

MSc in Bioinformatics

Master Thesis

**Detection of obesity susceptibility GENOMIC variants in Spanish population using sequencing data**

Isaac David De la Hoz Saltaren

*Supervised by*

Juan R González, ISGlobal

*Academic tutor*

Raquel Egea, UAB



July 2019

# Acknowledgments

# Approval and signature

# Abstract

# Table of contents

[Acknowledgments 4](#_Toc9867901)

[Approval and signature 5](#_Toc9867902)

[Abstract 6](#_Toc9867903)

[Table of contents 7](#_Toc9867904)

[1. Introduction 8](#_Toc9867905)

[1.1. Obesity 8](#_Toc9867906)

[1.1.1. Obesity in Spanish population 8](#_Toc9867907)

[1.2. Causes of obesity 8](#_Toc9867908)

[1.2.1. Body weight control 8](#_Toc9867909)

[1.2.2. Pathway of energy homeostasis 9](#_Toc9867910)

[1.2.3. Disorder of energy homeostasis 10](#_Toc9867911)

[1.3. Importance of genetics in obesity 10](#_Toc9867912)

[1.4. Obesity susceptibility variants 10](#_Toc9867913)

[2. Objectives 12](#_Toc9867914)

[3. Material and Methods 13](#_Toc9867915)

[3.1. Data description 13](#_Toc9867916)

[3.1.1. Samples 13](#_Toc9867917)

[3.1.2. Methodology of DNA extraction and sequencing 13](#_Toc9867918)

[3.2. SNV detection 14](#_Toc9867919)

[3.2.1. Variant Calling 14](#_Toc9867920)

[3.2.2. Significant variants detection 14](#_Toc9867921)

[3.3. CNV Analysis 14](#_Toc9867922)

[4. Results 15](#_Toc9867923)

[5. Discussion 16](#_Toc9867924)

[6. Conclusion 17](#_Toc9867925)

[7. Bibliography 18](#_Toc9867926)

[Appendix 20](#_Toc9867927)

# Introduction

## Obesity

Obesity is a medical condition that is defined as excessive accumulation of fat that is sufficient to adversely affect health1,2. According World Health Organization (WHO), people with a body mass index (BMI; weight in kg/height in m2) higher than 30 kg/m2 are considered obese and higher than 25 kg/m2 are considered overweight. The 30% of Americans and 10%–20% of Europeans are classified as obese, with the prevalence rising in many developing countries1. Being overweight or obese can have a serious impact on health. Carrying extra fat leads to serious health consequences such as cardiovascular disease (mainly heart disease and stroke), type 2 diabetes, musculoskeletal disorders like osteoarthritis, and some cancers (endometrial, breast and colon). These conditions cause premature death and substantial disability3.

### Obesity in Spanish population

The 26.6% of adult population in Spain are considered as obese and 62% are overweight. Furthermore, in Spanish childhood population, the prevalence of obesity has been increasing in recent years to such a stent that, the prevalence has increase until 18.3% among children (0-9 years of age) and 30% among adolescents (10-19 years of age). This data has made Spain be considered by WHO as one of the countries with the highest prevalence of obesity and overweight4.

