2* and *PEG3*.

Another limitation of the work presented here is that many individuals have patchy medical information, including missing weight information post-surgery: this has caused some difficulties in drawing conclusions in terms of the longer term implications of these rare variants for treatment outcomes.

In addition, the proposed methodology here in detecting rare variants through the customised array is very limited to known rare variants in databases and literature and cannot detect novel variants. We also have a mix of ethnicities in the PMMO study and cannot be certain that some of the variants we have identified are as rare may actually be more frequent in specific sub-populations.

9.5 Future work

From the analyses performed in this thesis, there are several potential findings that need to be addressed in future work. These can be classified as follows:

A) Remaining analysis of the array data

The custom genotyping array can be used for many analyses which could not all be done during my PhD due to the timeline, and some of them are beyond my PhD scope. This includes analysing the rare variants in diabetes genes, analysing the mouse model genes of obesity and diabetes in more detail, analysing the copy number variation through discovery analysis. Sufficient common SNPs were included in the array design to generate genetic risk scores for BMI, T2D and a range of relevant phenotypes. These can be used by future researchers to investigate the interactions of rare variants with common variants in determination of baseline phenotypes and treatment

outcomes. Some of these analyses have already been initiated by Dale Handley, another of the PhD students in Alex Blakemore's research group.

B) Retrieving clinical data and genetic counselling

There is a large number of subjects with identified possible causative variants that have missing post-surgery information and the latest follow weight and diabetes status, due to missing some clinical follow-up appointments. This information needs to be retrieved by checking their electronic clinical records from both hospital and GP sources. This will be essential to come to a conclusion in terms of the longer-term implications of rare variants for surgery outcomes.

Furthermore, on the other hand, some patients with variants in syndromic obesity genes will need to have a careful and detailed clinical evaluation by clinicians to confirm the other features that are related to the syndrome. This is vital to discover if the genes could be found in obese individuals that lack a cognitive phenotype. Also, many of the subjects with a potential causative variant require genetic counselling to inform the person of their condition, and to discuss future family plans and whether screening of other family members is indicated/desired.

C) Functional studies

As described in the limitations section, many of the identified variants here lack functional studies. The lack of functional studies represents a lack of supportive evidence, which

dramatically influences the evaluation of the variants according to the clinical genetics criteria, ACMG guideline. Functional studies are considered very demanding and have been forming a big challenge due to a lack of an efficient, reliable and non-time-consuming approach that can explore the functional impact of a variant quickly. Thus, functional studies on number of potential variants are planned to be performed through collaborative work of or through a commercial company. There are several kinds of functional studies that we plan to perform, ranging from analysis of RNA expression and protein to CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/Cas9) technology. We are currently initiating a functional study in collaboration with Dr Mieke Van Haelst (Amsterdam) using CRISPR/Cas9 technology in cell lines to investigate the effect of MC3R mutations (including those discovered in this project). This can lead to a better understanding of the variants' impacts and underlying mechanisms of the genes causing Mendelian obesity.

D) Applying the array in other ethnicities and re-design a new version of the array

As the studies conducted in this thesis focus largely on European ancestries, additional studies addressing other ethnicities are equally important to enhance our understanding of genetics architecture of obesity and diabetes underlying those ethnicities and how they differ from what we know. Furthermore, a new re-design of the array to include new genes or variants that were not in the current version will be worth doing to further solidify the approach.

9.6 Overall Summary

This thesis presents a body of work contributing to understanding of the genetic architecture of severe obesity and laying the groundwork for future research in an area of significant unmet clinical need. A new potential diagnostic screening tool was designed and applied to severely affected study participants. At least 220 of study participants carried rare deleterious variants in genes known to be implicated in human obesity, and new “obesity genes” were identified by screening of genes highlighted by rodent models. The results indicate potential clinical utility of this approach and provide a basis for future research in this important area.

REFERENCES

1. WHO. *Health topics, obesity*. World Health Organization 2019; Available from: https://www.who.int/health-topics/obesity.
2. Helaine E Resnick, P.V., Jeffrey B Halter, Xihong Lin, *Relation of weight gain and weight loss on subsequent diabetes risk in overweight adults.* J Epidemiol Community Health, 2000. **54**: p. 596–602.
3. Mokdad, A.H.F., Earl S ; Bowman, Barbara A ; Dietz, William H ; Vinicor, Frank ; Bales, Virginia S ; Marks, James S, *Prevalence of Obesity, Diabetes, and Obesity-Related Health Risk Factors, 2001.* Jama, 2003. **289**(1): p. 76-79.
4. Alter, D.W.a.D.A., *The economic burden of obesity worldwide: a systematic review of the direct costs of obesity.* Obesity reviews : an official journal of the International Association for the Study of Obesity, 2011. **12**(2): p. 131-41.
5. El-Sayed Moustafa, J.S. and P. Froguel, *From obesity genetics to the future of personalized obesity therapy.* Nat Rev Endocrinol, 2013. **9**(7): p. 402-13.
6. Flannick, J., S. Johansson, and P.R. Njolstad, *Common and rare forms of diabetes mellitus: towards a continuum of diabetes subtypes.* Nat Rev Endocrinol, 2016. **12**(7): p. 394-406.
7. Roser, H.R.a.M. *Obesity*. OurWorldInData.org 2017; Available from: https://ourworldindata.org/obesity.
8. Burgio, E., Lopomo, A. & Migliore, L. , *Obesity and diabetes: from genetics to epigenetics.* Mol Biol Rep, 2015. **42**: p. 799-818.
9. Hill JO, W.H., Peters JC. , *Energy balance and obesity.* Circulation, 2012. **126**(1): p. 126-132.
10. Romieu I, D.L., Barquera S, et al. , *Energy balance and obesity: what are the main drivers?* Cancer Causes Control, 2017. **28**(3): p. 247-258.
11. Hill JO, W.H., Peters JC. , *The Importance of Energy Balance.* Eur Endocrinol, 2013. **9**(2): p. 111-115.
12. Schneeberger M, G.R., Claret M. , *Hypothalamic and brainstem neuronal circuits controlling homeostatic energy balance.* . J Endocrinol, 2014. **220**(2): p. 25-46.
13. MM., A., *Central and peripheral control of food intake.* . Endocr Regul., 2017. **51**(1): p. 52-70.
14. Lee, A., Cardel, M., & Donahoo, W. T. , *Social and Environmental Factors Influencing Obesity.* In K. R. Feingold (Eds.) et. al., Endotext. MDText.com, Inc., 2019.
15. Hermine H. M. Maes, M.C.N., and Lindon J. Eaves, *Genetic and Environmental Factors in Relative Body Weight and Human Adiposity.* Behavior Genetics, 1997. **7**: p. 325-351.
16. Jenny van Dongen, P.E.S., Harmen H. M. Draisma, Nicholas G. Martin and Dorret I. Boomsma, *The continuing value of twin studies in the omics era.* Nature, 2013. **13**: p. 640-653.
17. Yates, W.R., *Adoption Studies.* John Wiley & Sons, Ltd., 2011: p. 1-5.
18. A. J. Stunkard, T.I.A.S., and G. Hanis, *An adoption study of human obesity," The New England Journal of Medicine.* The New England Journal of Medicine, 1986. **314**(4): p. 193-198.
19. T. I. Sørensen, R.A.P., A. J. Stunkard, and F. Schulsinger, *Genetics of obesity in adult adoptees and their biological siblings.* BMJ, 1989. **298**(6666): p. 87-90.

20. Albert J. Stunkard, M.D., Jennifer R. Harris, Ph.D., Nancy L. Pedersen, Ph.D., and Gerald E. McClearn, Ph.D., *The Body-Mass Index of Twins Who Have Been Reared Apart.* N Engl J Med 1990(322): p. 1483-1487.
21. Jacob v.B. Hjelmborg, C.F., Karri Silventoinen, Matt McGue, Maarit Korkeila, Kaare Christensen, Aila Rissanen3 and Jaakko Kaprio, *Genetic Influences on Growth Traits of BMI: A Longitudinal Study of Adult Twins.* Obesity, 2008. **16**(4): p. 847–852.
22. Bouchard, R.J.F.L.C., *Obesity: is it a genetic disorder?* Journal of Internal Medicine, 2003. **254**: p. 401–425.
23. Coleman DL, H.K., *The influence of genetic background on the expression of the obese.* Diabetologia, 1973. **9**: p. 287–293.
24. Yazdi FT, C.S., Meyre D. , *Obesity genetics in mouse and human: back and forth, and back again.* PeerJ., 2015. **3**(856): p. 1-69.
25. Yiying Zhang, R.P., Margherita Maffei, Marisa Barone, Lori Leopold, Jeffrey M. Friedman, *Positional cloning of the mouse obese gene and its human homologue.* Nature, 1994. **374**(6521): p. 479.
26. Chen H, C.O., Tartaglia LA, et al. , *Evidence That the Diabetes Gene Encodes the Leptin Receptor: Identification of a Mutation in the Leptin Receptor Gene in db/db Mice.* Cell., 1996. **84**(3): p. 491-495.
27. Montague CT1, F.I., Whitehead JP, Soos MA, Rau H, Wareham NJ, Sewter CP, Digby JE, Mohammed SN, Hurst JA, Cheetham CH, Earley AR, Barnett AH, Prins JB, O'Rahilly S., *Congenital leptin deficiency is associated with severe early-onset obesity in humans.* Nature Genetics, 1997. **387**: p. 903-908.
28. Ozata, M.O., I C ; Licinio, J, *Human Leptin Deficiency Caused by a Missense Mutation: Multiple Endocrine Defects, Decreased Sympathetic Tone, and Immune System Dysfunction Indicate New Targets for Leptin Action, Greater Central than Peripheral Resistance to the Effects of Leptin, and Spontaneous Correction of Leptin-Mediated Defects.* Clinical Endocrinology & Metabolism, 1999. **84**(10): p. 3686-3695.
29. Andreas Strobel, T.I., Luc Camoin, Metin Ozata & A. Donny Strosberg, *A leptin missense mutation associated with hypogonadism and morbid obesity.* Nature Genetics, 1998. **18**: p. 213 - 215.
30. Farooqi, I.S., et al., *Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency.* Journal of Clinical Investigation, 2002. **110**(8): p. 1093-1103.
31. families, H.p.o.l.a.m.-r.g.m.i.c.w.s.o.f.P.c., *Sadia Saeed, Taeed A. Butt, Mehwish Anwer , Muhammad Arslan, Philippe Frogue.* Molecular Genetics and Metabolism, 2012. **106**: p. 121-126.
32. Farooqi IS, O.R.S., *20 years of leptin: human disorders of leptin action.* . J Endocrinol., 2014. **223**(1): p. 63-70.
33. S Margetic, C.G., GG Pegg & RA Hill *Leptin: a review of its peripheral actions and interactions.* International Journal of Obesity volume 2002. **26**: p. 1407–1433.
34. Robert V. Considine, P.D., Madhur K. Sinha, Ph.D., Mark L. Heiman, Ph.D., Aidas Kriauciunas, Ph.D., Thomas W. Stephens, Ph.D., Mark R. Nyce, M.S., Joanna P. Ohannesian, B.S.N., Cheryl C. Marco, R.D., Linda J. McKee, M.H.S., Thomas L. Bauer, M.D., and José F. Caro, M.D., *Serum immunoreactive-leptin concentrations in normal-weight and obese humans.* The New England journal of medicine, 1996. **334**(5): p. 292-295.

