

A Genetic Analytics Approach for Risk Variant Identification to Support Intervention Strategies for People Susceptible to Polygenic Obesity and Overweight

C. Aday Curbelo Montañez¹, P. Fergus¹, A. Hussain¹, D. Al-Jumeily¹, B. Abdulaimma¹, Haya Al-Askar²

¹Liverpool John Moores University, Applied Computing Research Group, Faculty of Engineering and Technology, Byrom Street, Liverpool, L3 3AF, UK.
C.A.Curbelomontanez@2015.ljmu.ac.uk
{P.Fergus, A.Hussain, D.Aljumeily}@ljmu.ac.uk
B.T.Abdulaimma@2015.ljmu.ac.uk

²College of Computer Engineering and Science, Sattam Bin Abdulaziz University, KSA
sun_2258@hotmail.com

Abstract. Obesity is a growing epidemic that has increased steadily over the past several decades. It affects significant parts of the global population and this has resulted in obesity being high on the political agenda in many countries. It represents one of the most difficult clinical and public health challenges worldwide. While eating healthy and exercising regularly are obvious ways to combat obesity, there is a need to understand the underlying genetic constructs and pathways that lead to the manifestation of obesity and their susceptibility metrics in specific individuals. In particular, the interpretation of genetic profiles will allow for the identification of Deoxyribonucleic Acid variations, known as Single Nucleotide Polymorphism, associated with traits directly linked to obesity and validated with Genome-Wide Association Studies. Using a robust data science methodology, this paper uses a subset of the TwinsUK dataset that contains genetic data from extremely obese individuals with a $BMI \geq 40$, to identify significant obesity traits for potential use in genetic screening for disease risk prediction. The paper posits an approach for methodical risk variant identification to support intervention strategies that will help mitigate long-term adverse health outcomes in people susceptible to obesity and overweight.

Keywords: Genetics, SNPs, Obesity, Data Science, R, Bioconductor, gwascat

1. Introduction

According to the World Health Organization (WHO) ¹, the occurrence of obesity and overweight worldwide has doubled since 1980. More recently, the figures suggest

¹<http://www.who.int/>

that in 2014 more than 1.9 billion adults were overweight and 600 million were obese. The condition was initially recognized as a disease in 1948 by the WHO [1] and since then its prevalence has continued to increase making it a global phenomenon and one of the main contributors to poor health. Today it is considered one of the most difficult clinical and public health challenges worldwide [2]–[4]. It is the leading cause of Type 2 Diabetes, cardiovascular disease, premature death, hypertension, osteoarthritis, stroke and certain cancers [2],[3],[5]. Consequently, it is high on the political agenda of many countries. While North America is considered the most obese continent, the UK is ranked as one of the most obese nations in Europe².

The predisposition to obesity in humans is referred to as polygenic obesity and is considered a complex and multifactorial disease that manifests itself through interactions between genetic, behavioural and environmental factors. While obesity tends to exist within families, the way it is inherited does not correspond to known patterns. We know that it is dependent on environmental factors and numerous studies have shown that an individual's predisposition to obesity is more similar among genetically related individuals than those that are not. The phenotypes associated with obesity exhibit significant additive heritability (the proportion of the variability of a trait that is attributable to genetic factors) [6]. In the particular case of Body Mass Index (BMI), family and twin studies have shown that between 40% and 70% of the inter-individual variation in susceptibility to obesity can be attributed to genetic differences in the population [7]–[9]. The remaining percentage is associated with other factors, such as lifestyle. It is clear that the prediction of obesity based solely on genetic information will never be completely accurate. Thus, the ideal genetic predictor, should complement, but never replace traditional predictors.

Building on existing works and tools for genetic analysis, this paper explores polygenic obesity and associated risk factors. Using a robust data science methodology, data extracted from Genome-Wide Association Studies (GWAS) is utilised to identify significant Single Nucleotide Polymorphism's (SNPs) associated with obesity traits. This study serves as a genetic analytics framework for the interpretation of genetic profiles for the discovery of significant biomarkers and their associated genes for use in early detection and intervention strategies.