## Causes of obesity

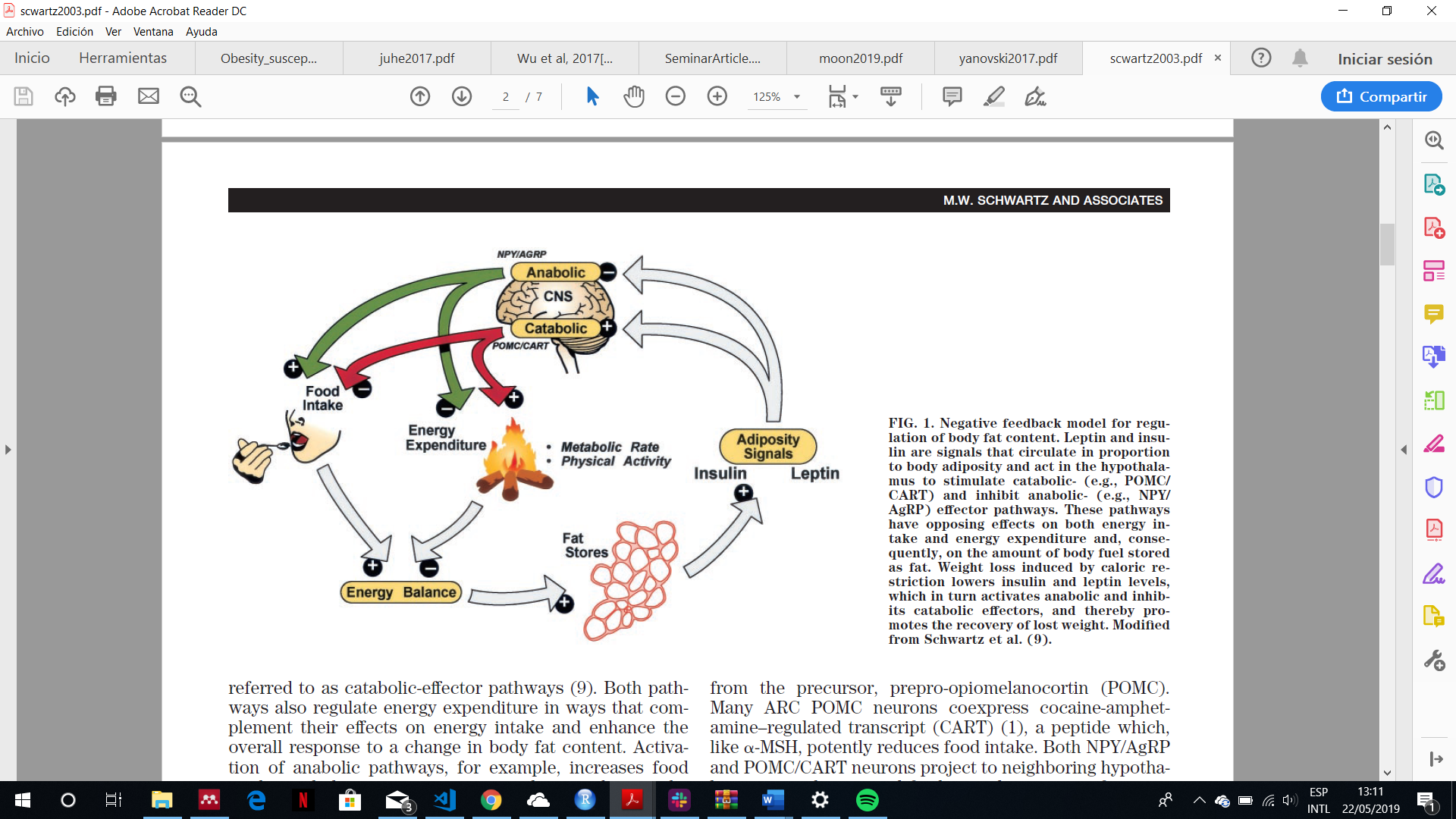
### Body weight control

The humans are able to regulate their body weight over long periods of time despite day-to-day variation in the number of calories consumed and in levels of energy expenditure1. The maintenance of constant body weight and body composition requires two conditions be met. [1] An even energy balance must be attained, i.e. energy expenditure must on average be equal to energy intake. [2] There must be an even balance for each individual substrate, i.e. protein, carbohydrate and fat oxidation must be equal to protein, carbohydrate and fat intakes respectively If this state were not present, body composition would inevitably change, even during isoenergetic feeding. If energy intake continuously exceeds energy expenditure, the excess energy ingested has to be deposited in order to increase the nutrients stores.

There is clearly a ‘hierarchy’ in substrate oxidation during overfeeding. Any increase in protein intake will rapidly lead to stimulation of protein oxidation, restoring a steady protein balance. The same is true for carbohydrates, the oxidation of which increases over 1–3 d to match any increase in carbohydrate intake. The result of this hierarchy is that excess energy intake leads essentially to fat storage, mainly in subcutaneous and visceral adipose tissue. In contrast, a period of hypoenergetic feeding will lead to a negative fat balance and a loss of adipose tissue1,5.