35. Nobuhiro Wada , S.H., Fumiko Takenoya , Haruaki Kageyama, Mai Okabe , Seiji Shioda, *Leptin and its receptors.* Journal of Chemical Neuroanatomy, 2014. **61-62**: p. 191-199.
36. May-Yun Wang, Y.T.Z., Christopher B. Newgard, Roger H. Unger, *A novel leptin receptor isoform in rat.* Federation of European Biochemical Societies., 1996. **392**: p. 87-90
37. Gwo-Hwa Lee, R.P., J. M. Montez, K. M. Carroll, J. G. Darvishzadeh, J. I. Lee & J. M. Friedman *Abnormal Splicing of the Leptin Receptor in Diabetic Mice.* Nature, 1996. **379**: p. 632–635.
38. Michael W. Schwartz, S.C.W., Daniel Porte Jr, Randy J. Seeley & Denis G. Baskin, *Central nervous system control of food intake.* Nature, 2000. **404**: p. 661-671.
39. G. J. Morton, D.E.C., D. G. Baskin, G. S. Barsh & M. W. Schwartz *Central nervous system control of food intake and body weight.* Nature, 2006. **443**: p. 289-295.
40. Cone, R.D., *Anatomy and regulation of the central melanocortin system.* Nature, 2005. **8**: p. 571–578.
41. C C Cheung, D.K.C., R A Steiner, *Proopiomelanocortin neurons are direct targets for leptin in the hypothalamus.* Endocrinology., 1997. **138**(10): p. 4489-4492.
42. Tao, Y.-X., *Molecular mechanisms of the neural melanocortin receptor dysfunction in severe early onset obesity.* Molecular and Cellular Endocrinology, 2005. **239**(1-2).
43. Aaron J. Mercer, S.T.H., Charles K. Meshul, and Malcolm J. Low, *Unraveling the central proopiomelanocortin neural circuits.* Frontiers in Neuroscience, 2013. **7**.
44. Jr., M.B.A.a.M.G.M., *20 years of leptin: connecting leptin signaling to biological function.* The Journal of Endocrinology,, 2014. **223**(1): p. 25-35.
45. Sarah H. Bates, W.H.S., Trevor A. Dundon, Markus Schubert, Annette W. K. Tso, Yongping Wang, Alexander S. Banks, Hugh J. Lavery, Asma K. Haq, Eleftheria Maratos-Flier, Benjamin G. Neel, Michael W. Schwartz & Martin G. Myers Jr *STAT3 signalling is required for leptin regulation of energy balance but not reproduction.* Nature, 2003. **421**: p. 856-859.
46. Scott A. Robertson, G.M.L., 1 and Martin G. Myers, Jr., *Molecular and neural mediators of leptin action.* Physiol Behav, 2008. **94**(5): p. 637-642.
47. Kevin D. Niswender, G.J.M., Walter H. Stearns, Christopher J. Rhodes, Martin G. Myers Jr & Michael W. Schwartz, *Intracellular signalling. Key enzyme in leptin-induced anorexia.* Nature, 2001. **413**: p. 794–795.
48. Marc Claret, M.A.S., Rachel L Batterham, Colin Selman, Agharul I Choudhury, Lee G D Fryer, Melanie Clements, Hind Al-Qassab, Helen Heffron, Allison W Xu, John R Speakman, Gregory S Barsh, Benoit Viollet, Sophie Vaulont, Michael L J Ashford, David Carling, Dominic J Withers, *AMPK Is Essential for Energy Homeostasis Regulation and Glucose Sensing by POMC and AgRP Neurons.* The Journal of Clinical Investigation, 2007. **117**(8): p. 2235-2336.
49. Horvatha, L.V.a.T.L., *Leptin and Insulin Pathways in POMC and AgRP Neurons That Modulate Energy Balance and Glucose Homeostasis.* European molecular biology organization, 2012. **13**(12): p. 1079-1086.
50. Theodore Kelesidis, I.K., Sharon Chou, and Christos S. Mantzoros,, *Narrative Review: The Role of Leptin in Human Physiology: Emerging Clinical Applications.* Ann Intern Med, 2010. **152**(2): p. 93-100.

51. Walley, A., Asher, J. & Froguel, P. , *The genetic contribution to non-syndromic human obesity.* . Nat Rev Genet, 2009. **10**: p. 431-442.
52. Tao, Y.-X., *The melanocortin-4 receptor: physiology, pharmacology, and pathophysiology.* Endocrine Reviews, 2010. **31**(4): p. 506 –543.
53. Toshiro Kishi, C.J.A., Charlotte E Lee, Kathleen G Mountjoy, Clifford B Saper, Joel K Elmquist, *Expression of melanocortin 4 receptor mRNA in the central nervous system of the rat.* Journal of Comparative Neurology, 2003. **457**(3): p. 213-235.
54. W Fan, B.A.B., R A Kesterson, V J Hruby, R D Cone, *Role of Melanocortinergic Neurons in Feeding and the Agouti Obesity Syndrome.* Nature, 1997. **165**(8): p. 165-168.
55. YX., T., *Constitutive activity in melanocortin-4 receptor: biased signaling of inverse agonists.* . Adv Pharmacol, 2014. **70**: p. 135-154.
56. G S Yeo, I.S.F., S Aminian, D J Halsall, R G Stanhope, S O'Rahilly, *A frameshift mutation in MC4R associated with dominantly inherited human obesity.* Nature, 1998. **20**(2): p. 111-112.
57. D Huszar , C.A.L., V Fairchild-Huntress, J H Dunmore, Q Fang, L R Berkemeier, W Gu, R A Kesterson, B A Boston, R D Cone, F J Smith, L A Campfield, P Burn, F Lee, *Targeted disruption of the melanocortin-4 receptor results in obesity in mice.* Cell, 1997. **88**(1): p. 131-141.
58. O'Rahilly, S.F.S., *Mutations in ligands and receptors of the leptin–melanocortin pathway that lead to obesity.* Nature, 2008. **4**: p. 569–577.
59. Cecile Lubrano-Berthelier 1, B.D., Jean-Marc Lacorte, Franck Picard, Astrid Shapiro, Sumei Zhang, Sandrine Bertrais, Serge Hercberg, Arnaud Basdevant, Karine Clement, Christian Vaisse, *Melanocortin 4 Receptor Mutations in a Large Cohort of Severely Obese Adults: Prevalence, Functional Classification, Genotype-Phenotype Relationship, and Lack of Association With Binge Eating.* Clinical Endocrinology & Metabolism, 2006. **91**(5): p. 1811-1118.
60. Tinh-Hai Collet, B.D., Jacek Mokrosinski, Hillori Connors, Julia M. Keogh, Edson Mendes de Oliveira, Elana Henning, Christine Poitou-Bernert, Jean-Michel Oppert,3,4 Patrick Tounian, Florence Marchelli,3 Rohia Alili, Johanne Le Beyec, Dominique Pépin, Jean-Marc Lacorte, Andrew Gottesdiener,Rebecca Bounds, Shubh Sharma, Cathy Folster, Bart Henderson, Stephen O'Rahilly, Elizabeth Stoner, Keith Gottesdiener, Brandon L. Panaro, Roger D. Cone, Karine Clément, I. Sadaf Farooqi, and Lex H.T. Van der Ploeg, *Evaluation of a melanocortin-4 receptor (MC4R) agonist (Setmelanotide) in MC4R deficiency.* Molecular metabolism, 2017. **6**(10): p. 1321-1329.
61. I. Sadaf Farooqi, J.M.K., Giles S.H. Yeo, Emma J. Lank, Tim Cheetham, and Stephen O'Rahilly, *Clinical Spectrum of Obesity and Mutations in the Melanocortin 4 Receptor Gene.* The New England Journal of Medicine, 2003. **348**(12): p. 1085-1095.
62. Niamh X Cawley, Z.L.a.Y.P.L., *60 YEARS OF POMC: Biosynthesis, trafficking, and secretion of pro-opiomelanocortin-derived peptides.* Journal of Molecular Endocrinology, 2016. **56**(4): p. 77-97.
63. A I Smith, J.W.F., *Proopiomelanocortin processing in the pituitary, central nervous system, and peripheral tissues.* Endocrine Reviews, 1988. **9**(1): p. 59-179.
64. L Yaswen, N.D., M B Brennan, U Hochgeschwender, *Obesity in the mouse model of pro-opiomelanocortin deficiency responds to peripheral melanocortin.* Nature, 1999. **5**(9): p. 1066-1070.

65. Heiko Krude, H.B., Werner Luck, Rüdiger Horn, Georg Brabant & Annette Grüters *Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans.* . Nature, 1998. **19**: p. 155-157.
66. I. Sadaf Farooqi, S.D., Agnes Clements, Julia M. Keogh, Joanna Biernacka4, Sarah Lowenbein, Benjamin G. Challis and Stephen O'Rahilly, *Heterozygosity for a POMC-Null Mutation and Increased Obesity Risk in Humans.* Diabetes 2006. **55**(9): p. 2549-2553.
67. Sadia Saeed, M.A., and Philippe Froguel, *Genetics of Obesity in Consanguineous Populations: Toward Precision Medicine and the Discovery of Novel Obesity Genes.* Obesity, 2018. **26**(3): p. 474-484.
68. Stephen O'Rahilly, I.S.F., Giles S. H. Yeo, Benjamin G. Challis, *Minireview: Human Obesity—Lessons from Monogenic Disorders.* Endocrinology, 2003. **144**(9): p. 3757–3764.
69. Gray J, Y.G., Hung C, Keogh J, Clayton P, Banerjee K, McAulay A, O'Rahilly S, Farooqi IS, *Functional characterization of human NTRK2 mutations identified in patients with severe early-onset obesity.* International Journal of Obesity volume, 2007. **31**: p. 359–364.
70. Park H, P.M., *Neurotrophin regulation of neural circuit development and function.* Nature, 2013. **14**(1): p. 7-23.
71. Baoji Xu, E.H.G., Keling Zang, David Cepoi, Roger D Cone, Kevin R Jones, Laurence H Tecott, and Louis F Reichardt, *Brain-derived neurotrophic factor regulates energy balance downstream of melanocortin-4 receptor.* Nat Neurosci, 2003. **6**(7): p. 736–742.
72. Steven G. Kernie, D.J.L., and Luis F. Parada, *BDNF regulates eating behavior and locomotor activity in mice.* The EMBO Journal, 2000. **19**(6): p. 1290-1300.
73. Ranadive SA, V.C., *Lessons from Extreme Human Obesity: Monogenic Disorders.* Endocrinol Metab Clin North Am., 2008. **37**(3): p. 733-751.
74. An, J.J., Kinney, C.E., Tan, J. et al. , *TrkB-expressing paraventricular hypothalamic neurons suppress appetite through multiple neurocircuits.* Nat Commun, 2020. **11**(1729): p. 1-16.
75. Farooqi IS, K.J., Kamath S, Jones S, Gibson WT, Trussell R, Jebb SA, Lip GY, O'Rahilly S, *Partial leptin deficiency and human adiposity.* Nature, 2001. **1**(414): p. 34-35.
76. Farooqi IS, W.T., Collins S, et al., *Clinical and molecular genetic spectrum of congenital deficiency of the leptin receptor.* N Engl J Med., 2007. **356**(3): p. 237-247.
77. Lubrano-Berthelier C, L.S.C., Bougnères P, Vaisse C., *A homozygous null mutation delineates the role of the melanocortin-4 receptor in humans.* J Clin Endocrinol Metab., 2004. **89**(5): p. 2028-2032.
78. Farooqi IS, K.J., Yeo GS, Lank EJ, Cheetham T, O'Rahilly S., *Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene.* . N Engl J Med., 2003. **348**(12): p. 1085-1095.
79. Dubern B, B.S., Talbaoui H, Le Beyec J, Tounian P, Lacorte JM, Clément K, *Homozygous null mutation of the melanocortin-4 receptor and severe early-onset obesity.* J Pediatr, 2007. **150**(6): p. 613-617.
80. Kaur Y, d.S.R., Gibson WT, Meyre D., *A systematic review of genetic syndromes with obesity.* Obesity Reviews, 2017. **18**(6): p. 603-634.
81. Katie Weihbrecht, W.A.G., Thomas Pak, Janelle E. Garrison, Adam P. DeLuca, Edwin M. Stone, Todd E. Scheetz, and Val C. Sheffield1,, *Keeping an Eye on Bardet-Biedl*