2. Background

The central dogma of molecular biology describes how cells replicate their Deoxyribonucleic Acid (DNA) in daughter cells via cell division. It also focuses on how proteins are synthesized from DNA, according to the flows of information from DNA to ribonucleic Acid (RNA) and RNA to protein. Genes encode amino acid chains that compose the proteins, which have specific functions in the organism [10], [11]. Understanding the function of each gene, and the proteins they synthesize can help to identify the cause(s) of different diseases. This has the real potential to find better ways to treat and prevent diseases through early detection and prevention strategies, personalized drugs and tailored therapies.

² <http://www.who.int/>

In the context of polygenic obesity, this disease is complex and does not exhibit a typical Mendelian pattern of transmission. There is evidence derived from GWAS that suggest single nucleotide polymorphisms (SNPs) in certain genes are associated with obesity risk factors and BMI. Examples of these genes include those associated with fat mass and obesity (FTO) and the Melanocortin 4 receptor gene (MC4R) [12]–[15]. Additional studies have reported that certain genes, including those mentioned, have a strong link with energy consumption in the nervous system when the hypothalamus part of the brain is stimulated [16].

While in [17], a meta-analysis of 340,000 people was conducted and 97 Genome-wide Significance (GWS) loci associated with BMI were reported, of which 56 were novel. At least 52 of these have been previously identified and are reported to be associated with a predisposition to obesity. However, their predictive capability is poor, because it only explains a fraction of the total variance. Therefore, it cannot compete with the predictability of traditional risk factors such as, obesity in parents and children. Nonetheless, some of these loci are currently being used in direct-to-consumer (DTC) personal genomic profiles to estimate the risk of obesity in the lives of individuals [15].

The predictive capability of loci is however likely to improve as the cost of genetic sequencing decreases. In 2003, the cost was \$2.7 billion, while in 2013 it was \$5000 [18]. This has resulted in a plethora of sequencing methods designed to determine the order of nucleotide bases that make up a molecule of DNA or RNA [19]. The first generation of sequencing, known as Sanger sequencing, made the Human Genome Project possible. More recently, these have been replaced with next-generation sequencing (NGS) techniques that provide cheaper alternatives [20], [21]. In particular, they allow for faster discovery of genes and regulatory elements associated with diseases, thus significantly increasing the clinical diagnosis performed with the nucleic acid sequence.

This is further supported by the fact that a considerable number of base pairs (nucleotides) in the human genome are identical for all humans [22]. Hence, in genetic association studies, bases where there is variation between humans are only considered. These differences are commonly referred to as SNPs. Initially, studies were commissioned to investigate the basic genetics of common obesity based on candidate gene approaches and linkage analyses. However, other types of more effective studies utilizing hypothesis-free methodologies have been used in obesity studies to identify many obesity related loci. These studies are termed GWAS and permit the analysis of a larger number of genetic variants for association with traits of interest.

However, despite these advances, sequencing the whole genome is still expensive; therefore, researchers tend to focus on coding regions of the human genome known as exons. It is estimated that 85% of disease-causing mutations are located in coding and functional regions of the genome, so whole-exome sequencing (WES) methods are a potential alternative to unravel the cause of rare genetic disorders [23].

This paper adopts this approach where 68 exonerated sequenced samples from extremely obese people are used and compared with a publicly available and manually curated collection of published GWASs.

3. Methodology

In this paper, we compare the genetic profiles of extremely obese people, obtained from the UK10K_OBESITY_TWINSUK REL-2013-04-20 dataset hosted by Sanger Institute³, with GWAS results indexed in the National Human Genome Research Institute (NHGRI) Catalog⁴. The dataset represents 68 cases of extremely obese twins. These samples are a subset of the TwinsUK cohort samples that have a BMI \geq 40 and were exome sequenced by the UK10K Obesity group [24], [25].

The UK10K_OBESITY_TWINSUK REL-2013-04-20 dataset contains SNPs and associated metadata variables that include *chromosome, position, identifier or reference (REF) and alternate (ALT) bases*. The samples in the VCF format were sequenced using GRCh37 and accessed in NHGRI Catalog data (hg19 genome build) based on this assembly.

