The Genetic Mechanisms Underlying Human Obesity

Stephen James Clark

Department of Genomics of Common Disease

Imperial College London

This thesis is submitted for the degree of Doctor of Philisophy

December 2012

I herby declare that this thesis describes work carried out soley by myself. Information derived from the work of others has been acknowledged in the text and a list of references is given in the bibliography.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Abstract

Obesity is becoming one the leading causes of mortality in the western world. Although environment is a factor in its development, it is highly heritable and despite a number of genes that have been found to be associated to the disease its genetics are still poorly understood. Discovery of genetic pathways that influence obesity risk can provide a better understanding of the pathophysiology of the disease and identify possible pharmacological targets for its treatment.

This project was designed to investigate possible genetic associations between five candidate genes and severe obesity in both adult and child French Caucasians (n=2,822). Tag SNPs were chosen along with a selection of common SNPs not in the HapMap database and genotyped using Sequenom iPLEX assays. Putative associations were discovered to obesity in three genes, *SIRT1* (corrected p-values: 0.034, 0.019), *APLN* (corrected p-value: 0.017) and *IL11* (corrected p-value: 0.016), although associations do not withstand genome-wide correction. *SIRT1* and *IL11* SNPs were subsequently genotyped within a family cohort for which transcription data in adipose tissue was available. In this cohort, *SIRT1* genotypes were nominally associated with BMI (corrected p-values: 0.014, 0.019, 0.014) and a significant difference in expression levels of the gene was observed between lean and obese individuals ($p=1.6x10^{-35}$) providing suggestive evidence of a role of this gene in the development of obesity. Expression and genotypes of another gene, *IRS1* were also analysed and although no significant associations to obesity were found, an association between SNP variation and gene expression was discovered (corrected p-value: $1.0x10^{-5}$).

Another aim of this project was to investigate the possibility that DNA methylation influences obesity risk. Firstly, a method for the measurement of the quantitative trait of DNA methylation status at individual CpG sites was developed using direct sequencing. Next, methylation in the leptin gene CpG island was measured in a subset of 184 case-control samples and a nominal association was discovered between the quantitative measurement of methylation at a single CpG site and obesity ($p=0.013$).

In summary, putative associations to obesity have been discovered with genetic variants as well as transcription levels and CpG methylation. Replication in other populations is required in order to confirm these associations.

Acknowledgements

First and foremost, I would like to offer thanks to my supervisor, Dr Andrew Walley for giving me the opportunity to work on this project. This thesis would not have been possible without his continued support and expertise. I am also indebted to Prof. Philippe Froguel who, as well as my second supervisor and the head of the department, was responsible for establishing the collaborations that provided the DNA samples from the French case-control cohorts.

I would also like to thank the members of the Section of Genomic Medicine for their help and guidance throughout. In particular Mario Falchi for his statistical expertise and for his help in analysing transcription data. Jo Andersson for her help with the analysis of the Swedish data set and M.Sc. student researcher, Hang Yee Wong for her contribution to this project. Also to the department secretary, Pat Murphy for helping the department run smoothly and for assisting me in many different ways.

I would also like to recognise the contribution of our collaborators to this project. Firstly the group at the Pasteur Institute in Lille, France for providing DNA from the adult and child obesity case-control cohorts. Secondly, the Department of Molecular and Clinical Medicine at Gothenburg University, Sweden for providing DNA from their obesity sib-pair cohort together with the corresponding transcription data. I also like to thank the participants who took part in these studies as subjects. Without these DNA samples this project would not have been possible.

Lastly, I am forever indebted to my wife Emma and to my parents for their understanding, patience and encouragement when I needed it the most.

Table of Contents

Abstract....4

Acknowledgements....6

List of Figures and Tables....14

Chapter 1

Introduction....19

1.1 Obesity overview....20

1.2 Energy regulation in mammals....21

1.2.1 Rodent models of obesity....21

1.2.2 Central nervous system control of energy regulation....23

1.2.2.1 Orexigenic signals in the CNS....23

1.2.2.2 Anorexigenic signals in the CNS....27

1.2.3 Insulin as an adiposity signal....30

1.2.4 Leptin as an adiposity signal....31

1.2.5 Other adiposity signals....32

1.2.6 Gastrointestinal signals....33

1.2.7 Lipid metabolism....35

1.3 Heritability of obesity....37

1.4 The genetics of obesity....40

1.4.1 Monogenic obesity....40

1.4.2 Syndromic obesity....43

1.4.3 Common polygenic obesity....44

1.4.3.1 Linkage studies....44

1.4.3.2 Candidate gene association studies....46

1.4.3.3 Genome-wide association studies (GWAS)....50

1.4.4 Missing heritability......54
1.4.4.1 Copy number variation (CNV) 55
1.4.4.2 Epigenetics......56
1.5 Aims......61
1.5.1 Candidate gene association studies......62
1.5.1.1 Sirtuin-1 (SIRT1)......63
1.5.1.2 Apelin (APLN)......64
1.5.1.3 Interleukin 11 (IL-11)......65
1.5.1.4 Adiponutrin (PNPLA3)......65
1.5.1.5 Nesfatin-1 (NUCB2)......66
1.5.1.6 Insulin Receptor Substrate 1 (IRS-1)......67
1.5.2 Epigenetic variation in the Leptin gene (LEP)......68
Chapter 2
Materials and Methods......69
2.1 Subjects......70
2.1.1 Obese cases......70
2.1.2 Non-obese controls......70
2.1.3 Families......71
2.2 Protocols......71
2.2.1 SNP identification strategy......71
2.2.2 Genotyping......72
2.2.3 PCR......73
2.2.4 Agarose gel electrophoresis......73
2.2.5 Sequencing of PCR product......74
2.2.6 CpG island identification and primer design......74
2.2.7 Bisulfite conversion of DNA......74

2.2.8 Bisulfite conversion of DNA using Qiagen EpiTect 96-well format....75

2.2.9 Positive control nested PCR....77

2.3 Statistical Analysis....77

2.3.1 Qualitative trait analysis....77

2.3.2 Quantitative trait analysis....78

2.3.3 Family-based association analysis....78

2.3.4 Linkage disequilibrium (LD)....78

2.3.5 Imputation....78

2.3.6 Analysis of transcription data....79

2.3.7 Statistical power calculations....79

2.3.8 Analysis of sequence data....81

2.3.9 Interpretation of bisulfite sequence data....82

2.3.10 Statistical analysis of CpG methylation data....82

Chapter 3

Investigation of genetic variants within the Sirtuin-1 gene for association to common polygenic obesity....83

3.1 Introduction....84

3.1.1 SIRT1 regulates energy metabolism....84

3.1.2 SIRT1 activity is regulated by metabolic factors....85

3.1.3 SIRT1 mouse models....86

3.1.4 SIRT1 as a pharmacological target for treating metabolic disease....86

3.1.5 SIRT1 genetic variation and obesity....87

3.2 Results....89

3.2.1 SNP Selection and Case-Control Genotyping....89

3.2.2 Genotyping within the Swedish sib-pair cohort....95

3.2.3 Analysis of SIRT1 transcription....97

3.2.4 Sequencing of SIRT1 ... 102
3.2.5 Case-control genotyping in promoter ... 104
3.2.6 Imputation ... 107
3.2.7 Statistical power calculations ... 108
3.2.8 Thesis-wide multiple testing correction ... 112
3.3 Discussion ... 113
Chapter 4
Investigation of genetic variants within the apelin gene for association to common polygenic obesity ... 119
4.1 Introduction ... 120
4.1.1 Apelin is upregulated in obesity ... 120
4.1.2 Effects of apelin on energy regulation ... 121
4.1.3 Apelin mouse models ... 121
4.1.4 Apelin genetic variation and obesity ... 122
4.2 Results ... 123
4.2.1 SNP Selection and Case-Control Genotyping ... 123
4.2.2 Sequencing of APLN ... 128
4.2.3 Thesis-wide multiple testing correction ... 130
4.2.4 Power calculations ... 130
4.3 Discussion ... 133
Chapter 5
Investigation of genetic variants within the Interleukin 11 gene for association to common polygenic obesity ... 137
5.1 Introduction ... 138
5.1.1 IL11 inhibits adipogenesis ... 138

5.1.2 IL11 variation and obesity......139
5.2 Results......141
5.2.2 Genotyping within Swedish sib-pair cohort......146
5.2.3 Analysis of IL-11 transcription......148
5.2.4 Thesis-wide multiple testing correction......153
5.2.5 Power calculations......154
5.3 Discussion......156
Chapter 6
Investigation of genetic variants within the adiponutrin gene for association to common polygenic obesity......159
6.1 Introduction......160
6.1.1 Adiponutrin has a role in lipid metabolism......160
6.1.2 Adiponutrin is regulated by metabolic factors......160
6.1.3 Adiponutrin gene variation and non-alcoholic fatty liver disease......161
6.1.4 Adiponutrin gene variation and obesity......162
6.2 Results......164
6.2.1 SNP Selection and Case-Control Genotyping......164
6.2.2 Power calculations......170
6.3 Discussion......171
Chapter 7
Investigation of genetic variants within the Nesfatin gene for association to common polygenic obesity......172
7.1 Introduction......173
7.1.1 Central action of nesfatin-1 to reduce food intake......173
7.1.2 NUCB2 expression is regulated by metabolic factors......174

7.1.3 Peripheral nesfatin-1 174

7.1.4 Nesfatin-1 as a pharmacological target for the treatment of obesity 175

7.1.5 NUCB2 variation and obesity 175

7.2 Results 178

7.2.2 Power calculations 186

7.3 Discussion 189

Chapter 8

Investigation of genetic variation within the IRS-1 gene and transcription levels for association to BMI 191

8.1 Introduction 192

8.1.1 IRS-1 and glucose metabolism 192

8.1.2 IRS-1 and obesity 192

8.2 Results 194

8.3 Discussion 197

Chapter 9

Investigation of DNA methylation for associations with human polygenic obesity 199

9.1 Introduction 200

9.2 Results 203

9.2.1 Developing a Protocol for Measuring CpG Methylation 203

9.2.2 Locating CpG Islands within Obesity-Associated Genes 206

9.2.3 Investigation of DNA methylation in leptin gene 210

9.2.3.1 Low volume bisulfite sequencing protocol 212

9.2.4 Investigation of leptin gene DNA methylation levels for possible association to human polygenic obesity 213

9.2.5 Analysis of LEP transcript levels....217
9.3 Discussion....219
Chapter 10
Discussion....222
10.1 Summary of thesis results....223
10.2 Future work....227
References....236
Appendix....262
A1. Tagging SNPs genotyped and the corresponding captured SNPs....263
A2. Primer Sequences Used in Sequenom Genotyping Assays....265
A3. Sequencing PCR primers....273
A4. Primer Sequences Used in DNA Methylation Study....275
A 6. R scripts....276
A 6.1. Correction of transcript data for covariate effects using linear regression....276
A 6.2 Correction of BMI or transcript data for effects of relatedness using clustering....276
A 7. Statistical power calculations for locating SNPs in LD with associated variants....277

List of Figures and Tables

Figure 1.1 Energy regulation in the hypothalamus	29
Table 1.1. Hormones released from the gastrointestinal tract and their effects on energy regulation	35
Table 1.2. Monogenic forms of obesity in humans	42
Table 1.3. Candidate genes with associations to obesity reported in meta-analyses	49
Table 1.4. Genes significantly associated with obesity phenotypes in genome-wide association studies	53
Figure 2.1. Statistical power using adult case-control subjects	80
Figure 2.2. Statistical power using child case-control subjects	80
Figure 2.3. Statistical power for analysing BMI in the French controls	81
Figure 3.1. SNPs analysed for association to BMI in SIRT1 region in the GIANT GWAS	88
Figure 3.2. Schematic diagram of the SIRT1 gene showing the positions of SNPs successfully genotyped in this study	89
Table 3.1. SNPs genotyped within the SIRT1 gene and minor allele frequencies observed in the case-control subjects	90
Table 3.2. Allelic association analysis of SIRT1 SNPs to obesity in children and adult French case-controls	92
Table 3.3. Association analysis of SIRT1 SNPs to the quantitative trait of BMI in French controls using age and sex as covariates	93
Figure 3.3. Linkage disequilibrium (r^2) plot of SIRT1 SNPs genotyped in the French control samples	94
Table 3.4. Haplotype association analysis to obesity in adults using the four SNPs found to be nominally associated with adult obesity	95
Table 3.5. Association analysis of SIRT1 SNPs with BMI in the Swedish families	96
Table 3.6. Haplotype association analysis of SIRT1 SNPs with BMI in the Swedish families	97
Table 3.7. Association analysis of SIRT1 SNPs with SIRT1 expression corrected for age, sex and BMI in the Swedish families	98
Figure 3.4. Box-plot of SIRT1 transcript levels in obese and non-obese subjects from Swedish sib-pairs	100
Table 3.8. T-test of SIRT1 transcription in obese and non-obese subjects from Swedish families	100
Figure 3.5. SIRT1 transcription levels plotted against BMI in non-obese siblings	101

Figure 3.6. SIRT1 transcription levels plotted against BMI in obese siblings 102

Figure 3.7. SNPs discovered within the SIRT1 gene promoter region in samples that carried the minor allele variants associated with obesity 104

Table 3.9. SIRT1 promoter SNPs successfully genotyped with the minor allele frequencies observed in the case-control subjects 105

Table 3.10. Association analysis of SIRT1 promoter SNPs in French case-controls 106

Figure 3.8. LD (r^2) plot within the French control samples containing all SNPs genotyped and analysed within SIRT1 and its promoter region 107

Figure 3.9. LocusZoom plot of imputed SIRT1 SNPs tested for association in the French adult case-control cohort. 108

Table 3.11. Power calculations carried out using MAF and OR observed for each SNP in the SIRT1 gene investigated 109

Figure 3.10. Sample size plotted with statistical power for a SNP with MAF=0.13 110

Figure 3.11. Sample size required for varying odds ratios in order to achieve a statistical power of 95% for a SNP with MAF=0.13 110

Figure 3.12. Statistical power against effect size in French controls assuming a MAF of 0.13 and SD of 2.1 111

Table 3.12 P-values after correcting for multiple testing using permutations and taking into account all SNPs investigated in this thesis 112

Figure 4.1. Map of APLN gene with positions of SNPs genotyped in this project 123

Table 4.1. APLN SNPs successfully genotyped with minor allele frequencies observed 124

Table 4.2. Allelic association analysis of Apelin SNPs to obesity in children and adult French case-controls 125

Figure 4.2. Linkage disequilibrium (r^2) plot of APLN SNPs genotyped in the French female control samples 126

Table 4.3 Haplotype analysis in the child case-controls using the two SNPs nominally associated with obesity 127

Table 4.4. Analysis of APLN SNPs for association with the quantitative trait of BMI in French controls using linear regression with age and sex as covariates 127

Figure 4.3. Map of APLN gene showing positions of PCR products used to sequence the promoter region and the two translated exons 129

Table 4.5. P-values after correcting for multiple testing using permutations and taking into account all SNPs investigated in this thesis 130

Table 4.6. Power calculations carried out using MAF and OR observed for each SNP investigated in the APLN gene 130

Figure 4.4. Sample size plotted with statistical power for a SNP with MAF=0.06 131

Figure 4.5. Sample size required for varying odds ratios in order to achieve a statistical 132

power of 95% for a SNP with MAF=0.06

Figure 4.4 LD plot (D') of APLN gene and the surrounding region 136

Figure 5.1. SNPs analysed for association to BMI in IL11 region in the GIANT GWAS 140

Figure 5.1. Schematic diagram of the IL-11 gene showing the positions of SNPs successfully genotyped in this study. 141

Table 5.1. IL-11 SNPs successfully genotyped with minor allele frequencies observed 142

Table 5.2. Allelic association analysis of IL-11 SNPs to obesity in children and adult French case-controls 143

Table 5.3. Association analysis of IL-11 SNPs to the quantitative trait of BMI within the French controls using linear regression and age and sex as covariates 144

Figure 5.2. Linkage disequilibrium (r^2) plot of IL-11 SNPs genotyped in the French control samples 145

Table 5.4. Haplotype association analysis with obesity in adults using the four SNPs found to be nominally associated with adult obesity 146

Table 5.5. Association analysis of IL-11 SNPs with BMI in the Swedish families using the QFAM test and the within-families model of association correcting for age and sex 147

Table 5.6. Association analysis of IL-11 SNPs with IL-11 transcript expression in subcutaneous adipose tissue 149

Figure 5.3. Box-plot of IL-11 transcript levels in subcutaneous adipose tissue with obese and non-obese subjects from Swedish sib-pairs displayed separately 150

Table 5.7. T-test of subcutaneous adipose IL-11 transcription in obese and non-obese subjects from Swedish sib-pair families 151

Figure 5.4. IL-11 transcription levels in subcutaneous adipose tissue plotted against BMI in non-obese siblings 152

Figure 5.5. IL-11 transcription levels in subcutaneous adipose tissue plotted against BMI in obese siblings 153

Table 5.8. Thesis-wide correction of p-values for muliple testing using 1 million permutations from association analysis of IL11 SNPs and obesity 154

Table 5.9. Power calculations carried out using MAF and OR observed for each SNP investigated in the IL-11 gene 154

Figure 5.6. Sample size plotted with statistical power for a SNP with MAF= 0.30 155

Figure 5.7. Sample size required for varying odds ratios in order to achieve a statistical power of 95% for a SNP with MAF= 0.30 155

Figure 6.1. SNPs analysed for association to BMI in the PNPLA3 region in the GIANT GWAS 163

Figure 6.1. Map of adiponutrin gene with positions of SNPs genotyped in this project 164

Table 6.1. SNPs successfully genotyped with minor allele frequencies observed 165

Table 6.2. Allelic association analysis of Adiponutrin SNPs to obesity in children and adult French case-controls 166

Figure 6.2. Linkage disequilibrium (r^2) plot of Adiponutrin SNPs genotyped in the French control samples 167

Table 6.3. Association analysis of Adiponutrin SNPs to BMI in the French controls 169

Table 6.5. Power calculations carried out using MAF and OR observed for each SNP investigated in the adiponutrin gene. 170

Figure 7.1. SNPs analysed for association to BMI in NUCB2 region in the GIANT GWAS 176

Figure 7.2. Map of NUCB2 gene and surrounding region with positions of SNPs genotyped in this project 178

Table 7.1. NUCB2 SNPs successfully genotyped with observed minor allele frequencies 179

Table 7.2. Allelic association analysis of NUCB2 SNPs to obesity in children and adult French case-controls 181

Table 7.3. Allelic association analysis of NUCB2 SNPs to obesity in non-diabetic children and adult French case-controls 182

Table 7.4. Association analysis of NUCB2 SNPs to BMI using linear regression with age and sex as covariates 183

Figure 7.3. Linkage disequilibrium (r^2) plot of SNPs genotyped in NUCB2 and the surrounding region in the French control samples 185

Table 7.6. Power calculations carried out using MAF and OR observed for each SNP investigated in the NUCB2 gene region 186

Figure 7.4. Sample size plotted with statistical power for a SNP with MAF=0.06 187

Figure 7.5. Sample size required for varying odds ratios in order to achieve a statistical power of 95% for a SNP with MAF=0.06 188

Table 8.1. Association analysis of IRS-1 SNP, rs2943641 with BMI corrected for age and sex in the Swedish families 194

Table 8.2. Association analysis of IRS-1 SNP, rs2943641 with IRS-1 transcript level corrected for age, sex and BMI in the Swedish families 195

Figure 8.1. Box-plot of IRS-1 transcript levels in subcutaneous adipose tissue with obese and non-obese subjects from Swedish sib-pairs displayed separately 196

Table 8.3. T-test of subcutaneous adipose IRS-1 transcription in obese and non-obese subjects from Swedish sib-pairs 196

Figure 9.1. Electropherogram showing the sequencing of the CpG island within the POMC promoter 204

Figure 9.2. Bisulfite sequencing of POMC CpG island in two samples. CpG sites are highlighted and differences in methylation status can be observed at CpG 3. 205

Figure 9.3. CpG plots of the region covering the promoter region and start of the leptin gene. 206

Figure 9.4. CpG plots of the region covering the promoter region and start of the FTO gene 207

Figure 9.5. CpG plots of the region covering the promoter region and start of the PPARγ gene. 208

Figure 9.6. CpG plots of the region covering the promoter region and start of the Adiponutrin 209

Figure 9.7. The region of the leptin gene analysed by bisulfite sequencing with CpG dinucleotides numbered and shown in bold 210

Figure 9.8. Electropherograms showing differing levels of methylation at CpG site 6 from the bisulfite sequencing of Leptin within three obese individuals 211

Table 9.1. Variation in levels of DNA methylation at six CpG sites sequenced in the Leptin promoter CpG island between different DNA samples 212

Table 9.2. Association of obesity and BMI with the percentage methylation at the three variably-methylated CpG sites in the Leptin gene CpG island 214

Table 9.3. Association analysis of SNPs in the Leptin gene region with BMI in French controls using linear regression and age and sex as covariates 215

Table 9.4. Association analysis of SNPs in the Leptin gene region with CpG 5 methylation in French controls using linear regression and age and sex as covariates 216

Figure 9.9. Box-plot of LEP transcript levels in obese and non-obese subjects from Swedish sib-pairs 218

Table 9.5. T-test of LEP transcription in obese and non-obese subjects from Swedish families 218

Chapter 1

Introduction

1.1 Obesity overview

According to the World Health Organization (WHO), global obesity rates have doubled since 1980 and in 2008 there were 1.4 billion adults (age > 20 years) classed as overweight and at least 500 million classed as obese[1]. This means that more than one in ten of the world's adult population is obese. Body weight is classified using body mass index (BMI), which is calculated as a persons weight in kilograms divided by their height in metres squared (kg/m^2). Overweight individuals are those whose BMI is between 25 kg/m^2 and 30 kg/m^2. Obese are those whose BMI is greater than 30 kg/m^2. Those whose BMI is between 35 and 40 kg/m^2 are classified as severely obese and anyone with a BMI greater than 40 kg/m^2 is morbidly obese. These classifications are based on the likelihood of a person developing one of the many serious conditions associated with increased body fat. Of all associated diseases, type 2 diabetes (T2D) has the strongest correlation to obesity[2] and according to the WHO, obesity is the cause of 44% of all cases of T2D[1]. Other comorbidities include cardiovascular disease, hypertension, dyslipidemia (abnormal blood lipid or cholesterol levels) and coronary heart disease along with a number of cancers that affect the oesophagus, pancreas, colon, rectum, breast, endometrium and kidney[2]. Additionally, obesity is associated with respiratory disorders (reviewed by Murugan and Sharma[3]) chronic kidney disease (reviewed by Ting *et al.*[4]), musculoskeletal disorders (reviewed by Wearing *et al.*[5]) and liver disease (reviewed by Batty *et al.*[6]).

Because of these associated diseases, obesity is one of the leading preventable causes of death worldwide[1] and the related health care costs make it one of the most serious public health problems in the modern world. Currently these costs are estimated to be as high as 1.2% of gross domestic product (GDP) in the USA[7] and between 0.09% to 0.61% of GDP in European countries[8].

1.2 Energy regulation in mammals

Obesity is caused by excessive fat mass resulting from an increased storage of surplus energy in the form of lipids in adipose tissue. It develops as a result of a disruption of the homeostatic system that regulates energy intake and energy expenditure. In humans, body weight is a highly regulated trait. When normal-weight individuals eat a caloric surplus, they are found to increase their energy expenditure such that they maintain adiposity levels[9]. Conversely when subjects are fed a semistarvation diet they are found to reduce their energy expenditure as well as experience extreme hunger[10]. To achieve this regulation a series of signals are produced in the adipose tissue and digestive system that relay nutritional status to the central nervous system (CNS) which then adjusts food intake and energy expenditure according to requirements. Dysregulation of this system can lead to obesity and the study of obesity genetics has led to an increased understanding of the way energy is regulated in mammals which is not only physiologically important but also has pharmacological implications as the basis for the design of therapeutics for treatment of obesity.

1.2.1 Rodent models of obesity

Many of the advancements in understanding energy regulation began with mouse and rat models of the disease. Many of the mutations that cause obese phenotypes in rodents have turned out to be in genes involved in the leptin-melanocortin pathway that regulates appetite and energy expenditure in both rodents and humans.

The first rodent model to be reported was the *agouti* mutation in mice. Originally discovered in 1902, the gene was cloned 90 years later[11]. The mutation that causes obesity is known as lethal yellow mutant mouse (A^{y}) and these mice exhibit obesity, T2D and yellow coat colour[11]. The yellow coat colour results from *agouti* overexpression in the skin which blocks alpha melanocyte-stimulating hormone (α-MSH) signalling at melanocortin-1 receptors in the hair follicle. Studies of this mouse strain led to the discovery of the melanocortin pathway, which is central in the CNS control of energy balance in mammals.

The autosomal recessive *obese (ob)* mutation was discovered in 1950[12] which causes affected animals to reach a body weight three times that of control mice. The gene responsible, *leptin*, was identified using methods of linkage mapping followed by positional cloning in 1994[13,14]. Leptin is a hormone that is secreted from adipose tissue and is important in signalling the level of stored energy to the CNS (see section 1.2.4). The *db/db* (diabetes) mouse, discovered in 1966[15], has a similar phenotype to *ob/ob* mice but also develops hyperglycaemia by 8 weeks of age which has resulted in their use as a model for the study of T2D. The mutation responsible was found to be in the leptin receptor gene (*LEPR*)[16,17].

The Zucker fatty rat was discovered in 1961 and is caused by an autosomal recessive variant in the gene denoted *fatty (fa)*[18]. These rats have early-onset obesity, which appears at 5 weeks of age. The *fa* gene was subsequently cloned and shown to be *LEPR*[19]. Another rat model of obesity is the Otsuka Long-Evans Tokushima Fatty (OLETF), discovered in 1984 and characterised by obesity, adult-onset hyperglycaemia and chronic non-insulin dependent diabetes[20]. This was found to be caused by a mutation in the gene that encodes the receptor of cholecystokinin (CCK), a gut hormone that is released after feeding and suppresses appetite[21].

A high-fat diet (HFD) has also been used to induce obesity in otherwise lean animals. The C57BL/6J strain of mouse is often used for this purpose because they exhibit symptoms similar to human metabolic syndrome when fed the HFD (review by Collins *et al.*[22]).

1.2.2 Central nervous system control of energy regulation

Early lesion experiments in animals revealed the importance of the CNS in controlling body weight. These studies identified the hypothalamus as the area of the brain responsible and in particular the ventromedial hypothalamic nucleus (VMH)[23] and the paraventricular hypothalamic nucleus (PVN)[24]. Lesions in these regions result in animals which become hyperphagic (an abnormally increased appetite) and obese. Lesions in the lateral hypothalamus (LH) on the other hand, produce aphagia (the inability or refusal to swallow) and weight loss[25].