Syndrome: A Comprehensive Review of the Role of Bardet-Biedl Syndrome Genes in the Eye. Med Res Arch., 2017. **5**(9): p. 603-634.
82. Elizabeth Forsythe*, J.K., Chiara Bacchelli and Philip L. Beales, *Managing Bardet–Biedl Syndrome—Now and in the Future.* Frontiersin in pediatric, 2018. **6**(23): p. 1-8.
83. P L Beales, N.E., A S Woolf, D Parker, F A Flinter, *New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey.* . J Med Genet., 1999. **36**: p. 437-446.
84. Suzanne B. Cassidy MD, S.S.P., Jennifer L. Miller MD & Daniel J. Driscoll MD, PhD, *Prader-Willi syndrome.* Genetics in Medicine, 2012. **14**: p. 10–26.
85. M. A. Angulo, M.G.B., M. E. Cataletto, *Prader-Willi syndrome: a review of clinical, genetic, and endocrine findings.* J Endocrinol Invest 2015. **38**: p. 1249-1263.
86. Falco, M.e.a., *RAI1 gene mutations: mechanisms of Smith-Magenis syndrome.* The application of clinical genetics, 2017. **10**: p. 85-94.
87. Toulouse A, R.D., Roussel J, Joober R, Rouleau GA, *Molecular cloning and characterization of human RAI1, a gene associated with schizophrenia.* Genomics, 2003. **82**(2): p. 162-71.
88. Walters RG, J.S., Valsesia A, de Smith AJ, Martinet D, et al., *A new highly penetrant form of obesity due to deletions on chromosome 16p11.2.* . Nature, 2010. **463**: p. 671-675.
89. Jacquemont S, R.A., Zufferey F, Harewood L, Walters RG, et al. , *Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus.* . Nature, 2011. **478**: p. 97-102.
90. Pulst, S.M., *Genetic Linkage Analysis.* Archives of neurology, 1999. **56**(6): p. 667-672.
91. Loos, R.J.F., *Recent progress in the genetics of common obesity.* British Journal of Clinical Pharmacology, 2009. **68**(6): p. 811-829.
92. Helen N Lyon, J.N.H., *Genetics of common forms of obesity: a brief overview* The American Journal of Clinical Nutrition, 2005. **82**(1): p. 215S–217S.
93. Radhika Patnala, J.C.a.J.B., *Candidate gene association studies: a comprehensive guide to useful in silico tools.* BMC Genetics, 2013. **14**(39): p. 1-11.
94. Rao, K.R.e.a., *Genetic & epigenetic approach to human obesity.* The Indian journal of medical research, 2014. **140**(5): p. 589-603.
95. Rankinen T, Z.A., Chagnon YC, et al. , *The Human Obesity Gene Map: The 2005 Update.* Obesity, 2006. **14**(4): p. 529-644.
96. Yáñez M, V.G., PérezMorales R, et al., *Monogenic, Polygenic and Multifactorial Obesity in Children: Genetic and Environmental Factors.* Austin Journal of Nutrition & Metabolism, 2017. **4**(3): p. 1-12.
97. Hinney, A., Anna-Lena Volckmar, and Jochen Antel., *Genes and the Hypothalamic Control of Metabolism in Humans.* Best Practice & Research Clinical Endocrinology & Metabolism, 2014. **28**(5): p. 635-647.
98. Venkatesan R, V.M., *Obesity – Are we continuing to play the genetic "blame game"?* Advances in Genomics and Genetics 2016. **6**: p. 11-23.
99. Dina, C., et al., *Variation in FTO contributes to childhood obesity and severe adult obesity.* Nat Genet, 2007. **39**(6): p. 724-6.
100. 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.

101. 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.
102. Bradfield, J.P., et al., *A genome-wide association meta-analysis identifies new childhood obesity loci.* Nat Genet, 2012. **44**(5): p. 526-31.
103. Wang, H., et al., *Genetic variants in FTO associated with metabolic syndrome: a meta- and gene-based analysis.* Mol Biol Rep, 2012. **39**(5): p. 5691-8.
104. Meyre, D., et al., *Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations.* Nat Genet, 2009. **41**(2): p. 157-9.
105. Wei, J.e.a., *Differential m6A, m6Am, and m1A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm.* Molecular cell, 2018. **71**(6): p. 973-985.
106. Lan N, L.Y., Zhang Y, et al., *FTO - A Common Genetic Basis for Obesity and Cancer.* Front Genet., 2020. **11**: p. 1-12.
107. Deng X, S.R., Stanford S, Chen J. , *Critical Enzymatic Functions of FTO in Obesity and Cancer. .* Front Endocrinol., 2018. **9**(396): p. 1-7.
108. Loos RJ, L.C., Li S, et al., *Common variants near MC4R are associated with fat mass, weight and risk of obesity.* Nature Genetics, 2008. **40**(6): p. 768-775.
109. Willer, C.J.e.a., *Six new loci associated with body mass index highlight a neuronal influence on body weight regulation.* Nat. Genet, 2009. **41**: p. 25–34.
110. Locke AE, K.B., Berndt SI, et al. , *Genetic studies of body mass index yield new insights for obesity biology.* Nature, 2015. **518**(7538): p. 197-206.
111. Yengo L, S.J., Kemper KE, et al. , *Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. .* Hum Mol Genet. , 2018. **27**(20): p. 3641-3649.
112. Akiyama M, O.Y., Kanai M, et al., *Genome-wide association study identifies 112 new loci for body mass index in the Japanese population.* Nature, 2017. **49**(10): p. 1458-1467.
113. Yang, J.e.a., *Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index.* Nature genetics, 2015. **47**(10): p. 1114-1120.
114. Conrad, D.F.e.a., *Origins and functional impact of copy number variation in the human genome.* Nature, 2010. **464**: p. 704-712.
115. Rodriguez-Revenga, L., Montserrat Mila, Carla Rosenberg, Allen Lamb, and Charles Lee. , *Structural Variation in the Human Genome: The Impact of Copy Number Variants on Clinical Diagnosis.* Genetics in Medicine. **2007**(9): p. 600-606.
116. Aouiche, C., Xuequn Shang, and Bolin Chen. , *Copy Number Variation Related Disease Genes.* Quantitative Biology, 2018. **6**(2): p. 99-112.
117. Willer, C., Speliotes, E., Loos, R. et al. , *Six new loci associated with body mass index highlight a neuronal influence on body weight regulation.* Nature, 2009. **41**: p. 25–34.
118. Sha BY, Y.T., Zhao LJ, et al., *Genome-wide association study suggested copy number variation may be associated with body mass index in the Chinese population.* J Hum Genet, 2009. **54**(4): p. 199-202.
119. Jarick, I.e.a., *Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis.* Human molecular genetics, 2011. **20**(4): p. 840-852.
120. Ingelsson E, M.M., *Human Genetics of Obesity and Type 2 Diabetes Mellitus: Past, Present, and Future.* Circ Genom Precis Med., 2018. **11**(16): p. 1-21.

121. Ramachandrappa S, F.I., *Genetic approaches to understanding human obesity.* . J Clin Invest., 2011. **121**(6).
122. Yang, J., Benyamin, B., McEvoy, B. et al., *Common SNPs explain a large proportion of the heritability for human height.* Nat Genet, 2010. **42**(7): p. 565–569.
123. Behjati S, T.P., *What is next generation sequencing?* . Arch Dis Child Educ Pract Ed, 2013. **98**(6): p. 236-238.
124. JR., D., *Whole-Genome Sequencing: Opportunities and Challenges for Public Health, Food-borne Outbreak Investigations, and the Global Food Supply.* . J Infect Dis., 2016. **213**: p. 499-501.
125. Voight BF, K.H., Ding J, et al. , *The Metabochip, a Custom Genotyping Array for Genetic Studies of Metabolic, Cardiovascular, and Anthropometric Traits.* PLoS genetics, 2012. **8**(8): p. 1-12.
126. HGMD. *The Human Gene Mutation Database (HGMD)*. 2019; Available from: http://www.hgmd.cf.ac.uk/.
127. gnomAD. *The Genome Aggregation Database* 2020; Available from: https://gnomad.broadinstitute.org/.
128. R., W.H., *Update on treatment strategies for obesity.* The Journal of clinical endocrinology and metabolism, 2013. **98**(4): p. 1299–1306.
129. AMA Catenacci VA, W.H., *The role of physical activity in producing and maintaining weight loss.* Nat Clin Pract Endocrinol Metab., 2007. **3**(7): p. 518-529.
130. Jeffery RW , e.a., *Physical activity and weight loss: does prescribing higher physical activity goals improve outcome?* Am J Clin Nutr., 2003. **78**: p. 684–689.
131. Kumar RB, A.L., *Pharmacologic Treatment of Obesity.* Endotext, 2017.
132. Sjöström L, R.A., Andersen T, et al., *Randomised placebo-controlled trial of orlistat for weight loss and prevention of weight regain in obese patients. European Multicentre Orlistat Study Group.* European Multicentre Orlistat Study Group, 1998. **352**(9123): p. 167-172.
133. R, G., *Mode of action of orlistat.* Int J Obes Relat Metab Disord, 1997. **21**(3): p. S12-S23.
134. HR, W., *Update on treatment strategies for obesity.* J Clin Endocrinol Metab. , 2013. **98**(4): p. 1299-1306.
135. Blüher M, M.C., *From leptin to other adipokines in health and disease: Facts and expectations at the beginning of the 21st century.* Metabolism, 2015. **64**(1): p. 131-145.
136. Neff LM, K.R., *Emerging role of GLP-1 receptor agonists in the treatment of obesity.* . Diabetes Metab Syndr Obes. , 2010. **3**(263-273).
137. Müller, T.D., et al., *Glucagon-like peptide 1 (GLP-1).* Molecular Metabolism, 2019. **30**: p. 72-130.
138. Christou GA, K.N., Blundell J, Fruhbeck G, Kiortsis DN. , *Semaglutide as a promising antiobesity drug.* Obes Rev., 2019. **20**(6): p. 805-815.
139. Plodkowski RA, M.M., Reisinger-Kindle K, et al., *Obesity Management: Clinical Review and Update of the Pharmacologic Treatment Options.* . Fed Pract., 2016. **1**(3): p. 6-16.
140. Sjöström L, L.A., Peltonen M, et al., *Lifestyle, diabetes, and cardiovascular risk factors 10 years after bariatric surgery.* The New England Journal of Medicine (NEJM), 2004. **351**: p. 2683-2693.
141. Sjöström L, L.A., Peltonen M, et al., *Lifestyle, diabetes, and cardiovascular risk factors 10 years after bariatric surgery.* The New England Journal of Medicine (NEJM) 2004. **351**(26): p. 2683-2693.

142. Adams, T.D. and S.C. Hunt, *Cancer and obesity: effect of bariatric surgery.* World J Surg, 2009. **33**(10): p. 2028-33.
143. Pories WJ, M.J., Staton KM, *The surgical treatment of type two diabetes mellitus.* Surg Clin North Am 2011. **91**: p. 821–836.
144. Yska, J.P., et al., *Remission of Type 2 Diabetes Mellitus in Patients After Different Types of Bariatric Surgery: A Population-Based Cohort Study in the United Kingdom.* JAMA Surg, 2015. **150**(12): p. 1126-33.
145. Hatoum, I.J., et al., *Heritability of the weight loss response to gastric bypass surgery.* J Clin Endocrinol Metab, 2011. **96**(10): p. E1630-3.
146. Aslan, I.R., et al., *Bariatric surgery in a patient with complete MC4R deficiency.* Int J Obes (Lond), 2011. **35**(3): p. 457-61.
147. Jelin, E.B., et al., *Melanocortin-4 receptor signaling is not required for short-term weight loss after sleeve gastrectomy in pediatric patients.* Int J Obes (Lond), 2016. **40**(3): p. 550-3.
148. Valette M, P.C., Le Beyec J, Bouillot JL, Clement K, Czernichow S., *Melanocortin-4 receptor mutations and polymorphisms do not affect weight loss after bariatric surgery.* PLoS One, 2012. **7**(11): p. e48221.
149. Sarzynski, M.A., et al., *Associations of markers in 11 obesity candidate genes with maximal weight loss and weight regain in the SOS bariatric surgery cases.* Int J Obes (Lond), 2011. **35**(5): p. 676-83.
150. Kuipers, F., Bloks, V. & Groen, A. , *Beyond intestinal soap—bile acids in metabolic control.* . Nat Rev Endocrinol, 2014. **10**: p. 488-498.
151. Hatoum, I.J.e.a., *HWeight loss after gastric bypass is associated with a variant at 15q26.1.* American journal of human genetics, 2013. **92**(2): p. 827-834.
152. Dixon, J.B., le Roux, C. W., Rubino, F., & Zimmet, P., *Bariatric surgery for type 2 diabetes.* Lancet (London, England), 2012. **379**(9833): p. 2300–2311.
153. Hattersley, A.T., *Beyond the beta cell in diabetes.* Nature Genetics, 2006. **38**: p. 12-13.
154. Mohlke, K.L. and M. Boehnke, *Recent advances in understanding the genetic architecture of type 2 diabetes.* Hum Mol Genet, 2015. **24**(R1): p. R85-92.
155. Bonnefond, A. and P. Froguel, *Rare and common genetic events in type 2 diabetes: what should biologists know?* Cell Metab, 2015. **21**(3): p. 357-68.
156. Thomas, P.M.e.a., *Switching from Insulin to Oral Sulfonylureas in Patients with Diabetes Due to Kir6.2 Mutations.* The New England Journal of Medicine, 2006. **355**.
157. Flanagan, S.E., et al., *Mutations in KCNJ11, which encodes Kir6.2, are a common cause of diabetes diagnosed in the first 6 months of life, with the phenotype determined by genotype.* Diabetologia, 2006. **49**(6): p. 1190-7.
158. Glaysher MA, M.A., Prechtl CG, et al., *A randomised controlled trial of a duodenal-jejunal bypass sleeve device (EndoBarrier) compared with standard medical therapy for the management of obese subjects with type 2 diabetes mellitus.* BMJ Open, 2017. **7**(11).
159. Rafey MF, M.C., Abdalgwad R, et al. , *Effects of a Milk-Based Meal Replacement Program on Weight and Metabolic Characteristics in Adults with Severe Obesity.* . Diabetes Metab Syndr Obes., 2020. **13**(197-205).
160. Mohlke KL, B.M., *Recent advances in understanding the genetic architecture of type 2 diabetes.* . Hum Mol Genet. , 2015. **24**(1): p. 85-92.
161. Paolacci S, P.G., Paolini B, et al., *Mendelian non-syndromic obesity.* Acta Biomed, 2019. **90**: p. 87-89.