The primary focus in this study is to identify the frequency of SNPs or their variants of interest in the 68 examined profiles (only Europeans considered) and which of these are indexed in the NHGRI Catalog. This mapping allows us to link SNP-traits in our samples with risk alleles indexed in the NHGRI Catalog. This process results in a set of traits that best represent phenotypes related to obesity.

3.1 Data Pre-Processing

The dataset contains 5,827,333 observations (records) and 4 variables that describe the chromosome number, position, SNP ID and REF, and ALT bases which are combined to represent the genotype, as shown in in Table 1. These are the variables used in this study to identify SNPs and risk alleles previously reported in GWAS and indexed in the NHGRI Catalog. Column names were edited for simplicity and convenience.

#Chrom	Position	RSID	Genotype
1	14907	Rs79585140	A/G
1	63671	Rs116440577	G/A
1	63697	Rs201582574	T/C
1	69511	Rs75062661	A/G

Table 1: Columns extracted from the UKTwins dataset and used in this study

In the initial data pre-processing stages observations with missing values (NA values) were removed including those observations with no SNP IDs. Duplicate SNPs in all the samples were summed and utilised as a feature to represent SNP frequency. This initial data pre-processing step reduced the number of observations from 5,827,333 to 361,899.

³ <http://www.sanger.ac.uk/>

⁴ <https://www.ebi.ac.uk/>

3.2 NHGRI Catalog comparison

Using metadata from the NHGRI Catalog (“ebicat37”) the SNP ID and the Genotype contained in the samples were linked with SNPs and risk alleles reported in the Catalog. This was achieved by using the *RSID* and *Genotype* values in the samples to index *SNPS* and *STRONGEST.SNP.RISK.ALLELE* in the NHGRI Catalog. Those SNPs with risk alleles marked as “?” in the Catalog were discarded from our samples.

This resulted in 1,304 matching SNPs from our samples. However, the potential risk alleles identified in the NHGRI Catalog were obtained in studies conducted using people from different populations and ethnic backgrounds. Therefore, since our samples belonged to UK twins, only those variants identified in studies conducted in European populations were used. This was achieved using a filtering process to select SNPs where the column *INITIAL.SAMPLE.DESCRPTION* contained the word “European”. This reduced the number of SNPs further to 1,041 – 11 variables were included as shown in Table 2.

Variables
#chrom
position
rsid
genotype
SNPCount
STRONGEST.SNP.RISK.ALLELE
DISEASE.TRAIT
P.VALUE
REPORTED.GENE.S.
INITIAL.SAMPLE.DESCRPTION
JOURNAL

Table 2: Variables selected after European filtering process.

Following the identification of risk variants associated with European populations, traits linked with obesity were selected as shown in Table 3.

Variables	Abbreviations
Body mass index	BMI
Visceral adipose tissue adjusted for BMI	VATaBMI
Obesity	O
Obesity(early onset extreme)	OEOE
Weight	W
Weight loss	WL
Birth weight	BW
Height	H
Fat body mass	FBM
Eating disorders	ED
Waist-hip ratio	WHR
Waist Circumference	WC

Table 3: Obesity traits selected.

Figure 1 shows the 1041 SNPs contained in our dataset that are indexed in the NHGRI Catalog. The x axis represents the chromosomes, while the y value shows each of the SNPs – the strongest associations have the smallest P-values, therefore negative logarithms will be the greatest. So while all of the SNPs in the study are considered to be statistically significant [26], SNPs with higher y values are deemed to be the most significant. In other words, all the SNPs represented in Figure 1 have been significantly associated with obesity-related traits in European populations, according to GWAS publications included in the NHGRI Catalog. The SNPs highlighted in green correspond to the obesity-related traits previously listed in Table 3. The $-\log_{10}(p)$ represents the statistical significance of the association reported in the GWAS.