The specific types of neurons within the hypothalamus that are responsible for energy regulation have since been characterised (see Figure 1.1). Neurotransmitters in the CNS that effect changes in energy balance can be divided into two groups: Those that promote feeding and inhibit energy expenditure, which are known as orexigenic and those that inhibit feeding and promote energy expenditure, which are known as anorexigenic.

1.2.2.1 Orexigenic signals in the CNS

The most well studied orexigenic neurons are the neuropeptide tyrosine (NPY) and agouti-related peptide (AgRP) expressing cells of the arcuate nucleus (ARC) of the hypothalamus. The ARC is positioned in the brain adjacent to the median eminence, an area of the brain that is devoid of the blood-brain barrier[26]. Thus the ARC is able to receive signals from circulating hormones that have been released from the periphery[27].

NPY belongs to a family of structurally related proteins that include the peripheral hormones pancreatic polypeptide (PP) and peptide tyrosine-tyrosine (PYY)[28,29]. NPY is expressed throughout the CNS but is highly concentrated in the ARC[30]. Its orexigenic effects have been demonstrated by infusing the peptide into the brains of rodents which results in significantly increased feeding followed by weight gain[31]. In diet induced obese mice, NPY expression levels in the ARC are observed to increase when animals are placed on a low-calorie diet and lose their excess fat, which is congruent with a role for NPY in stimulating feeding[32].

Another important orexigenic signal is AgRP, a protein that shares sequence homology to the agouti protein responsible for the A^{y} obese mouse described above[33]. It is primarily expressed in the ARC and co-localises with NPY[33]. As with NPY, the orexigenic effects of AgRP have been demonstrated by infusing the protein into the brains of rodents which potently stimulates feeding and results in weight gain[34]. Interestingly, administration of the N-terminal domain of AgRP also results in increased body weight but without the associated hyperphagia[35]. Additionally, reduction of AgRP expression using RNA interference results in increased energy expenditure and loss of weight without changes in appetite[36] suggesting that regulation of energy expenditure is important in AgRPs orexigenic action. AgRP is the endogenous antagonist of the melanocortin 3 and 4 receptors (MC3R and MC4R) which, along with proopiomelanocortin (POMC), comprise the melanocortin system (detailed in section 1.2.2.2)[37]. Furthermore, synthetic MC3R/MC4R antagonists produce a similar effect to AgRP when injected into rodent brains[38].

At times of negative energy balance, expression of AgRP and NPY are upregulated. This has been demonstrated during periods of fasting where both AgRP and NPY levels rise and this is primarily due to a reduction in levels of the peripheral hormones insulin and leptin along with a rise in ghrelin, all of which are important regulators of adiposity (see sections 1.2.3, 1.2.4 and 1.2.6)[39,40]. At times of positive energy balance, expression is downregulated: Both AgRP and NPY gene expression is lower in rats on a high-calorie diet compared to a low-calorie diet[41].

Deletion of *AgRP* and/or *NPY* genes in mice does not lead to the expected changes in body-weight[42]. It has been hypothesised that animals with knockouts in AgRP or NPY induce compensation mechanisms during embryonic development. To test this, experiments have been conducted which involve ablation of AgRP and NPY genes in adult mice. These studies resulted in significant decreases in food intake and body weight demonstrating that AgRP and NPY are important for proper energy control in the normally developed mouse[43].

NPY and AgRP neurons project into the PVN as well as other hypothalamic areas and areas of the brain outside of the hypothalamus[44]. Neurons expressing NPY and AgRP in the ARC are found to express high concentrations of leptin receptor[45] and leptin is a hormone that is important in signalling current adiposity levels (see section 1.2.3). NPY and AgRP neurons in the ARC also express the inhibitory neurotransmitter, gamma-Aminobutyric acid (GABA) and projections that release GABA have been found to project onto POMC neurons in the ARC to inhibit POMC neuron action, the anorexigenic peptide of the melanocortin system (see section 1.2.2.2)[46]. There is also evidence that GABA action from these neurons is necessary for the proper maintenance of feeding[47].

NPY and AgRP are the most extensively studied orexigenic neurotransmitters but there are a number of others that have been characterised.

Melanin-concentrating hormone (MCH) was originally discovered in rodents where it was found to be overexpressed in response to fasting and also in genetically obese (*ob/ob*) mice[48,49]. MCH is expressed exclusively in the lateral hypothalamus and zona incerta[50]. The role of MCH as an orexigenic molecule has been demonstrated using mouse knockouts which are lean and hypophagic[51]. Additionally, centrally administered MCH has been shown to stimulate feeding[48].

Orexin A and orexin B are two peptides derived from the same gene, orexin neuropeptide precursor (*HCRT*), which also act as orexigenic neurotransmitters[52]. Orexin expressing neurons are found in the dorsal, lateral and posterior hypothalamic areas[52,53]. As with NPY, AgRP and MCH, centrally administered orexin peptides result in increased food consumption in rodents[52] and their expression is upregulated during fasting[54].

Galanin is another neurotransmitter known to stimulate feeding and reduce energy expenditure. It is expressed in the gut as well as in the hypothalamus, particularly the PVN, ARC and LH[55]. As with other orexigenic peptides, when galanin is infused into the brain it leads to increased feeding[56]. Both insulin[57] and leptin[58] reduce galanin expression, and both these hormones are known to induce anorexigenic responses (see Sections 1.2.3 and 1.2.4). Galanin has been implicated in regulating preference for high fat foods and GAL-null mice consume less food than wild type when given a high fat diet[59]. Galanin-like peptide (GALP) is structurally related to galanin. It is expressed in the ARC[60] and injections of GALP into the brains of rats stimulates feeding which suggests a role as an orexigenic peptide[61]. Its role in feeding may be more complex, however, since in mice it has been reported to have the opposite effect[62].

1.2.2.2 Anorexigenic signals in the CNS

POMC is the gene that codes for the anorexigenic proteins of the melanocortin system, which also includes AgRP and the melanocortin receptors. POMC is expressed in the ARC[63] where it is translated into multiple peptides including alpha melanocyte stimulating hormone (α-MSH), β-MSH, γ-MSH, adrenocorticotropic hormone$_{1-24}$ (ACTH(1-24)), and ACTH(1-13)NH2 (desacetyl-α-MSH)[64]. α-MSH is the endogenous agonist of melanocortin receptors and when a melanocortin agonist is applied to rodent brains the result is inhibition of feeding and a reduction in body-weight. Melanocortin antagonists are found to produce the opposite result[65]. Deletion of the *POMC* gene in mice causes obesity and hyperphagia, which demonstrates that functional POMC peptides are crucial for proper energy regulation[66]. POMC-null mice are not diabetic, however they do have reduced glucagon secretion in response to experimental hypoglycaemia, indicating a role of POMC peptides in regulating blood glucose[67]. POMC expression is downregulated after fasting and is significantly reduced in genetically obese (*ob/ob*) mice compared to wild type[68].

There are a number of different melanocortin receptors, of which MC3R and MC4R are the most important for the regulation of energy homeostasis. Studies in mice indicate that MC4R is involved in the control of both feeding and energy expenditure[69,70]. It is also involved in the control of glucose and lipid metabolism, independent from it's role in controlling food intake and energy expenditure[71]. *MC4R* is widely expressed throughout the brain and spinal cord[72] whereas *MC3R* is localised to the ARC and other regions of the hypothalamus[73]. MC3R then appears to function in negative feedback control between AgRP/NPY and POMC/CART neurons which have projections between one-another in the ARC [46,73]. *MC3R* knockout mice are normal-weight but display other symptoms of metabolic syndrome: they have increased levels of fat mass and decreased levels of lean mass, they are hyperleptinaemic and have decreased levels of energy expenditure[74,75]. The MC4R gene knockout on the other hand results in mice that are

hyperphagic and obese[69]. The various actions of MC4R have been located to different parts of the brain using knockout studies in specific neuron populations in mice. For example, MC4R expressed by SIM1 neurons in PVH and amygdala control food intake but not energy expenditure or glucose metabolism[76] and MC4R expressed outside of the hypothalamus are implicated in the control of energy expenditure, hyperinsulinemia and hyperglycaemia[77].

AgRP/NPY neurons seem to primarily sense energy deficits since *NPY* is upregulated by starvation but not affected by overfeeding[78]. On the other hand hypothalamic *POMC* mRNA in rodents is downregulated by starvation[79] and upregulated by overfeeding[80] suggesting that POMC neurons respond to both stimuli. These findings may suggest a reason why it is easier for individuals to gain weight than to lose weight.

Another important anorexigenic signal is cocaine and amphetamine-regulated transcript (CART), which highly expressed in the hypothalamic neurons and co-localises with *POMC*[81,82]. CART has been implicated in feeding, drug reward behaviour, stress, and development (reviewed by Kuhar *et al.*[83]). Its role in regulating feeding was demonstrated using intracerebrovasular injections in rodents which result in significantly reduced food intake[84]. Mice engineered without functional *CART* gene are obese, indicating that functional CART is necessary for regulation of energy balance[85]. When rodents are deprived of food, CART mRNA in the ARC, decreases[86]. As with NPY/AgRP neurons, leptin receptor has been found to co-localise with *POMC* and *CART* in the ARC[87]. The mouse model of obesity, *ob/ob,* which lacks functional versions of leptin have significantly reduced *CART* expression in the ARC and other hypothalamic regions. Injections of leptin into these animals results in normal levels of *CART* mRNA thus providing evidence that leptin's anorexigenic effects act, in part, via CART neurons[86].

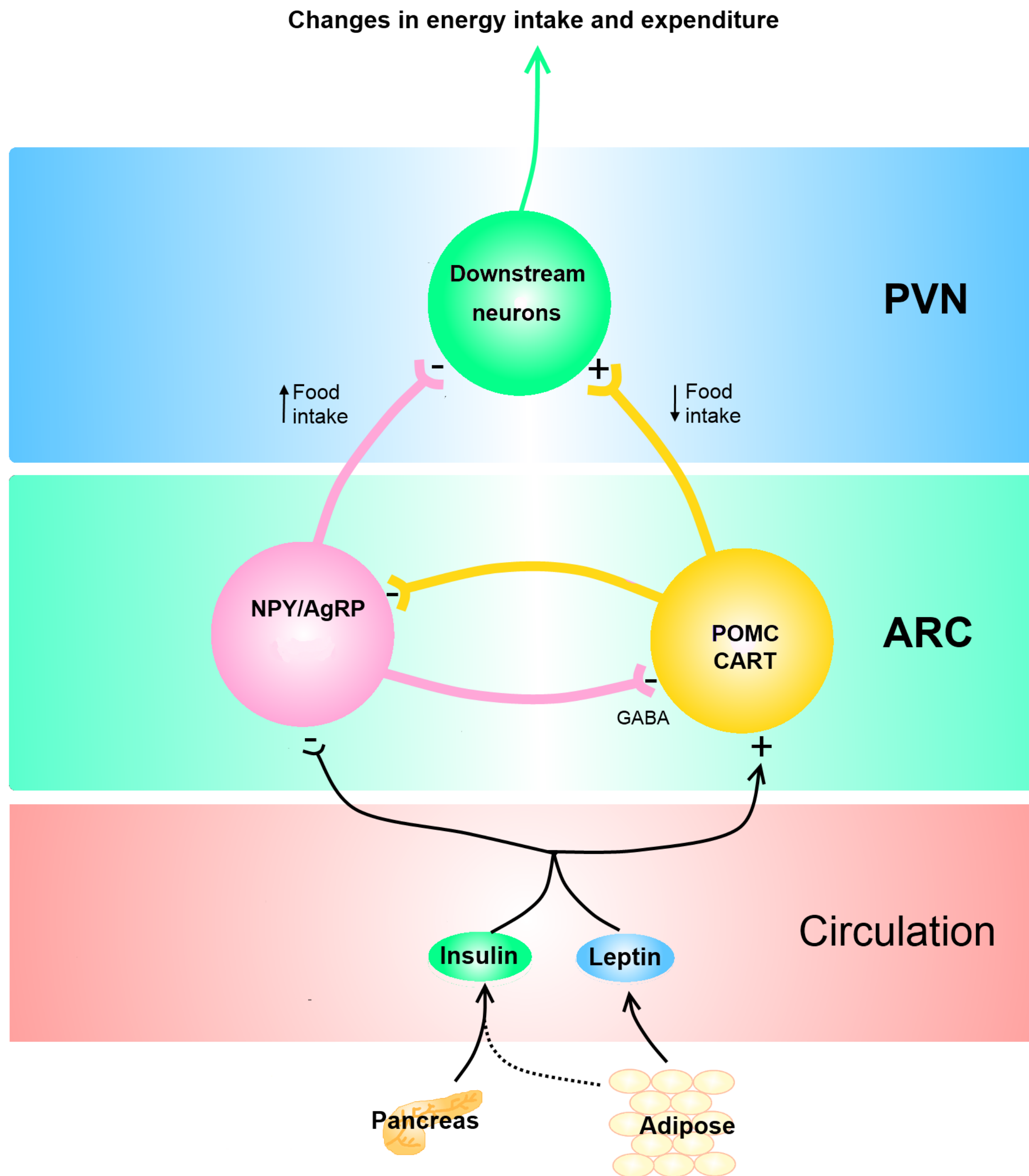

Figure 1.1 Energy regulation in the hypothalamus. Circulating insulin and leptin, which are released in proportion to adiposity levels, stimulate and suppress activity of the arcuate nucleus neurons POMC/CART and NPY/AgRP respectively. These neurons respond by increasing or decreasing the expression of their neuropeptides which then act on downstream neurons in the paraventricular nucleus to decrease or increase feeding and increase or decrease energy expenditure respectively.

1.2.3 Insulin as an adiposity signal

There are a number of signal peptides whose serum concentrations are altered by adiposity levels, thus acting to relay the level of energy storage to the CNS. Insulin was the first such molecule to be implicated. Whilst its primary role is in controlling serum concentrations of nutrients, insulin levels in the blood are found to be proportional to the total adipose mass of an organism[88,89]. This has been shown to be true after fasting, after feeding and after an intravenous glucose infusion[90]. Thus insulin is secreted in response to blood glucose in proportion to body fat such that a person with greater fat mass secretes proportionally more insulin to a given glucose concentration[90]. Additionally, endogenous insulin is found in the cerebrospinal fluid of mammals[91] and it has been shown to be able to cross the blood brain barrier which demonstrates the possibility of it having a signalling role in the brain[92].

Injection of insulin into the brain decreases food intake and body weight in rodents and primates[93,94,95]. Furthermore in humans, intranasal insulin has been shown to reduce body fat[96]. Conversely administration of insulin antibodies results in hyperphagia and weight gain[97,98]. Transgenic mice with a neuron and glial cell specific knockout of insulin receptor are hyperphagic and obese[99] and selective decrease in hypothalamic insulin receptor protein following administration of insulin receptor antisense oligonucleotides produces the same result[100]. Histological studies have shown that this insulin-mediated control of feeding and body weight is located in the ARC, PVN, dorsomedial and VMH of the hypothalamus[98] and insulin receptor is highly expressed in the ARC[101].

1.2.4 Leptin as an adiposity signal

Leptin is a protein expressed and secreted by adipocytes that has a serum concentration directly proportional to the fat mass of the individual[102,103]. Intraperitoneal injections of leptin into mice decreases feeding and result in weight loss and reduced fat mass which is evidence that leptin is an anorexigenic hormone[104]. These experiments produce the same result in rats and leptin is found to work additively together with insulin[105]. As noted previously, mice lacking functional leptin genes *(ob/ob)* are obese and so are mice that lack a functional leptin receptor gene (*db/db).* This is also true in humans who are have congenital leptin deficiency and it has been shown that leptin injections significantly reduce fat mass and reverse obesity in such patients[106]. Its use as a therapeutic in common obesity, however is limited due to obesity being associated with leptin resistance[107]. Mice with brain-specific leptin receptor knockouts are obese whereas peripheral-specific knockouts are not suggesting that leptin's effects are controlled in the brain[108]. Thus leptin acts as a hormone to signal adiposity levels to the CNS in order that appetite and energy expenditure can be adjusted such that energy stores are kept in balance.

Leptin's action in the CNS has been well characterised. It inhibits hypothalamic expression of *NPY* and *AgRP*[109] and upregulates *POMC* and *CART*[82]. Leptin receptor is found to be expressed in the majority of NPY/AgRP and POMC/CART neurons in the ARC[110,87]. POMC neurons are depolarized by leptin treatment thus producing an anorexigenic effect[46]. Knockdown of leptin receptor in POMC as well as AGRP/NPY neurons lead to increased body weight in mouse models which demonstrates that leptin's action in the ARC is necessary for the proper control of fat storage[111,112].

Thus leptin is a crucial hormone that works to create a feedback loop that ensures adiposity levels are kept within the correct range. Any increase in fat mass causes an increase in circulating leptin which starts a signal cascade which then reduces fat mass. Thus stored fat levels are kept within a narrow range.

1.2.5 Other adiposity signals

Amylin is a hormone that is cosecreted with insulin by the beta cells of the pancreas. As such it is a hormone that is secreted in proportion to adiposity levels[113]. Administration of amylin into rats reduces feeding and results in lower body weight indicating a role as an anorexigenic signal[114]. Its effects are thought to be mediated mainly by decreasing meal size[115]. Conversely, administration of the amylin receptor antagonist AC187 stimulates eating by increasing meal size[116].

Adiponectin is a peptide that is expressed and secreted by adipocytes. It has been found to stimulate food intake and deletion of the adiponectin gene in mice leads to a decrease in feeding[117]. Additionally, adiponectin receptors colocalize with POMC and NPY/AgRP neurons in the ARC which supports a role in regulating energy homeostasis[118].

1.2.6 Gastrointestinal signals

In addition to hormones that signal current levels of stored fat and thereby act to regulate long term energy needs there are also a number of molecules that regulate food intake in the short term. For the most part they comprise hormones secreted by the gastrointestinal tract to signal the presence of nutrients as they enter various parts of digestive system. The stomach produces signals in response mainly to mechanical distention, whereas the intestine releases signals in response to chemical content of the food (reviewed by Powley *et al.*[119]).

CCK is produced in the mucosa cells of the intestine[120] and has been shown to decrease food intake in rats and humans[121] and rats without CCK receptor are obese[21]. CCK is rapidly released after protein or fat-rich food enters the duodenum[122]. It inhibits gastric emptying and gastric acid secretion and mediates digestion in the duodenum which in turn stimulates meal termination[123]. The inhibitory effect on feeding is short-lived, lasting less than 30 minutes and so CCK works to reduce meal size and duration but does not effect the onset of the next meal[124,125].

Peptide tyrosine tyrosine (PYY) is secreted from intestinal L-cells immediately after feeding and remain elevated for up to six hours[126]. Administration of PYY inhibits feeding in rodents and humans[127,128,129] and as such PYY has been trialled as a potential therapeutic for treating obesity[130]. PYY knockout mice have a reduced food intake and body weight compared to controls[131]. PYY seems to respond more to protein intake than other nutrients: PYY levels are higher after consumption of a high-protein meal compared to a high-fat or high-carbohydrate meal with the same calorie content[131]. Fasting PYY levels correlated negatively with BMI indicating a role in long-term regulation of feeding[127,131].

Ghrelin is a hormone that is unique amongst gut signals in that it is a stimulator of food intake rather than an inhibitor[132]. It is secreted from the stomach and its levels are correlated with hunger[133], rising shortly before feeding and falling sharply during food intake[134]. Carbohydrates are the most potent macronutrient suppressors of ghrelin, while fat suppresses ghrelin levels to a lesser extent[135]. Administration of ghrelin into normal weight subjects results in a stimulation of appetite[132] and mice with ghrelin or ghrelin receptor knockouts are resistant to diet-induced obesity[136,137]. In addition, ghrelin is thought to be involved in long-term energy regulation as its circulating levels are negatively correlated with BMI[138], are increased following weight loss[139] and its receptors are found in ARC NPY/AGRP neurons[140].

There are a number of other hormones secreted from areas of the digestive system, the major ones are listed in Table 1.1 along with their site of production and known effect on appetite.

Hormone	**Main site of production**	**Primary actions**
Cholecystokinin (CCK)	Duodenum[141]	Inhibits gastric emptying[142], and reduces meal size and duration[143]
Ghrelin	Stomach[144]	Stimulates food intake[132], is released just before a meal[134]
Glucagon-like peptide-1 (GLP-1)	Ileum and colon[145]	Reduces food intake[146], induces glucose-dependent insulin secretion and inhibits glucagon secretion[147]
Obestatin	Stomach[148]	Encoded by the same gene as ghrelin. Acts in opposition to ghrelin to reduce appetite[149]
Oxyntomodulin	Ileum and colon[150]	Suppresses appetite and increases energy expenditure[151]
Pancreatic polypeptide (PP)	Pancreas[152]	Decreases food intake[153].
Peptide tyrosine tyrosine (PYY)	Ileum and colon[126]	Induces satiety by inhibition of NPY/AGRP neurons[127]

Table 1.1. Hormones released from the gastrointestinal tract and their effects on energy regulation.

1.2.7 Lipid metabolism

Since obesity is caused by increased quantity of stored energy in adipose tissue, the study of lipid metabolism is a crucial part of understanding the disease.

Most of the energy reserves in the human body are stored in the form of triacylglycerols (TAG) within lipid droplets in adipocytes. There are two important processes relating to lipid metabolism within the adipocyte: TAG storage and TAG hydrolysis. When energy intake is higher than expenditure, storage occurs. Within the adipocyte, free fatty acids are esterified into TAG and partitioned into lipid droplets. When energy demands increase, adipose tissue acts to

provide nutrients to other tissues via lipolysis which breaks down TAG into free fatty acids and glycerol. Lipid droplets are composed of a core of TAG and cholesterol esters, surrounded by a phospholipid monolayer, which contains lipid droplet associated proteins[154,155]. These proteins are thought to control the accessibility of lipases to their lipid substrates.

The perilipin family of proteins represent the most abundant of the lipid droplet associated proteins[156]. They localise specifically to lipid droplets and have not been found in any other subcellular compartment[157]. Under fed conditions, perilipin inhibits lipolysis which in turn increases TAG storage[158]. Conversely, during fasting, phosphorylated perilipin results in increased lipolysis[159]. Perilipin knockout mice are lean and resistant to diet induced obesity[160].

Desnutrin is considered to be the major TAG lipase in white adipose tissue (WAT). It is expressed in adipocytes and is found to be downregulated in *ob/ob* mice[161]. Furthermore, transgenic mice that overexpress desnutrin in adipose have increased lipolysis and fatty acid oxidation in WAT, resulting in higher energy expenditure and resistance to diet-induced obesity[162]. Adipose-specific phospholipase A2 (AdPLA) is a membrane-associated lipase expressed exclusively in adipocytes[163] and AdPLA knockout mice exhibit unrestrained adipocyte lipolysis and are extremely lean[164]. These mouse models do not have elevated serum fatty acid levels but instead have increased fatty acid oxidation in adipocytes. AdPLA null mice have significantly increased expression of uncoupling protein 1 (UCP-1) in WAT and brown adipose tissue (BAT)[164] and this is thought to result in an increased metabolic rate and consequently protects against obesity. UCP-1 overexpression in mouse adipocytes increases mitochondrial respiratory uncoupling, (the generation of heat energy instead of ATP) and subsequently, results

in a significantly leaner mouse[165]. In mice, inhibiting lipolysis results in promotion of diet-induced obesity[166]. These results demonstrate the possibility that impaired lipolysis in adipose tissue may be a factor in the development of obesity and thus increased lipolysis may be a potential therapeutic target for the treatment of obesity.

1.3 Heritability of obesity

The recent obesity epidemic is assumed to be due to changing environments such that calorie content of meals has increased while at the same time physical activity has decreased (reviewed by French *et al.*[167]), although recent evidence suggests that total energy expenditure has not changed between hunter-gatherer and Western societies[168]. However there is also strong evidence of a heritable influence to obesity.

Heritability (h^2) is the proportion of phenotypic variation between individuals in a population that can be attributed to genetic rather than environmental factors. It is measured by estimating the contributions of genetic and non-genetic differences to the total variation of a phenotype in a sample from a population. There are three main types of heritability studies that have been carried out to try to determine the genetic component of traits such as BMI. These are adoption studies, family studies and twin studies. All three types of study have produced estimates for the heritability of BMI to be in the range of 40% to 80%. For a review of heritability studies in body weight see Andersson and Walley[169].

Adoption studies work by measuring the correlation between adoptees and their biological parents. The advantages of this method are that they resolve the effects of genes from the effects of shared environment. This is because adopted children share genetics with their biological parents but not their adoptive parents and share environment with their adoptive

parents but not their biological parents. One disadvantage of this method is that the effects of age are not accounted for. There is also an assumption that adoptees are not selectively placed into families based on any shared characteristics and an assumption that there is no significant effect of the prenatal environment. An additional limitation of this study design is the relatively low sample sizes for which data is available for both adoptive and biological parents. One notable exception was a study published by Stunkard *et al.* using the Danish Adoption Register which contained BMI data for 3580 adoptees. A strong correlation between adoptees and their biological parents was discovered but no correlation to their adoptive parents indicating a strong degree of heritability[170].

The family study design most commonly uses parent-offspring or sibling-sibling data sets. The main advantage is the ease with which cohorts can be recruited. It suffers though, from the fact that environmental effects cannot be separated from genetic effects. Correcting the data using an environmental index is possible but this will not be able to separate all factors. One large study of BMI in Norway reported an average parent-offspring correlation of 0.20 and siblings ranged from 0.21 in brothers to 0.26 in sisters. Second degree relatives did not correlate significantly[171]. This study estimated h^2 of BMI to be0.39.

Twin studies were the first heritability studies to be performed and have been used extensively for studying the heritability of traits including BMI and other body-weight phenotypes. They usually rely on using samples of identical twins and samples of fraternal twins. Fraternal or dizygotic (DZ) twins share on average half their genetic variants. Identical or monozygotic (MZ) twins share effectively 100% so are twice as similar, genetically, than DZ twins. If a trait is heritable then MZ twins should be more similar than DZ twins. The genetic contribution to variation (h^2) can thus can be calculated as twice the difference in correlation between MZ and DZ twins. Because twins are matched for age, any effects that age has on variation of a trait are

the same for both twins. Furthermore they are statistically powerful: One thousand twin pairs provide 95% power to reject a purely environmental model with a false-negative rate of 0.05 if additive genetic effects explain 20% and shared environment effects explain 50%[172]. Power increases if the actual genetic effects are higher. The main criticism of this method is that twins are not a random sample of the population, they differ in developmental environment and so may not be representative. However, in the case of BMI, twins have not been found to be significantly different from singletons[173]. Another criticism is that twin studies often use self-reported zygosity measures which may not always be correct. MZ twins that are discrepant in height and weight could be incorrectly viewed as being DZ and this bias would lead to increased heritability estimates[174]. Twin studies also assume that the shared environment is identical between MZ twin pairs and DZ twin pairs and this may not be true[175].