162. Alsters SI, G.A., Buxton JL, et al. , *Truncating Homozygous Mutation of Carboxypeptidase E (CPE) in a Morbidly Obese Female with Type 2 Diabetes Mellitus, Intellectual Disability and Hypogonadotrophic Hypogonadism.* . Plos One, 2015. **10**(6).
163. Hendricks, A.E., Bochukova, E.G., Marenne, G. et al., *Rare Variant Analysis of Human and Rodent Obesity Genes in Individuals with Severe Childhood Obesity.* Sci Rep, 2017. **7**(4394): p. 1-14.
164. Gilissen C, H.A., Brunner HG, Veltman JA. , *Disease gene identification strategies for exome sequencing.* Eur J Hum Genet., 2012. **258**: p. 1-8.
165. Rentzsch P, W.D., Cooper GM, Shendure J, Kircher M. , *CADD: predicting the deleteriousness of variants throughout the human genome.* . Nucleic Acids Res. , 2019. **47**: p. 886-894.
166. Ng PC, H.S., *SIFT: Predicting amino acid changes that affect protein function.* Nucleic Acids Res., 2003. **31**(13): p. 3812-3814.
167. Adzhubei I, J.D., Sunyaev SR. , *Predicting functional effect of human missense mutations using PolyPhen-2.* . Curr Protoc Hum Genet, 2013. **7**(20).
168. Rogers MF, S.H., Mort M, Cooper DN, Gaunt TR, Campbell C., *FATHMM-XF: accurate prediction of pathogenic point mutations via extended features.* . Bioinformatics., 2018. **34**(3): p. 511-513.
169. Stessman, H.A.F., et al., *Disruption of POGZ Is Associated with Intellectual Disability and Autism Spectrum Disorders.* Am J Hum Genet, 2016. **98**(3): p. 541-552.
170. White, J., et al., *POGZ truncating alleles cause syndromic intellectual disability.* Genome Med, 2016. **8**(1): p. 3.
171. Fitzgerald, T.W., et al., *Large-scale discovery of novel genetic causes of developmental disorders.* Nature, 2014. **519**(7542): p. 223-228.
172. Nozawa, R.-S., et al., *Human POGZ modulates dissociation of HP1α from mitotic chromosome arms through Aurora B activation.* Nature Cell Biology, 2010. **12**(7): p. 719-727.
173. Ye, Y., et al., *De novo POGZ mutations are associated with neurodevelopmental disorders and microcephaly.* Cold Spring Harb Mol Case Stud, 2015. **1**(1): p. a000455.
174. Laget S, J.M., Le Masson F, et al., *The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA.* PLoS One, 2010. **5**(8): p. e11982.
175. Talkowski ME, M.S., Rosenfeld JA, et al., *Assessment of 2q23.1 microdeletion syndrome implicates MBD5 as a single causal locus of intellectual disability, epilepsy, and autism spectrum disorder.* Am J Hum Genet, 2011. **89**(4): p. 551-563.
176. Mullegama, S.V. and S.H. Elsea, *Clinical and Molecular Aspects of MBD5-Associated Neurodevelopmental Disorder (MAND).* Eur J Hum Genet, 2016. **24**(9): p. 1235-43.
177. Hodge JC, M.E., Pillalamarri V, et al., *Disruption of MBD5 contributes to a spectrum of psychopathology and neurodevelopmental abnormalities.* Mol Psychiatry, 2014. **19**(3): p. 368-379.
178. Mullegama SV, P.L., Burns B, et al., *MBD5 haploinsufficiency is associated with sleep disturbance and disrupts circadian pathways common to Smith-Magenis and fragile X syndromes.* Eur J Hum Genet, 2015. **23**(6): p. 781-789.
179. Shichiji M, I.Y., Shimojima K, et al., *A cryptic microdeletion including MBD5 occurring within the breakpoint of a reciprocal translocation between chromosomes 2 and 5 in a patient with developmental delay and obesity.* Am J Med Genet A, 2013. **161A**(4): p. 850-855.

180. Marco EJ, A.A., Nair VP, et al., *Burden of de novo mutations and inherited rare single nucleotide variants in children with sensory processing dysfunction.* BMC Med Genomics, 2018. **11**(1): p. 50.
181. Han JY, J.W., Park J, Kim M, Kim Y, Lee IG., *Diagnostic approach with genetic tests for global developmental delay and/or intellectual disability: Single tertiary center experience.* Ann Hum Genet, 2018.
182. Wang Y, D.X., Bin R, et al. , *Genetic Variants Identified from Epilepsy of Unknown Etiology in Chinese Children by Targeted Exome Sequencing.* Sci Rep, 2017. **7**: p. 40319.
183. Han JY, J.J., Park J, Lee IG. , *Targeted Next-Generation Sequencing of Korean Patients With Developmental Delay and/or Intellectual Disability.* Front Pediatr, 2018. **6**: p. 391.
184. Tadros S, W.R., Waters JJ, et al., *Inherited 2q23.1 microdeletions involving the MBD5 locus.* Mol Genet Genomic Med, 2017. **5**(5): p. 608-613.
185. Woodbury-Smith M, N.R., Zarrei M, et al. , *Variable phenotype expression in a family segregating microdeletions of the NRXN1 and MBD5 autism spectrum disorder susceptibility genes.* NPJ Genom Med, 2017. **2**.
186. Ishizuka K, K.H., Yoshimi A, et al., *Investigation of single-nucleotide variants in MBD5 associated with autism spectrum disorders and schizophrenia phenotypes.* Nagoya J Med Sci, 2016. **78**(4): p. 465-474.
187. Pettersson M, V.H., Loid P, et al., *Copy Number Variants Are Enriched in Individuals With Early-Onset Obesity and Highlight Novel Pathogenic Pathways.* J Clin Endocrinol Metab, 2017. **102**(8): p. 3029-3039.
188. Ren D, L.M., Duan C, Rui L, *Identification of SH2-B as a key regulator of leptin sensitivity, energy balance, and body weight in mice.* Cell Metab, 2005. **2**: p. 95-104.
189. Ren D, Z.Y., Morris D, Li M, Li Z, Rui L, *Neuronal SH2B1 is essential for controlling energy and glucose homeostasis. .* J Clin Invest 2007. **117**: p. 397-406.
190. Morris DL, C.K., Rui L, *Critical role of the Src homology 2 (SH2) domain of neuronal SH2B1 in the regulation of body weight and glucose homeostasis in mice. .* Endocrinology, 2010. **151**(3643-3651).
191. Doche, M.E., et al., *Human SH2B1 mutations are associated with maladaptive behaviors and obesity.* J Clin Invest, 2012. **122**(12): p. 4732-6.
192. Pearce, L.R., et al., *Functional characterization of obesity-associated variants involving the alpha and beta isoforms of human SH2B1.* Endocrinology, 2014. **155**(9): p. 3219-26.
193. Bachmann-Gagescu, R., et al., *Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity.* Genet Med, 2010. **12**(10): p. 641-7.
194. Beckers S, Z.D., Van Gaal LF, Van Hul W., *Replication of the SH2B1 rs7498665 association with obesity in a Belgian study population. .* Obes Facts., 2011. **4**(6): p. 473-477.
195. Volckmar AL, B.F., Jarick I, et al. , *Mutation screen in the GWAS derived obesity gene SH2B1 including functional analyses of detected variants. .* BMC Med Genomics., 2015. **5**(65): p. 1-10.
196. Holzapfel C, G.H., Huth C, et al. , *Genes and lifestyle factors in obesity: results from 12,462 subjects from MONICA/KORA.* Int J Obes (Lond), 2010. **34**(10): p. 1538-1545.
197. Teams, G., *Hard-filtering germline short variants.* GATK, 2018.

198. Alsters SI, G.A., Buxton JL, et al., *Truncating Homozygous Mutation of Carboxypeptidase E (CPE) in a Morbidly Obese Female with Type 2 Diabetes Mellitus, Intellectual Disability and Hypogonadotrophic Hypogonadism.* . PLoS One., 2015. **10**(6).
199. An JY, C.A., Zhao Q, et al. , *Towards a molecular characterization of autism spectrum disorders: an exome sequencing and systems approach. Transl Psychiatry.* . Transl Psychiatry. , 2014. **4**(6): p. e394.
200. Macneil, D.J., *The role of melanin-concentrating hormone and its receptors in energy homeostasis.* Front Endocrinol (Lausanne), 2013. **4**: p. 49.
201. Pissios, P., et al., *Melanin-concentrating hormone receptor 1 activates extracellular signal-regulated kinase and synergizes with G(s)-coupled pathways.* Endocrinology, 2003. **144**(8): p. 3514-23.
202. Chee, M.J., et al., *Conditional deletion of melanin-concentrating hormone receptor 1 from GABAergic neurons increases locomotor activity.* Mol Metab, 2019. **29**: p. 114-123.
203. Lalonde, R. and S. Qian, *Exploratory activity, motor coordination, and spatial learning in Mchr1 knockout mice.* Behav Brain Res, 2007. **178**(2): p. 293-304.
204. Glier, M.B., et al., *The metabolic phenotype of SCD1-deficient mice is independent of melanin-concentrating hormone.* Peptides, 2010. **31**(1): p. 123-9.
205. Gibson WT, P.P., Trombly DJ, Luan J, Keogh J, Wareham NJ, Maratos-Flier E, O'Rahilly S, Farooqi IS., *Melanin-concentrating hormone receptor mutations and human obesity: functional analysis. Obes Res.* Obes Res., 2004. **12**(5): p. 743-749.
206. Wermter, A.-K.e.a., *Mutation analysis of the MCHR1 gene in human obesity.* European journal of endocrinology, 2005. **152**: p. 851-862.
207. Gibson WT, P.P., Trombly DJ, et al., *Melanin-concentrating hormone receptor mutations and human obesity: functional analysis.* Obes Res., 2004. **12**(5): p. 743-749.
208. Fleury C, N.M., Collins S, et al., *Uncoupling protein-2: a novel gene linked to obesity and hyperinsulinemia.* . Nat Genet, 1997. **15**(3): p. 269-272.
209. Boss, O., et al., *Uncoupling protein-3: a new member of the mitochondrial carrier family with tissue-specific expression.* FEBS Letters, 1997. **408**(1): p. 39-42.
210. Antonio Vidal-Puig, G.S., 1 Danica Grujic,1 Jeffrey S. Flier, and Bradford B. Lowell2 *UCP3 an uncoupling protein homologue expressed preferentially and abundantly in skeletal muscle and brown adipose tissue.* Biochemical and Biophysical Research Communications, 1997. **235**(1): p. 79-82.
211. M. Liebig, C.v.P., G. Heldmaier, M. Klingenspor, *Absence of UCP3 in Brown Adipose Tissue Does Not Impair Nonshivering Thermogenesis.* Physiol. Biochem. Zool, 2004. **77**(1): p. 116-126.
212. Edward Saltzman, M.D., Susan B. Roberts, Ph D *The Role of Energy Expenditure in Energy Regulation: Findings from a Decade of Research.* Nutrition Reviews, 1995. **53**(8): p. 209–220.
213. Acosta, A., et al., *Association of UCP-3 rs1626521 with obesity and stomach functions in humans.* Obesity (Silver Spring), 2015. **23**(4): p. 898-906.
214. Busiello, R.A., S. Savarese, and A. Lombardi, *Mitochondrial uncoupling proteins and energy metabolism.* Front Physiol, 2015. **6**: p. 36.
215. Lombardi, A., et al., *UCP3 translocates lipid hydroperoxide and mediates lipid hydroperoxide-dependent mitochondrial uncoupling.* J Biol Chem, 2010. **285**(22): p. 16599-605.