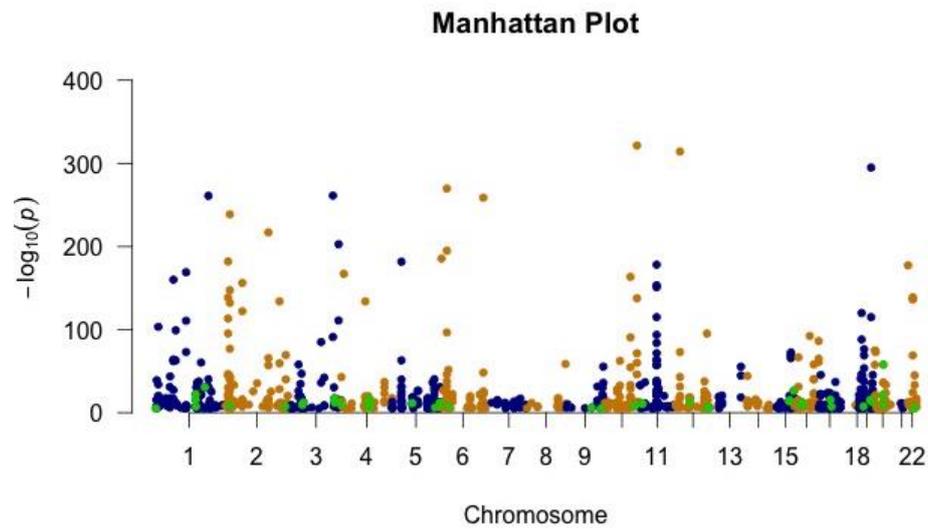


Figure 1: Manhattan plot with obesity traits highlighted in green.

Obesity-related SNPs were identified in chromosomes 1, 2, 3, 4, 5, 6, 9, 10, 11, 12, 15, 16, 17, 19, 20 and 22 and highlighted in green. However, the chromosomes of specific interest are 1, 3 and 19 which each have four obesity SNPs. It should be noted that not all the SNPs included in the three chromosomes are significant as will be discussed in more detail in the results section. Additionally, the absence of obesity-related traits in chromosomes 7, 8, 13, 14, 18 and 21 can also be noticed. Chromosomes X and Y were not included in the representation as there were no identifiable SNPs indexed from the NHGRI Catalog for these chromosomes in our data.

Using our 1041 SNPs, we filtered the dataset to only include those SNPs related to obesity – this reduced the dataset to 34 obesity related SNPs. Figure 2 describes the 34 obesity related SNPs that appear in the initial 68 samples of extremely obese subjects.