Many twin studies have been performed over the years to investigate the heritability of body-weight (for a review see[169]). When analysing BMI in adult twin samples, monozygotic twins are consistently more similar in BMI than dizygotic twins (h^2=0.64[176] h^2=0.80[177] h^2=0.77[178]). When analysing BMI in adolescent twin samples heritability is also very high (h^2=0.87 at age 11[179]). When skinfold thickness is analysed in adolescent twins the story is the same (h^2=0.98 for trunk fat and 0.46 for limb fat[180]).

Studies that used twins reared apart are the most powerful study design since they can be used to easily separate the effects of genotype from post-natal environment. The correlation of MZ twin reared apart is s a direct estimate of the heritability. They are limited by the relatively small numbers of twins that are raised apart. Additionally they rely on the same assumptions of non-

selective adoption and no effect of the prenatal environment, which may not be valid. In practise such studies have corroborated the results of studies of twins reared together with h^2 estimates in the same range. One such study reported heritability estimates of 0.70 for men and 0.66 for women[181].

1.4 The genetics of obesity

Obesity can be classified into three types based on the underlying genetic basis. In the first instance there is monogenic obesity which is characterised by severe obesity without other abnormalities and has a Mendelian pattern of inheritance. Secondly, there is syndromic obesity which includes additional phenotypes, often mental retardation and dysmorphic features. Last is common polygenic obesity which affects the general population.

1.4.1 Monogenic obesity

Monogenic diseases are caused by variants within a single gene and mutations in many of the genes involved in the regulation of appetite detailed above have been discovered to underlie monogenic forms of obesity in humans. Most of these were discovered following the development of mouse models of the disease which used linkage analysis to locate disease loci. These were crucial in the identification of the leptin-melanocortin signalling pathway that regulates metabolism in mammals and provided candidate genes for human monogenic forms of obesity. Table 1.2 lists the known monogenic forms of obesity.

The first monogenic form of obesity to be reported was found in two severely obese cousins from a Pakistani family. A frameshift mutation in the leptin gene was found to be responsible, which when translated produced a truncated form of the leptin protein that is not secreted[182]. A similar mutation has also been discovered in a Turkish family in which subjects are severely obese and hyperphagic[183]. Since their obesity is caused by a lack of leptin in the blood stream it was hypothesised that infusions of recombinant leptin could help with reducing body weight. Indeed, daily injections of human recombinant leptin for a period of one year was shown to produce a vastly reduced food intake and fat mass and a reversal of obesity in such leptin-deficient patients[106,184]. This is currently the only genetic form of obesity that successfully responds to pharmacological intervention.

The discovery of the first Mendelian form of obesity caused by leptin mutations was quickly followed by the identification of other forms of monogenic obesity caused by variants in genes in the leptin-melanocortin pathway. These include a mutation in the leptin receptor gene (*LEPR*) that was found to cause hyperphagia and obesity in individuals homozygous for the variant. These subjects do not have functional leptin receptors[185]. There are also variants in *POMC* that cause hyperphagia and early-onset obesity in homozygotes[186,187,188,189]. These subjects also have pale skin and red hair pigmentation due to lack of melanocyte stimulating hormone (MSH) action at melanocortin 1 receptors in skin and hair follicles.

A number of studies have also found mutations in *MC4R* which cause severe early-onset obesity and hyperphagia[190,191]. Patients that are heterozygous for this mutation are obese but those that are homozygous are more severely obese, indicating a codominant model of inheritance and this fits with mouse knockout studies[69]. MC4R is the most common form of monogenic obesity known, present in an estimated 1-6% of all early onset severely obese cases[192,193,194].

A mutation in the gene encoding brain-derived neurotrophic factor (*BDNF)* is responsible for another form of monogenic obesity[195]. BDNF is expressed in the VMH and is thought to have a role in controlling energy regulation downstream of the melanocortin system[196]. *NTRK2* encodes the receptor for BDNF and a mutation in this gene has been reported which causes early-onset obesity[197].

A compound heterozygous mutation in the prohormone convertase gene *(PSCK1)* has been found which causes severe early-onset obesity along with hypergonadotropic hypogonadism, abnormal glucose homeostasis, and increased plasma concentrations of proinsulin and POMC[198]. Prohormone convertase is expressed in neuroendocrine tissues and cleaves certain prohormones to their intermediate or final forms including some that are involved in energy regulation such as POMC and insulin[199].

Gene	**Position**	**Gene function**	**Inheritance**	**References**
BDNF	11p14.1	Neuron growth factor expressed in hypothalamus	R	195
LEP	7q32.2	Adiposity signal	R	182,183,200
LEPR	1p31.3	Leptin receptor	R	185
MC4R	18q21.32	Key receptor of the melanocortin system	C	190,191,192
POMC	2p24.1	Precursor for the anorexigenic neurotransmitter α-MSH	R	186,188,189
NTRK2	9q22.1	BDNF receptor	R	197
PCSK1	5q15	Cleaves POMC and proinsulin into active hormones	R	198

Table 1.2. Monogenic forms of obesity in humans. R = autosomal recessive, C = autosomal codominant.

1.4.2 Syndromic obesity

There are around 30 rare syndromes caused by genetic mutations or chromosomal abnormalities that are currently known to include obesity as a phenotype (reviewed by Farooqi *et al.*[201]). Phenotypes usually include mental retardation, dysmorphic features and developmental abnormalities. They often include hyperphagia indicating dysfunction of pathways in the hypothalamus.

The most common form of syndromic obesity is Prader-Willi syndrome with an estimated prevalence of 1 in every 25,000 births[202]. It is characterised by low birth-weight and feeding difficulties in infancy, followed by hyperphagia and obesity in childhood along with short stature, hypogonadism and mental retardation[203]. It is normally caused by a deletion or translocation of the paternal segment at 15q11-q13[204] or by uniparental disomy of chromosome 15 where a patient inherits two copies of the maternal chromosome 15 and zero copies of the paternal chromosome 15[205]. Alternatively it can sometimes be caused by defects in the imprinting of the paternal chromosome 15[206]. A smaller deletion (187kb) in the same region has been reported in a patient with hyperphagia, severe obesity, mild learning difficulties and hypogonadism but who lacked other symptoms of Prader-Willi[207]. This work has helped identify the genes responsible for the Prader-Willi phenotype, which comprise a family of non-coding RNAs.

Bardet-Biedl is another example of syndromic obesity, which along with increased body weight, includes retinal degeneration, hypogonadism, polydactylism, renal dysfunction, and mental retardation[208]. It is genetically heterogeneous with causative mutations having been located in severeal distinct regions of the genome[209,210]. Its inheritance is usually autosomal recessive, however some cases of digenic and tri-allelic inheritance have also been reported[211,212].

A translocation between 1p22.1 and 6q16.2 that disrupts the single-minded homolog 1 (SIM1) gene has been reported to be the cause of a syndrome of obesity, hypotonia and developmental delay[213]. *SIM1* encodes a transcription factor that is expressed in the PVN of the hypothalamus[214] and overexpression of the gene in *Agouti* mice and diet-induced obese mice has been shown to eliminate hyperphagia and result in a reduction in fat mass[215]. Other obesity syndromes include Cohen syndrome, an autosomal recessive disorder caused by a mutation at 8q22[216] and Alström syndrome, which is caused by mutations in the ALMS1 gene at 2p13.1[217].

1.4.3 Common polygenic obesity

Polygenic traits are those for which inheritance of a phenotype is caused by two or more genes in combination with the influence of environment. These phenotypes are often quantitative and vary on a continuous scale distributed normally. BMI is an example of such a quantitative trait with obesity found at the top end of the scale. If an individual carries many genetic variants that increase BMI then obesity may result. Any such genetic variant will be found in a higher frequency within obese populations than in normal and lean populations. The search for variants that influence susceptibility to common obesity have taken three approaches: Linkage studies, candidate gene association studies and genome-wide association studies.

1.4.3.1 Linkage studies

Linkage studies have been used to identify genetic markers that are shared by descent among relatives with the same disease or phenotype. It relies on the fact that markers that are located close to one-another on a chromosome are more likely to be inherited together than markers that are far apart due to crossing over events in meiosis. Thus the more often an allele at a particular site is shared in relatives who also share the phenotype in question, the more likely it

is that the site is in close proximity to a disease causing locus. As discussed previously, this method has been employed successfully to locate the genes responsible for monogenic forms of obesity in mice. However, it is less suited to the investigation of complex diseases. This is because variants affecting polygenic obesity only cause a small change in risk and so some relatives may carry the risk allele being studied but still be non-obese. Conversely, some relatives may not carry the risk allele being studied yet still be obese. This reduces the chances of detecting regions harbouring disease loci.

The first genome-wide linkage study specifically designed for obesity phenotypes was published in 1997 and reported a susceptibility locus on chromosome 2[218]. This was followed shortly in 1998 by a study that used nuclear families ascertained specifically for obesity. This too found a susceptibility locus on chromosome 2 and others on chromosomes 5 and 10[219]. Both the chromosome 2 locus[220,221] and the chromosome 10 locus[222,223] were replicated in subsequent reports. Since then over 250 QTLs from more than 60 linkage scans have been published with around 50 of these regions supported by two or more studies[224].The strongest evidence of linkage came from a cohort of Utah pedigrees at chromosome 4 with a logarithm of odds (LOD) score of 9.2[225]. However, despite numerous published studies, no linkage regions have been unambiguously replicated, even after meta-analysis of 37 of such studies[226]. This demonstrates the limitations of the linkage method for investigation of polygenic traits such as common obesity.

1.4.3.2 Candidate gene association studies

In a candidate gene association study, a gene is chosen based on possible functional links to the disease being investigated. Genetic markers within the gene are then measured in a set of samples after which associations to the disease are analysed. Samples can be either cases and controls or samples of a population. Associations may be detected if one of the markers analysed is a variant that directly effects disease risk or if a marker is within linkage disequilibrium (LD) of a variant that effects disease risk.

LD is a measurement of the correlation of alleles at two or more loci. It is the occurrence of sets of alleles found together in a population more often than would be expected by chance given their respective frequencies. On the other hand, when combinations of alleles can be found in their expected proportions they are said to be in linkage equilibrium. LD is usually higher between alleles in close proximity to one another due to linkage; that is, crossing over during meiosis makes it less likely that markers that are far apart will be inherited together. LD is also higher between alleles that have arisen relatively recently in evolution since there have been less opportunities for crossing over events to separate them and between alleles where there is a low rate of recombination. LD is measured using D, D' or r^2. D is calculated as the difference between the observed frequency (x) and the expected frequency (p multiplied by q) of a combination of markers (A and B):

$$D_{AB} = x_{AB} - p_A q_B$$

D suffers as a measurement because it depends on the frequencies of the alleles. Thus if an allele has a frequency of 1 or 0 then D=0. D' corrects for this problem by dividing D by the theoretical maximum of D, D_{max}:

$$D' = \frac{D}{D_{max}}$$

Finally r^2 is the correlation coefficient between a pair of alleles:

$$r^2 = \frac{D^2}{p_A p_B q_A q_B}$$

A D' or r^2 value of 0 means that the examined loci are independent of one another whereas a D' or r^2 value of 1 means that they are completely dependent. The measurement of r^2 has a more strict interpretation than that of D′ and $r^2=1.0$ only when both loci also have identical allele frequencies. Thus the allele at one locus will always be predicted by the allele at the other locus. LD is a useful tool in association studies because it allows a researcher to select a subset of genetic markers within a region of interest that can predict the genotypes of other markers within the region without having to directly measure them. These are called tagging markers or, since the most commonly used genetic marker is single nucleotide polymorphisms (SNPs), tagging SNPs.

A catalogue of SNPs and corresponding LD data is available to researchers from the International HapMap project which has genotyped several million SNPs in 269 individuals from four populations[227,228,229]. Thus tagging SNPs can be selected such that association studies can be carried out without having to genotype every SNP within a gene.

The main advantage of the candidate gene approach is the greater power to detect variants with small effect sizes that comes from using case-control cohorts when compared to linkage analysis using trios[230]. This statistical power is further increased by the fact that it is easier to recruit large numbers of the unrelated samples used in case-control analysis than it is to recruit families[230]. When candidate gene association studies were first undertaken, many associations to complex diseases, including obesity, were reported (reviewed by Rankinen *et al.*[224]). A large proportion of these, however, were not replicated suggesting that many were not real associations. This is probably because of the fact that on average, one in twenty statistical tests will result in a nominal significance of $p<0.05$ simply by chance, so failure to properly account for multiple testing meant that many associations were likely to be false-positives. Publication bias aggravated the problem since studies reporting statistical association are much more likely to be published than those that do not[231].

Another issue with early association studies is the presence of confounding factors such as ethnicity, age and sex. If cases and controls are not properly matched then differences in allele frequencies between the two groups may arise because of one of these other variables. For example if there is a difference in allele frequencies between two ethnic groups and one ethnic group is more common in the cases than in the controls then it might appear as if there is a genetic association to the disease being examined when in fact what has been detected might be a genetic association with ethnicity. This is known as population stratification[232].

It is also possible that failure to replicate associations may be down to genetic differences between study populations. This can happen if, for example, a tested SNP is in high LD with a causative SNP in one population but not in another. The tested SNP may show association in populations where it tags the causative SNP but not in populations where it doesn't. However it should be noted that this is an unlikely explanation since LD structures within European

populations have been found to be very similar when examining common variation such that genuine associations, such as those reported by GWAS have replicated well[233]. Therefore the most likely explanation for the lack of replication is that the original association is a false-positive[234].

Several genes have been reported to be associated to obesity in ten or more studies including many genes of the leptin-melanocortin system (reviewed by Rankinen *et al.*[224]), however this might reflect the fact that these genes present some of the more likely candidates so have been investigated the most. In fact, meta-analyses that have been carried out have found no evidence of association between obesity and leptin[235,236] or in leptin receptor[236,237]. Table 1.3 lists genes that are associated to obesity or BMI in meta-analyses. This includes *ENPP1,* associated after one meta-analysis[238] but not conclusively supported in a previous meta-analysis[239].

Gene	**Location**	**Gene function**	**References**
ADIPOQ	3q27.3	Adipose expressed hormone	[236]
ADRB3	8p11.23	β_3 adrenergic receptor, stimulates lipolysis[240]	[241]
ENPP1	6q22-q23	A transmembrane glycoprotein that is expressed in adipose	[238]
MC4R	18q22	Key receptor of the melanocortin system	[242]
IL1β	2q14	Proinflammatory cytokine	[236]
IL6	7p15.3	Proinflammatory cytokine	[236]
TNF-α	6p21.33	Proinflammatory cytokine	[236,243]

Table 1.3. Candidate genes with associations to obesity reported in meta-analyses.

1.4.3.3 Genome-wide association studies (GWAS)

At the beginning of this PhD project, GWAS had recently become technologically feasible. This was made possible by array based genotyping platforms developed by Affymatirx[244] and Ilumina[245] that are able to measure hundreds of thousands of genotypes in a single DNA sample at high accuracy and relatively low cost per genotype. Using LD data gathered from the HapMap project[227], SNPs can be eliminated from investigation so that a reduced number of variants can be used to capture the majority of the common variation (minor allele frequencies >5%), within the human genome.

Typically, GWAS involve a discovery phase in which hundreds of thousands of SNPs are genotyped in a subset of samples using genome-wide arrays. This is then followed by a replication stage in which the most significant markers from the discovery phase are genotyped in a second set of samples. The two datasets can then be meta-analysed together in order to report more accurate odds ratios.

The advantages of this method are that studies are not based on prior data on the function of genes thought to be involved in disease, i.e. they are hypothesis free, yet they can still use the large number of unrelated samples used in candidate gene association studies which increases the power to detect modest effect sizes.

The first obesity gene to be discovered using a genome-wide association study (GWAS) was *FTO* (fat mass and obesity associated gene) which was reported simultaneously by two research groups around the same time this project was started in 2007[246,247]. The function of FTO is still unclear but it is widely expressed in the brain, particularly the hypothalamus, suggesting a possible role in appetite regulation[248].

This was followed in 2008 by two studies that reported associations between obesity phenotypes and *MC4R* variation. One study was designed to look for associations to insulin resistance in an Indian population[249], the other investigated BMI and obesity risk in Europeans[250]. The later of these also replicated the *FTO* association.

The next obesity GWAS to be published came at the end of 2008 from the Genetic Investigation of ANthropometric Traits (GIANT) consortium, a collaboration of multiple study groups each performing genome wide scans in different cohorts[251]. Data that included over two million SNPs in almost 40,000 subjects was meta-analysed and the most significant results followed up in a further set of almost 60,000 individuals. As well as replicating the *FTO* and *MC4R* associations, six new obesity loci were reported. Published shortly afterwards, in early 2009, a meta-analysis which included around 35,000 samples from Iceland, the Netherlands, North America (European and African descent) and Scandinavia, with a follow-up of a further 35,000, reported eleven significant associations including seven novel loci[252]. This was followed by a case-control study that used 1,380 Europeans with childhood and morbid adult obesity and 1,416 age-matched normal-weight controls, followed-up in around 15,000 subjects and reported associations to three novel loci[253].

Many of the associated variants are located within or near genes that are highly expressed in the CNS and in particular, the hypothalamus and are thought to be involved in the regulation of energy balance. This supports the theory, first put forward by O'Rahilly and Farooqi, that the heritable aspects of obesity are largely neurological in origin[254].

Some time after the practical work from this PhD project was finished, in late 2010, the GIANT consortium published a new meta-analysis which included almost 250,000 subjects and discovered a further eighteen novel associations to BMI[233]. Since then there has been one meta-analysis published which analysed associations to body-fat percentage and found novel associations at two loci[255] and another which analysed associations to childhood obesity and discovered another two novel associated loci[256]. The results of these GWAS are summarised in Table 1.4.

Date published	Total number of subjects	Phenotype	Genes	Ref.
May 2007	28,587 adults	BMI	*FTO*	[246]
	10,172 children			
July 2007	6,148 relatives	BMI	*FTO*	[257]
June 2008	77,228 adults	BMI	*FTO, MC4R*	[250]
	10,583 children			
June 2008	11,955	WC	*MC4R*	[249]
June 2008	1,000 adults	BMI	*CTNNBL1*	[258]
	3,812 adults	Obesity		
December 2008	91,469	BMI	*FTO, MC4R, TMEM18, KCTD15, GNPDA2, SH2B1, MTCH2, NEGR1*	[251]
January 2009	69,593	BMI	*FTO, MC4R, TMEM18, KCTD15, SH2B1, NEGR1, SEC16B, ETV5, BDNF, AIF1 /BCDIN3D*	[252]
February 2009	8,128 adults	Obesity	*FTO, MC4R, NPC1, MAF, PTER*	[253]
	8,855 children			
April 2009	8,842	BMI, WHR	*FTO, MC4R, CTNNBL1, PTPN11*	[259]
June 2009	118,691 adults	WC, WHR	*TFAP2B, MSRA*	[260]
	47,633 women	WHR	*LYPLAL1*	
April 2010	36,581	Obesity	*FTO, MC4R, TMEM18, SDCCAG8, TNKS/MSRA*	[261]
November 2010	249,796	BMI	*RBJ/POMC, GPRC5B, MAP2K5, QPCTL/GIPR, TNNI3K, SLC39A8, FLJ35779/HMGCR, LRRN6C, TMEM160, FANCL, CADM2, PRKD1, LRP1B, PTBP2, MTIF3, ZNF608, RPL27A, NUDT3 **	[233]
November 2010	190,803	WHR (adjusted for BMI)	*RSPO3, VEGFA, TBX15-WARS2, NFE2L3, GRB14, DNM3-PIGC, ITPR2-SSPN, LY86, HOXC13, ADAMTS9, ZNRF3-KREMEN1, NISCH-STAB1, CPEB4, LYPLAL1 **	[262]
June 2011	76,202	Body fat percentage	*FTO, IRS1, SPRY2*	[255]
May 2012	25,637	Childhood obesity	*OLFM4 , HOXB2*	[256]

Table 1.4. Genes significantly associated with obesity phenotypes in genome-wide association studies. BMI = body mass index, WHR = waist to hip ratio, WC = waist circumference. * These studies also found significant associations to variants previously reported by Thorleifsson *et al*.[252] and Willer *et al*.[263]

1.4.4 Missing heritability

Between them, these associations only explain a small proportion of the heritability of obesity. Speliotes *et al.* report that together the 32 SNPs associated with BMI account for 1.45% of the inter-indivudual variation in the trait with *FTO*, the variant with the largest effect size, accounting for 0.34%. This suggests that many more genetic associations are yet to be discovered. The situation is similar with other complex diseases and there is some debate as to where this missing heritability might be found[264].

Due to the large amount of statistical tests being performed in GWAS, very stringent requirements to correct for multiple testing are required in order to reduce the false-positive rate. As such many associations of small effect could end up being screened out. For example the leptin gene has been reported to be associated to obesity in at least 13 candidate gene studies[224] yet it has not found to be significant in any genome wide association study to date. In the most recent GIANT BMI publication[233] a SNP close to *LEP* resulted in a p-value of 3.1×10^{-3}. This is clearly well below the cut-off for a significant finding in a GWAS but it is possible that there is a weak association that is mistakenly being rejected.

Furthermore, since genome wide association studies initially relied on using only markers present in the HapMap, not all variants were being investigated. Phase two of HaMap contains 3.1 million SNPs, which is estimated to represent just one quarter to one third of all common (>5%) SNPs within the human genome[229]. More recently, imputation using genotypes from the 1000 genomes project[265] combined with higher density arrays such as the Illumina HumanOmni-5 (n=4.3million SNPs) has meant that coverage has been expanded to include the vast majority of common variation. But this also means that rare variants (<5%) are not covered by current GWA studies. Genotyping only common variants is an approach that is supported by the

common disease-common variant hypothesis, which proposes that a small number common variants are responsible for the genetic component of common disease rather than large numbers of rare variants. However, this hypothesis is now coming under challenge and it is thought that rare variants may also contribute to common diseases[266].

Another issue with obesity studies is that many GWAS are not designed primarily for investigating obesity but use BMI in *post-hoc* analysis. This means that they are only reporting variants that control weight variation in the non-obese population and it is not known how much overlap there is with these and genes that predispose to the disease state of obesity. Better subject selection could improve results.

For these reasons, the candidate gene approach to investigating genetics of common polygenic obesity can still be considered valid. Additionally, other forms of heritable variation, such as copy number variants (CNVs) and epigenetics, may also have a significant influence on the development of obesity.

1.4.4.1 Copy number variation (CNV)

CNVs are defined as deletions, duplications or tandem repeats of 1kb or more in size[267]. In the last few years their importance in the pathophysiology of heritable diseases, including obesity has become increasingly recognised[268].

A large deletion at 16p11.2 has been reported in patients with obesity and cognitive deficits. Similar deletions at the same location were subsequently found to be significantly associated with common polygenic obesity in case-control analysis[269] and a duplication in this region has also been found to associate with reduced BMI, demonstrating how gene dosage can be linked

to opposite body composition phenotypes[270]. A recent study in patients with early-onset obesity found that large (>500kb), rare (<1%) deletions were present in significantly higher frequencies in obese compared to controls[271]. A genome-wide scan in European American childhood obese case-control subjects found seventeen rare CNV loci that were present in at least three cases but no controls[272]. Eight of these associations were replicated in an African American cohort. Another CNV at 10q11.22 has been reported to be associated with BMI in a Han Chinese study[273].

1.4.4.2 Epigenetics

There is growing evidence that epigenetics can contribute to the development of obesity as well[274]. Epigenetics is the study of modifications to genes that do not involve changes to the DNA sequence and include DNA methylation and histone modifications. They are associated with the control of expression of genes (reviewed by Bird[275]).

Epigenetic variation between individuals can arise in three ways. Firstly, epigenetic changes can arise due to errors during cell programming. Secondly, some epigenetic markers have been found to be heritable so variation can be passed on from one generation to the next. Finally, epigenetic markers can be altered due to environmental factors.

Epigenetic markers are mitotically heritable, which enables maintenance of tissue-specific gene expression. There is also evidence that they can be meiotically heritable, which means epigenetic markers can be inherited transgenerationally (review by Chong *et al.*[276]). The most established case of transgenerational epigenetic inheritance in mammals is imprinting in which a gene is expressed preferentially from either the maternal or the paternal allele only (reviewed by Reik *et al.*[277]). Non-imprinting epigenetic inheritance has been observed in *Agouti* (A^{vy}/a)

mice[278] in the form of a transposable element. This mouse is also a model of epigenetically controlled obesity since mice with reduced methylation at the A^{vy} allele develop obesity along with yellow fur colour. Obesity is amplified through multiple generations of A^{vy}/a mice, but supplementing the diet with methyl donors prevents this amplification[279].

There is evidence from animal studies that prenatal and early postnatal environmental factors can result in alterations of epigenetic markers. These environmental factors include nutrition[280,281], behavioural cues[282,283], chemicals[284,285], reproductive factors[286,287] and radiation[288]. Furthermore these alterations may be inherited transgenerationally[278,289,290]. A recent study in rats found that overfeeding led to hypermethylation of CpG sites in the promoter region of *POMC* in the hypothalamus. This resulted in a reduction of transcription of the gene which normally produces an anorexigenic signal to decrease food intake in response to leptin and insulin. Thus overfeeding resulted in an altered methylation pattern which subsequently modified energy regulation systems leading to an increased disposition to obesity[291].

DNA methylation is the best-characterized epigenetic mechanism. It is a modification in which a methyl group is added to the 5 position of the cytosine nucleotide creating 5-methylctosine. In mammalian, adult cells it occurs at CpG dinucleotides, although non-CpG methylation is found in embryonic stem cells[292].

As DNA is replicated during cell-division, the enzyme DNA methyltransferase 1 (DNMT1) is the principal enzyme required to copy DNA methylation patterns into the daughter strands. It does this by recognising hemi-methylated CpG sites, CpGs that are methylated on just one of the two DNA strands, and catalysing the addition of a methyl group to the unmethylated cytosine. DNA methyltransferases 3a and 3b (DNMT3a and DNMT3b) are also thought to be involved in this process, particularly in regions dense in methylated CpG sites. DNMT3a and DNMT3b are also

known to be important enzymes in the process of *de novo* methylation. This requires another homologue, DNA methyltransferase 3L (DNMT3L), which is lacking in methyltransferase activity and acts to recruit DNMT3a and DMNT3b to the DNA strand. *De novo* methylation occurs during cell differentiation in order to establish methylation patterns particular to a specific cell type (review by Jones *et al.*[293]). The DNA methylation pattern is erased in the early embryo and then re-established at approximately the time of implantation[294,295]. Differential methylation is established by *de novo* methylation by DNMT3a and DNMT3b[296] and by a means of ensuring that CpG islands remain unmethylated. This second step is thought to be mediated by active demethylation[297], although it may also involve blocking of DNMT3L by the methylated histone, H3K4 [298,299,300]. Once the methylation pattern is established in the embryo it is subject to specific alterations during development and these can include both *de novo* methylation and demethylation[299,301].