216. Pohl, E.E., et al., *Important Trends in UCP3 Investigation.* Front Physiol, 2019. **10**: p. 470.
217. Jia, J.J., et al., *The polymorphisms of UCP2 and UCP3 genes associated with fat metabolism, obesity and diabetes.* Obes Rev, 2009. **10**(5): p. 519-26.
218. Schrauwen, P. and M.K. Hesselink, *The role of uncoupling protein 3 in fatty acid metabolism: protection against lipotoxicity?* Proc Nutr Soc, 2004. **63**(2): p. 287-92.
219. Herrmann, S.M., et al., *Uncoupling protein 1 and 3 polymorphisms are associated with waist-to-hip ratio.* J Mol Med (Berl), 2003. **81**(5): p. 327-32.
220. Liu, Y., et al.,, *Linkage and association analyses of the UCP3 gene with obesity phenotypes in Caucasian families.* Physiol Genomics 2005. **22**: p. 197–203.
221. Brondani, L.A., et al., *Association of the UCP polymorphisms with susceptibility to obesity: case-control study and meta-analysis.* Mol Biol Rep, 2014. **41**(8): p. 5053-67.
222. Mazzarella R, P.G., Jones J, Jones C, Schlessinger D. , *Cloning and expression of an immunoglobulin superfamily gene (IGSF1) in Xq25. .* Genomics., 1998. **48**(I 2): p. 157-162.
223. Robakis T, B.B., Lin SH, Bernard DJ, Scheiffele P., *An internal signal sequence directs intramembrane proteolysis of a cellular immunoglobulin domain protein.* J Biol Chem, 2008. **283**(52): p. 36369-36376.
224. Tajima, T., et al., , *Neonatal screening and a new cause of congenital central hypothyroidism.* Ann Pediatr Endocrinol Metab, 2014. **19**(3): p. 117-121.
225. Bernard, D.J., et al., *From Consternation to Revelation: Discovery of a Role for IGSF1 in Pituitary Control of Thyroid Function. .* J Endocr Soc, 2018. **2**(3): p. 220-231.
226. Tajima, T., et al., *Recent advances in research on isolated congenital central hypothyroidism.* Clinical Pediatric Endocrinology, 2019. **28**(3): p. 69-79.
227. Romm, E., et al., *Myt1 family recruits histone deacetylase to regulate neural transcription.* J Neurochem, 2005. **93**(6): p. 1444-53.
228. Blanchet, P., et al., *MYT1L mutations cause intellectual disability and variable obesity by dysregulating gene expression and development of the neuroendocrine hypothalamus.* PLoS Genet, 2017. **13**(8): p. e1006957.
229. Vierbuchen, T., et al., *Direct conversion of fibroblasts to functional neurons by defined factors.* Nature, 2010. **463**(7284): p. 1035-41.
230. Vasconcelos, F.F., et al., *MyT1 Counteracts the Neural Progenitor Program to Promote Vertebrate Neurogenesis.* Cell Rep, 2016. **17**(2): p. 469-483.
231. De Rocker, N., et al., *Refinement of the critical 2p25.3 deletion region: the role of MYT1L in intellectual disability and obesity.* Genet Med, 2015. **17**(6): p. 460-6.
232. Doco-Fenzy, M., et al., *Early-onset obesity and paternal 2pter deletion encompassing the ACP1, TMEM18, and MYT1L genes.* Eur J Hum Genet, 2014. **22**(4): p. 471-9.
233. Stevens, S.J., et al., *MYT1L is a candidate gene for intellectual disability in patients with 2p25.3 (2pter) deletions.* Am J Med Genet A, 2011. **155A**(11): p. 2739-45.
234. Maria Clara Bonaglia, R.G., and Sergio Zanini, *A new patient with a terminal de novo 2p25.3 deletion of 1.9 Mb associated with early-onset of obesity, intellectual disabilities and hyperkinetic disorder.* Molecular Cytogenetics, 2014. **7**(53): p. 1-5.
235. Elliott, D.J., Dalgliesh, Caroline, Hysenaj, Gerald, and Ehrmann, Ingrid., *RBMX Family Proteins Connect the Fields of Nuclear RNA Processing, Disease and Sex Chromosome Biology.* The International Journal of Biochemistry & Cell Biology., 2019. **108**: p. 1-6.

236. Tsend-Ayush E, O.S.L., Grützner FS, et al., *RBMX gene is essential for brain development in zebrafish.* . Developmental dynamics : an official publication of the American Association of Anatomists, 2005. **234**(3): p. 682-688.
237. Shashi V, X.P., Schoch K, et al., *The RBMX gene as a candidate for the Shashi X-linked intellectual disability syndrome.* . Clin Genet., 2015. **88**(4): p. 386-390.
238. Izumi K, N.R., Zhang Z, et al. , *Germline gain-of-function mutations in AFF4 cause a developmental syndrome functionally linking the super elongation complex and cohesin.* Nat Genet, 2015. **47**(4): p. 338-344.
239. Raible SE, M.D., Bettale C, et al., *Clinical and molecular spectrum of CHOPS syndrome.* Am J Med Genet A, 2019. **179**(7): p. 1126-1138.
240. Chen, Y., and Patrick Cramer. , *Structure of the super-elongation complex subunit AFF4 C-terminal homology domain reveals requirements for AFF homo- and heterodimerization.* The Journal of biological chemistry, 2019. **294**(27): p. 10663-10673.
241. Carmona-Mora, P., et al., *RAI1 transcription factor activity is impaired in mutants associated with Smith-Magenis Syndrome.* PLoS One, 2012. **7**(9): p. e45155.
242. Vilboux T, C.C., Blancato JK, et al. , *Molecular analysis of the Retinoic Acid Induced 1 gene (RAI1) in patients with suspected Smith-Magenis syndrome without the 17p11.2 deletion.* PLoS One, 2011. **6**(8): p. 1-14.
243. Carmona-Mora P, C.C., Cao L, et al. , *RAI1 transcription factor activity is impaired in mutants associated with Smith-Magenis Syndrome.* . PLoS One., 2012. **7**(9): p. 45155.
244. Carmona-Mora P, E.C., Canales CP, Cao L, Molina J, et al., *Functional and cellular characterization of human Retinoic Acid Induced 1 (RAI1) mutations associated with Smith-Magenis Syndrome.* BMC Mol Biol, 2010. **11**(63).
245. Bi W, O.T., Nakamura H, et al. , *Inactivation of Rai1 in mice recapitulates phenotypes observed in chromosome engineered mouse models for Smith-Magenis syndrome.* . Hum Mol Genet., 2005. **14**(8): p. 983-995.
246. Bi W, Y.J., Shi X, et al. , *Rai1 deficiency in mice causes learning impairment and motor dysfunction, whereas Rai1 heterozygous mice display minimal behavioral phenotypes.* . Hum Mol Genet., 2007. **16**(15): p. 1802-1813.
247. Girirajan S, V.C., Szomju BB et al, *Genotype-phenotype correlation in Smith–Magenis syndrome: evidence that multiple genes in 17p11.2 contribute to the clinical spectrum.* . Genet Med 2006. **8**: p. 417-427.
248. Slager RE, N.T., Vlangos CN, Finucane B, Elsea SH, *Mutations in RAI1 associated with Smith–Magenis syndrome.* . Nat Genet, 2003. **33**: p. 466–468.
249. Edelman EA, G.S., Finucane B, et al., *Gender, genotype, and phenotype differences in Smith-Magenis syndrome: a meta-analysis of 105 cases.* Clin Genet. , 2007. **71**(6): p. 540-550.
250. Girirajan, S., Vlangos, C., Szomju, B. et al., *Genotype–phenotype correlation in Smith-Magenis syndrome: Evidence that multiple genes in 17p11.2 contribute to the clinical spectrum.* Genet Med 2006. **8**: p. 417-427.
251. Alaimo JT, B.L., Mullegama SV, Wills RD, Foster RH, Elsea SH. , *Individuals with Smith-Magenis syndrome display profound neurodevelopmental behavioral deficiencies and exhibit food-related behaviors equivalent to Prader-Willi syndrome.* Res Dev Disabil., 2015: p. 27-38.
252. Luscan A, L.I., Malan V, et al. , *Mutations in SETD2 cause a novel overgrowth condition.* . J Med Genet., 2014. **51**(8): p. 512-517.

253. Marzin P, R.S., Aldinger KA, et al. , *SETD2 related overgrowth syndrome: Presentation of four new patients and review of the literature.* Am J Med Genet C Semin Med Genet., 2019. **181**(4): p. 509-518.
254. Rana, K., R.A. Davey, and J.D. Zajac, *Human androgen deficiency: insights gained from androgen receptor knockout mouse models.* Asian J Androl, 2014. **16**(2): p. 169-77.
255. Grossmann, R.A.D.a.M., *Androgen Receptor Structure, Function and Biology From Bench to Bedside.* The Clinical Biochemist Reviews, 2016. **37**(1): p. 1-15.
256. Fagman, J.B., et al., *The androgen receptor confers protection against diet-induced atherosclerosis, obesity, and dyslipidemia in female mice.* FASEB J, 2015. **29**(4): p. 1540-50.
257. Fan, W., Yanase, T., Nomura, M., Okabe, T., Goto, K., Sato, T., Kawano, H., Kato, S., and Nawata, H. , *Androgen Receptor Null Male Mice Develop Late-Onset Obesity Caused by Decreased Energy Expenditure and Lipolytic Activity but Show Normal Insulin Sensitivity With High Adiponectin Secretion >.* Diabetes 2005. **54**: p. 1000–1008
258. Shiina, H., Matsumoto, T., Sato, T., Igarashi, K., Miyamoto, J., Takemasa, S., Sakari, M., Takada, I., Nakamura, T., Metzger, D., Chambon, P., Kanno, J., Yoshikawa, H., and Kato, S. , *Premature ovarian failure in androgen receptor-deficient mice.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(1): p. 224–229
259. Walters, K.A., et al., *Female mice haploinsufficient for an inactivated androgen receptor (AR) exhibit age-dependent defects that resemble the AR null phenotype of dysfunctional late follicle development, ovulation, and fertility.* Endocrinology, 2007. **148**(8): p. 3674-84.
260. Lin, H.Y., Xu, Q., Yeh, S., Wang, R. S., Sparks, J. D., and Chang, C. , *Insulin and Leptin Resistance With Hyperleptinemia in Mice Lacking Androgen Receptor.* Diabetes, 2005. **54**: p. 1717-1725
261. Yeh, S., et al., *Abnormal mammary gland development and growth retardation in female mice and MCF7 breast cancer cells lacking androgen receptor.* J Exp Med, 2003. **198**(12): p. 1899-908.
262. Lund, A., et al., *A novel sequence variation in the transactivation regulating domain of the androgen receptor in two infertile Finnish men.* Fertility and Sterility, 2003. **79**: p. 1647-1648.
263. Akcay, T., et al., *AR and SRD5A2 gene mutations in a series of 51 Turkish 46,XY DSD children with a clinical diagnosis of androgen insensitivity.* Andrology, 2014. **2**(4): p. 572-8.
264. Kalfa, N., et al., *Minor hypospadias: the "tip of the iceberg" of the partial androgen insensitivity syndrome.* PLoS One, 2013. **8**(4): p. e61824.
265. Imbert G, S.F., Yvert G, Devys D, Trottier Y, Garnier JM, Weber C, Mandel JL, Cancel G, Abbas N, Dürr A, Didierjean O, Stevanin G, Agid Y, Brice A, *Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats.* Nature Genetics, 2017. **8**(6): p. 1-21.
266. Pulst SM, N.A., Nechiporuk T, Gispert S, Chen XN, Lopes-Cendes I, Pearlman S, Starkman S, Orozco-Diaz G, Lunkes A, DeJong P, Rouleau GA, Auburger G, Korenberg JR, Figueroa C, Sahba S, *Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2* Nature Genetics, 1996. **14**(3): p. 269-276.