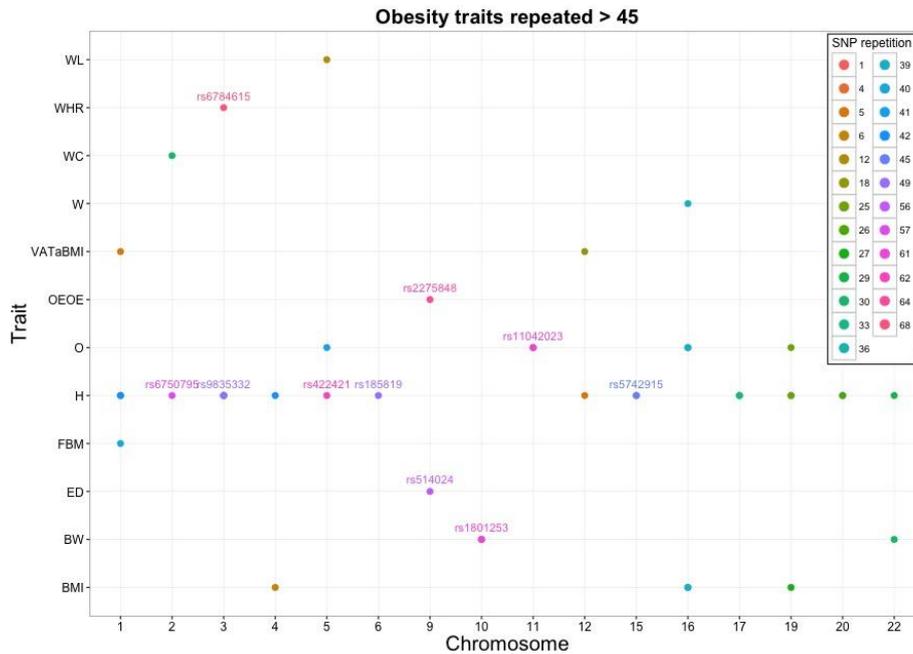


Figure 2: Obesity related traits and their frequency in different chromosomes

Figure 2 represents the SNPs and their obesity associated traits for each chromosome – colours show the frequency of the variants. The full trait names described on the y axis are listed in Table 3. The SNPs that are repeated more than 45 times are labelled in Figure 2. In some cases, SNPs overlap as they are related to the same traits in the same chromosome. The trait height was identified nineteen times in different chromosomes. Obesity was also found in 5 chromosomes. Waist-hip ratio and obesity (early onset extreme) traits are represented in pink and labelled as they were repeated 68 and 64 times respectively. Surprisingly, one of the SNPs associated with body mass index, rs13107325 (chromosome 4), was only repeated in six samples and for that reason it was represented as an orange-green colour.

4. Results

In this study we investigated the occurrence of SNPs in 68 samples of cases with extreme obesity. The results in Table 4 provide a brief description of the obesity related SNPs that were identified in the samples. The table identifies the chromosome number, position, SNP id, SNP frequency, obesity trait, p-value and reported genes. From all the traits identified, the most prominent, given its association with 19 identified SNPs, is *height*. However, this trait would not represent a risk as in most cases it corresponded to the REF base on the genotype, not the ALT. Nonetheless, we do not have any clinical data to confirm this.

#Chrom	Position	SNP ID	Freq	Trait	P.Value	Reported Genes
1	149906413	rs11205303	42	Height	4e-23	MTMR11

1	150727539	rs2230061	40	Fat body mass	4e-08	CTSS
1	184023529	rs1046934	39	Height	2e-31	TSEN15
1	3649562	rs9662633	5	Visceral adipose tissue adjusted for BMI	6e-06	TP73
2	232378231	rs6750795	57	Height	2e-08	C2orf52
2	27801759	rs1919128	30	Waist Circumference - Triglycerides (WC-TG)	2e-09	C2orf16
3	52506426	rs6784615	68	Waist-hip ratio	4e-10	NISCH, STAB1
3	56667682	rs9835332	49	Height	5e-13	C3orf63
3	172165727	rs572169	41	Height	3e-18	GHSR
3	172165727	rs572169	41	Height	1e-12	GHSR
4	1701317	rs2247341	42	Height	2e-11	SLBP/FGFR3
4	103188709	rs13107325	6	Body mass index	2e-13	SLC39A8
5	176517326	rs422421	62	Height	1e-12	FGFR4/NSD1
5	75003678	rs2307111	41	Obesity	3e-12	C5orf37
5	159820931	rs10515808	12	Weight loss (gastric bypass surgery)	4e-07	C1QTNF2
6	32050067	rs185819	49	Height	3e-08	HLA class III
9	95887320	rs2275848	64	Obesity (early onset extreme)	1e-06	NINJ1
9	130504070	rs514024	56	Eating disorders (purging via substances)	5e-06	PKN3, SET, WDR34, ZDHHC1, ZER1
10	115805056	rs1801253	62	Birth weight	4e-09	ADRB1
11	8662516	rs11042023	62	Obesity	1e-11	RPL27A
12	124399550	rs1316952	18	Visceral adipose tissue adjusted for BMI	5e-06	DNAH10
12	56740682	rs2066807	5	Height	1e-13	STAT2
15	74336633	rs5742915	45	Height	1e-15	PML
15	89388905	rs16942341	4	Height	4e-27	ACAN
16	28883241	rs7498665	39	Obesity	3e-13	SH2B1
16	28883241	rs7498665	39	Body mass index	5e-11	SH2B1
16	28883241	rs7498665	39	Body mass index	3e-10	ATP2A1, SH2B1
16	28883241	rs7498665	39	Obesity	5e-12	SH2B1
16	28883241	rs7498665	39	Weight	1e-09	ATP2A1, SH2B1
17	43216281	rs4986172	33	Height	2e-16	ACBD4
17	47390014	rs2072153	29	Height	4e-08	ZNF652
19	41937095	rs17318596	36	Height	5e-16	ATP5SL
19	46202172	rs2287019	27	Body mass index	2e-16	GIPR, QPCTL
19	46181392	rs1800437	25	Obesity	3e-14	GIPR
19	17283303	rs2279008	25	Height	3e-08	MYO9B
20	34025756	rs143384	39	Height	1e-58	GDF5
20	32333181	rs7274811	26	Height	6e-22	ZNF341
22	42336172	rs5758511	30	Birth weight	3e-06	CENPM