### Pathway of energy homeostasis

Pathways that stimulate food intake and promote weight gain are referred to here as anabolic-effector pathways, whereas those that promote anorexia and depletion of body fat are referred to as catabolic-effector pathways6. Both pathways also regulate energy expenditure in ways that complement their effects on energy intake and enhance the overall response to a change in body fat content. Activation of anabolic pathways, for example, increases food intake and decreases energy expenditure, whereas the reverse is true for catabolic pathways6. Anabolic and catabolic pathways are generally regulated in a reciprocal manner, such that increases in the activity of one are often accompanied by decreases in the other6. The anabolic and catabolic pathways sense changes in energy balance due to the hormones leptin and insulin that circulate on blood proportionate to body fat mass and enter into the brain, where they bind to an and activate their respective receptors on the plasma membrane of targets neurons6. Low concentrations of leptin and insulin increase energy intake and reduce energy expenditure. Hence, the reciprocal nature of the neuronal response to an energy deficit (activation of anabolic pathways and inhibition of catabolic pathways) may be accounted for, at least in part, by reduced levels of these two hormones (Figure 1)1,6,7.



**Figure 1: energy balance pathway.** This model explains the body fat mass storage mechanism. When the food intake increases, the excess of energy is stored at adipocytes. Once the adipocytes have stored energy in form of fat, they produce leptin that activate central nervous system pathways which it stimulates the decrease of energy intake (satiety signals) and the increase in energy expenditure. Figure adapted from Schwartz et.al 20036.

### Disorder of energy homeostasis

Disorders of energy homeostasis are fundamentally due to factors that disrupt the balance between energy intake and expenditure over time, the utilization of substrates (fat, protein, carbohydrate), and/or nutrient partitioning (storage of excess calories). The environmental influences of weight gain such as the adoption of sedentary lifestyles due to reduced physical activity at work and in leisure time coupled with an abundance of easily available, energy-rich, highly palatable foods represents a nutrition transition that, according with the WHO, is now one of the greatest factors to ill health worldwide1.

Other factor that disrupt the energy homeostasis is the genetic. In any environment, either energy rich or energy lacking, there are considerable distribution of different body weight among people. This evidence says that not only environmental factors cause obesity but also genetics factors1,2,7–9.

## Importance of genetics in obesity

The genetic contribution to body weight has been established through family studies, investigations of parent-offspring relationships, and the study of twins and adopted children10,11. These studies estimate a heritability (Fraction of the total phenotypic variance of a quantitative trait attributable to genes in a specified environment) of 40-70%. This genetic predisposition has been widely recognized in the human evolutionary history. Obesity stem from natural selection on our ancient ancestors favouring “thrifty genes”, defined as conferring a phenotype of being extremely efficient keeping all extra energy during periods of food abundance in order to deal with large famine periods. In modern society, however, with plentiful and continuous food, this thrifty phenotype process deleterious because it promotes efficient storage of fat for a famine period that never comes1,12.

## Obesity susceptibility variants

As explained before, genetics factors have big genetics influences in the appearance of the obesity. These genetic influences are likely to operate across the weight spectrum but may be more penetrant when studying childhood-onset obesity and at both extremes of the BMI distribution (thinness and severe obesity). Genetic variance of obesity depends on 5 factors1:

* The nature and amount of mutational variance in a population
* The segregation and frequency of the alleles that influence a trait in a particular population. The lower MAF the worse phenotype.
* The effect size of the variant. The effect that a variant can produce could be additive or non-additive.
* The mode of gene action
* The degree of genetic control of phenotypic variance of the trait in question.

Until now, 97 genetic loci have been discovered associated with BMI through Genome-wide association studies (GWAS) approaches13. Nonetheless, these loci only explain 2.7% of the variances of BMI. Several monogenic drivers of isolated early-onset obesity have been identified, emphasizing the importance of energy homeostasis (LEP, LEPR, POMC, MC4R) and cilia function (CEP19)9,14. In addition, the gene that encodes the fat mass associated protein (FTO) has unequivocally been associated with obesity by the existence of single nucleotide polymorphism (SNPs) in both childhood and adult obesity populations.