It is estimated that over 60% of all CpG sites in the human genome are methylated[302]. Methylated cytosine residues are less stable than non-methylated residues and can spontaneously deaminate to form thyamine residues[303]. This means that over the course of mammalian evolution CpG dinucleotides have become under-represented in the genome, occurring at around only 20% of the frequency expected by chance[304]. However, there are also regions of the genome that contain a much higher frequency of CpGs, probably due to a lack of cytosine methylation in these regions throughout evolution. These are called CpG islands and the promoters of around 60% of all known human genes contain such regions[305,306]. Hypermethylation in gene promoters results in transcriptional silencing of the gene which is inherited by daughter cells and is understood to be an important influence in cancer development (reviewed by Jones *et al.*[307]).

During early years (0-3 years) monozygotic twins have identical DNA methylation and histone acetylation profiles, however differences begin to appear later in life[308]. Furthermore, twin pairs who had spent less of their lifetime together were those that had the greatest differences. This demonstrates the environmental effect on DNA methylation.

A recent twin study revealed a significant difference in the DNA methylation patterns between dizygotic twins compared to monozygotic twins which cannot be explained by underlying genetic differences[309]. Methylation at 12,000 CpG sites within gene promoters were measured in white blood cells and buccal epithelial cells of 20 sets of MZ and 20 sets of DZ twins matched for age and sex. A significantly higher epigenetic difference was observed in both cell types between DZ co-twins compared to MZ co-twins. This demonstrates the heritability of DNA methylation and opens the possibility that it may contribute to complex diseases such as obesity.

The most established example of an epigenetic disorder that results in obesity in humans is Prader-Willi Syndrome (PWS), which is characterised by obesity as well as learning disabilities and hypotonia. Its genetic causes have been isolated to chromosomal region 15q11-13, which contains a group of imprinted genes which are expressed either from the paternal or maternal chromosome only (reviewed by Buiting *et al.*[310]). As well as deletions, translocations and uniparental disomy it can be caused by deficiencies in the imprinting of genes in this region of the paternal chromosome.

Another example of DNA methylation being associated with obesity comes from studies of adults who were exposed to poor nutrition *in utero* due to the Dutch famine of 1944 to 1945. These individuals have an increased prevalence of glucose intolerance, dyslipidemia, early coronary heart disease, and obesity[311,312]. In a recent study, such subjects were found to have

reduced DNA methylation levels in the imprinted gene IGF2 when compared to unexposed siblings demonstrating that environmental conditions can cause lasting epigenetic changes and the that these epigenetic changes might influence obesity risk[313]. It is thought that this is due to a decreased supply of methyl donors such as folate and methionine during development.

Genomic imprinting has also been reported to influence common obesity at three loci in a study that used genome-wide parent-of-origin linkage analysis. A maternal effect was discovered for obesity at 10p12, a maternal effect for BMI was discovered at 12q24 and a paternal effect for BMI at 13q32[314]. In another study, *TNFα* promoter CpG methylation in blood mononuclear cell DNA has been demonstrated to predict response to a weight loss diet in humans[315].These results suggest that DNA methylation and possibly other epigenetic mechanisms can influence susceptibility to common polygenic obesity.

Many studies have chosen to focus on CpG island methylation but recent evidence suggests that methylation in gene body regions may be critical to tissue-specific gene expression. DNA methylation in gene body regions has been found to correlate with increased expression rather than decreased expression that is associated with promoter methylation[316,317]. The active X chromosome has been found to be hypermethylated in gene body regions whilst being hypomethylated in gene promoter regions, supporting a role of gene-body methylation in stimulating expression[318]. Another study reported differential expression of a nitric oxide inhibitor gene due to differential methylation patterns in the gene body[319]. A more recent meta-analysis found that the relationship between gene-body methylation and expression levels follows a bell-shaped distribution such that mid-level expressed genes have the highest levels of gene-body methylation, whereas the lowest and most highly expressed sets of genes both have low levels of methylation[320].

These results support a role of epigenetic variation in the control of gene expression and therefore phenotype and disease risk that can be inherited transgenerationally. As such epigenetic variation could explain some of the missing heritability of common obesity.

1.5 Aims

The aim of this thesis was to add to the understanding of the genetic and epigenetic mechanisms that contribute to the development of common polygenic obesity in humans by studying candidate genes. When this project was started, GWAS had only just become feasible and so a candidate gene approach to studying the genetics of common disease was still considered a valid choice. As this project continued it became clear that GWAS had not elucidated all of the genes responsible for the heritable causes of obesity and other common diseases and so a candidate gene method could still potentially find associations missed by the genome-wide method.

The specific aims were:

1. Investigation of common genetic variation within the following genes for association with human polygenic obesity:

   a) Sirtuin-1 (SIRT1)(see section 1.5.1.1)

   b) Apelin (APLN)(see section 1.5.1.2)

   c) Interleukin 11 (IL11)(see section 1.5.1.3)

d) Adiponutrin (PNPLA3)(see section 1.5.1.4)

e) Nesfatin (NUCB2)(see section 1.5.1.5)

f) Insulin Receptor Substrate 1 (IRS1)(see section 1.5.1.6)

2. Investigation of epigenetic variation in the Leptin gene for association to human polygenic obesity (see section 1.5.2).

1.5.1 Candidate gene association studies

Subjects used were 1533 obese cases and 1237 non-obese controls, all French Caucasians. Cases were 896 obese adults and 637 obese children. These were analysed separately but used the same control samples and as such are not independent data sets used for replicating results. Additional subjects included 732 Swedish Caucasians from 154 families that contained siblings discordant for obesity and for which genome-wide transcription data within adipose tissue was available (see Materials and Methods for details). SNPs that covered all common variation in each candidate gene were genotyped using Sequenom iPlex assays. It is important to note that while rare variants have been highlighted as a potential source of genetic variation that could contribute to obesity risk, this project only investigated common variation due to statistical power limitations. Each candidate gene was chosen based on having functional links to appetite and metabolism. Descriptions of the selected genes are given below.

1.5.1.1 Sirtuin-1 (SIRT1)

Sirtuin 1 is a nicotinamide adenine dinucleotide (NAD^+)-dependent protein deacetylase that regulates transcription of a variety of genes including a number that have been implicated in the control of metabolism in humans. As SIRT1 activity is under the control of cellular NAD^+ concentration it acts as an energy sensing molecule that alters transcription of certain genes in response to the energy status of the cell[321].

Most notably, SIRT1 inhibits transcription of PPARγ, thereby promoting lipolysis and mobilisation of free fatty acids from WAT as well as inhibiting adipogenesis[322]. Inhibition of SIRT1 in 3T3 L1 adipocyte cell culture increases differentiation and overexpression of SIRT1 reduces it[322]. It also stimulates mitochondrial activity and thus increases glucose metabolism in the liver through the transcriptional regulator PGC-1α[323]. SIRT1 has also been shown to deacetylate the transcription factor forkhead box protein O1 (FoxO1), a key regulator of lipid metabolism[324]. In pancreatic β cells SIRT1 enhances insulin secretion in response to glucose via regulation of uncoupling protein 2[325]. SIRT1 has also been discovered to alter expression of adiponectin, although the two studies reported opposite effects[326,327]. Another mechanism by which SIRT1 is linked to obesity is an association with lower levels of nuclear factor-κB (NF-κB) activity. NF-κB is a key regulator of many genes involved in inflammation and its inhibition in mouse models have been shown to protect from diabetes and fatty liver[328,329].

Transgenic mice that overexpress SIRT1 weigh less than control mice, have a lower WAT mass and have lower plasma levels of free fatty acids, leptin and adiponectin[330]. They are also protected from glucose intolerance induced by ageing, diet-induced obesity and genetically induced obesity[324,328]. Conversely, SIRT1 knockout mice have increased body weight, increased fatty liver disease, increased insulin resistance, increased glucose intolerance and inflammation in WAT and liver[331,332].

Activators of *SIRT1* including resveratrol protect against obesity and insulin resistance in mice fed a high fat diet[333,334,335,336]. These activators increase levels of metabolism and lipolysis via transcriptional activation of PPARγ and PCG1α. Resveratrol has also been reported to increase lipolysis in human adipose tissue culture and this is mediated via SIRT1[337].

1.5.1.2 Apelin (APLN)

The *APLN* gene is expressed in adipocytes in both mice and humans and its transcription is inhibited by fasting[338]. Expression of *APLN* in subcutaneous adipose tissue has been reported to be significantly elevated in mice and rats fed a high fat diet and has been correlated with serum leptin levels[338,339,340]. Intravenous injection of apelin stimulates glucose utilisation in muscle and adipose tissue in mice and lowers blood glucose levels[341]. Genetically engineered apelin-knockout mice have reduced insulin sensitivity and are hyperinsulinemic[342]. Administration of pyroglutamated apelin-13 was found to improve insulin sensitivity in both apelin-deficient mice and *db/db* diabetic mice[342].

The apelin receptor, *APJ*, is expressed in hypothalamic neurons in rats[343,344] and injection of apelin-13 and apelin-12 into the brains of rats has been reported to decrease food intake[345,346]. Treatment with apelin has been reported to lower the mass of WAT in both non-obese and obese mice (on average a ~30% reduction in epididymal WAT mass was reported in both non-obese and obese mice) as well as lowering serum levels of insulin and triglycerides[347]. Apelin treatment has been reported to increase expression of *UCP1* in BAT, a gene involved in fatty acid activated uncoupling of respiration[347,348]. In the same report, apelin treatment also increased expression of *UCP3* in skeletal muscle[347]. UCP3 regulates fatty acid export and genetic variants in its gene have been reported to be putatively associated with obesity, although these associations have not been confirmed in any GWAS to-date and so have not been robustly validated[349,350].

Plasma apelin levels have been reported to be significantly increased in obese subjects compared to non-obese controls and were significantly correlated with BMI[351]. A similar result has been observed between in subjects with T2D or impared fasting glucose and this association is caused by a correlation of apelin levels to BMI[339,352]. In insulin-resistant high-fat fed mice, apelin and APJ expression in adipose tissue was significantly increased compared to controls[339]. Another study reported a correlation of plasma apelin levels with BMI in normal and underweight individuals and this correlation was even more pronounced in obese subjects[353].

1.5.1.3 Interleukin 11 (IL-11)

IL-11 has also been reported to inhibit adipogenesis and suppress lipoprotein lipase activity in mouse 3T3-L1 preadipocyte cell culture[354,355,356]. Treatment of fully differentiated mouse 3T3-L1 adipocytes with IL-11 has been found to lower transcription of leptin[357]. In human bone marrow tissue cultures, IL-11 inhibits fat accumulation in the adherent cells via a decrease in the number of vacuoles in macrophages and inhibition of preadipocyte differentiation[358],[359].

1.5.1.4 Adiponutrin (PNPLA3)

Adiponutrin, also known as Patatin-like phospholipase domain-containing protein 3 (PNPLA3) is a cell membrane protein that is expressed in adipose cells[360]. It is a triacylglycerol lipase that has been shown to catalyse the hydrolysis of triglycerides in vitro[361,362]. In addition to adipocytes, it is highly expressed in liver and transcription of the gene appears to be connected to liver fat content[363,364,365,366]. Expression of *PNPLA3* is nutritionally controlled. In mice, adipose transcription levels are reported to be downregulated during fasting and upregulated when fasted mice are fed a high-calorie diet[360,361]. Furthermore, expression of the gene is fifty-fold higher in obese compared to normal rats[360].

In obese women, a three-week very low calorie diet has been reported to significantly reduce expression of the gene in subcutaneous adipose tissue[367]. A clamp study in non-obese, non-diabetic humans found that subcutaneous adiponutrin expression was induced by both insulin and glucose[368]. Expression of Adiponutrin in subcutaneous and visceral adipose tissue has also been reported to be increased in obese compared to non-obese subjects in a Swedish study[369].

1.5.1.5 Nesfatin-1 (NUCB2)

In the brain, nesfatin-1 protein is found in several areas of the hypothalamus including the PVN, the supraoptic nucleus, the ARC and the LH as well as the nucleus of the solitary tract, a brainstem area with known roles in feeding[370,371,372,373]. Neurons in which nesfatin-1 is located have also been found to express MCH, CART, α-MSH, POMC and NPY supporting a role for nesfatin-1 in appetite control [370,374,371,375]. Additionally nesfatin-1 has been shown to inhibit NPY neurons[376].

Injection of nesfatin-1 into the brains of both rats and mice has been shown to reduce feeding consistent with a function as an anorexigenic regulator of food intake[372,377]. Furthermore, knockdown of *NUCB2* in the rat hypothalamus using an antisense oligonucleotide increases food intake and body weight[372]. The nesfatin-1 protein can cross the blood-brain barrier in both directions[378,379] and intraperitoneal injection of nesfatin-1 has been reported to lower food intake in mice[380].

NUCB2 transcription in the rat hypothalamus is significantly reduced after twenty-four hours fasting and this is restored after re-feeding[372,373]. In humans, a putative association has been reported between fasting levels of plasma nesfatin-1 and obesity in male subjects and BMI in the non-obese group[381]. A recent study using female anorexic patients reported a significant increase in circulating nesfatin-1 compared to age-matched healthy controls[382].

1.5.1.6 Insulin Receptor Substrate 1 (IRS-1)

IRS-1 is an intracellular signalling molecule that is expressed in all insulin-sensitive tissues[383,384]. It is a key molecule in the signal cascade, critical to energy regulation, that is triggered when insulin binds to its receptor. In addition to its role in insulin signalling, IRS-1 is also a substrate for insulin-like growth factor 1 (IGF-1) receptor which is thought to be an important regulator of cell differentiation[384]. IRS-1 stimulation in response to insulin and IGF-1 leads to uptake of glucose by muscle and fat tissue and inhibition of hepatic glucose production. It also stimulates cell differentiation and growth, promotes lipogenesis and inhibits lipolysis[385].

IRS-1 knockout mice are hyperinsulinemic and show mild insulin resistance but do not have altered glucose tolerance[384]. Interestingly body weight in knockout mice is reduced by 50% suggesting a role of the gene in somatic growth. Furthermore, adipocyte cell lines from these mice were found to have a significant reduction in their differentiation and lipid accumulation compared to wild type suggesting a role in controlling adiposity[386].

Variants within the *IRS-1* gene have be reported to influence T2D risk[387] including a recent GWAS which found associations between rs2943641 and T2D, insulin resistance and hyperinsulinemia[388]. There have also been associations reported between *IRS-1* SNPs and obesity and since insulin levels are correlated with BMI, dysregulation of the insulin signalling pathways could be a factor in the development of obesity[387,389]. More recently, a GWAS found a significant association between a variant close to *IRS-1* and body fat percentage[255].

1.5.2 Epigenetic variation in the Leptin gene (LEP)

Methylation in the *LEP* CpG island has been shown to reduce activity of the promoter in human adipose cell lines using luciferase assays[390]. The *LEP* CpG island is differentially methylated across tissue types, which could indicate a role of methylation in the regulation of this gene[391,392].

Rats that are fed a high fat diet and become overweight have been found have increased methylation at one CpG site in the leptin promoter when DNA from adipose was examined. This methylation was associated with lower blood levels of leptin[393]. These results indicate the possibility of a role of DNA methylation in the *LEP* gene in the regulation of BMI.

Chapter 2

Materials and Methods

2.1 Subjects

2.1.1 Obese cases

Two groups of case subjects were used. The first set consisted of 896 unrelated morbidly obese (BMI over 40kg/m^2) adults (mean BMI = 47.5kg/m^2±7.5kg/m^2; mean age = 44.3±11.9; 689 females and 207 males). The second set consisted of 637 unrelated obese children (BMI over 97th percentile for age and sex) (mean BMI = 29.6kg/m^2±6.5kg/m^2; mean zBMI=4.27±1.2; mean age =11±3.2; 341 females and 296 males). These were all French Caucasians recruited through a multimedia campaign run by the Centre National de la Recherche Scientifique (CNRS), Hotel Dieu Hospital, the Pasteur Institute, Lille and the Department of Paediatric Endocrinology of Jeanne de Flandres Hospital. Ethical approval for the studies was given by local ethical committees.

2.1.2 Non-obese controls

Two groups of control subjects were used, both unrelated non-obese French Caucasians. The first contained 532 individuals from the Haguenau cohort (mean BMI = 21.3 kg/m^2 ±2.0 kg/m^2; mean age 22.7 ±3.5; 243 males, 289 females)[394] The second was 705 subjects from the Epidemiology Study on Insulin Resistance (D.E.S.I.R)(mean BMI = 23.3 ±1.8 kg/m^2; mean age 53.9 ±5.6; 282 males, 423 females)[395]. Control participants were pooled for case control analyses, but prior to pooling χ^2 tests for each SNP were performed to make sure there was no significant difference ($p>0.05$) in genotype or allele frequencies between the control groups. BMI was calculated as the subject's weight (in kilograms) divided by the square of their height (in metres).

2.1.3 Families

154 nuclear families containing 732 Swedish Caucasians were identified as part of the Swedish Obesity Study cohort (SOS)[396] on the basis of an obese proband with at least one sibling discordant for obesity (BMI difference >10kg/m^2). Mean BMI in the lean siblings was 23.4±2.8, mean BMI in the obese siblings was 37.7±5.3. Siblings included 132 males and 292 females.

Gene expression data from subcutaneous adipose tissue was available for 359 out of 424 of the siblings[396]. This data was obtained from an eQTL study which used an Affymetrix Human Genome U133 Plus 2.0 gene expression array to measure genome wide transcript levels and had been normalised using the Robust Multiarray Average method[396].

2.2 Protocols

2.2.1 SNP identification strategy

In all cases where a tag SNP approach was used, tagging SNPs were selected from the HapMap database (http://www.hapmap.org) with $r^2 \geq 0.8$ and a minor allele frequency ≥0.05 using the pairwise tagger algorithm in the Haploview 4.0 software[397]. Additional SNPs not present in the HapMap were selected from NCBI database, each with a reported minor allele frequency ≥0.05. Coding SNPs were included regardless of frequencies.

2.2.2 Genotyping

Genotyping was carried out on the Sequenom MassArray platform (Sequenom Inc., USA)[398]. PCR and extension primers were designed using MassARRAY Assay Design 3.1 software (Sequenom Inc. USA). For a multiplex PCR, 2µl (2.5ng/µl) of each DNA sample were mixed with 2.18µl H_2O, 0.5µl Hot Star Buffer (Qiagen Ltd.), 0.2µl $MgCl_2$ (25mM), 0.02µl Hot Star Taq (5 units/µl), 0.1µl of 10mM dNTP mix (Bioline Ltd., UK) and 0.5µl of 1µM primer mix (See appendix for primer sequences). Reaction conditions were as follows: 95°C for 15 minutes, then five cycles of 95°C for 20s, 65°C for 30s and 72°C for 1 min, followed by five cycles of 95°C for 20s, 58°C for 30s and 72°C for 1 minute, followed by 38 cycles of 95°C for 20s, 53°C for 30s and 72°C for 1 minute, ending with 72°C for 3 minutes. To dephosphorylate dNTPs and any remaining primers, 1.53 µl of dH_2O, 0.17µl SAP Buffer (Sequenom) and 0.30µl shrimp alkaline phosphatase (SAP) enzyme (Sequenom) was added to each reaction. This was incubated using a thermocycler at 37°C for 40 minutes followed by 85°C for 5 minutes. 0.74µl of dH_2O, 0.2µl iPLEX Buffer Plus (Sequenom), 0.2µl iPLEX Termination Mix (Sequenom), 0.041 µl iPLEX enzyme (Sequenom) were added to each reaction. Extension reactions were cycled as follows: 94°C for 30 seconds followed by 40 cycles of 94°C for 5 seconds and 5 nested cycles of 52°C for 5 seconds and 80°C for 5 seconds, ending with a final incubation of 72°C for 3 minutes.

For each reaction, 15–25nl of sample was dispensed onto a SpectroCHIP using the Samsung Nano-dispenser and the chip was then analysed using the Bruker Autoflex MALDI-TOF mass spectrometer. Mass Array Typer 3.4 software was then used to call the genotypes based on the calculated mass of the extension products and the experimental mass-spectrum result. Automated genotype calls were manually checked.

Genotyping was considered satisfactory if the success rate for the SNP was ≥85% and the genotype distribution did not depart significantly from Hardy-Weinberg equilibrium ($p>0.05$ for chi-squared test between expected and observed values). Hardy-Weinberg p-values were calculated using the online Finetti HWE calculator (http://ihg.gsf.de/cgi-bin/hw/hwa1.pl).

2.2.3 PCR

Primers were designed using Primer3 software (http://frodo.wi.mit.edu/). A standard PCR was carried out as follows (see Appendix for individual primer sequences and conditions). 5.2µl H_2O, 1µl Amplitaq Gold Buffer (Applied Biosystems, Life Technologies Inc., USA), 0.8µl dNTPs (10mM), 0.8µl $MgCl_2$ (25mM), 1µl Primer Mix (5µM of each primer), 0.2µl AmpliTaq Gold (Applied Biosystems)(5U/µl) was mixed with 1µl of template DNA. PCR was then carried out at 94°C for 10 minutes followed by 40 cycles of 94°C for 30 seconds, 50°C for 30 seconds and 72°C for 30 seconds ending with an incubation at 72°C for 5 minutes. Annealing temperatures and extension times were altered based on requirements of the specific amplicon (See Appendix for details). GC-rich amplicons often required the addition of 2µl Q Solution (Qiagen) in a 10µl PCR (see Appendix for amplicons that required Q Solution).

2.2.4 Agarose gel electrophoresis

To confirm the presence of the correct sized PCR product, 5µl was mixed with 1µl 6X loading buffer (Fermentas GmbH, Germany). A 150ml 2.5% (w/v) agarose (Invitrogen) gel was made up with 1X Tris Borate EDTA buffer (Sigma-Aldrich Inc.) and 7.5 µl Safeview (NBS Biologicals Ltd., UK). A 100bp ladder (Fermentas GmbH, Germany) was loaded alongside samples as a size marker. DNA was viewed using a UV transilluminator and photographed using LabWorks Image Acquisition Software (Ultra Violet Products Ltd, UK).

2.2.5 Sequencing of PCR product

5µl of PCR product was purified by mixing with 2µl ExoSAPit (USB) and placed in a thermocycler at 37°C for 15 minutes followed by 80°C for 15 minutes. This was then diluted fivefold by adding 43µl dH_2O. The concentration of the PCR product was measured using a NanoDrop ND-100 spectrometer (NanoDrop Technologies, USA). 3.2pmol of sequencing primer was mixed with 90ng/kb PCR product and made up to 10µl with dH_2O. Samples were processed by the Imperial College Clincial Sciences Centre Genomics Core Laboratory (Hammersmith Campus) using the Big Dye terminator mix v3.1 and capillary gel electrophoresis was carried out to separate the sequencing products on a 3730XL DNA Analyzer (Applied Biosystems).

2.2.6 CpG island identification and primer design

Putative CpG islands were identified using Methyl Primer Express version 1.0 software (Applied Biosystems) and CpGPlot was used to display C+G composition, CpG frequency and position of putative CpG islands (http://www.ebi.ac.uk/Tools/emboss/cpgplot/)[305]. CpG islands were defined as being at least 200bp in length with C+G composition>50% and a ratio of observed to expected CpG dinucleotides >0.6[399]. Primers were designed to PCR within the CpG islands of bisulfite treated DNA using Methyl Primer Express. Sequencing primers were designed to sequence the PCR product using the online software Primer3 (http://frodo.wi.mit.edu/).

2.2.7 Bisulfite conversion of DNA

Bisulfite conversion was carried out using the MethylDetector kit (Active Motif Inc., USA). Conversion Buffer was prepared by resuspending Conversion Reagent with 700µl dH_2O, and then mixing with 350µl Buffer A and 175µl Denaturation Reagent at room temperature for 10 minutes with intermittent vortexing. Hydroquinone was prepared by resuspending in 100µl

dH_2O. Desulfonation Buffer was prepared by combining 22µl Buffer B with 88µl dH_2O and 110µl 100% isopropanol. These three buffers were prepared fresh to be used on the same day. 200ng of genomic DNA was mixed with 7µl Hydroquinone and 120µl Conversion Buffer and made up to 140µl with dH_2O. This was placed in a thermocycler for three minutes at 94°C followed by nine hours at 50°C.

500 µl DNA Binding Buffer was mixed with the converted DNA which was then pipetted into a DNA purification column. This was spun at 10,000 rpm for 30 seconds in a microcentrifuge (Centrifuge 5415 C, Eppendorf Ltd., UK), and the collection tube emptied. 200µl DNA Wash Buffer was added to the column which was then span at 10,000 rpm for 30 seconds. 200µl Desulfonation Buffer was added to the column and incubated at room temperature for 20 minutes before spinning at 10,000rpm for 30 seconds. 200µl of DNA Wash Buffer was then added and the column spun at 10,000rpm for 30 seconds. The column was then placed in a new 1.5ml microcentrifuge tube and 50 µl of DNA Elution buffer was pipetted onto the surface of the column filter. This was incubated at room temperature for 3 minutes before spinning down at 10,000rpm for 30 seconds. The eluted solution was the purified converted DNA ready for PCR.

2.2.8 Bisulfite conversion of DNA using Qiagen EpiTect 96-well format

9 ml of dH_2O was added to Bisulfite Mix (Qiagen) and mixed by vortexing until fully dissolved. 2 µl DNA (5 ng/µl) was added to each well of an EpiTect Conversion Plate (Qiagen). 38 µl dH_2O, 85 µl Bisulfite Mix and 15 µl of DNA Protect Buffer (Qiagen) were added to each well and mixed by vortexing. The plate was span in a centrifuge (Sorvall Legend RT, Thermo Scientific) at 650 g to collect the liquid at the bottom of the wells. The conversion reaction was placed in a thermal cycler using the following programme: 95°C for 5 minutes, 60°C for 25 minutes, 95°C for 5 minutes, 60°C for 1 hour and 25 minutes, 95°C for 5 minutes and finally 60°C for 2 hours and 55 minutes.

The converted DNA was spun in a centrifuge at 650 g to collect the liquid at the bottom of the wells. An EpiTect 96 Plate (Qiagen) was placed onto a QIAvac vacuum manifold (Qiagen) and 560 µl freshly prepared Buffer BL containing 10 µg/ml carrier RNA (Qiagen) was pipetted into each well of the EpiTect 96 Plate. The converted DNA was then added to the EpiTect 96 Plate and mixed with the Buffer BL by pipetting 4 times. The vacuum source was switched on and once the liquid had passed through the membrane, switched off. 500 µl of Buffer BW (Qiagen) was added to each well of the EpiTect 96 Plate. The vacuum source was switched on and once the liquid had passed through the membrane, switched off. 250 µl of Buffer BD (Qiagen) was added to each well of the EpiTect 96 Plate and incubated at room temperature (15–25°C) for 15 minutes. The vacuum source was switched on and once the liquid had drained through the membrane, switched off. 500 µl of Buffer BW (Qiagen) was added to each well of the EpiTect 96 Plate. The vacuum source was switched on and once the liquid had passed through the membrane, switched off. 500 µl of Buffer BW (Qiagen) was added to each well of the EpiTect 96 Plate. The vacuum source was switched on and once the liquid had passed through the membrane, switched off. 250 µl of ethanol (100% v/v, Sigma-Aldrich, Inc.) was added to each well of the EpiTect 96 Plate. The vacuum source was switched on and left running for 10 minutes after the ethanol had passed through the membrane. The EpiTect 96 Plate was removed from the vacuum manifold and tapped onto absorbent paper in order to remove any remaining ethanol. The waste tray of the vacuum manifold was removed and replaced with an EpiTect Elution Plate (Qiagen). 70 µl of Buffer EB (Qiagen) and 10 µl of Top Elute Fluid (Qiagen) were dispensed into each well of the EpiTect Plate. The vacuum source was switched on for 1 minute to elute the DNA.