22	40820151	rs5757949	29	Height	4e-06	MKL1, SGSM3
----	----------	-----------	----	--------	-------	----------------

Table 4: Obesity related SNPs identified in the 68 extremely obese samples.

Our findings show that the SNP associated with the waist-hip ratio trait, *rs6784615*, was present across all samples, followed by the SNP *rs2275848*, associated with obesity (early onset extreme) which is repeated 64 times. Other outstanding SNPs such as *rs1801253*, *rs11042023* and *rs514024* were linked to birth weight, obesity and eating disorders and were present in 61, 61 and 56 of the samples respectively.

The SNP *rs7498665* in chromosome 16 has been associated to several obesity-related traits according to the information included in the NHGRI Catalog. This particular SNP has been identified in 39 of our samples and can be related to body mass index, obesity or weight traits. It has also been associated with genes, such as SH2B1 and ATP2A1.

The SNP *rs572169* is listed twice as it can be observed in Table 4. In the NHGRI Catalog the same SNP was reported twice with different p-values from different studies. After performing the comparison between the 68 samples of extremely obese people and the NHGRI Catalog, both results were included in the table. The SNP *rs572169* is not the only one identified with different p-values. *rs7498665* was also included in the catalog and reported in different studies. This time, *rs7498665* was reported within different genes: ATP2A1 and SH2B1, as well as different p-values.

Although this study has focused on risk variants associated with obesity traits, other SNPs with other traits within the 10 most repeated have also been identified. In addition to *rs6784615*, the SNP *rs37370* was also present in all samples which is linked to *Asymmetrical dimethylarginine levels*. Other SNPs were associated with *psoriasis*, *parkinson's disease*, *osteosarcoma*, *Insulin-like growth factors*, *coronary heart disease*, *myocardial infarction (early onset)*, *inflammatory bowel disease*, *Ulcerative colitis*, *blood metabolite levels*, *Type 2 diabetes* as well as some type of cancer among other diseases. In future works these SNPs and their association with obesity, including causation factors will be studied further.

5. Discussion

Understanding obesity as a complex disease is an arduous task. In an attempt to better understand how obesity is developed, we sought to identify the main genetic variants present in samples of people with extreme obesity. This was achieved using information contained in the NHGRI Catalog as a reference for the identification of risk variants in our samples. The methodology proposed in this paper investigated a solid foundation for identifying SNPs and their associated risks and in this sense provides a foundation for much more complex genetic analytics.

However, the study has a number of limitations, the major one being the size of the dataset – 68 samples are not representative of a population. However, we did analyse samples belonging to twins in the UK which is part of a more limited and specific population group.

The lack of contextual information about the samples also proved to be an issue. Although we knew that the samples belonged to British twins with extreme obesity

(BMI > 40), we did not have any other information regarding their age, gender, lifestyle or clinical history, which greatly limited the scope of this study. For this reason, using manipulation and data visualization techniques, the objective of this study was to identify statistically significant obesity related variants from GWAS. This has provided a robust foundation for future research directions, particularly in the area of detection and prediction and the use of advanced machine learning algorithms, which will be the focus of our future work.

We also compared whole-exome sequencing data (WES data) against whole genome results data (GWAS data) and this means that we could be ignoring other regions of the DNA that could provide more insights to our study which will be considered in future work.

Starting from an initial number of 5827333 observations, we managed to reduce the dataset using standard data science tools and techniques, such as R⁵ and Bioconductor⁶ allowing us to identify 39 SNPs associated with obesity risks listed in Table 4. Among all the SNPs associated with obesity the most prominent one in the 68 samples was rs6784615. This SNP is associated with waist-hip ratio (WHR), a trait commonly used when it comes to quantifying obesity [15].