267. Takano H, S.K., et al., *Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT.* Nature Genetics, 1996. **14**(3): p. 277-284.
268. Huynh DP, D.B.M., Ho DH, Pulst SM, *Expression of ataxin-2 in brains from normal individuals and patients with Alzheimer's disease and spinocerebellar ataxia 2.* American Neurological Association, 2001. **45**(2): p. 232-241.
269. Ostrowski, L.A., A.C. Hall, and K. Mekhail, *Ataxin-2: From RNA Control to Human Health and Disease.* Genes (Basel), 2017. **8**(6).
270. Gwinn-Hardy K, C.J., Liu HC, Liu TY, Boss M, Seltzer W, Adam A, Singleton A, Koroshetz W, Waters C, Hardy J, Farrer M, *Spinocerebellar ataxia type 2 with parkinsonism in ethnic Chinese.* Neurology 2000. **55**(6): p. 800-805.
271. Elden, A.C., et al., *Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS.* Nature, 2010. **466**(7310): p. 1069-75.
272. Wilusz, C.J. and J. Wilusz, *Eukaryotic Lsm proteins: lessons from bacteria.* Nat Struct Mol Biol, 2005. **12**(12): p. 1031-6.
273. Jiménez-López D, G.P., *Insights into the evolution and domain structure of ataxin-2 proteins across eukaryotes.* BMC Research Notes 2014. **7**(453): p. 1-12.
274. Jimenez-Lopez, D., J. Bravo, and P. Guzman, *Evolutionary history exposes radical diversification among classes of interaction partners of the MLLE domain of plant poly(A)-binding proteins.* BMC Evol Biol, 2015. **15**: p. 195.
275. Bar, D.Z., et al., *Cell size and fat content of dietary-restricted Caenorhabditis elegans are regulated by ATX-2, an mTOR repressor.* Proc Natl Acad Sci U S A, 2016. **113**(32): p. E4620-9.
276. Lastres-Becker, I., et al., *Mammalian ataxin-2 modulates translation control at the pre-initiation complex via PI3K/mTOR and is induced by starvation.* Biochim Biophys Acta, 2016. **1862**(9): p. 1558-69.
277. Meierhofer, D., et al., *Ataxin-2 (Atxn2)-Knock-Out Mice Show Branched Chain Amino Acids and Fatty Acids Pathway Alterations.* Mol Cell Proteomics, 2016. **15**(5): p. 1728-39.
278. Carmo-Silva, S., et al., *Unraveling the Role of Ataxin-2 in Metabolism.* Trends Endocrinol Metab, 2017. **28**(4): p. 309-318.
279. Scoles, D.R., et al., *ETS1 regulates the expression of ATXN2.* Hum Mol Genet, 2012. **21**(23): p. 5048-65.
280. Lastres-Becker, I., et al., *Insulin receptor and lipid metabolism pathology in ataxin-2 knock-out mice.* Hum Mol Genet, 2008. **17**(10): p. 1465-81.
281. Kiehl, T.R., et al., *Generation and characterization of Sca2 (ataxin-2) knockout mice.* Biochem Biophys Res Commun, 2006. **339**(1): p. 17-24.
282. Auburger, G., et al., *12q24 locus association with type 1 diabetes: SH2B3 or ATXN2?* World J Diabetes, 2014. **5**(3): p. 316-27.
283. van de Loo, S., et al., *Ataxin-2 associates with rough endoplasmic reticulum.* Exp Neurol, 2009. **215**(1): p. 110-8.
284. Yokoshi, M., et al., *Direct binding of Ataxin-2 to distinct elements in 3′ UTRs promotes mRNA stability and protein expression.* Mol Cell, 2014. **55**(2): p. 186-98.
285. Koyano, S., et al., *Neuropathological staging of spinocerebellar ataxia type 2 by semiquantitative 1C2-positive neuron typing. Nuclear translocation of cytoplasmic 1C2*

underlies disease progression of spinocerebellar ataxia type 2. Brain Pathol, 2014: p. 599-606.
286. Haas, H. and P. Panula, *The role of histamine and the tuberomamillary nucleus in the nervous system.* Nat Rev Neurosci, 2003. **4**(2): p. 121-30.
287. Arrang, J.M.e.a., *Highly potent and selective ligands for histamine H3-receptors.* Nature, 1987. **327**: p. 117–123.
288. Panula, P. and S. Nuutinen, *The histaminergic network in the brain: basic organization and role in disease.* Nat Rev Neurosci, 2013. **14**(7): p. 472-87.
289. Toyota, H., et al.,, *Behavioral characterization of mice lacking histamine H(3) receptors.* Molecular Pharmacology 2002. **62**(2): p. 389 - 397.
290. Takahashi, K., et al., *Targeted disruption of H3 receptors results in changes in brain histamine tone leading to an obese phenotype.* Journal of Clinical Investigation, 2002. **110**(12): p. 1791-1799.
291. Li H, Z.Y., Wu Q. , *Role of corin in the regulation of blood pressure. .* Curr Opin Nephrol Hypertens., 2017. **26**(6): p. 67-73.
292. Wang W, L.X., Fukuda K, et al. , *Corin variant associated with hypertension and cardiac hypertrophy exhibits impaired zymogen activation and natriuretic peptide processing activity. .* Circ Res., 2008. **103**(5): p. 502-508.
293. Chan JC, K.O., Wu F, Morser J, Dole WP, Wu Q., *Hypertension in mice lacking the proatrial natriuretic peptide convertase corin.* Proc Natl Acad Sci U S A. , 2005. **102**(3): p. 785-790.
294. Motomura, Y., et al., *The transcription factor E4BP4 regulates the production of IL-10 and IL-13 in CD4+ T cells.* Nat Immunol, 2011. **12**(5): p. 450-9.
295. Li, L., Keverne,E.B.,Aparicio,S.A.,Ishino,F.,Barton,S.C.,andSurani, M.A., *Regulation of maternal behavior and offspring growth by paternally expressed Peg3.* Science, 1999. **284**: p. 623-629
296. Relaix , F., et al.,, *Pw1, a novel zinc finger gene implicated in the myogenic and neuronal lineages.* Developmental Biology, 1996. **177**(2): p. 383-396.
297. Yamaguchi, A., et al., *Peg3/Pw1 is involved in p53-mediated cell death pathway in brain ischemia/hypoxia.* J Biol Chem, 2002. **277**(1): p. 623-9.
298. Johnson, M.D., et al., *Peg3/Pw1 is a mediator between p53 and Bax in DNA damage-induced neuronal death.* J Biol Chem, 2002. **277**(25): p. 23000-7.
299. Wu, Y.D.a.X., *Peg3yPw1 promotes p53-mediated apoptosis by inducing Bax translocation from cytosol to mitochondria.* pnas, 2000. **97**(22): p. 12050-12055.
300. J. P. Curley, S.B.P., S. L. Dickson, R. Thresher, N. Miyoshi, M. A. Surani, and E. B. Keverne, *Increased body fat in mice with a targeted mutation of the paternally expressed imprinted gene Peg3.* The FASEB Journal 2005. **10**: p. 1-20.
301. LC., C., *The Phosphoinositide 3-Kinase Pathway.* Science, 2002. **296**: p. 1655–1657.
302. Braccini, L., et al., *PI3K-C2gamma is a Rab5 effector selectively controlling endosomal Akt2 activation downstream of insulin signalling.* Nat Commun, 2015. **6**: p. 7400.
303. Rozycka M, L.Y., Brown RA, Lau MR, Shipley JM, Fry MJ., *cDNA Cloning of a Third Human C2-Domain-Containing Class II Phosphoinositide 3-Kinase, PI3K-C2 g , and Chromosomal Assignment of This Gene (PIK3C2G) to 12p12.pd.* Genomics, 1998. **54**(13).
304. Margaria, J.P., et al., *Class II PI3Ks at the Intersection between Signal Transduction and Membrane Trafficking.* Biomolecules, 2019. **9**(3).

305. Daimon, M., et al., *Association of the PIK3C2G gene polymorphisms with type 2 DM in a Japanese population.* Biochem Biophys Res Commun, 2008. **365**(3): p. 466-71.
306. Saeed, M., *Locus and gene-based GWAS meta-analysis identifies new diabetic nephropathy genes.* Immunogenetics, 2018. **70**(6): p. 347-353.
307. Shia, W.C., et al., *Genetic copy number variants in myocardial infarction patients with hyperlipidemia.* BMC Genomics, 2011. **12 Suppl 3**: p. S23.
308. Anderson, D., et al., *First genome-wide association study in an Australian aboriginal population provides insights into genetic risk factors for body mass index and type 2 diabetes.* PLoS One, 2015. **10**(3): p. e0119333.
309. Gifford CA, R.S., Samarakoon R, et al. , *Oligogenic inheritance of a human heart disease involving a genetic modifier.* Science., 2019. **364**(6443): p. 865-870.
310. Schaaf CP, S.A., Sakai Y, et al., *Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders.* Hum Mol Genet. , 2011. **20**: p. 3366–3375.
311. Hoefele J, W.M., O'Toole JF, Otto EA, Schultheiss U, Dêschenes G, Attanasio M, Utsch B, Antignac C, Hildebrandt F, *Evidence of oligogenic inheritance in nephronophthisis.* J Am Soc Nephrol., 2007. **18**(10): p. 2789-2795.
312. Manara, E., Paolacci, S., D'Esposito, F. et al., *Mutation profile of BBS genes in patients with Bardet–Biedl syndrome: an Italian study.* Ital J Pediatr, 2019. **45**(72).
313. Li J, G.K., Yan H, et al. , *Reanalysis of whole exome sequencing data in patients with epilepsy and intellectual disability/mental retardation. .* Gene., 2019. **700**: p. 168-175.
314. Ewans, L.J., Schofield, D., Shrestha, R. et al. , *Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders. .* **Genet Med**, 2018. **20**: p. 1564–1574.
315. Salfati, E.L., Spencer, E.G., Topol, S.E. et al. , *Re-analysis of whole-exome sequencing data uncovers novel diagnostic variants and improves molecular diagnostic yields for sudden death and idiopathic diseases.* Genome Me, 2019. **11**(83): p. 1-8.
316. Pedersen, H., Gudmundsdottir, V., Pedersen, M. et al. , *Ranking factors involved in diabetes remission after bariatric surgery using machine-learning integrating clinical and genomic biomarkers.* Genomic Med, 2016. **1**.
317. Adkins DE, A.K., McClay JL, et al., *Genomewide pharmacogenomic study of metabolic side effects to antipsychotic drugs. .* Mol Psychiatry, 2011. **16**(3): p. 321-332.
318. Chagnon YC, B.A., Gendron D, et al. , *Possible association of the pro-melanin-concentrating hormone gene with a greater body mass index as a side effect of the antipsychotic olanzapine.* Am J Med Genet B Neuropsychiatr Genet., 2007. **144**(8): p. 1063-1069.
319. al, Y.H.e., *Genome-Wide Association Study Suggested the PTPRD Polymorphisms Were Associated With Weight Gain Effects of Atypical Antipsychotic Medications.* Schizophr Bull. , 2016. **42**: p. 814-823.
320. Roffeei SN, M.Z., Reynolds GP, et al., *Association of FTO, LEPR and MTHFR gene polymorphisms with metabolic syndrome in schizophrenia patients receiving antipsychotics.* Pharmacogenomics, 2014. **15**(4): p. 477-485.
321. Zhang JP, L.T., Zhang RX, et al. Pharmacogenetic Associations of Antipsychotic Drug-Related Weight Gain: A Systematic Review and Meta-analysis., *Pharmacogenetic Associations of Antipsychotic Drug-Related Weight Gain: A Systematic Review and Meta-analysis.* Schizophrenia Bulletin, 2016. **42**(6): p. 1418-1437.