2.2.9 Positive control nested PCR

This was performed to confirm that the bisulfite conversion was successful and included two reactions: an outer PCR followed by an inner PCR. 4.6 µl H_2O, 1 µl Amplitaq Gold Buffer, 1 µl dNTPs (10mM), 1.2 µl $MgCl_2$ (25mM), 1 µl p16 Outer Primer Mix (10µM), 0.2 µl AmpliTaq Gold (5U/ µl) (Applied Biosystems) was mixed with 1 µl of converted DNA. 10ng of unconverted DNA was used as a negative control. PCR was then carried out at 94°C for 10 minutes followed by 40 cycles of 94°C for 30 seconds, 50°C for 30 seconds and 72°C for 30 seconds and ending with 72°C for 5 minutes.

This PCR reaction was then diluted 500-fold and used as a template in the following inner PCR reaction. 4.6 µl H_2O, 1 µl Amplitaq Gold Buffer, 1 µl dNTPs (10mM), 1.2 µl $MgCl_2$ (25mM), 1 µl p16 Inner Primer Mix (10µM), 0.2 µl AmpliTaq Gold (5U/ µl) was mixed with 1 µl of diluted outer PCR product. 10ng of unconverted DNA was used as a negative control. PCR was then carried out at 94°C for 10 minutes followed by 40 cycles of 94°C for 30 seconds, 50°C for 30 seconds and 72°C for 30 seconds and ending with 72°C for 5 minutes. The 10 µl PCR reaction was then mixed with 2 µl 6x DNA loading buffer and the sample loaded onto a 2.5%(w/v) agarose gel in 1x TBE. The gel was run until the dye front was near the end of the gel.

2.3 Statistical Analysis

2.3.1 Qualitative trait analysis

Analysis was performed using PLINK v1.07 (http://pngu.mgh.harvard.edu/purcell/plink/)[400]. Association to obesity was analysed by comparing allele frequencies using the chi-squared test and one million permutations to give accurate empirical p-values that did not need correction for multiple testing. Odds ratios are reported with a 95% confidence interval.

2.3.2 Quantitative trait analysis

Analysis for association to BMI was performed with PLINK v1.07 using the linear model with age and sex as covariates. One million permutations were used to correct for false-positives.

2.3.3 Family-based association analysis

Mendelian errors were detected using Pedstats[401] while genotypes that resulted in tight double recombinants were identified with Merlin and treated as missing data in the analyses[402]. Association to BMI was analysed using the QFAM procedure in PLINK v1.07 using the within-families model of association and one million permutations to correct for multiple testing.

2.3.4 Linkage disequilibrium (LD)

LD was calculated and displayed from the genotype data of the controls using Haploview v4.0[397]. Linkage disequilibrium blocks were identified using the confidence interval method described by Gabriel *et al.*[403] as implemented in Haploview 4.0. Haplotypes were constructed and analysed for association using PLINK v1.07. Individuals who had missing genotypes were excluded from this analysis.

2.3.5 Imputation

Genotypes were imputed using the 1000 genomes phase 1 integrated variant set release (NCBI build 37)[265]. IMPUTE2 software was used for imputation[404,405] and SNPTEST v1.1.5 for analysing the imputed data for association[406].

2.3.6 Analysis of transcription data

To correct for relatedness between siblings, regressions were evaluated with a clustering option within families using R v2.8.1 (http://www.r-project.org/; see Appendix for R commands used). Data was analysed for normality using a Q-Q plot in SPSS v17.0. An independent samples t-test was used to compare transcript levels between obese and non-obese siblings and data was displayed with box plots using R v2.8.1 (see Appendix for R commands used). BMI and transcript data were plotted using Excel 2007 (Microsoft Inc.). All r^2 calculations and their corresponding p-values were calculated using R v2.8.1 (see Appendix for R commands used).

2.3.7 Statistical power calculations

Power calculations were carried out using PSv3.0.14[407]. Assuming an alpha of 0.005, (the p-value threshold that would be used if testing ten independent SNPS) a MAF of 0.5 and an odds ratio of 1.2 then the null hypothesis that the odds ratio equals 1 can be rejected with a probability (i.e. power) of 23% in the French adult case-control subjects (see Figure 2.1) and 17% in the French child case-control subjects (see Figure 2.2). With a MAF of 0.05 this power is reduced to 3.3% in the adults and 2.9% in the children (see Figures 2.1 and 2.2). Statistical power will drop substantially if using a significance level of 5×10^{-8} (genome-wide significance).

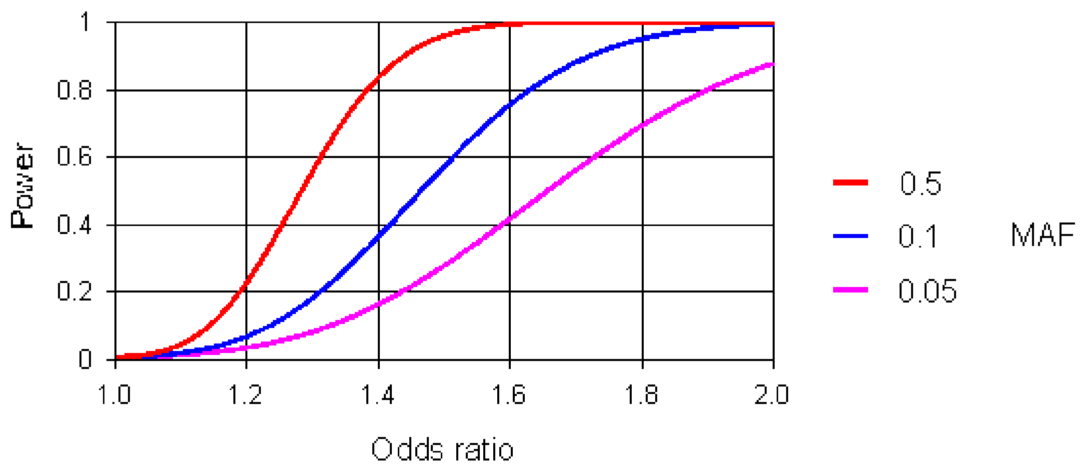

Figure 2.1. Statistical power using adult case-control subjects (896 cases and 1237 controls) plotted against odds ratio for three different minor allele frequencies (MAF) assuming an alpha of 0.005.

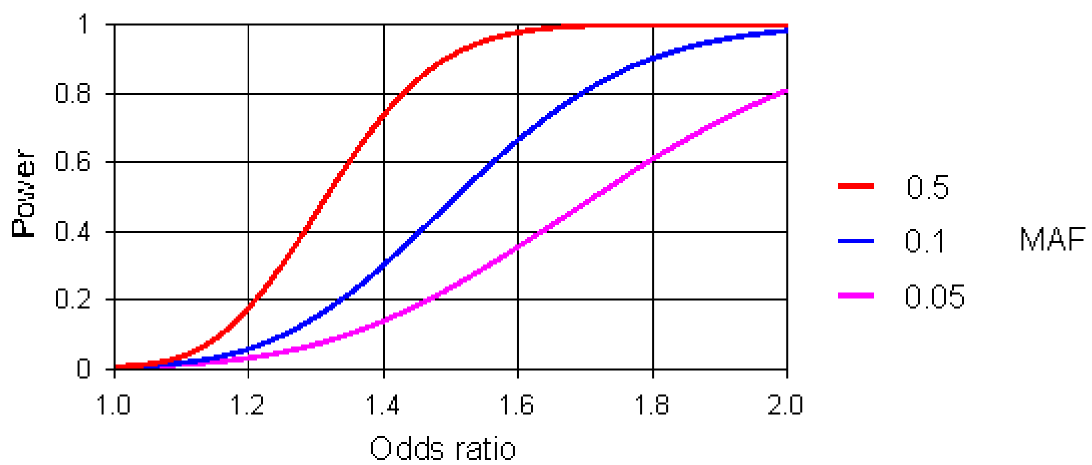

Figure 2.2. Statistical power using child case-control subjects (637 cases and 1237 controls) plotted against odds ratio for three different minor allele frequencies (MAF) assuming an alpha of 0.005.

For quantatitive triat analysis in the French controls, assuming a statistical significance of 0.005 (ten tests), a MAF of 0.5, a difference in mean BMI between alleles of 0.5 and an SD of 2.1 (as observed for BMI in all 1237 French controls) power is calculated to be 87%. With a MAF of 0.1 this drops to 34% and for MAF of 0.05, 15% (see Figure 2.3).

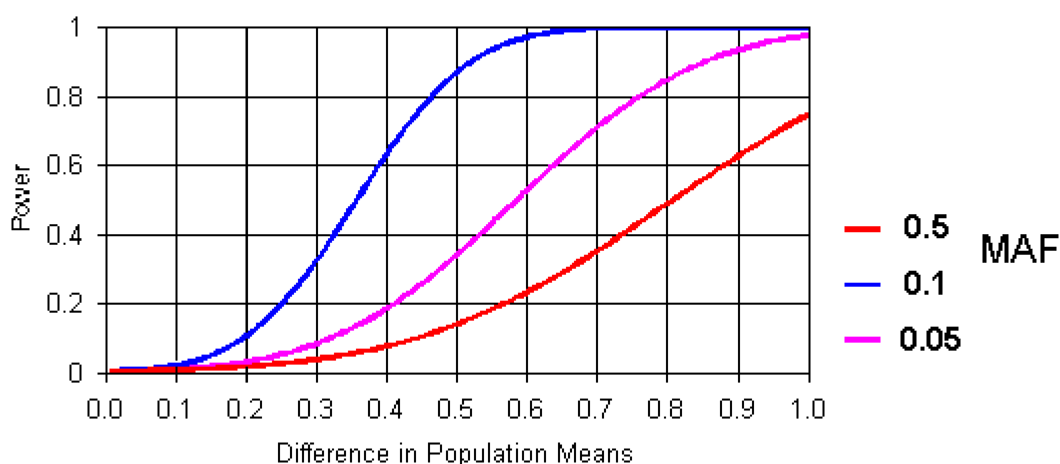

Figure 2.3. Statistical power for analysing BMI in the French controls (n=1237). Difference in population means is measured in BMI units. This assumes a SD of 2.1 (as observed for BMI in all 1237 French controls) and a statistical significance of 0.005 (ten tests).

Power calculations specific for each gene investigated in this thesis appear in their respective results chapters.

2.3.8 Analysis of sequence data

Sequence data was analysed using SeqScape v2.0 (Applied Biosystems). Reference sequences were imported in Genbank format from the NCBI website (http://www.ncbi.nlm.nih.gov/gene). Sequencing data was then aligned with the reference sequence to locate variants within the test sequences.

2.3.9 Interpretation of bisulfite sequence data

Sequences were viewed and peak heights measured using Sequence Scanner software version 1.0 (Applied Biosystems). Peak heights of cytosine residues within CpG sites were measured as a percentage of cytosine at that site (i.e. the height of the cytosine peak was divided by the sum of the height of the cytosine and thymine). This figure gives a conversion rate and indicates the amount of methylated DNA copies in a sample at that particular CpG site. When the reverse strand was sequenced, percentage of guanine to adenine was measured.

For quality control purposes two non-CpG cytosines (guanines in reverse strands) were measured and if either site was less than 95% converted the sequence was rejected.

2.3.10 Statistical analysis of CpG methylation data

Percentage methylation was treated as a quantitative trait and analysed for association to obesity using an independent samples t-test implemented by SPSS version 17.0 (SPSS Inc.). It was also analysed for association with BMI corrected for age and sex using linear regression implemented in R v2.8.1 (see Appendix for R commands).

Chapter 3

Investigation of genetic variants within the Sirtuin-1 gene for association to common polygenic obesity

3.1 Introduction

SIRT1 is the mammalian form of the yeast protein, silent information regulator 2 (SIR2)[408], a protein known to be crucial for the lifespan extending effects of caloric restriction in lower organisms[409]. The SIR2 family of enzymes catalyse deacetylation of acetylated lysine residues of certain proteins in a reaction that is dependent on the presence of nicotinamide adenine dinucleotide (NAD+)[410].

SIRT1 is localised to the nucleus and in particular, is associated with euchromatin[411]. It facilitates the formation of heterochromatin, the tightly formed chromatin that is associated with histone hypoacetylation and gene repression[412]. SIRT1 was originally thought to be a histone-specific deacetylase as the first targets to be identified were the histones H1, H3 and H4[412]. However, a number of non-histone targets of SIRT1 have since been identified including the transcription factors, nuclear factor kappa-light-chain-enhancer of activated B cells (NF-KB)[413], Forkhead box 1 (FOXO1)[414], FOXO3a[415] and p53[416,417]. Many of these transcription factor targets are involved in the regulation of energy metabolism.

3.1.1 SIRT1 regulates energy metabolism

SIRT1 stimulates lipid mobilisation and inhibits adipogenesis in WAT by repressing activity of the adipogenesis factor PPARg[418]. In skeletal muscle, SIRT1 stimulates fatty acid oxidation by activating PPARa and PPARg co-activator 1 alpha (PGC-1a)[419]. In the liver, SIRT1 suppresses glycolysis and stimulates hepatic glucose output and this also occurs *via* stimulation of PGC-1a[420]. It also suppresses lipid synthesis and stimulates lipid oxidation[421,422]. SIRT1 also activates

the mitochondrial enzymes acetyl-CoA synthetase 1 and 2[423,424] which is thought to result in reduction of the rate of fatty acid synthesis. Therefore SIRT1 acts in multiple tissue types to stimulate energy expenditure and promote a reduction in stored triglycerides thus potentially influencing the fat mass of the individual.

SIRT1 has been also found to act on insulin receptor substrate-1 (IRS1) and IRS2 indicating a possible role in the insulin signalling pathway[425]. Furthermore, in pancreatic β-cells, SIRT1 suppresses activity of UCP2 which results in reduced uncoupling of mitochondrial ATP production and an increased yield of ATP from glucose oxidation and ultimately stimulation of insulin secretion[426]. This means that SIRT1 can potentially regulate the uptake of glucose and lipids from the blood stream as well as appetite and energy expenditure in the hypothalamus.

3.1.2 SIRT1 activity is regulated by metabolic factors

SIRT1 activity is dependent on NAD+ which links it to the metabolic state of the cell[409]. SIRT1 converts NAD+ to nicotinamide, and nicotinamide can non-competitively bind with and inhibit SIRT1 activity thus providing negative feedback control to this mechanism[427]. NAD+ levels have been shown to rise in muscle, liver and white adipose tissue (WAT) during fasting, caloric restriction and exercise[428,429]. Conversely, a high-fat diet in mice reduces cellular NAD+ levels[430]. This then provides one mechanism by which energy levels can regulate SIRT1 activity.

As well as activity of the SIRT1 enzyme, its expression has also been found to be upregulated by fasting in fat, muscle, and brain tissues[431,432,433]. Additionally, *SIRT1* is downregulated in mice fed a high-fat diet[434]. Nutrient withdrawal appears to stimulate expression of *SIRT1 via* the transcription factors FOXO3a and p53[435] suggesting a positive feedback loop since SIRT1

regulates activity of these same transcription factors. PPARg has been reported to repress SIRT1 expression, which would suggest a negative feedback loop[436]. SIRT1 expression in the liver, however, appears to be regulated differently; in mice *SIRT1* is reduced by caloric restriction and stimulated by a high-calorie diet[433].

3.1.3 SIRT1 mouse models

Over-expression of the *SIRT1* gene in mice results in inhibition of adipogenesis, enhanced lipolysis and release of free fatty acids and this has been shown to occur *via* repression of PPARg[418]. Transgenic mice that over-express *SIRT1* weigh less than control mice and have a lower WAT mass[330]. Specific over-expression of SIRT1 in pancreatic β-cells, leads to enhanced glucose-stimulated insulin secretion and ATP production *via* repression of UCP2[437]. In *SIRT1* knockouts, the resulting mice have a higher body weight than controls and exhibit increased insulin resistance, glucose intolerance and inflammation in both WAT and liver[332,421]. These mouse models, therefore suggest an important role for the *SIRT1* gene in ensuring proper regulation of energy balance and body-weight.

3.1.4 SIRT1 as a pharmacological target for treating metabolic disease

Resveratrol, a naturally occurring chemical found in grapes and red wines, activates SIRT1 and its administration to several species of invertebrate results in extended life span[438,439]. In mice, resveratrol has been reported to improve health and survival in animals fed a high-calorie diet[440] and in rats it has been shown to reduce the symptoms of diabetes[441]. In diet-induced obese

mice, daily oral dosing of resvertrol has been shown to reduce blood glucose levels and improve insulin sensitivity[419]. This result has been confirmed in *fa/fa* rats with both resveretrol and synthetic SIRT1 activators[334]. As such SIRT1 activators have been the subject of research in to possible therapeutics for the treatment of T2D[334].

3.1.5 *SIRT1* genetic variation and obesity

In 2007 when this gene was selected for study, no reported associations between *SIRT1* variation and obesity were found in the literature. Three such studies have been published since. Firstly, in a Belgian chort, one SNP, rs7069102 was found to be associated to obesity in a case-control analysis of around 1400 subjects ($p=0.007$)[442]. Another SNP, rs2273773 was then reported to be associated with BMI in a Dutch population ($n=3575$, $p=0.001$)[443] and in a third publication that used two Dutch cohorts (n=6,251 from a population-based study and n=2347 from a family-based study) two SNPs, rs7895833 and rs1467568 were associated with both BMI ($p=0.02$ and 0.008) and obesity ($p=0.007$ and 0.0009)[444]. *SIRT1* mRNA expression has also been reported to be associated with obesity. In a Danish study, transcription levels in adipose were found to be significantly increased in lean compared to obese women ($n=24$, $p<0.02$)[337]. This study also demonstrated an increase in *SIRT1* expression after six days of fasting in humans (n=9).

SIRT1 variants have not been found to be significantly associated with obesity or BMI in any GWAS to date. In the French GWAS, which used some of the same sample as this study[253], no SNPs within the SIRT1 region were found to have a p-value of <0.05. The most recent GIANT GWAS[233] found a nominally significant p-value of 1.2×10^{-4} to rs471962, a SNP within the SIRT1

region (data acessed online[445], see Figure 3.1). The SIRT1 gene contains a number of SNPs validated in the NCBI database but not present in the HapMap, which means that it is possible that a portion of variation in the gene has not been investigated for association with obesity or BMI.

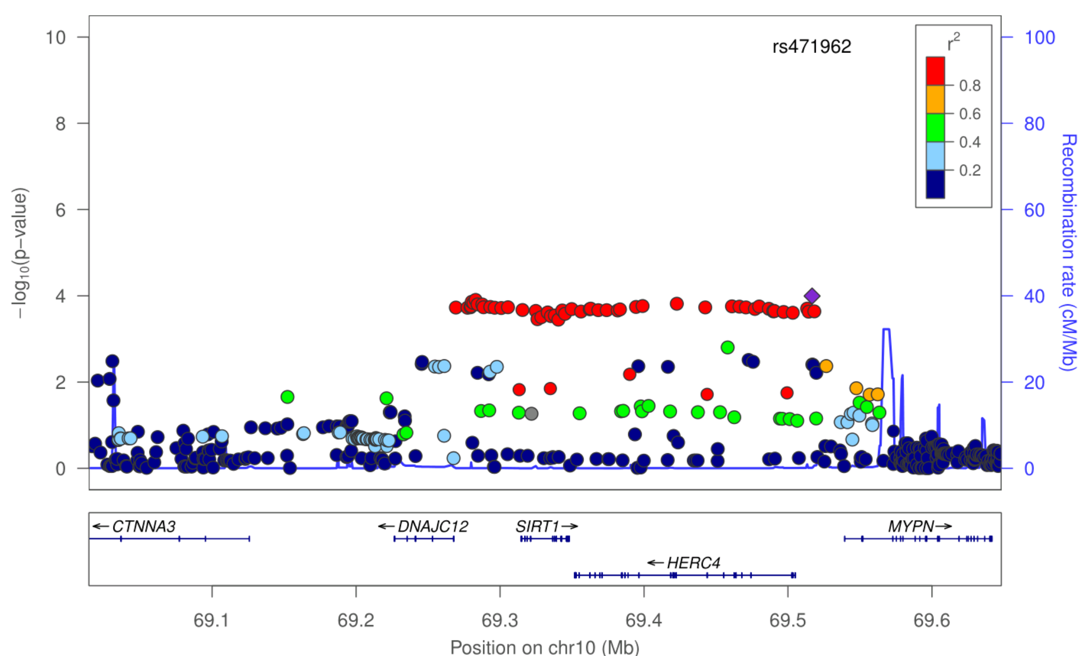

Figure 3.1. SNPs analysed for association to BMI in SIRT1 region in the GIANT GWAS[233].

In summary, SIRT1 has an important role in energy regulation and as such is a plausible candidate gene for polygenic obesity. This study was designed to investigate possible associations between variants in the SIRT1 gene and common polygenic obesity in French obese cases and non-obese controls. A Swedish sib-pair cohort was used for replication and for investigating possible associations between SIRT1 expression in adipose tisse and obesity (see Materials and Methods for details).

3.2 Results

3.2.1 SNP Selection and Case-Control Genotyping

Twenty-seven SNPs were genotyped within the SIRT1 gene including three tag SNPs which captured all twenty-one SNPs included in the HapMap database (mean r^2=0.975)(See Figure 3.2). Twenty-two SNPs achieved acceptable genotyping rates (>85%) with a mean call rate of 90.2%. Eleven of these had minor allele frequencies lower than the required minimum of 5% and were not taken forward for statistical analysis (See Table 3.1). Of the three tag SNPs genotyped, two passed QC, rs12413113 and rs11596401.

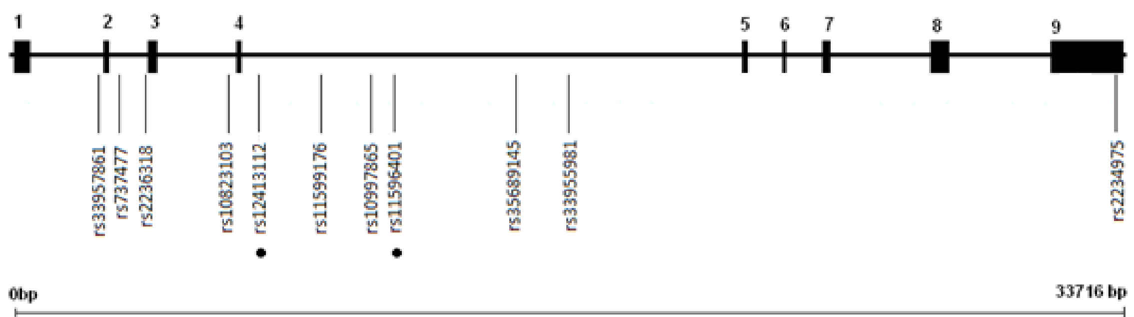

Figure 3.2. Schematic diagram of the SIRT1 gene showing the positions of SNPs successfully genotyped in this study. • tagging SNPs, █ SIRT1 exons.

SNP	Position on Chromosome 10 [a]	Position within SIRT1 gene	Minor Allele Frequency
rs33957861	69646976	Intron 1	0.12
rs737477	69647605	Intron 2	0.06
rs2236318	69648569	Intron 2	0.45
rs34122272	69650069	Intron 3	<0.01
rs10823103	69650969	Intron 3	0.31
rs12413112	69651866	Intron 4	0.12
rs11599176	69653775	Intron 4	0.12
rs10997865	69655315	Intron 4	0.35
rs11596401	69656037	Intron 4	0.34
rs2894057	69656790	Intron 4	<0.01
rs35689145	69659767	Intron 4	0.08
rs33955981	69661358	Intron 4	0.31
rs41274092	69666657	Exon 5	<0.01
rs36067477	69666666	Exon 5	<0.01
rs3818292	69666901	Intron 5	0.04
rs34414573	69667359	Intron 5	0.01
rs11594238	69668495	Intron 6	<0.01
rs7073231	69669607	Intron 7	<0.01
rs10997871	69672935	Intron 8	<0.01
rs35592342	69674262	Intron 8	0.02
rs35224060	69676297	Exon 9	<0.01
rs2234975	69678078	3′ Untranslated Region	0.09

Table 3.1. SNPs genotyped within the SIRT1 gene and minor allele frequencies observed in the case-control subjects. [a]Positions in bp correspond to NCBI build 37.1 measured from the top of the p arm of the chromosome.

One SNP, rs2234975 was found to be nominally associated with obesity in children (p=0.018, OR= 0.74 CI= 0.57-0.95) but this does not withstand correction for multiple tests using permutations (p=0.123). No association was observed between this SNP and obesity in the adult cohort. Four SNPs, rs33957861 (p= 0.006, OR= 0.75, CI= 0.61-0.92), rs12413112 (p= 0.013, OR= 0.77, CI= 0.63-0.95), rs11599176 (p= 0.003, OR= 0.74, CI= 0.61-0.90) and rs35689145 (p= 0.015, OR= 0.74, CI= 0.58-0.95) were nominally associated with severe obesity

in adults and two of these SNPs survive correction using one million permutations (rs33957861, p=0.034 and rs11599176, p=0.019). None of these SNPs were associated with obesity in children. All associated SNPs had a greater frequency of the minor allele in the controls compared to the cases, suggesting a protective effect (See Table 3.2).