6. Conclusions

The identification of biomarkers contained in the entire genome, or more specifically, those found in the protein-coding regions or exons, has helped us to identify genetic variations associated with particular traits in extremely obese samples.

This study used a manually curated and publicly available dataset with results from GWAS for the identification of obesity-related traits in 68 extremely obese samples. Metadata from the NHGRI Catalog was accessed using the R/Bioconductor package *gwascat* and the hg19 genome build *ebicat37* data. The dataset and the samples were analysed using rigorous data science techniques, which led to a significant reduction in the initial number of observations and the identification of SNPs that could help to explain the development of extreme obesity. Missing values and no relational data items were removed from the study. The variables *rsid* and *genotype* as well as *SNPS* and *STRONGEST.SNP.RISK.ALLELE* were the key for the identification of SNP-obesity related traits in our samples.

While the results show specific variants related to obesity traits present in all the samples, they are unlikely by themselves to be causative variants. The potential for variants identified in GWAS to predict the risk of complex diseases is problematic. At the moment, currently known variants do not fully explain the risk of disease occurrence to be of clinical use in predictive systems. However, the presence of specific obesity-related SNPs in extremely obese individuals could help us to gain a better understanding about what SNPs we need to look for. Of course, we need to consider the inclusion of contextual and clinical data and study the interaction between

⁵ <https://www.r-project.org>

⁶ <https://www.bioconductor.org/>

SNPs to gain a better understanding of obesity aetiology. The results do however raise further questions. First, it would be useful to see if the identified SNPs interact conjointly to cause extreme obesity. Second, it would also be useful to explore machine learning methods, such as support vector machines and other advanced artificial neural network architectures to study the non-linearity interaction between SNPs. While this study focused on UK twins, it would be interesting to look at other population groups as well.

References

- [1] W. P. T. James, "WHO recognition of the global obesity epidemic.," *Int. J. Obes. (Lond).*, vol. 32 Suppl 7, pp. S120–S126, 2008.
- [2] A. J. Walley, A. I. F. Blakemore, and P. Froguel, "Genetics of obesity and the prediction of risk for health," *Hum. Mol. Genet.*, vol. 15, no. SUPPL. 2, pp. 124–130, 2006.
- [3] W. Yang, T. Kelly, and J. He, "Genetic epidemiology of obesity," *Epidemiol. Rev.*, vol. 29, no. 1, pp. 49–61, 2007.
- [4] A. Ryley, "HEALTH SURVEY FOR ENGLAND 2012: Chapter 11, Children's BMI, overweight and obesity," England, 2013.
- [5] S. Vallgarda, M. E. J. Nielsen, M. Hartlev, and P. Sandoe, "Backward- and forward-looking responsibility for obesity: policies from WHO, the EU and England," *Eur. J. Public Health*, pp. 1–4, Apr. 2015.
- [6] N. F. Butte, G. Cai, S. a Cole, and A. G. Comuzzie, "Viva la Familia Study: genetic and environmental contributions to childhood obesity and its comorbidities in the Hispanic population.," *Am. J. Clin. Nutr.*, vol. 84, no. 3, pp. 646–654; quiz 673–674, 2006.
- [7] J. Wardle, S. Carnell, C. M. a Haworth, and R. Plomin, "Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment," *Am. J. Clin. Nutr.*, vol. 87, no. 2, pp. 398–404, 2008.
- [8] N. Zaitlen, P. Kraft, N. Patterson, B. Pasaniuc, G. Bhatia, S. Pollack, and A. L. Price, "Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits," *PLoS Genet.*, vol. 9, no. 5, 2013.
- [9] J. S. El-Sayed Moustafa and P. Froguel, "From obesity genetics to the future of personalized obesity therapy.," *Nat. Rev. Endocrinol.*, vol. 9, no. 7, pp. 402–13, 2013.
- [10] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, Sixth Edit. New York: Garland Science, Taylor & Francis Group, LLC, an informa business, 2014.
- [11] J. D. Watson, A. Gann, T. A. Baker, M. Levine, S. P. Bell, and R. Losick, *Molecular Biology of the Gene*, Seventh Ed., vol. 364, no. 6440. John Inglis, 2014.
- [12] K. Wang, W. Li, C. K. Zhang, Z. Wang, J. T. Glessner, S. F. A. Grant, H. Zhao, H. Hakonarson, and R. A. Price, "A Genome-Wide Association Study on Obesity and Obesity-Related Traits," *PLoS One*, vol. 6, no. 4, p. e18939, Apr. 2011.