In the other hand, several copy number variants (CNV) that contribute with the obesity heritability have been reported. including deletions upstream of the NEGR1 gene15, proximal and distal deletions at 16p11.216,17, gains at 10q26.6 containing the CYP2E1 gene (MIM 124040)18, and homozygous deletions at 11q11 encompassing olfactory receptor genes19, among others.

In short, both single nucleotides variants (SNPs and Indels) and copy number variants may be focus of study not only for their everyday growing causative variants discovery rate, but also because the variants already found only explain less than 3% of the heritability of BMI.

# Objectives

In order to contribute to the variant discovery, this master thesis has the following objectives:

* To develop two workflows capable of analysing exome sequencing data for obtaining single nucleotide variants and copy number variants, respectively.
* To develop appropriate statistical analyses in order to discover new single nucleotide variants (such as SNPs or INDELs) and copy number variants highly correlated with obesity in Spanish population.
* To find the biological significance of these variants in developing obesity analysing the genes and metabolic pathways affected.

The data used for performing the whole analyses come from Spanish individuals with extreme obesity. By this way, it is tried to increase the power for finding new significant variants.

# Material and Methods

## Data description

### Samples

The data come from 18 unrelated adult individuals with severe early-onset obesity of Iberian origin (their characteristics are exposed in **Table 1**). Of these, 10 individuals were recruited from the Endocrinology Service of the General Hospital of Valencia (Spain), 4 adult individuals from the Endocrinology Service of the Infanta Cristina Hospital in Badajoz (Spain), and 4 adult individuals from the Coimbra Hospital (Portugal) between May and December 2017. The selection criteria were the following:

* More than 3 Kg at be born
* To get the obesity being less than 6 years old
* BMI higher than 45 kg/m2 in adults and greater than the 99th percentile in children
* The existence, at least, other three cases of morbid obesity among first- or second-degree relatives
* To be free of hypertension, diabetes or any cardiovascular disease

**Table 1:** Anthropometrics characteristics of the sample

|  |  |  |  |
| --- | --- | --- | --- |
|  | Man | Woman | total |
| Sex | 9 | 9 | 18 |
| Years-old | 45.2 ±4.1 | 35.2 ±10.7 | 40.2 ±9.3 |
| Weigh(kg) | 175.7 ±31.6 | 130.2 ±19.9 | 153 ±34.6 |
| Shape (m) | 1.76 ±0.07 | 1.63 ±0.03 | 1.69 ±0.08 |
| BMI (kg/m2) | 56.96 ±11.3 | 49.4 ±8.5 | 53.2 ±10.2 |

***Important considerations:***

*The study protocol was approved by the Directorate General of Innovation and Curriculum Development and the Ethics Committee of the Ministry of Education (both of the Government of Portugal), and by the Ethics Committee of the General Hospital of Valencia and the Hospital Infanta Cristina de Badajoz (Spain). The study was conducted in accordance with the institutional and ethical requirements of the University of Coimbra, as well as the Declaration of Helsinki and its subsequent revisions. The written informed consent of all the patients was obtained before the participation in the study.*

### Methodology of DNA extraction and sequencing

The genomic DNA was extracted from peripheral blood mononuclear cells using the MagNA pure system (Roche Life Science, Barcelona, España), according with the manufacturer’s instructions. The quantification and purity of DNA was determined by the fluorometer Qubit 2.0 (Thermo Fisher Scientific Inc., Waltham, MA, EE. UU.) and the spectrophotometer NanoDrop (Thermo Fisher Scientific Inc., Waltham, MA, EE. UU.) respectively. The DNA integrity was observed using agarose gel electrophoresis.

From each sample, genomic DNA was broken into 150-200 base pair (bp) fragments using the focused-ultrasonicator Covaris S2 (Covaris, Brighton, Reino Unido). The exome capture was prepared using the instructions provided by the Agilent SureSelect Human All Exon V6 (Agilent Technologies, Santa Clara, CA). The sequencing was performed by the platform Illumina HiSeq 2500 (Illumina, Inc., San Diego, CA, EE. UU.) using v3.0 SBS with densities of flow grouping per cell of 700-800 K/mm2 on average.