322. Hatoum IJ, G.D., Cotsapas C, Daly MJ, Reitman ML, Kaplan LM. , *Weight loss after gastric bypass is associated with a variant at 15q26.1.* . Am J Hum Genet., 2013. **92**(5): p. 827-834.
323. Scott RA, S.L., Mägi R, et al. , *Scott RA, Scott LJ, Mägi R, et al. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. Diabetes. 2017;66(11):2888-2902. doi:10.2337/db16-1253.* Diabetes., 2017. **66**(11): p. 2888-2902.
324. Weiss LA, S.Y., Korn JM, et al. , *Association between microdeletion and microduplication at 16p11.2 and autism.* N Engl J Med. , 2008. **358**(7): p. 667-675.
325. Marshall CR, N.A., Vincent JB, et al. , *Structural variation of chromosomes in autism spectrum disorder.* . Am J Hum Genet., 2008. **82**(2): p. 477-488.
326. Hasstedt SJ, X.Y., Mao R, Lewis T, Adams TD, Hunt SC. , *A Copy Number Variant on Chromosome 20q13.3 Implicated in Thinness and Severe Obesity.* J Obes., 2015. **20**: p. 1-8.
327. Goberdhan DC, W.C., *PTEN: tumour suppressor, multifunctional growth regulator and more.* Hum Mol Genet, 2003. **12**: p. 239-248.
328. Poluri RTK, A.-W.É., *Genomic Deletion at 10q23 in Prostate Cancer: More Than PTEN Loss?.* Front Oncol. , 2018. **8**(246): p. 1-7.
329. C., E., *PTEN: one gene, many syndromes.* Hum Mutat., 2003. **22**(3): p. 183-198. .
330. Kleinendorst L, M.M., Cooiman MI, et al., *Genetic obesity: next-generation sequencing results of 1230 patients with obesity.* . J Med Genet., 2018. **55**(9): p. 578-586.
331. Foucan L , L.L., Durand E , Rambhojan C , Armand C , Michel CT , Billy R , Dhennin V , De Graeve F , Rabearivelo I , et al., *High Prevalence of Rare Monogenic Forms of Obesity in Obese Guadeloupean Afro-Caribbean Children. J Clin Endocrinol Metab 2018;103:539-545.* J Clin Endocrinol Metab, 2018. **103**: p. 539-545.
332. Loid P, M.T., Mäkitie RE, et al. , *Rare Variants in Genes Linked to Appetite Control and Hypothalamic Development in Early-Onset Severe Obesity.* Front Endocrinol (Lausanne), 2020. **11**(81): p. 1-9.
333. Creemers JW, C.H., Stijnen P, et al. , *Heterozygous mutations causing partial prohormone convertase 1 deficiency contribute to human obesity.* Diabetes, 2012. **61**(2): p. 383-390.
334. Severin R, S.A., Mahmoud AM, Arena R, Phillips SA. , *Precision Medicine in Weight Loss and Healthy Living.* Prog Cardiovasc Dis., 2019. **62**(1): p. 1-15.
335. Christensen P, B.H., Riecke BF, Leeds AR, Astrup A, Christensen R. , *Comparison of a low-energy diet and a very low- energy diet in sedentary obese individuals: a pragmatic randomized controlled trial.* . Clin Obes, 2011. **1**(1): p. 31-40.
336. Lean ME, L.W., Barnes AC, et al. , *Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. Lancet.* Lancet., 2018. **391**: p. 541-551.
337. Hector AJ, M.G., Churchward-Venne TA, et al. , *Whey protein supplementation preserves postprandial myofibrillar protein synthesis during short-term energy restriction in overweight and obese adults.* . J Nutr, 2015. **145**(2): p. 246–252.
338. van Meijl LE, M.R., *Effects of milk and milk constituents on postprandial lipid and glucose metabolism in overweight and obese men.* Br J Nutr, 2013. **110**(3): p. 413-419.
339. Dent R, M.R., Harper ME., *Factors affecting weight loss variability in obesity.* Metabolism, 2020. **113**: p. 1-13.

340. Hatoum IJ, G.D., Cotsapas C, et al. , *Heritability of the weight loss response to gastric bypass surgery.* . J Clin Endocrinol Metab. , 2011. **96**(10): p. 1603-1633.
341. Hainer V, S.A., Kunesova M, et al., *Intrapair resemblance in very low calorie diet-induced weight loss in female obese identical twins.* Int J Obes Relat Metab Disord, 2000. **24**(8): p. 1051-1056.
342. Dinesh Nagi, C.H., Roy Taylor, *Remission of type 2 diabetes: a position statement from the Association of British Clinical Diabetologists (ABCD) and the Primary Care Diabetes Society (PCDS).* The British Journal of Diabetes 2019. **19**(1): p. 73-76.
343. McCombie L, L.W., Taylor R, Kennon B, Sattar N, Lean MEJ. , *Beating type 2 diabetes into remission.* . BMJ., 2017. **13**: p. 358.
344. Loos RJ, R.T., *Gene-diet interactions on body weight changes.* . J Am Diet Assoc, 2005: p. 29-34.
345. Moore BS, M.U., Yost EA, et al., *Long-term weight-loss in gastric bypass patients carrying melanocortin 4 receptor variants.* . PLoS One. , 2014. **9**(4): p. 1-8.
346. Censani M, C.R., Deng L, et al. , *Weight loss after bariatric surgery in morbidly obese adolescents with MC4R mutations.* Obesity (Silver Spring). 2014. **22**(1): p. 225-231.
347. Aslan IR, C.G., Calton MA, Evans DS, Merriman RB, Vaisse C. , *Weight loss after Roux-en-Y gastric bypass in obese patients heterozygous for MC4R mutations.* Obesity surgery, 2011. **21**: p. 7.
348. Valette M, P.C., Le Beyec J, Bouillot JL, Clement K, Czernichow S. , *Melanocortin-4 receptor mutations and polymorphisms do not affect weight loss after bariatric surgery. PLoS One. 2012;7(11):e48221. doi:10.1371/journal.pone.0048221.* PLoS One., 2012. **7**(11): p. 1-5.
349. Li, Y., Zhang, H., Tu, Y. et al. , *Monogenic Obesity Mutations Lead to Less Weight Loss After Bariatric Surgery: a 6-Year Follow-Up Study.* OBES SURG, 2019. **29**: p. 1169-1173.
350. Cooiman, M., Kleinendorst, L., Aarts, E. et al. , *Genetic Obesity and Bariatric Surgery Outcome in 1014 Patients with Morbid Obesity.* . OBES SURG, 2020. **30**: p. 470-477.
351. Zechner JF, M.U., Satapati S, Berglund ED, Rossi J, et al., *Weight-independent effects of roux-en-Y gastric bypass on glucose homeostasis via melanocortin-4 receptors in mice and humans.* Gastroenterology, 2013. **144**: p. 580-590.
352. Morton GJ, C.D., Baskin DG, et al. , *Central nervous system control of food intake and body weight.* Nature, 2006. **443**: p. 289-295.
353. Hofsø D, F.F., Borgeraas H, et al. , *Gastric bypass versus sleeve gastrectomy in patients with type 2 diabetes (Oseberg): a single-centre, triple-blind, randomised controlled trial.* . Lancet Diabetes Endocrinol. , 2019. **7**(12): p. 912-924.
354. Tsilingiris D, K.C., Kokkinos A., *Remission of Type 2 Diabetes Mellitus after Bariatric Surgery: Fact or Fiction?.* Int J Environ Res Public Health., 2019. **16**(17): p. 1-22.
355. Pérez-Pevida B, E.J., Miras AD, Frühbeck G. , *Mechanisms Underlying Type 2 Diabetes Remission After Metabolic Surgery.* . Front Endocrinol (Lausanne), 2019. **10**(641): p. 1-9.
356. Clément, K.e.a., *Efficacy and safety of setmelanotide, an MC4R agonist, in individuals with severe obesity due to LEPR or POMC deficiency: single-arm, open-label, multicentre, phase 3 trials.* The lancet. Diabetes & endocrinology 2020. **8**(12): p. 960-970.
357. Katsanis N, A.S., Badano JL, et al. , *Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder.* Science., 2001. **293**(5538): p. 2256-2259.

APPENDIX

Appendix 1. Human and mouse obesity and Diabetes mellitus (T2D) genes

Table S1.1 List of obesity and syndromic obesity genes

GENE SYMBOL	GENE NAME
ADCY3	Adenylate Cyclase 3
ADIPOQ	Adiponectin, C1Q And Collagen Domain Containing
AFF4	AF4/FMR2 Family Member 4
ALMS1	ALMS1 Centrosome And Basal Body Associated Protein
ARL14EP	ADP Ribosylation Factor Like GTPase 14 Effector Protein
ARL6	ADP Ribosylation Factor Like GTPase 6
BBIP1	BBSome Interacting Protein 1
BBS1	Bardet-Biedl Syndrome 1
BBS10	Bardet-Biedl Syndrome 10
BBS12	Bardet-Biedl Syndrome 12
BBS2	Bardet-Biedl Syndrome 2
BBS4	Bardet-Biedl Syndrome 4
BBS5	Bardet-Biedl Syndrome 5
BBS7	Bardet-Biedl Syndrome 7
BBS9	Bardet-Biedl Syndrome 9
BDNF	Brain Derived Neurotrophic Factor
CADM2	Cell Adhesion Molecule 2
CARTPT	CART Prepropeptide
CEP19	Centrosomal Protein 19
CEP290	Centrosomal Protein 290
CLMP	CXADR Like Membrane Protein
COA3	Cytochrome C Oxidase Assembly Factor 3
CPE	Carboxypeptidase E
DIA1R	Divergent Protein Kinase Domain 2B
DNAAF1	Dynein Axonemal Assembly Factor 1
DYRK1B	Dual Specificity Tyrosine Phosphorylation Regulated Kinase 1B
FAAH	Fatty Acid Amide Hydrolase
FAM3C	FAM3 Metabolism Regulating Signaling Molecule C
FOXA3	Forkhead Box A3
GHSR	Growth Hormone Secretagogue Receptor
GNB3	G Protein Subunit Beta 3
HDAC8	Histone Deacetylase 8
IFT27	Intraflagellar Transport 27
IGSF1	Immunoglobulin Superfamily Member 1
INPP5E	Inositol Polyphosphate-5-Phosphatase E

KSR2	Kinase Suppressor Of Ras 2
LEP	Leptin
LEPR	Leptin Receptor
LRP2	LDL Receptor Related Protein 2
LZTFL1	Leucine Zipper Transcription Factor Like 1
MAGEL2	MAGE Family Member L2
MANF	Mesencephalic Astrocyte Derived Neurotrophic Factor
MBD5	Methyl-CpG Binding Domain Protein 5
MC3R	Melanocortin 3 Receptor
MC4R	Melanocortin 4 Receptor
MCHR1	Melanin Concentrating Hormone Receptor 1
MKKS	McKusick-Kaufman Syndrome
MKS1	MKS Transition Zone Complex Subunit 1
MRAP2	Melanocortin 2 Receptor Accessory Protein 2
MYT1L	Myelin Transcription Factor 1 Like
NMB	Neuromedin B
NPY2R	Neuropeptide Y Receptor Y2
NPY4R	Neuropeptide Y Receptor Y4
NR0B2	Nuclear Receptor Subfamily 0 Group B Member 2
NTRK2	Neurotrophic Receptor Tyrosine Kinase 2
NUCB2	Nucleobindin 2
PCSK1	Proprotein Convertase Subtilisin/Kexin Type 1
POGZ	Pogo Transposable Element Derived With ZNF Domain
POMC	Proopiomelanocortin
POU3F2	POU Class 3 Homeobox 2
PRKD1	Protein Kinase D1
PTEN	Phosphatase And Tensin Homolog
PYY	Peptide YY
RAB23	RAB23, Member RAS Oncogene Family
RAI1	Retinoic Acid Induced 1
RBMX	RNA Binding Motif Protein X-Linked
RETN	Resistin
RORA	RAR Related Orphan Receptor A
SDCCAG8	SHH Signaling And Ciliogenesis Regulator SDCCAG8
SETD2	SET Domain Containing 2, Histone Lysine Methyltransferase
SH2B1	SH2B Adaptor Protein 1
SIM1	SIM BHLH Transcription Factor 1
SLC35D3	Solute Carrier Family 35 Member D3
TAOK2	TAO Kinase 2
TMEM18	Transmembrane Protein 18

TRIM32	Tripartite Motif Containing 32
TTC8	Tetratricopeptide Repeat Domain 8
TUB	TUB Bipartite Transcription Factor
UCP1	Uncoupling Protein 1
UCP3	Uncoupling Protein 3
VPS13B	Vacuolar Protein Sorting 13 Homolog B
WDPCP	WD Repeat Containing Planar Cell Polarity Effector
WNT10B	Wnt Family Member 10B

Table S1.2: Diabetes mellitus (T2D) genes list.