- [13] B. Xi, C. Wang, L. Wu, M. Zhang, Y. Shen, X. Zhao, X. Wang, and J. Mi, "Influence of physical inactivity on associations between single nucleotide polymorphisms and genetic predisposition to childhood obesity," *Am. J. Epidemiol.*, vol. 173, no. 11, pp. 1256–1262, 2011.
- [14] D. Corella, et al., "Statistical and Biological Gene-Lifestyle Interactions of MC4R and FTO with Diet and Physical Activity on Obesity: New Effects on Alcohol Consumption," *PLoS One*, vol. 7, no. 12, p. e52344, Dec. 2012.
- [15] R. J. F. Loos, "Genetic determinants of common obesity and their value in prediction," *Best Pract. Res. Clin. Endocrinol. Metab.*, vol. 26, no. 2, pp. 211–226, 2012.
- [16] C. J. Willer, E. K. Speliotes, R. J. F. Loos, S. Li, C. M. Lindgren, I. M. Heid, S. I. Berndt, a. L. Elliott, a. U. Jackson, C. Lamina, and Others, "Six new loci associated with body mass index highlight a neuronal influence on body weight regulation," *Nat. Genet.*, vol. 41, no. 1, p. 25, 2009.
- [17] A. E. Locke, et al., "Genetic studies of body mass index yield new insights for obesity biology," *Nature*, vol. 518, no. 7538, pp. 197–206, 2015.
- [18] O. Shaer, O. Nov, J. Okerlund, M. Balestra, E. Stowell, L. Ascher, J. Bi, C. Schlenker, and M. Ball, "Informing the Design of Direct-to-Consumer Interactive Personal Genomics Reports," *J. Med. Internet Res.*, vol. 17, no. 6, p. e146, Jun. 2015.
- [19] A. Grada and K. Weinbrecht, "Next-Generation Sequencing: Methodology and Application," *J. Invest. Dermatol.*, vol. 133, no. 8, pp. e11–4, 2013.
- [20] B. Rabbani, M. Tekin, and N. Mahdieh, "The promise of whole-exome sequencing in medical genetics.," *J. Hum. Genet.*, vol. 59, no. 1, pp. 5–15, 2014.
- [21] J. Shendure and H. Ji, "Next-generation DNA sequencing," *Nat. Biotechnol.*, vol. 26, no. 10, pp. 1135–1145, Oct. 2008.
- [22] T. Fall and E. Ingelsson, "Genome-wide association studies of obesity and metabolic syndrome," *Mol. Cell. Endocrinol.*, vol. 382, no. 1, pp. 740–757, 2014.
- [23] J. Majewski, J. Schwartzentruber, E. Lalonde, A. Montpetit, and N. Jabado, "What can exome sequencing do for you?," *J. Med. Genet.*, vol. 48, no. 9, pp. 580–589, 2011.
- [24] A. Moayyeri, C. J. Hammond, A. M. Valdes, and T. D. Spector, "Cohort Profile: TwinsUK and Healthy Ageing Twin Study," *Int. J. Epidemiol.*, vol. 42, no. 1, pp. 76–85, Feb. 2013.
- [25] A. Moayyeri, C. J. Hammond, D. J. Hart, and T. D. Spector, "The UK Adult Twin Registry (TwinsUK Resource)," *Twin Res. Hum. Genet.*, vol. 16, no. 01, pp. 144–149, Feb. 2013.
- [26] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff, and H. Parkinson, "The NHGRI GWAS Catalog, a curated resource of SNP-trait associations," *Nucleic Acids Res.*, vol. 42, no. D1, pp. D1001–D1006, Jan. 2014.