	Genotype Counts														
	Controls			Obese Adults			Obese Children			Odds Ratio Adults (95% CI)	P-values Adults		Odds Ratio Children (95% CI)	P-values Children	
SNP	1,1[a]	1,2	2,2	1,1	1,2	2,2	1,1	1,2	2,2		Empirical	Corrected		Empirical	Corrected
rs33957861	874	254	21	663	133	14	482	121	8	0.75 (0.61-0.91)	**0.0057**	**0.034**[b]	0.85 (0.69-1.1)	0.15	0.66
rs737477	1051	132	6	449	56	1	504	69	5	0.94 (0.69-1.3)	0.72	1.0	1.14 (0.86-1.5)	0.38	0.95
rs2236318	319	568	223	249	388	146	176	282	124	0.91 (0.8-1.0)	0.17	0.68	0.99 (0.86-1.2)	0.94	1.0
rs10823103	503	475	111	378	311	66	286	223	62	0.88 (0.76-1.0)	0.086	0.43	0.93 (0.79-1.1)	0.35	0.92
rs12413112	876	238	22	671	149	7	457	119	7	0.77 (0.63-0.95)	**0.013**	0.085	0.91 (0.73-1.1)	0.40	0.96
rs11599176	886	261	25	684	156	8	463	122	9	0.74 (0.61-0.9)	**0.0025**	**0.019**[b]	0.87 (0.71-1.1)	0.22	0.77
rs10997865	480	511	129	354	323	81	249	264	71	0.9 (0.78-1.0)	0.14	0.60	1.02 (0.88-1.2)	0.80	1.0
rs11596401	468	520	123	339	329	70	254	255	72	0.89 (0.77-1.0)	0.083	0.44	0.99 (0.86-1.2)	0.94	1.0
rs35689145	948	187	9	693	107	1	526	80	4	0.74 (0.58-0.94)	**0.015**	0.10	0.79 (0.61-1.0)	0.076	0.39
rs33955981	518	456	115	389	333	76	280	243	60	0.95 (0.83-1.1)	0.47	0.98	0.98 (0.84-1.2)	0.83	1.0
rs2234975	952	220	7	720	132	9	499	75	6	0.87 (0.7-1.1)	0.19	0.73	0.74 (0.57-0.95)	**0.018**	0.12

Table 3.2. Allelic association analysis of SIRT1 SNPs to obesity in children and adult French case-controls. [a] 1 denotes the common allele and 2 denotes the rare allele, [b] Statistically significant p-values. Corrected p-values result from one million permutations.

SIRT1 SNPs were analysed for association to BMI within the control groups using age and sex as covariates. No SNPs were found to be associated (see Table 3.3).

	Mean BMI (Standard Deviation)				
SNP	**1,1[a]**	**1,2**	**2,2**	**Empirical P-value**	**Corrected P-value**
rs33957861	21.8 (2.1)	22.6 (2.1)	22.4 (2.2)	0.22	0.91
rs737477	23.4 (1.1)	22.3 (2.2)	22.4(2.1)	0.87	1.0
rs2236318	22.2 (2.1)	22.4 (2.1)	22.4 (2.2)	0.97	1.0
rs10823103	22.2 (2.1)	22.4 (2.2)	22.4 (2.1)	0.64	1.0
rs12413112	22.0 (2.0)	22.5 (2.1)	22.4 (2.1)	0.24	0.92
rs11599176	22.0 (2.2)	22.6 (2.1)	22.4 (2.1)	0.36	0.99
rs10997865	22.2 (2.0)	22.4 (2.2)	22.4 (2.1)	0.79	1.0
rs11596401	22.1 (2.1)	22.4 (2.2)	22.4 (2.1)	0.61	1.0
rs35689145	21.8 (2.1)	22.6 (2.2)	22.4 (2.1)	0.11	0.68
rs33955981	22.4 (2.0)	22.5 (2.2)	22.4 (2.1)	0.46	1.0
rs2234975	22.6 (1.3)	22.2 (2.2)	22.4 (2.1)	0.41	0.99

Table 3.3. Association analysis of *SIRT1* SNPs to the quantitative trait of BMI in French controls using age and sex as covariates. [a] 1 denotes the common allele, 2 denotes the minor allele.

A high degree of linkage disequilibrium (LD) was found throughout the gene especially between the four SNPs nominally associated with adult obesity with r^2 values ranging from 0.63 to 0.96 (See Figure 3.3). The two SNPs that were significantly associated with obesity after permutation correction were in strong LD with one-another with r^2=0.96. One LD block was identified using the confidence interval method described by Gabriel *et al.*[403] between rs2236318 and rs33955981.

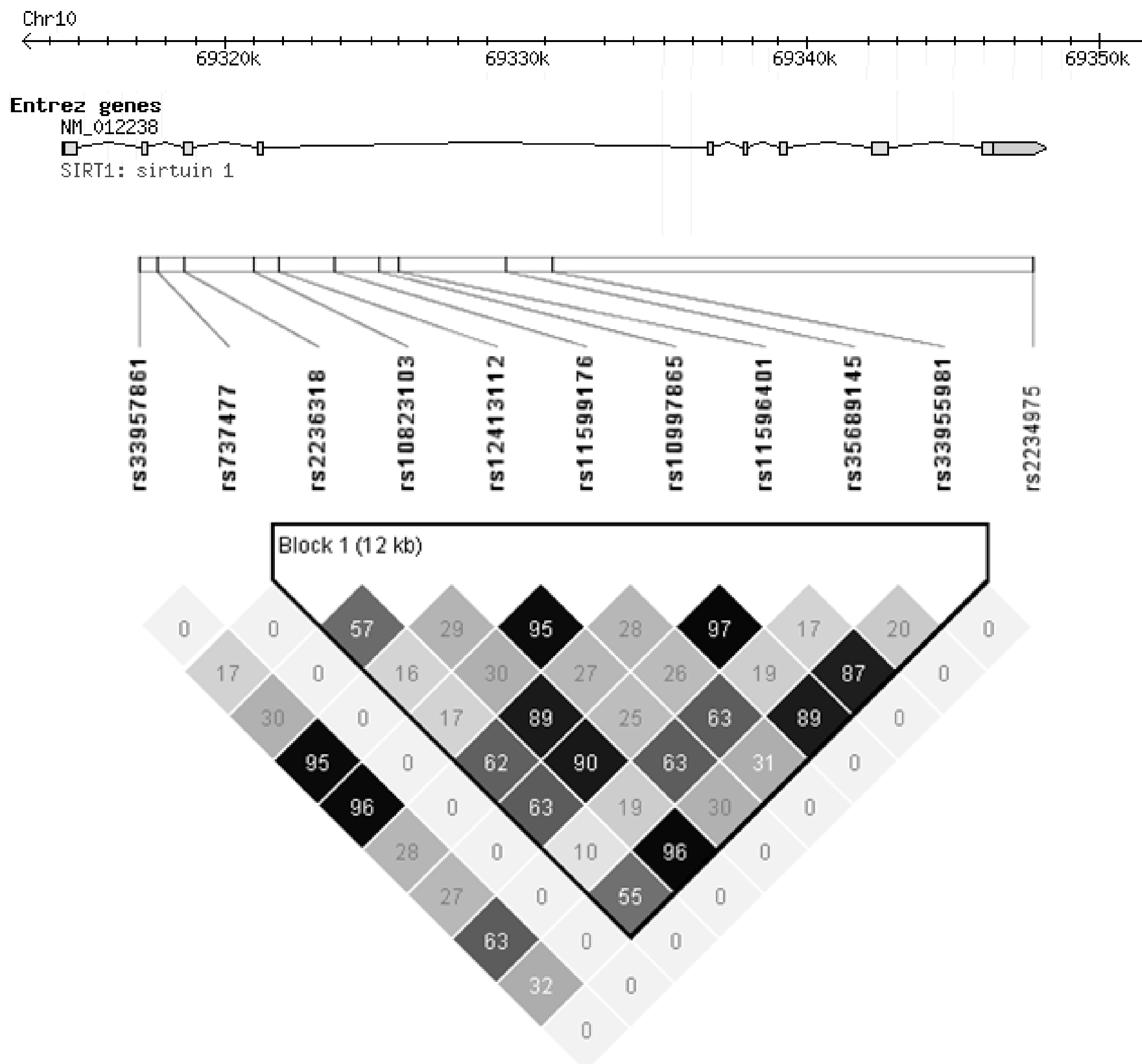

Figure 3.3. Linkage disequilibrium (r^2) plot of *SIRT1* SNPs genotyped in the French control samples. Black squares indicate 100% LD and white squares indicate zero LD. The percentage LD level (r^2) within grey squares are displayed.

Haplotype analysis using the four nominally associated SNPs revealed a significant association to obesity (p=0.028 after one million permutations). Haplotype analysis performed on the two significantly associated SNPs, rs33957861 and rs11599176 revealed a more significant association to obesity than was found for any individual SNP (p=0.001 after 1 million permutations) (See Table 3.4).

SNP 1	SNP 2	SNP 3	SNP 4	Empirical p-value	Corrected p-value
rs33957861	rs12413112	rs11599176	rs35689145	**0.0083**	**0.028**[a]
rs33957861	rs12413112	rs11599176	-	**0.022**	0.067
rs33957861	rs12413112	rs35689145	-	**0.027**	0.081
rs33957861	rs11599176	rs35689145	-	**0.019**	0.059
rs12413112	rs11599176	rs35689145	-	**0.049**	0.14
rs33957861	rs12413112	-	-	**0.0011**	**0.0030**[a]
rs33957861	rs11599176	-	-	**0.00040**	**0.0010**[a]
rs33957861	rs35689145	-	-	**0.020**	0.062
rs12413112	rs35689145	-	-	**0.010**	**0.034**[a]
rs12413112	rs11599176	-	-	**0.0029**	**0.0089**[a]
rs11599176	rs35689145	-	-	**0.0054**	**0.014**[a]

Table 3.4. Haplotype association analysis to obesity in adults using the four SNPs found to be nominally associated with adult obesity. [a] Statistically significant p-values. Corrected p-values result from one million permutations.

3.2.2 Genotyping within the Swedish sib-pair cohort

Genotyping was carried out in the Swedish families using the same set of Sequenom iPlex assays. Genotyping success rate averaged 94% for the five SNPs nominally associated within the French case control analysis. These SNPs were analysed for association with BMI using the QFAM test within PLINK. Four SNPs, rs11599176, rs12413112, rs33957861 and rs35689145 were nominally associated with BMI (p=4.5×10^{-4}, 5.7×10^{-4}, 4.5×10^{-4} and 0.057 respectively, see Table 3.5). These are the same four SNPs nominally associated with adult obesity. At each SNP, the minor allele was associated with a reduced BMI with a difference of up to 2.8kg/m^2 between the two homozygote genotypes. After correction using one million permutations, three SNPs remained significant, rs11599176 (p=0.014), rs12413112 (p=0.019) and rs33957861 (p=0.014)(see Table 3.5).

	Genotype Numbers			Mean BMI ± 95% CI (kg/m^2)				
SNP	**1,1**[a]	**1,2**	**2,2**	**1,1**	**1,2**	**2,2**	**Empirical p-value**	**Corrected p-value**
rs33957861	487	133	9	29.2 ± 0.55	28.4 ± 1.2	26.4 ± 3.0	**0.0045**	**0.014**[b]
rs12413112	478	129	8	29.3 ± 0.56	28.4 ± 1.2	26.4 ± 3.3	**0.0057**	**0.019**[b]
rs11599176	481	132	9	29.3 ± 0.56	28.4 ± 1.2	26.8 ± 3.0	**0.0045**	**0.014**[b]
rs35689145	534	82	2	29.1± 0.54	28.3 ± 1.5	31.6 ± 8.9	**0.057**	0.088
rs2234975	522	91	11	29.2 ± 0.57	28.8 ± 1.1	26.8 ± 2.5	0.99	0.99

Table 3.5. Association analysis of *SIRT1* SNPs with BMI in the Swedish families using the QFAM test and the within-families model of association correcting for age and sex as implemented by PLINK. [a] 1 denotes the common allele and 2 denotes the rare allele, [b] statistically significant p-values. Corrected p-values result from one million permutations.

Haplotypes of the four associated SNPs were imputed and tested for association to BMI. After correction, five haplotypes were significantly associated; however p-values were higher than for the individual SNPs with the most significant haplotype of the three significantly associated SNPs (rs33957861, rs12413112 and rs11599176; p=0.017; see Table 3.6.)

SNP 1	SNP 2	SNP 3	Empirical P-value	Corrected P-value
rs33957861	rs12413112	rs35689145	**0.045**	0.073
rs33957861	rs12413112	rs11599176	**0.0053**	**0.017**[a]
rs35689145	rs12413112	rs11599176	**0.069**	0.099
rs33957861	rs12413112	-	**0.018**	**0.048**[a]
rs33957861	rs11599176	-	**0.011**	**0.033**[a]
rs33957861	rs35689145	-	0.14	0.19
rs12413112	rs35689145	-	0.11	0.15
rs11599176	rs35689145	-	**0.018**	**0.050**[a]
rs12413112	rs11599176	-	**0.011**	**0.034**[a]

Table 3.6. Haplotype association analysis of SIRT1 SNPs with BMI in the Swedish families using the QFAM test and the within-families model of association implemented by PLINK. [a] Statistically significant p-values. Corrected p-values result from one million permutations.

3.2.3 Analysis of *SIRT1* transcription

Gene expression data from subcutaneous adipose tissue of the Swedish sib-pairs was available for analysis. This data was obtained from an eQTL study which used an Affymetrix Human Genome U133 Plus 2.0 gene expression array to measure genome wide transcript levels[396].

There was no significant association between *SIRT1* SNPs and *SIRT1* transcript levels after correcting for the effects of age, sex and BMI (see Table 3.7).

	Genotype Numbers			Mean SIRT1 Transcription (95% CI)				
SNP	**1,1[a]**	**1,2**	**2,2**	**1,1**	**1,2**	**2,2**	**Empirical P-value**	**Corrected P-value**
rs33957861	256	73	6	0.18 (-0.21 – 0.59)	-0.44 (-1.2 - 0.28)	-3.2 (-5.29 - -1.03)	0.90	0.89
rs12413112	254	71	5	0.19 (-0.20 – 0.60)	-0.26 (-1.0 -0.48)	-3.7 (-6.01 - -1.31)	0.66	0.66
rs11599176	254	72	6	0.19 (-0.20 – 0.60)	-0.17 (-0.91 – 0.57)	-3.2 (-5.3 - -1.0)	0.78	0.79
rs35689145	280	48	2	0.010 (-0.36 – 0.38)	-1.1 (-2.1 – -0.11)	-2.5 (-6.0 – 1.0)	0.060	0.072
rs2234975	285	42	8	0.0073 (-0.37 – 0.39)	0.16 (-0.84 – 1.1)	2.1 (-0.15 – 4.4)	0.63	0.57

Table 3.7. Association analysis of *SIRT1* SNPs with *SIRT1* expression corrected for age, sex and BMI in the Swedish families using the QFAM test and the within-families model of association implemented by PLINK. [a]1 denotes the common allele and 2 denotes the rare allele.

When siblings were split into two groups of non-obese (BMI<30 kg/m^2) and obese (BMI>30kg/m^2) and the transcription levels corrected for age, sex and relatedness, the two groups were normally distributed as assessed using a normal Q-Q plot. Using an independent samples t-test, a significant difference in the level of *SIRT1* transcription was found between the two groups ($p=1.56x10^{-35}$, see Figure 3.4 and Table 3.8). When corrected *SIRT1* transcription was plotted against BMI, a negative correlation was observed in the lean group ($r^2=0.13$, $p=3.37x10^{-7}$, see Figure 3.5), however this was not the case in the obese group ($r^2= 0.012$, p=0.17, see Figure 3.6).

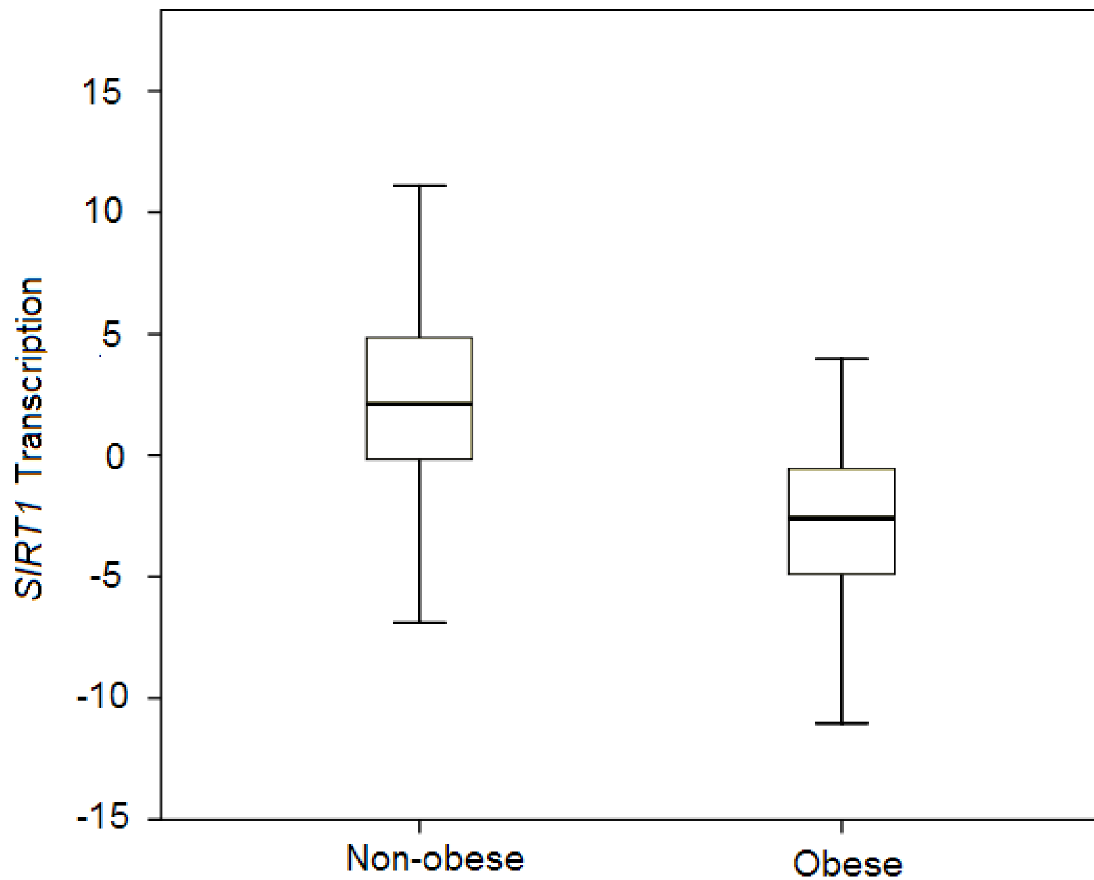

Figure 3.4. Box-plot of *SIRT1* transcript levels in obese and non-obese subjects from Swedish sib-pairs. *SIRT1* transcription is the microarray signal value corrected for age, sex and relatedness. Mean transcription ± standard deviation were 2.26 ± 4.08 in the non-obese and -2.93 ± 3.13 in the obese group.

Number of Non-obese	Number of Obese	Mean Difference in *SIRT1* transcription (95% CI)	P-value
190	156	5.2 (4.5-5.9)	1.6×10^{-35}

Table 3.8. T-test of *SIRT1* transcription in obese and non-obese subjects from Swedish families. *SIRT1* transcription is the microarray signal value corrected for age, sex and relatedness.

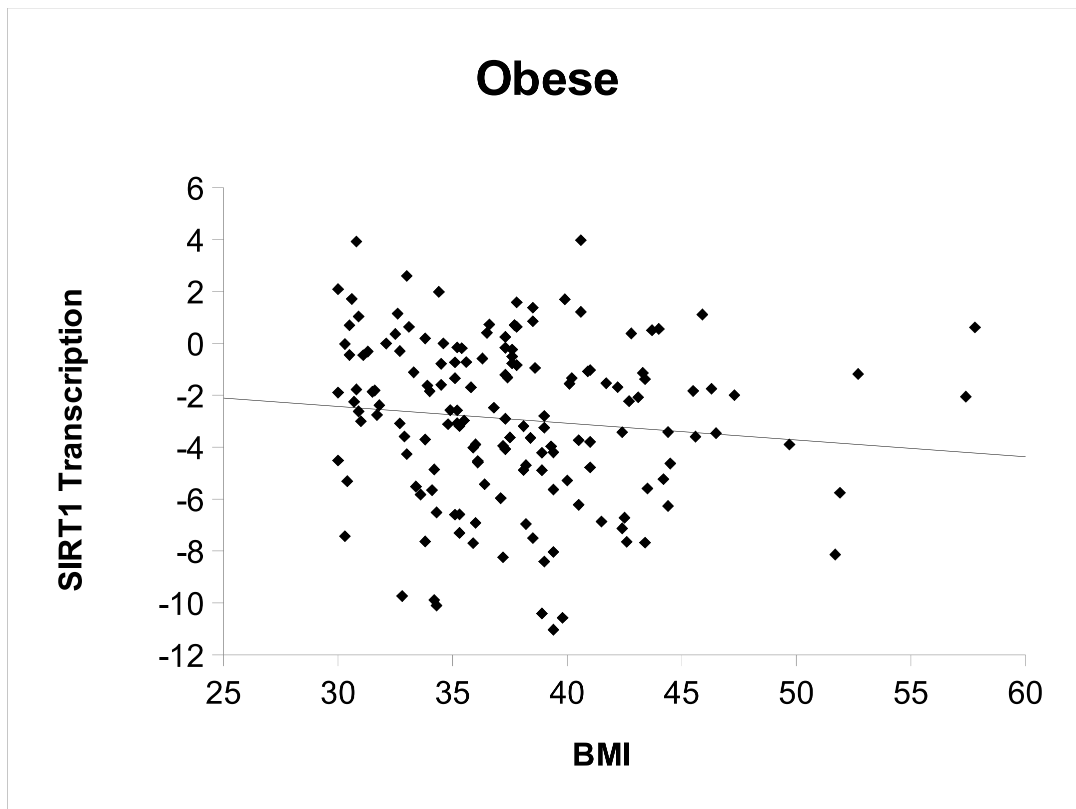

Figure 3.5. SIRT1 transcription levels plotted against BMI in non-obese siblings (n=196, r^2 = 0.13, p=$3.4x10^{-7}$). SIRT1 transcription was the DNA microarray signal value corrected for age, sex and relatedness.

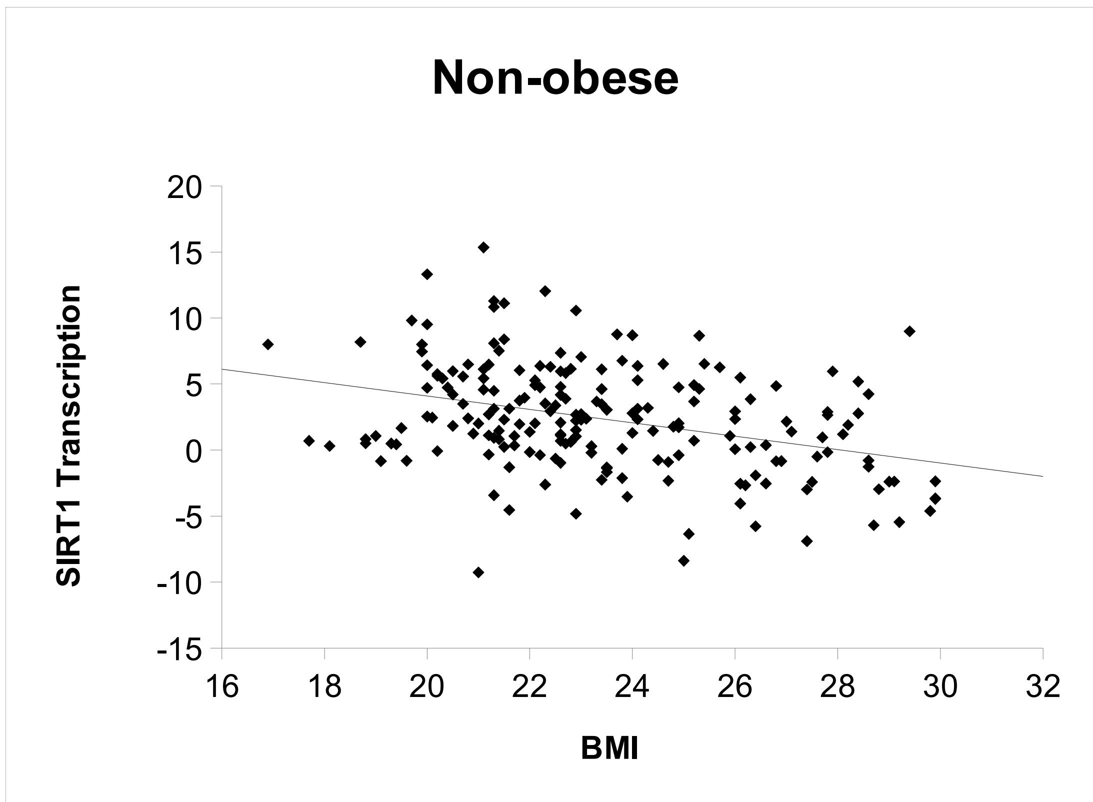

Figure 3.6. SIRT1 transcription levels plotted against BMI in obese siblings (n=156 ,r^2= 0.012, p=0.17) . SIRT1 transcription was the DNA microarray signal value corrected for age, sex and relatedness.

3.2.4 Sequencing of *SIRT1*

Sequencing was carried out in order to find variants in LD with associated SNPs in an attempt to locate the causative mutation. Ten subjects were chosen that were homozygous for each of the four minor alleles associated with adult obesity. A further nine subjects were sequenced that were homozygous for the minor allele of rs2234975 (associated with obesity in children). Subjects that had the minor variant for adult associated SNPs were chosen such that they had the common variant for the child associated SNP and *vice versa*. One sample that was

homozygous for each minor allele associated with both adult and childhood obesity was also included. Eleven PCRs were carried out covering the nine exons in *SIRT1* along with ten PCRs covering the 3kbp region upstream of the transcription start site. This was followed by direct sequencing.

One SNP was discovered within the promoter of the samples that carried the minor allele variant associated with obesity in the adult cohort. This was located 1348bp upstream of the *SIRT1* transcription start site and identified as rs12778366. Four SNPs were discovered within the promoter of the samples that contained the minor allele of the variant associated with obesity in the child smples. These were located 86bp, 210bp, 1085bp and 2473bp upstream of the transcription start site and were identified as rs2394443, rs932658, rs3758391 and rs12250285. One SNP was discovered within the promoters of all samples. This was located 1759bp upstream of the transcription start site and was identified as rs10740280. No SNPs were found within any of the *SIRT1* exons (see Figure 3.7).

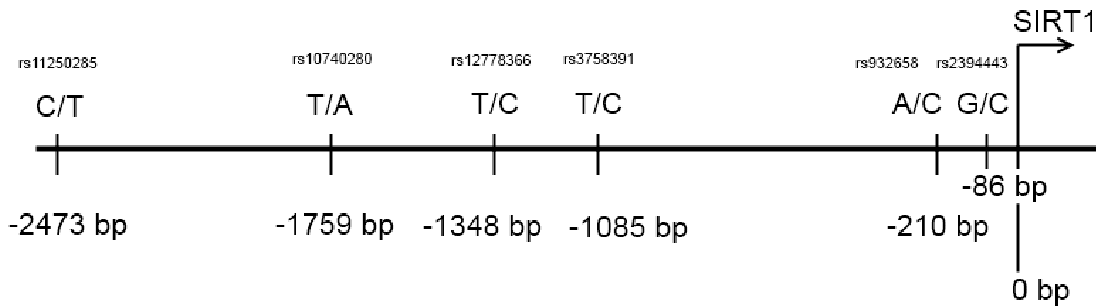

Figure 3.7. SNPs discovered within the *SIRT1* gene promoter region in samples that carried the minor allele variants associated with obesity. Positions are relative to the transcriptional start of *SIRT1.*

3.2.5 Case-control genotyping in promoter

Genotyping of the *SIRT1* promoter SNPs was then carried out in the French case-control subjects. Each of the six SNPs discovered above were genotyped along with four additional SNPs, which were chosen from the four kilobase promoter region of the gene in the NCBI database. Of these, five passed quality control, which included three of the six discovered SNPs including the one SNP that was found to be in LD with SNPs associated with adult obesity, rs12778366 (see Table 3.9).