GENE SYMBOL	GENE NAME
ABCC8	ATP Binding Cassette Subfamily C Member 8
ACACB**	Acetyl-CoA Carboxylase Beta
AKR1B1	Aldo-Keto Reductase Family 1 Member B
AKT2	AKT Serine/Threonine Kinase 2
APOE	Apolipoprotein E
APPL1	Adaptor Protein, Phosphotyrosine Interacting With PH Domain And Leucine Zipper 1
BLK	BLK Proto-Oncogene, Src Family Tyrosine Kinase
CASR	Calcium Sensing Receptor
CAT	Catalase
CDKAL1	CDK5 Regulatory Subunit Associated Protein 1 Like 1
CEL	Carboxyl Ester Lipase
CISD2	CDGSH Iron Sulfur Domain 2
DNAJC3	DnaJ Heat Shock Protein Family (Hsp40) Member C3
EIF2AK3	Eukaryotic Translation Initiation Factor 2 Alpha Kinase 3
FN3K	Fructosamine 3 Kinase
FOXA2	Forkhead Box A2
FOXP3	Forkhead Box P3
GATA6	GATA Binding Protein 6
GCGR	Glucagon Receptor
GCK	Glucokinase
GH1	Growth Hormone 1
GIPR	Gastric Inhibitory Polypeptide Receptor
GLP1R	Glucagon Like Peptide 1 Receptor
GLUD1	Glutamate Dehydrogenase 1
GPD2	Glycerol-3-Phosphate Dehydrogenase 2
HADH	Hydroxyacyl-CoA Dehydrogenase
HFE	Homeostatic Iron Regulator
HK2	Hexokinase 2
HMGA1	High Mobility Group AT-Hook 1
HNF1A	HNF1 Homeobox A
HNF1B	HNF1 Homeobox B
HNF4A	Hepatocyte Nuclear Factor 4 Alpha
IAPP	Islet Amyloid Polypeptide
IER3IP1	Immediate Early Response 3 Interacting Protein 1
INS	Insulin
INSR*	Insulin Receptor
IRS1	Insulin Receptor Substrate 1

IRS2	Insulin Receptor Substrate 2
KCNJ11	Potassium Inwardly Rectifying Channel Subfamily J Member 11
KLF11	Kruppel Like Factor 11
KRT17	Keratin 17
LIPE	Lipase E, Hormone Sensitive Type
MAPK8IP1	Mitogen-Activated Protein Kinase 8 Interacting Protein 1
MNX1	Motor Neuron And Pancreas Homeobox 1
MTNR1B	Melatonin Receptor 1B
MTTP*	Microsomal Triglyceride Transfer Protein
NEUROD1	Neuronal Differentiation 1
NEUROG3	Neurogenin 3
NKX2.2	NK2 Homeobox 2
PASK	PAS Domain Containing Serine/Threonine Kinase
PAX4	Paired Box 4
PCBD1	Pterin-4 Alpha-Carbinolamine Dehydratase 1
PDX1	Pancreatic And Duodenal Homeobox 1
PLAGL1	PLAG1 Like Zinc Finger 1
PPARA*	Peroxisome Proliferator Activated Receptor Alpha
PPARG*	Peroxisome Proliferator Activated Receptor Gamma
PPP1R3A	Protein Phosphatase 1 Regulatory Subunit 3A
PTF1A	Pancreas Associated Transcription Factor 1a
PTPRD	Protein Tyrosine Phosphatase Receptor Type D
RFX6	Regulatory Factor X6
SLC19A2	Solute Carrier Family 19 Member 2
SLC2A2	Solute Carrier Family 2 Member 2
SLC2A4	Solute Carrier Family 2 Member 4
SREBF1	Sterol Regulatory Element Binding Transcription Factor 1
TBC1D4	TBC1 Domain Family Member 4
TRMT10A	TRNA Methyltransferase 10A
UCP2	Uncoupling Protein 2
WFS1	Wolframin ER Transmembrane Glycoprotein
ZFP57	ZFP57 Zinc Finger Protein

Table S1.3: Obesity mouse model genes list.

GENE SYMBOL	GENE NAME
ACOT11	Acyl-CoA Thioesterase 11
ADIPOR2	Adiponectin Receptor 2
ADRB2	Adrenoceptor Beta 2
ADRB3	Adrenoceptor Beta 3
AGRP	Agouti Related Neuropeptide
AHSG	Alpha 2-HS Glycoprotein
ANGPTL6	Angiopoietin Like 6
ANKRD26	Ankyrin Repeat Domain 26
APOC3	Apolipoprotein C3
AR	Androgen Receptor
ASIP	Agouti Signaling Protein
ATXN2	Ataxin 2
BRS3	Bombesin Receptor Subtype 3
CCL2	C-C Motif Chemokine Ligand 2
CLOCK	Clock Circadian Regulator
CNR1	Cannabinoid Receptor 1
CORIN	Corin, Serine Peptidase
CRTC1	CREB Regulated Transcription Coactivator 1
ESR1	Estrogen Receptor 1
ESRRA	Estrogen Related Receptor Alpha
FABP2	Fatty Acid Binding Protein 2
FABP4	Fatty Acid Binding Protein 4
FABP5	Fatty Acid Binding Protein 5
FEN1	Flap Structure-Specific Endonuclease 1
FOXO1	Forkhead Box O1
FTO	FTO Alpha-Ketoglutarate Dependent Dioxygenase
GPR12	G Protein-Coupled Receptor 12
GPR45	G Protein-Coupled Receptor 45
GRB10	Growth Factor Receptor Bound Protein 10
GUCY2C	Guanylate Cyclase 2C
HCRT	Hypocretin Neuropeptide Precursor
HDC	Histidine Decarboxylase
HMGA2	High Mobility Group AT-Hook 2
HRH1	Histamine Receptor H1
HRH3	Histamine Receptor H3
ICAM1	Intercellular Adhesion Molecule 1
IL18	Interleukin 18

IL6	Interleukin 6
IRX3	Iroquois Homeobox 3
KDM3A	Lysine Demethylase 3A
MYD88	MYD88 Innate Immune Signal Transduction Adaptor
NCOA1	Nuclear Receptor Coactivator 1
NEIL1	Nei Like DNA Glycosylase 1
NMU	Neuromedin U
NMUR2	Neuromedin U Receptor 2
NPY1R	Neuropeptide Y Receptor Y1
NPY5R	Neuropeptide Y Receptor Y5
NTSR1	Neurotensin Receptor 1
OTP	Orthopedia Homeobox
PAM	Peptidylglycine Alpha-Amidating Monooxygenase
PEG3	Paternally Expressed 3
PIK3C2G	Phosphatidylinositol-4-Phosphate 3-Kinase Catalytic Subunit Type 2 Gamma
PPARGC1A	PPARG Coactivator 1 Alpha
PPARGC1B	PPARG Coactivator 1 Beta
PRKAA2	Protein Kinase AMP-Activated Catalytic Subunit Alpha 2
PRKAB2	Protein Kinase AMP-Activated Non-Catalytic Subunit Beta 2
PRKAG1	Protein Kinase AMP-Activated Non-Catalytic Subunit Gamma 1
PRLH	Prolactin Releasing Hormone
PROX1	Prospero Homeobox 1
ROCK1	Rho Associated Coiled-Coil Containing Protein Kinase 1
ROCK2	Rho Associated Coiled-Coil Containing Protein Kinase 2
RSC1A1	Regulator Of Solute Carriers 1
SCD	Stearoyl-CoA Desaturase
SDC3	Syndecan 3
SOCS3	Suppressor Of Cytokine Signaling 3
SST	Somatostatin
STAT3	Signal Transducer And Activator Of Transcription 3
TLR5	Toll Like Receptor 5
TYK2	Tyrosine Kinase 2
UBB	Ubiquitin B

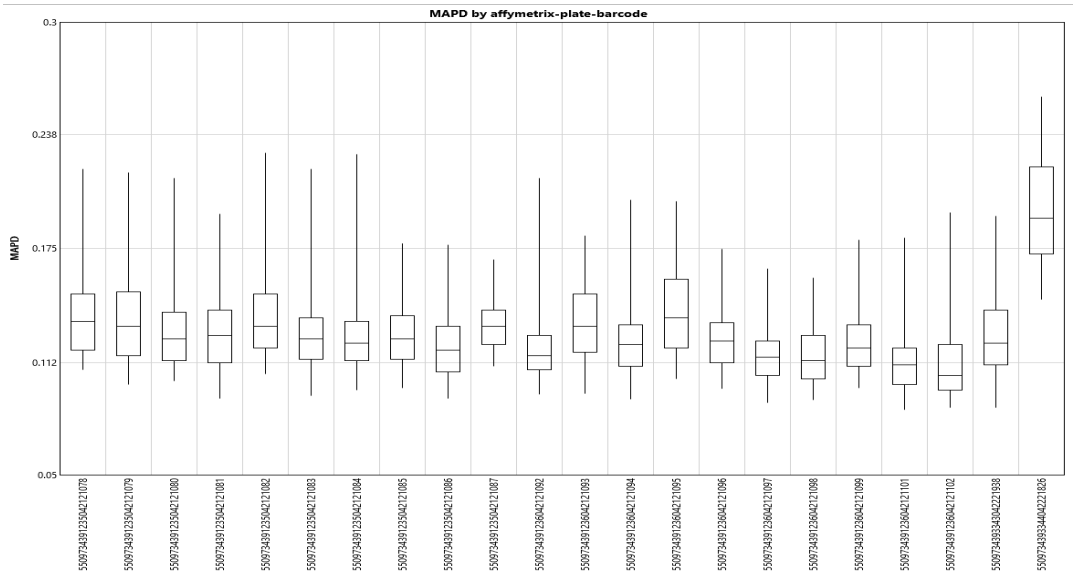
Table S1.4: Diabetes mellitus (T2D) mouse model genes list.

GENE SYMBOL	GENE NAME
ARHGEF11	Rho Guanine Nucleotide Exchange Factor 11
ARNTL	Aryl Hydrocarbon Receptor Nuclear Translocator Like
CAPN10	Calpain 10
CTF1	Cardiotrophin 1
CYB5R4	Cytochrome B5 Reductase 4
ENPP1	Ectonucleotide Pyrophosphatase/Phosphodiesterase 1
FEM1B	Fem-1 Homolog B
FOXM1	Forkhead Box M1
GADD45GIP1	GADD45G Interacting Protein 1
GHR	Growth Hormone Receptor
HMOX1	Heme Oxygenase 1
HTR2C	5-Hydroxytryptamine Receptor 2C
IFNGR2	Interferon Gamma Receptor 2
IGF2	Insulin Like Growth Factor 2
IGF2BP2	Insulin Like Growth Factor 2 MRNA Binding Protein 2
IL1R1	Interleukin 1 Receptor Type 1
LIPC	Lipase C, Hepatic Type
MADD	MAP Kinase Activating Death Domain
MAFA	MAF BZIP Transcription Factor A
NFE2L1	Nuclear Factor, Erythroid 2 Like 1
OAS1A	Oligoadenylate Synthetase 1
PHOX2A	Paired Like Homeobox 2A
PPARD	Peroxisome Proliferator Activated Receptor Delta
PPP1R3C	Protein Phosphatase 1 Regulatory Subunit 3C
PRCP	Prolylcarboxypeptidase
PRKCI	Protein Kinase C Iota
PTPN1	Protein Tyrosine Phosphatase Non-Receptor Type 1
PTPN22	Protein Tyrosine Phosphatase Non-Receptor Type 22
SIRT1	Sirtuin 1
SLC30A8	Solute Carrier Family 30 Member 8
SLC5A2	Solute Carrier Family 5 Member 2
SNAP25	Synaptosome Associated Protein 25
TBC1D1	TBC1 Domain Family Member 1
TCF7L2	Transcription Factor 7 Like 2
TGM2	Transglutaminase 2
TP53INP1	Tumor Protein P53 Inducible Nuclear Protein 1

VDR	Vitamin D Receptor
GENE SYMBOL	Gene Name
ARHGEF11	Rho Guanine Nucleotide Exchange Factor 11
ARNTL	Aryl Hydrocarbon Receptor Nuclear Translocator Like
CAPN10	Calpain 10
CTF1	Cardiotrophin 1
CYB5R4	Cytochrome B5 Reductase 4
ENPP1	Ectonucleotide Pyrophosphatase/Phosphodiesterase 1
FEM1B	Fem-1 Homolog B
FOXM1	Forkhead Box M1
GADD45GIP1	GADD45G Interacting Protein 1
GHR	Growth Hormone Receptor
HMOX1	Heme Oxygenase 1
HTR2C	5-Hydroxytryptamine Receptor 2C
IFNGR2	Interferon Gamma Receptor 2
IGF2	Insulin Like Growth Factor 2
IGF2BP2	Insulin Like Growth Factor 2 MRNA Binding Protein 2
IL1R1	Interleukin 1 Receptor Type 1
LIPC	Lipase C, Hepatic Type
MADD	MAP Kinase Activating Death Domain
MAFA	MAF BZIP Transcription Factor A
NFE2L1	Nuclear Factor, Erythroid 2 Like 1
OAS1A	Oligoadenylate Synthetase 1
PHOX2A	Paired Like Homeobox 2A
PPARD	Peroxisome Proliferator Activated Receptor Delta
PPP1R3C	Protein Phosphatase 1 Regulatory Subunit 3C
PRCP	Prolylcarboxypeptidase
PRKCI	Protein Kinase C Iota
PTPN1	Protein Tyrosine Phosphatase Non-Receptor Type 1
PTPN22	Protein Tyrosine Phosphatase Non-Receptor Type 22
SIRT1	Sirtuin 1
SLC30A8	Solute Carrier Family 30 Member 8
SLC5A2	Solute Carrier Family 5 Member 2
SNAP25	Synaptosome Associated Protein 25
TBC1D1	TBC1 Domain Family Member 1
TCF7L2	Transcription Factor 7 Like 2
TGM2	Transglutaminase 2
TP53INP1	Tumor Protein P53 Inducible Nuclear Protein 1
VDR	Vitamin D Receptor

Appendix 2. Boxplots of MAPD and WavinessSD by plate derived from CNV calling using universal reference (A and B)

A



B