### Genomic alignment

Aa

## SNV detection

### Variant Calling

### Significant variants detection

## CNV Analysis

# Results

# Discussion

# Conclusion

# Bibliography

1. Klaauw AA Van Der, Farooqi IS. Review The Hunger Genes : Pathways to Obesity. *Cell*. 2015;161(1):119-132. doi:10.1016/j.cell.2015.03.008

2. Moon S, Hwang MY, Jang HB, et al. Whole-exome sequencing study reveals common copy number variants in protocadherin genes associated with childhood obesity in Koreans. *Int J Obes*. 2017;41(4):660-663. doi:10.1038/ijo.2017.12

3. Adhanom Ghebreyesus T. WHO | What are the health consequences of being overweight? WHO. https://www.who.int/features/qa/49/en/. Published 2013. Accessed May 21, 2019.

4. Summers JB, Kaminski JM. Nutrition, physical activity, and obesity. *Lancet*. 2002;360(9341):1249. doi:10.1016/S0140-6736(02)11249-9

5. Tappy L, Binnert C, Schneiter P. Energy expenditure, physical activity and body-weight control. *Proc Nutr Soc*. 2004;62(03):663-666. doi:10.1079/pns2003280

6. Schwartz MW, Woods SC, Seeley RJ, Barsh GS, Baskin DG, Leibel RL. Is the Energy Homeostasis System Inherently Biased Toward Weight Gain ? 2003;52(February).

7. Wu Y, Wang W, Jiang W, Yao J, Zhang D. An investigation of obesity susceptibility genes in Northern Han Chinese by targeted resequencing. *Med (United States)*. 2017;96(7):1-6. doi:10.1097/MD.0000000000006117

8. Yanovski JA. Obesity: Trends in underweight and obesity — scale of the problem. *Nat Rev Endocrinol*. 2017;14(1):5-6. doi:10.1038/nrendo.2017.157

9. Serra-Juhé C, Martos-Moreno G, Bou de Pieri F, et al. Novel genes involved in severe early-onset obesity revealed by rare copy number and sequence variants. *PLoS Genet*. 2017;13(5):1-19. doi:10.1371/journal.pgen.1006657

10. Stunkard AJ, Harris JR, Pedersen NL, McClearn GE. The Body-Mass Index of Twins Who Have Been Reared Apart. *N Engl J Med*. 1990;322(21):1483-1487. doi:10.1056/NEJM199005243222102

11. Maes HHM, Neale MC, Eaves LJ. Genetic and Environmental Factors in Relative Body Weight and Human Adiposity. *Behav Genet*. 1997;27(4).

12. Speakman JR. Commentary A Nonadaptive Scenario Explaining the Genetic Predisposition to Obesity : The ‘“ Predation Release ”’ Hypothesis. 2007;(July):5-12. doi:10.1016/j.cmet.2007.06.004

13. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206. doi:10.1038/nature14177

14. Pettersson M, Viljakainen 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):3029-3039. doi:10.1210/jc.2017-00565

15. Thorleifsson G, Walters GB, Gudbjartsson DF, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. 2009;41(1):18-24. doi:10.1038/ng.274

16. Walters RG, Jacquemont S, Valsesia A, et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature*. 2010;463(7281):671-675. doi:10.1038/nature08727

17. Bochukova EG, Huang N, Keogh J, et al. Large , rare chromosomal deletions associated with severe early-onset obesity. *Nature*. 2010;463(7281):666-670. doi:10.1038/nature08689

18. Yang T-L, Guo Y, Shen H, et al. Copy Number Variation on Chromosome 10q26.3 for Obesity Identified by a Genome-Wide Study. *J Clin Endocrinol Metab*. 2013;98(1):E191-E195. doi:10.1210/jc.2012-2751

19. Jarick I, Vogel CIG, Scherag S, et al. Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis. *Hum Mol Genet*. 2011;20(4):840-852. doi:10.1093/hmg/ddq518

# Appendix