SNP	Position on Chromosome 10 (bp)[a]	Minor Allele Frequency
rs12246428	69640165	<0.01
rs12778366	69643079	0.12
rs3758391	69643342	0.31
rs35706870	69643617	0.13
rs3740051	69643959	0.06
rs932658	69644217	0.32

Table 3.9. *SIRT1* promoter SNPs successfully genotyped with the minor allele frequencies observed in the case-control subjects. [a] Positions in bp correspond to NCBI build 37.1 measured from the top of the p arm of the chromosome.

No SNPs were associated with obesity in either adults or children. Association analysis was carried out together with the previous set of SNPs in order to obtain corrected p-values that reflect the additional statistical tests being used. These results are presented in Table 3.10. After this correction was applied, two SNPs remain significantly associated with obesity in adults, rs33957861 (corrected p=0.048) and rs11599176 (corrected p=0.026). LD analysis within the control samples revealed some LD between rs12778366 and the adult associated SNPs (r^2 = 0.42 – 0.69), however very little LD was observed to any other promoter SNP. No LD was observed between the child associated SNP, rs2234975 and any promoter SNP (see Figure 3.8).

		Genotype Counts														
		Controls			Obese Adults			Obese Children			Obese Adults			Obese Children		
SNP	Position Within SIRT1 Gene	1,1[a]	1,2	2,2	1,1	1,2	2,2	1,1	1,2	2,2	Odds ratio (95% CI)	Empirical P-value	Corrected P-value	Odds ratio (95% CI)	Empirical P-value	Corrected P-value
rs12778366	Promoter	14	161	697	20	261	912	9	116	468	0.84 (0.69-1.0)	0.080	0.52	0.89 (0.72 - 1.1)	0.32	0.96
rs3758391	Promoter	76	370	420	121	526	555	69	245	285	0.92 (0.8-1.1)	0.22	0.87	1.0 (0.87 - 1.2)	0.90	1.0
rs35706870	Promoter	9	197	662	20	270	908	14	119	466	0.95 (0.79-1.2)	0.60	1.0	0.93 (0.76 - 1.2)	0.54	1.0
rs3740051	Promoter	4	78	644	5	122	1068	4	67	520	1.1 (0.81-1.4)	0.60	1.0	1.19 (0.89 - 1.6)	0.25	0.91
rs932658	Promoter	77	372	418	122	523	541	66	241	280	0.91 (0.8-1.0)	0.17	0.79	0.98 (0.85 - 1.1)	0.82	1.0
rs33957861	Intron 1	874	254	21	663	133	14	482	121	8	0.75 (0.61-0.91)	**0.0057**	**0.048**[b]	0.87 (0.7 - 1.1)	0.21	0.86
rs737477	Intron 2	1051	132	6	449	56	1	504	69	5	0.94 (0.69-1.3)	0.72	1.0	1.14 (0.85 - 1.5)	0.40	0.99
rs2236318	Intron 2	319	568	223	249	388	146	176	282	124	0.91 (0.8-1.0)	0.17	0.79	0.99 (0.86 - 1.2)	0.93	1.0
rs10823103	Intron 3	503	475	111	378	311	66	286	223	62	0.88 (0.76-1.0)	0.086	0.53	0.93 (0.8 - 1.1)	0.39	0.99
rs12413112	Intron 4	876	238	22	671	149	7	457	119	7	0.77 (0.63-0.95)	**0.013**	0.12	0.92 (0.74 - 1.2)	0.48	1.0
rs11599176	Intron 4	886	261	25	684	156	8	463	122	9	0.74 (0.61-0.90)	**0.0025**	**0.026**[b]	0.89 (0.71 - 1.1)	0.28	0.93
rs10997865	Intron 4	480	511	129	354	323	81	249	264	71	0.90 (0.78-1.0)	0.135	0.71	1.0 (0.88 - 1.2)	0.75	1.0
rs11596401	Intron 4	468	520	123	339	329	70	254	255	72	0.89 (0.77-1.0)	0.083	0.55	1 (0.86 - 1.2)	0.96	1.0
rs35689145	Intron 4	948	187	9	693	107	1	526	80	4	0.74 (0.58-0.94)	**0.015**	0.14	0.79 (0.61 - 1.0)	0.076	0.51
rs33955981	Intron 4	518	456	115	389	333	76	280	243	60	0.95 (0.83-1.1)	0.47	1.0	0.99 (0.85 - 1.2)	0.94	1.0
rs2234975	3' Untranslated Region	952	220	7	720	132	9	499	75	6	0.87 (0.7-1.1)	0.19	0.82	0.72 (0.56 - 0.94)	**0.018**	0.13

Table 3.10. Association analysis of *SIRT1* promoter SNPs in French case-controls analysed together with the original genotyped SNPs in order to take into account the additional statistical tests being performed. [a] 1 denotes the common allele and 2 denotes the rare allele, [b] P-values that are statistically significant after correction using one million permutations.

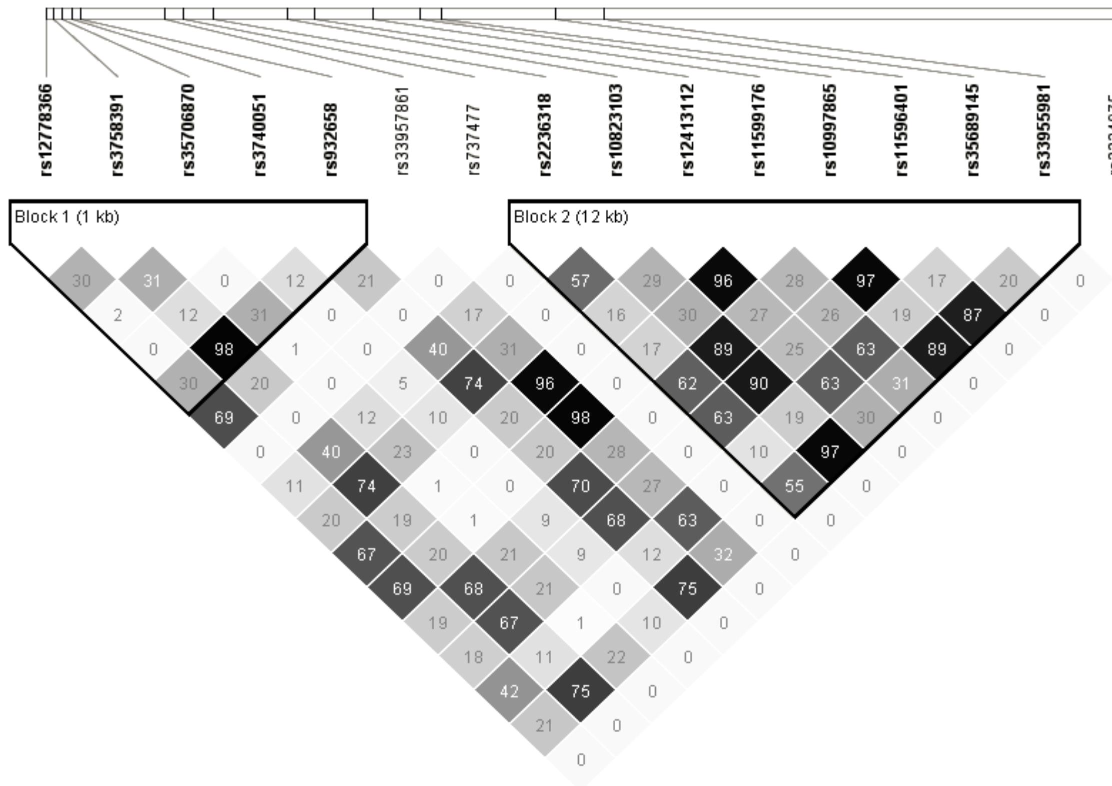

Figure 3.8. LD (r^2) plot within the French control samples containing all SNPs genotyped and analysed within *SIRT1* and its promoter region. Black squares indicate 100% LD and white squares indicate zero LD. The percentage LD level (r^2) within grey squares are displayed.

3.2.6 Imputation

An alternative method for fine-mapping the association involves imputation of genotypes from a reference population. Using the 1000 genomes data[265], genotypes were imputed 250kb either side of the *SIRT1* gene using IMPUTE2 which generated 7786 imputed SNPs[404]. SNPTEST was used to test case-control associations to the imputed SNPs taking into account the genotype uncertainty of imputed data. 414 SNPs out of 7786 tested were nominally associated with

obesity in the adult case-control cohort and 24 had p-values $<10^{-5}$ with the most significant SNP, rs12776134 $p=1.9\times10^{-6}$ (see Figure 3.9). In the child case-controls 249 were nominally associated but none had p-values $<10^{-5}$. None of the most significantly associated SNPs (those with p-values $<10^{-4}$) were in *SIRT1* exons.

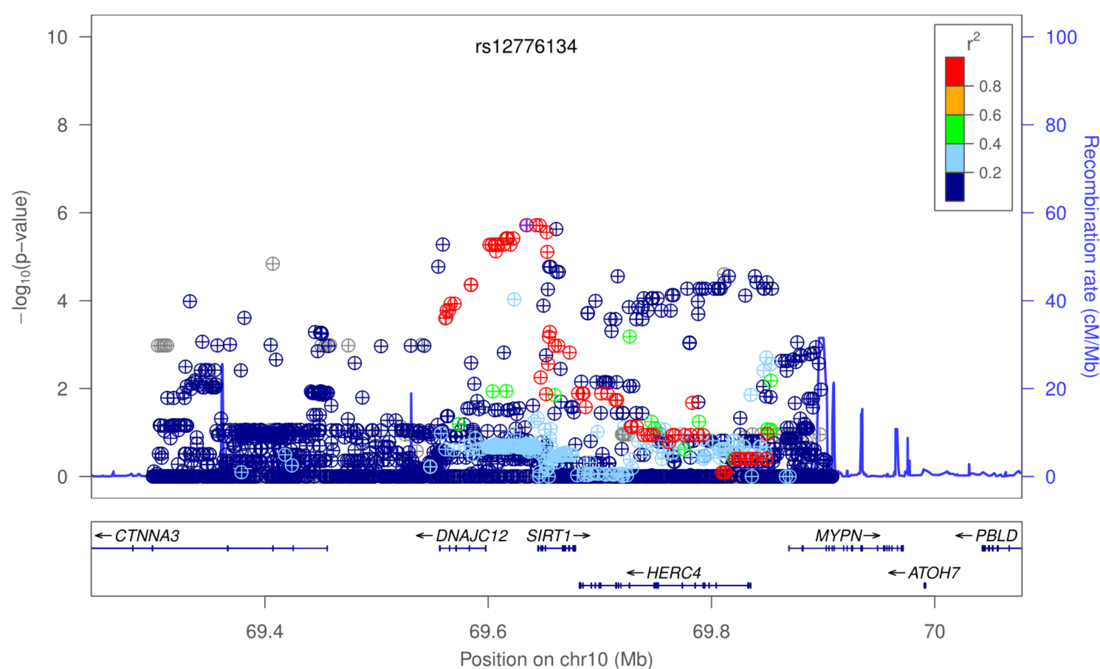

Figure 3.9. LocusZoom[445] plot of imputed SIRT1 SNPs tested for association in the French adult case-control cohort. LD values measured using r^2 are from 1000 genomes data and are displayed relative to rs12776134, the most significantly associated imputed SNP.

3.2.7 Statistical power calculations

Using an odds ratio of 0.74 (as observed for rs11599176, the most significantly associated SNP), a MAF of 0.13 (as observed for rs11599176) and a significance level of 0.0045 (0.05/11 tests), the power to detect an association was 25% in the adult cohort. In the child cohort, assuming an odds ratio of 0.87, power was 2.6% (see Table 3.11). In order to achieve a

statistical power of 95%, assuming an odds ration of 0.74, a MAF of 0.13 and a significance level of 0.0045, this study would need a sample size of 4332 cases and 4332 controls. With a significance level of 5.0×10^{-8} (genome-wide significance), this study would require 10842 cases and 10842 controls (see Figure 3.10). If the effect size was smaller (odds ratio greater than 0.74) then the sample size required to reach a statistical power of 95% increases (see Figure 3.11).

SNP	Significance level	MAF (controls)	OR (children)	OR (adults)	Power (children)	Power (adults)
rs33957861	0.0045	0.13	0.85	0.75	0.036	0.21
rs737477	0.0045	0.061	1.1	0.94	0.017	0.07
rs2236318	0.0045	0.46	0.99	0.91	<0.01	0.038
rs10823103	0.0045	0.32	0.93	0.88	0.015	0.066
rs12413112	0.0045	0.12	0.91	0.77	0.013	0.16
rs11599176	0.0045	0.13	0.87	0.74	0.026	0.25
rs10997865	0.0045	0.34	1.0	0.90	<0.01	0.043
rs11596401	0.0045	0.35	0.99	0.89	<0.01	0.055
rs35689145	0.0045	0.090	0.79	0.74	0.055	0.15
rs33955981	0.0045	0.32	0.98	0.95	<0.01	0.011
rs2234975	0.0045	0.099	0.74	0.87	0.117	0.026

Table 3.11. Power calculations carried out using MAF and OR observed for each SNP in the SIRT1 gene investigated.

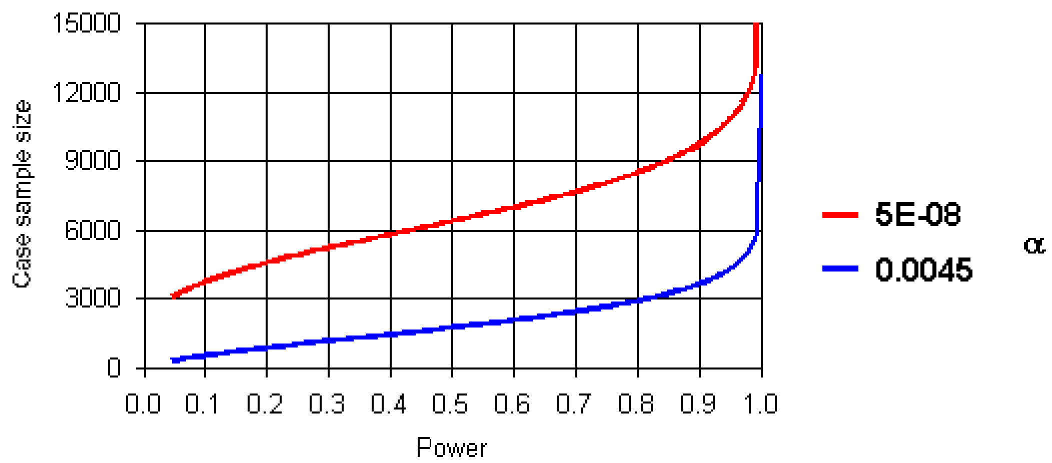

Figure 3.10. Sample size plotted with statistical power for a SNP with MAF=0.13 (as was observed for rs11599176). α is the significance level: 0.0045 for 11 statistical tests and $5x10^{-8}$ for genome-wide significance.

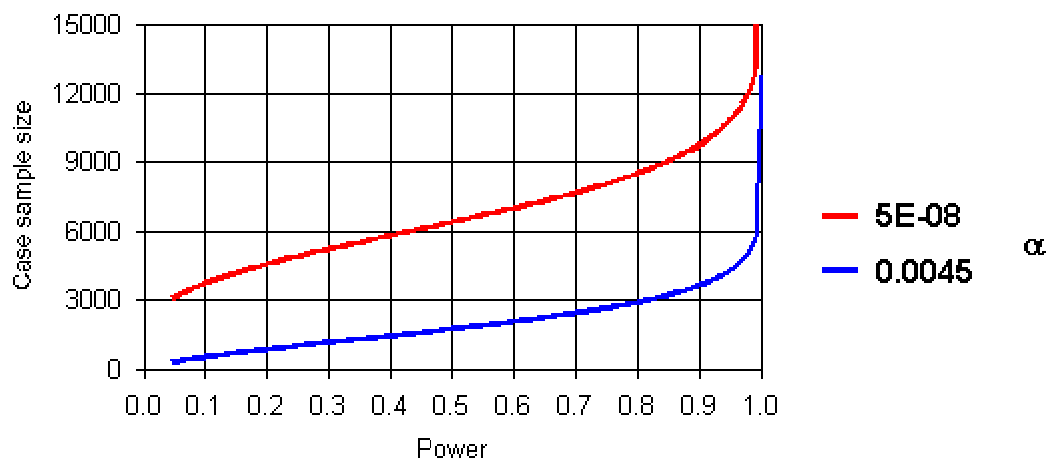

Figure 3.11. Sample size (number of cases, with number of controls=number of cases) required for varying odds ratios in order to achieve a statistical power of 95% for a SNP with MAF=0.13 (as was observed for rs11599176). α is the significance level: 0.0045 for 11 statistical tests and $5x10^{-8}$ for genome-wide significance.

Statistical power for quantitative trait analysis in the French control samples was 43% to detect an association with an effect size of 0.5 BMI units assuming a MAF of 0.13, a SD of 2.1 and a significance level of 0.0045. At genome-wide significance level this falls to <1%. Smaller effect sizes would result in less power (see Figure 3.12).

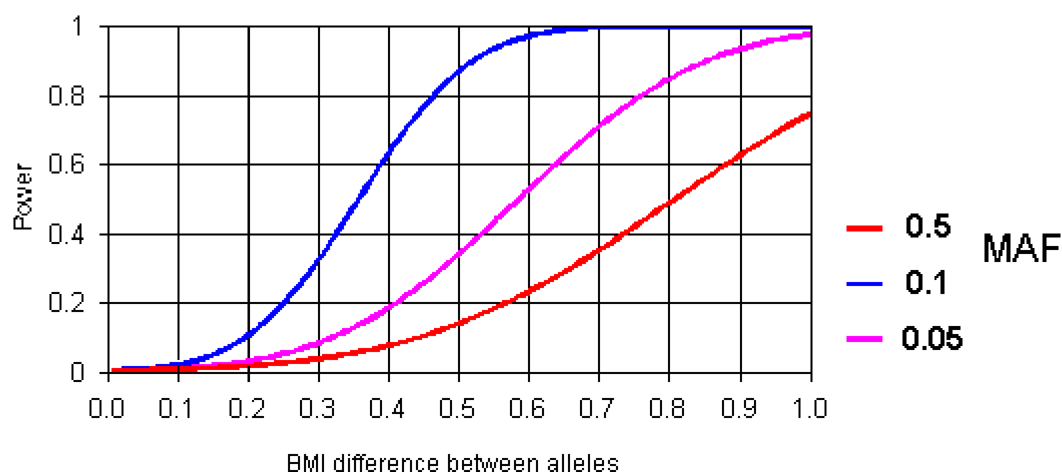

Figure 3.12. Statistical power against effect size in French controls assuming a MAF of 0.13 and SD of 2.1 BMI units and a significance level of 0.0045.

For the sequencing project, the probability of locating a common variant (MAF>5%) in high LD (r^2>0.8) with the associated SNPs using 10 samples was calculated to be >99%, assuming such a variant exists. The probability of locating a variant with MAF = 1% in moderate LD (r^2=0.5) was calculated to be 98%. The probability of locating a variant with MAF = 0.1% in moderate LD (r^2=0.5) was calculated to be 70% (see Appendix A7 for calculations).

3.2.8 Thesis-wide multiple testing correction

When *SIRT1* SNPs are analysed together with all other gene SNPs investigated in this thesis, no variants remain significantly associated to obesity in either adults or children after correction using permutations (see Table 3.12).

SNP	Empirical P-value Children	Corrected P-value Children	Empirical P-value Adults	Corrected P-value Adults
rs12778366	0.081	1.0	0.080	1.0
rs3758391	0.22	1.0	0.22	1.0
rs35706870	0.60	1.0	0.60	1.0
rs3740051	0.60	1.0	0.60	1.0
rs932658	0.17	1.0	0.17	1.0
rs33957861	0.15	1.0	**0.0057**	0.36
rs737477	0.38	1.0	0.72	1.0
rs2236318	0.94	1.0	0.17	1.0
rs10823103	0.35	1.0	0.086	1.0
rs12413112	0.40	1.0	**0.013**	0.64
rs11599176	0.22	1.0	**0.0025**	0.20
rs10997865	0.80	1.0	0.14	1.0
rs11596401	0.94	1.0	0.083	0.99
rs35689145	0.076	0.98	**0.015**	0.56
rs33955981	0.83	1.0	0.47	1.0
rs2234975	**0.018**	0.51	0.19	1.0

Table 3.12 P-values after correcting for multiple testing using permutations and taking into account all SNPs investigated in this thesis.

3.3 Discussion

These results suggest that the hypothesis that SIRT gene variation influences obesity risk merits further investigation in larger sample sizes. The lack of genome-wide significance and replication means that an association has not been conclusively proven. However, the observation of associations between *SIRT1* SNPs and obesity in French case-controls and BMI in Swedish families provides evidence of a possible association.

The p-values reported here do not withstand a genome-wide correction for multiple testing. Furthermore, permutation analysis using all SNPs tested within this thesis results in non-significance of all tests, which means that the possibility of a false positive cannot be ruled out. If this investigation is treated as an independent test assessing whether *SIRT1* variation is associated with obesity and permutation testing used to correct for multiple tests[397]. then two SNPs are found to be significantly associated with obesity in adults supporting the candidacy of SIRT1 as an obesity gene.

At each SNP the minor allele was present at a higher frequency within the controls compared to the cases indicating a protective association. In the Swedish siblings the minor allele was associated with a lower BMI, again indicating a protective association. In the most extreme case the minor homozygote had a BMI of 2.8 kg/m^2 lower than the common homozygote although the small numbers and large 95% confidence interval range means that the measurable effect may not be accurate. The same four SNPs that were nominally associated with severe obesity in adults were nominally associated in the families, which is evidence that this is a real genetic association. The failure to find an association between SNPs and BMI in the French control groups is likely due to the low level of variance in these cohorts (standard deviation = 2.1 kg/m^2), which results in a reduced power to detect association (43%). Alternatively it is possible that this is an association that effects obesity risk without modifying BMI in the non-obese population.

These results support and expand on the three previously reported SIRT1 association studies. The first of these reported a protective association between the *SIRT1* SNP rs7069102 (tagged by rs11596401 in our study with r^2=0.956 in HapMap) and obesity (p=0.007)[442]. In this thesis, rs11596401 was not significantly associated with obesity in either French adults (p=0.083) or children (p=0.94). The second discovered an association between minor genotypes in rs2273773 (genotyped but failed QC in this study) and reduced BMI (p=0.001)[443]. The third reported a protective association of *SIRT1* SNPs rs1467568 (p=0.0009) and rs7895833 (p=0.007) and obesity[444]. Rs7895833 is located 21kb upstream of *SIRT1* and was not tagged by any SNP investigated in this project. Rs1467568 is tagged by rs11596401 in this study (r^2=1.0 in HapMap) but this SNP was not found to be significantly associated with obesity in this study (p=0.073 in adults and p=1.000 in children).

Because the same SNPs are not associated in different studies, the association has not been replicated and so further work is required in order to exclude the possibility of a false-positive. Since each study reported the minor allele being associated with a reduced risk of obesity or a reduced BMI this is evidence that the same effect is being detected. However, the fact that the reported associations are to different markers means that the association has not been properly replicated. This could be due to the likely fact that the causative variant has not been identified and the variability between studies introduced by different recruitment criteria. Phenotype measurements and genotyping success rates may also add to the variation between studies. Again, the fact that none of these tests would be significant in a geneome-wide context means that the possibility of a false-positive cannot be ruled out and thus replication is required in other, larger cohorts. It should also be noted that the *SIRT1* gene appears to be in a region of low recombination (see Figure 3.1) so it is possible that any associated SNPs may be tagging variants in other nearby genes such as *HERC4, DNAJC12* and *CTNNA3*.

The absence of a significant association between *SIRT1* gene variation and obesity in any genome wide association study could be due to the stringent requirements to correct for multiple testing, which could result in associations with weak effect sizes being screened out. The most significant *SIRT1* variant in the most recent GIANT GWAS[233] was rs471962 with a p-value of $1.25x10^{-4}$. Whilst this p-

value does not reach genome-wide significance, it does reach nominal significance which might indicate an association that requires larger sample sizes in order to be detected. Rs471962 is located 169 kb downstream of the gene and so was not typed in this study. However it is in high LD (r^2>0.8 in the HapMap CEU population) with a number of other variants within the gene including rs10997860 which is tagged by rs10823103 (r^2=1.0) in this study. This SNP was not significantly associated with obesity (p=0.083 before correction). This might reflect the lower statistical power in the French case-controls compared to the GIANT study which used ~250,000 samples. The GIANT GWAS investigating associations to WHR analysed 17 SNPs within the SIRT1 gene and found no nominal associations[262].

Of the five nominally associated SNPs only rs12413112 and rs2234975 are present in the HapMap database. This means that other nominally associated SNPs, rs33957861, rs11599176 and rs35689145 are not tagged by any SNP using HapMap data. However, data presented in this thesis indicates that rs12413112 tags both rs33957861 and rs11599176 very well (r^2=0.95 and 0.95) and rs35689145 reasonably well (r^2=0.63). As such it is likely that coverage of common (>5%) SIRT1 SNPs in GWAS is sufficient.

The significant increase in *SIRT1* transcript in non-obese compared to obese ($p=1.56\times10^{-35}$) and the correlation of *SIRT1* ($p=3.37\times10^{-7}$) expression to BMI in non-obese sib-pairs supports the findings reported by a Danish study in which *SIRT1* transcription was significantly higher in lean compared to obese women[337] and provides further evidence for a role of SIRT1 in obesity. The lack of correlation in the obese sib-pairs may be due to the lower numbers and the lower variation in transcript levels in that group (n=196, SD=4.1 in non-obese; n=156, SD=3.1 in obese).

No significant association was found between *SIRT1* SNPs and expression of the gene. This might be due to the fact that the causative SNP was not genotyped. Alternatively, *SIRT1* variants might be affecting obesity risk without altering its transcription level. Instead, variants within the gene could be altering the activity of the Sirtuin 1 enzyme *via* a change in amino acid sequence or by modifying a splice site. Another possibility is that a mutation within the untranslated region of the mRNA might alter the efficiency of translation, perhaps by modifying its secondary structure and such mechanisms have

been observed[446]. This wouldn't affect levels of mRNA that were measured but would change the level of SIRT1 protein produced. SIRT1 protein then might influence susceptibility to obesity via its effects on the expression of genes important in lipid and glucose metabolism. *SIRT1* transcription levels might then be modified by BMI *via* other mechanisms. For example, expression of *SIRT1* is known to be increased by fasting in mice and humans[337,447] and more recently levels of SIRT1 protein have been demonstrated to be increased after fasting in mice[448].

Statistical power calculations indicate that this study was not sufficiently powered to detect statistically significant associations with the effect sizes that were observed. In order to achieve p-values that are significant in a genome-wide context, larger samples sizes are required, at least 10,000 cases and 10,000 controls assuming an odds ratio of 0.74 which was observed for rs11599176. If this odds ratio is actually lower, as might be expected of a variant with small effect size not detected in any large GWAS, then more samples are needed. Power was similarly low when analysing association to BMI in the French controls. This means that any attempt to replicate an association should be carried out in much larger cohorts. Another method of increasing the statistical power would be to meta-analyse these data together with previously published results.

Since only two of the three tag SNPs genotyped in this study passed QC, coverage of the *SIRT1* gene was not complete. The SNP that failed QC, rs2273773, was previously reported to be associated with BMI in a Dutch population[443] and as such it would be useful to know whether this association is replicated in the cohorts investigated in this thesis. Also, complete coverage was of course not achieved in this study because rare variants (<5%) were not investigated.

All associated SNPs are located within introns or in the case of rs2234975, the 3' untranslated region, which means they do not alter the amino acid sequence of the Sirtuin 1 protein. It is therefore unlikely that these SNPs are causative. That the haplotype analysis generated a lower p-value than any individual SNP could be evidence that these SNPs are tagging a more significantly associated variant, although the high LD between these SNPs (r^2=0.96) means that this difference in p-values may not be significant. This variant would most likely be located within an exon where it could alter the amino acid structure of the protein or within the promoter region where it could affect the expression of the *SIRT1*

messenger RNA. Sequencing was unable to locate any exonic SNPs within the same LD block of the obesity associated SNPs and of the five SNPs genotyped within the promoter region, none were found to be associated with obesity in either adults or children. Power to detect variants with MAF >1% and in high LD (r^2>0.8) with the associated SNPs was sufficiently high (>98%), however as MAF and LD values drop (MAF=0.1%, r^2=0.5) this power is reduced (70%) and if the causative SNP is even rarer (<0.1%) then this will drop further. Thus the possibility of a false-negative cannot be ruled out. Furthermore since only 3kb of the promoter region was sequenced, all regulatory regions have not been investigated.

An alternative method for searching for the causative mutation is by imputation. Using 1000 genomes data[265], around 7000 imputed SNPs in the *SIRT1* region were generated and tested for association to obesity in the French case-controls. 24 SNPs were found to be associated with obesity in the adults with p-values of $<10^{-5}$ with the most significant marker at $p=1.9 \times 10^{-6}$. Each of these 24 markers had MAF in the range of 1-2% and each had a higher frequency in the controls compared to the cases. None of these markers are within *SIRT1* exons so no clear candidates for functional mutations have been found. The fact that the most significant markers were not all in high LD with one-another is interesting since it suggests that if these results are genuine that there may be two or more independent associations to obesity within the *SIRT1* gene.

It is possible that using 1000 genomes data may not give accurate imputed SNPs since the ancestry of the two sample sets will not match. The IMPUTE2 software is designed to choose samples with matching haplotypes in order to get the most accurate imputation. However the small number of genotyped SNPs in this study might introduce errors into this process compared to using a GWAS dataset with many more variants. An alternative method would be to sequence a subset of the French case-controls and use this for imputation. Nevertheless, the fact that imputation resulted in more significantly associated SNPs could be evidence of a real genetic association and although these p-values still do not reach genome-wide significance, they would survive correction for 7000 statistical tests ($p<7 \times 10^{-6}$). In order to validate this result, the most significant hits would need to be genotyped to confirm their accuracy.

Therefore a mechanism by which SNPs could affect phenotype has not been established in this study. More sequencing of the *SIRT1* gene, including a substantial portion of the region upstream of the transcription start site should be carried out in a larger number of samples in order to locate the causative variant. Any SNPs found to be highly significantly associated with obesity in the promoter region can then be assessed for their affect on expression of the *SIRT1* transcript using luciferase assays. If a non-synonymous SNP within an exon is found to be significantly associated with obesity then activity of the Sirtuin 1 enzyme can be measured and compared to the wild type using assays that measure Sirtuin enzyme activity[449].

These results may have important clinical implications. An association between *SIRT1* variation and obesity is concordant with evidence in mice that activators of SIRT1 decrease fat mass and thus may provide pharmacological uses as a treatment for obesity or other metabolic diseases. Their use as a therapeutic for T2D has been proposed previously[334].

In conclusion, these results provide evidence of a possible association between variants of the *SIRT1* gene and the development of common obesity. This will require replication in other populations in order to confirm the association as well as resequencing and functional work to locate the causative variant underlying this genetic association.

Chapter 4

Investigation of genetic variants within the apelin gene for association to common polygenic obesity

4.1 Introduction

Apelin was discovered in 1998 as the endogenous ligand for APJ[450], a G protein coupled receptor known to be localised to neurons in the brain[451]. The apelin gene (*APLN*) is located on the X chromosome at Xq25-26.1[452].

APLN encodes a 77 amino acid pre-protein that is expressed in a wide range of tissues, including heart, lung, mammary gland and adipose[453]. The pre-protein is alternatively cleaved into one of four physiologically active signal peptides, apelin-36, apelin-17, apelin-13 and apelin-12. Apelin-13 may also undergo pyroglutamylation at the N-terminal glutamine residue before being secreted into the blood stream[453,454,455,456]. These peptides are all ligands for APJ, with the pyroglutamated 13 amino acid apelin being the most potent[457].

4.1.1 Apelin is upregulated in obesity

Apelin is expressed in adipose tissue and is secreted by adipocytes[458]. In mice, its expression in adipocytes is upregulated in obese animals and plasma levels of the apelin protein are significantly increased in obese compared to control mice[458]. APJ, the apelin receptor is also expressed in adipose tissue and is upregulated in obesity[459]. In humans, circulating apelin has also been shown to be increased in obese subjects showing that these phenomena are not isolated to rodents[460].

In human adipocyte cell culture, apelin is upregulated in reponse to insulin[458] suggesting that apelin expression in adipose is controlled by the hormone. In obese patients, plasma apelin and insulin levels have been found to be correlated, suggesting that the regulation of apelin by insulin could influence blood concentrations of apelin[461].

Other factors appear to regulate the expression of apelin in adipose. TNFa, an inflammatory cytokine that is upregulated in obesity has been shown to stimulate *APLN* expression in both mouse and human adipocyte culture[462]. Over-expression of PGC1a, a key regulator of cellular energy homeostasis, has been shown to stimulate both expression and secretion of apelin[463].

Apelin expression has also been reported to be upregulated during differentiation of pre-adipocyte into adipocyte cell culture indicating a role in development of adipose tissue[464,465].

4.1.2 Effects of apelin on energy regulation

Intracerebroventricular injections of apelin peptides result in a reduction in food intake in rats[466,467]. In mice, intravenous injections of apelin have been found to lower blood glucose and improve insulin sensitivity[341,342]. Daily apelin injection has also been shown to reduce adiposity in obese mice without influencing food intake[468]. This study also showed a reduction in serum levels of triglycerides and an increased expression of UCP1 and UCP3 along with increased body temperature and oxygen consumption which suggests that apelin acts to stimulate energy expenditure *via* mitochondrial uncoupling[468].

Apelin has also been found to stimulate angiogenesis and as such has been hypothesised to contribute to the development of new vasculature in the expanding adipose tissue[469].

4.1.3 Apelin mouse models

Apelin-null mice are hyperinsulinemic and insulin resistant[342]. These mice also exhibit increased adiposity and increased serum concentrations of free fatty acids[470]. Conversely, mice engineered to over-express *APLN* are resistant to diet-induced obesity and have increased oxygen consumption and body temperature without any difference in feeding behaviour[471].

These mouse models, therefore suggest a role of apelin in controlling body weight *via* glucose metabolism and regulation of energy expenditure and thus variation in the apelin gene may influence obesity risk.

4.1.4 Apelin genetic variation and obesity

No significant associations have been reported between variants in the apelin gene and obesity phenotypes to-date. A nominal association to fasting plasma glucose levels was discovered in a cohort of unrelated male Han Chinese but this study did not find any association to BMI or obesity[472].

No associations have been reported in any GWAS to-date, however most of these studies do not analyse variants on the X chromosome. This means that it is possible that there are variants in the apelin gene that influence obesity risk that haven't been investigated

A Finnish genome-wide linkage study reported a linkage peak with obesity at Xq24, which in linkage terms is relatively close to the Apelin gene located around 8Mbp downstream at Xq25-Xq26.3[473].

Given its role in glucose and lipid metabolism and the linkage peak reported to obesity, apelin is a plausible candidate gene for genetic association to obesity and so this study was designed to investigate possible associations between variants in the *APLN* gene and common polygenic obesity in French obese cases and non-obese controls (see Materials and Methods for details).

4.2 Results

4.2.1 SNP Selection and Case-Control Genotyping

All fourteen common SNPs (MAF > 5%) within the Apelin gene that were in the NCBI database in July 2007 were genotyped in the French case-control subjects using Sequenom iPlex assays (see Figure 4.1). Out of thirteen, eleven SNPs had acceptable genotyping success rates (> 85%, actual mean success rate = 94.5%) and were within HWE (p > 0.05). HWE was analysed in female controls only as *APLN* is located on the X chromosome and so males only have one allele at each SNP. Five SNPs had low minor allele frequencies (< 5%) so were not taken forward for analysis (see Table 4.1).

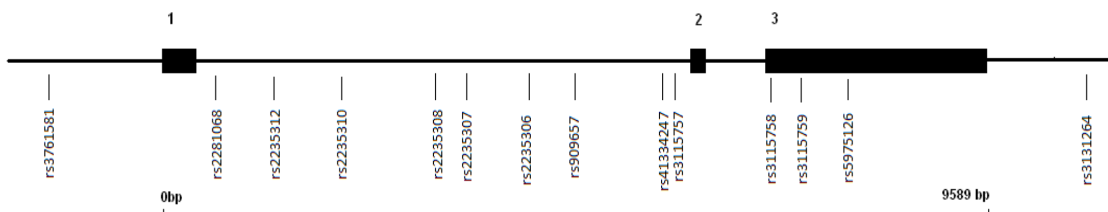

Figure 4.1. Map of *APLN* gene with positions of SNPs genotyped in this project. █ *APLN* exons.

SNP	Position on X Chromosome[a]	Position within APLN gene	Minor Allele Frequency
rs3131264	128777626	Downstream	0.04
rs5975126	128780392	3' UTR	<0.01
rs3115759	128781516	3' UTR	0.08
rs3115758	128781864	3' UTR	0.09
rs41334247	128782557	Intron 2	<0.01
rs2235306	128784098	Intron 1	0.02
rs2235307	128784832	Intron 1	0.05
rs2235308	128785212	Intron 1	0.08
rs2235312	128787095	Intron 1	0.08
rs2281068	128787773	Intron 1	0.08
rs3761581	128789721	Upstream	0.02

Table 4.1. *APLN* SNPs successfully genotyped with minor allele frequencies observed. [a]Positions in bp correspond to NCBI build 37.1 measured from the top of the p arm of the chromosome.

Of the six remaining SNPs analysed for association to obesity, rs3115759 and rs2281068 were nominally associated with obesity in children (p = 0.043 and 0.005). No SNPs were nominally associated with adult obesity (See Table 4.2). After correcting the p-values using one million permutations, rs2281068 remained significantly associated with obesity in children (p = 0.017, OR = 1.53, 95% CI = 1.14-2.05).

	Allele Counts						Odds Ratio Children (95% CI)	P-values Children		Odds Ratio Adults (95% CI)	P-values Adults	
	Controls		Children		Adults							
	1[a]	2	1	2	1	2		Empirical	Corrected		Empirical	Corrected
rs3115759	1730	136	871	91	1404	133	1.3 (1.0-1.8)	**0.043**	0.14	1.2 (0.94-1.6)	0.15	0.36
rs3115758	1598	136	848	89	1392	132	1.2 (0.93-1.6)	0.14	0.37	1.1 (0.87-1.4)	0.40	0.77
rs2235307	1564	74	867	56	1421	86	1.4 (0.96-2.0)	0.088	0.24	1.3 (0.93-1.8)	0.14	0.34
rs2235308	1688	131	849	86	1359	132	1.3 (0.98-1.7)	0.065	0.19	1.3 (0.97-1.6)	0.083	0.23
rs2235312	1709	131	847	86	1373	134	1.3 (1.0-1.8)	0.051	0.16	1.3 (0.99-1.6)	0.061	0.17
rs2281068	1600	108	854	88	1388	123	1.5 (1.1-2.1)	**0.0046**	**0.017**	1.3 (1.0-1.7)	0.051	0.14

Table 4.2. Allelic association analysis of Apelin SNPs to obesity in children and adult French case-controls. Analysis was performed using PLINK. Empirical p-values result from a single chi-squared test, corrected p-values result from one million permutations. Allele numbers are presented instead of genotypes due to *APLNs* position on the X chromosome. [a] 1 denotes the common allele, 2 denotes the minor allele.

A high degree of LD was found throughout the gene with r^2 values ranging from 0.61 to 1.0 (see Figure 4.2). Haplotypes constructed from the two nominally associated SNPs were significantly associated to obesity in children after correction using permutations (p = 0.016, OR = 1.48)(see Table 4.3).

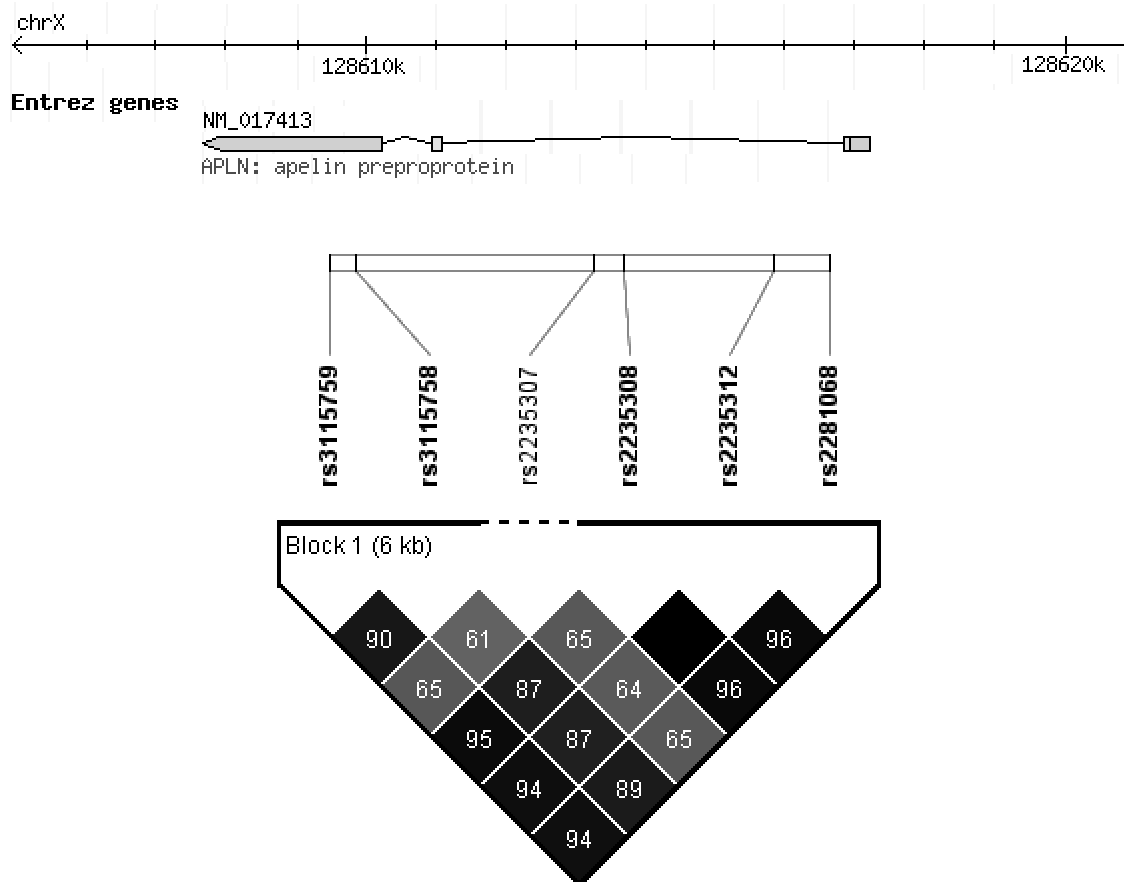

Figure 4.2. Linkage disequilibrium (r^2) plot of *APLN* SNPs genotyped in the French female control samples. Black squares indicate 100% LD. The percentage LD level (r^2) within grey squares are displayed.

SNP 1	SNP 2	Odds Ratio	Empirical P-value	Corrected P-value
rs2281068	rs3115759	1.48	**8.0×10^{-3}**	**0.016[a]**

Table 4.3 Haplotype analysis in the child case-controls using the two SNPs nominally associated with obesity. [a] Statistically significant p-values. Corrected p-values result from one million permutations.

APLN SNPs were analysed for association to BMI within the controls using age and sex as covariates, however no significant associations were discovered (see Table 4.4).

	Mean BMI (95% CI) (kg/m^2)			**P-Values**	
SNP	**1,1[a]**	**1,2**	**2,2**	**Empirical**	**Corrected**
rs3115759	32.1 (31.6-32.6)	33.7 (32.1-35.3)	31.4 (29.2-33.5)	0.19	0.44
rs3115758	32.5 (31.9-33.0)	33.0 (31.5-34.5)	31.6 (29.3-34.0)	0.40	0.76
rs2235307	32.6 (32.1-33.1)	34.8 (32.8-36.9)	31.6 (28.9-34.3)	0.37	0.72
rs2235308	32.0 (31.5-32.5)	33.8 (32.2-35.4)	31.4 (29.0-33.7)	0.14	0.34
rs2235312	31.9 (31.5-32.4)	33.9 (32.3-35.5)	31.4 (29.0-33.7)	0.11	0.28
rs2281068	32.4 (31.9-32.9)	34.3 (32.6-36.0)	31.7 (29.4-34.0)	0.060	0.17

Table 4.4. Analysis of *APLN* SNPs for association with the quantitative trait of BMI in French controls using linear regression with age and sex as covariates. Corrected p-vales result from one million permutations. [a] 1 denotes the common allele, 2 denotes the minor allele.

4.2.2 Sequencing of *APLN*

In order to locate the genetic variant or variants that are responsible for this association, sequencing was carried out within the *APLN* exons and promoter region. Ten subjects were chosen that were homozygous for both minor alleles associated with obesity, in rs3115759 and rs2281068, and two subjects were chosen as controls that were homozygous for the common alleles of rs3115759 and rs2281068. Six PCR reactions were designed to cover the 1.5kbp region upstream of the transcription start site as well as the two translated *APLN* exons. The PCR products were then directly sequenced.

One SNP was discovered within the 5' untranslated region of exon 1. This was previously reported as rs2281069 (MAF < 0.01 in Sub-Saharan African population, no data for Caucasians). No SNPs were discovered within exon 2 or within the promoter region (see Figure 4.3). Because it is untranslated, it is unlikely to be causative and so was not taken forward for genotyping.

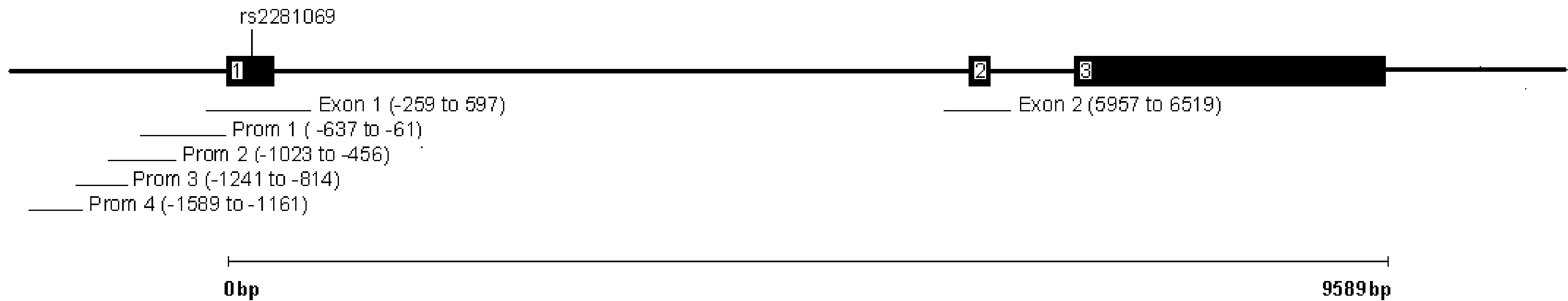

Figure 4.3. Map of *APLN* gene showing positions of PCR products used to sequence the promoter region and the two translated exons. Positions in bp correspond to NCBI build 37.1. █ *APLN* exons.

4.2.3 Thesis-wide multiple testing correction

When *APLN* SNPs were analysed together with all other SNPs investigated in this thesis, no significant associations remain after correction using permutations (see Table 4.5).

SNP	Epirical p-value children	Corrected p-value children	Epirical p-value adults	Corrected p-value adults
rs3115759	0.04	0.89	0.15	1.0
rs3115758	0.14	1.0	0.40	1.0
rs2235307	0.088	0.99	0.14	1.0
rs2235308	0.065	0.97	0.083	0.98
rs2235312	0.051	0.94	0.061	0.95
rs2281068	**0.0046**	0.22	0.051	0.90

Table 4.5. P-values after correcting for multiple testing using permutations and taking into account all SNPs investigated in this thesis.

4.2.4 Power calculations

Using an odds ratio of 1.53 (as observed in the most significantly associated SNP, rs2281068), a MAF of 0.063 (as observed in rs2281068 in the controls), and a significance level of 0.0083 (0.05 divided by 6 tests), the power to detect an association was 40% in the child cohort and 15% in the adult cohort (see Table 4.6).

SNP	Significance level	MAF in controls	Odds ratio in children	Odds ratio in adults	Power in children	Power in adults
rs3115759	0.0083	0.073	1.3	1.2	0.17	0.075
rs3115758	0.0083	0.078	1.2	1.1	0.08	0.025
rs2235307	0.0083	0.045	1.4	1.3	0.13	0.085
rs2235308	0.0083	0.072	1.3	1.3	0.14	0.11
rs2235312	0.0083	0.071	1.3	1.3	0.16	0.13
rs2281068	0.0083	0.063	1.5	1.3	0.40	0.15

Table 4.6. Power calculations. carried out using MAF and OR observed for each SNP investigated in the APLN gene.

To achieve a statistical power of 95%, assuming a significance level of 0.0083, odds ratio of 1.53 and a MAF of 0.06, 5,748 samples would be required (2874 cases and 2874 controls). Assuming a significance level of $5x10^{-8}$ (genome-wide significance) this rises to 16,532 (see Figure 4.4). If the real odds ratio is lower then the samples size requirements will be increased (see Figure 4.5).

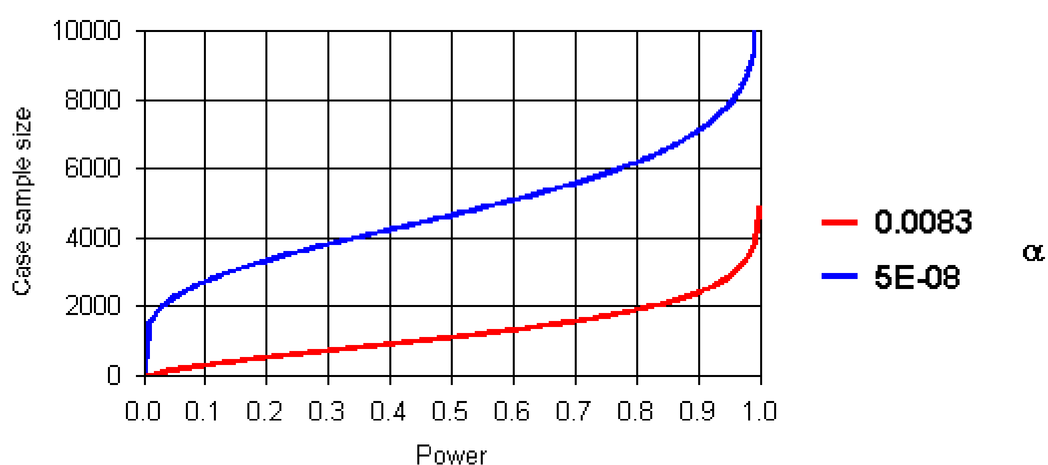

Figure 4.4. Sample size plotted with statistical power for a SNP with MAF=0.06 (as was observed for rs2281068). α is the significance level: 0.0083 for 6 statistical tests and $5x10^{-8}$ for genome-wide significance.

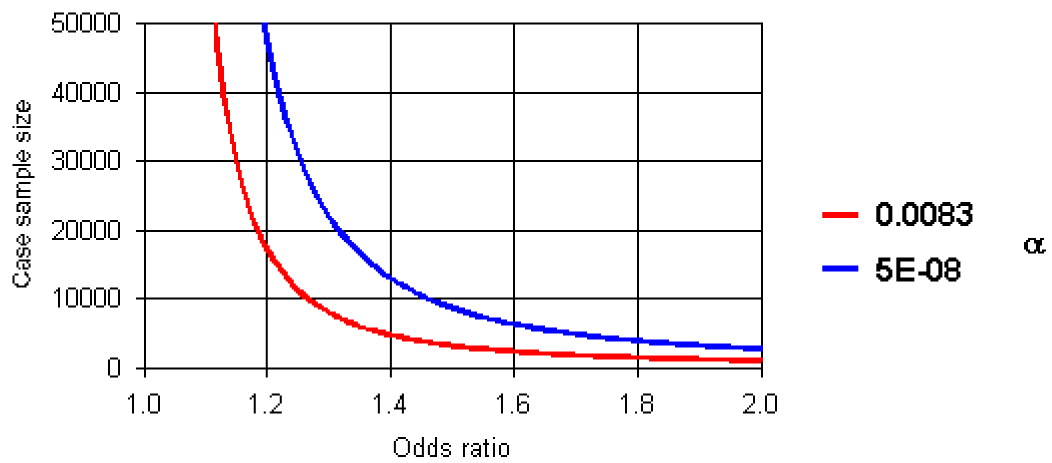

Figure 4.5. Sample size required for varying odds ratios in order to achieve a statistical power of 95% for a SNP with MAF=0.06 (as was observed for rs2281068). α is the significance level: 0.0083 for 6 statistical tests and $5x10^{-8}$ for genome-wide significance.

For the sequencing project, the probability of locating a common variant (MAF>5%) in high LD (r^2>0.8) with the associated SNP (rs2281068, MAF=6.3%) using 10 samples was calculated to be >99%, assuming such a variant exists. The probability of locating a variant with MAF = 0.1% in moderate LD (r^2=0.5) was calculated to be 84% (see Appendix A7 for calculations).

4.3 Discussion

These results provide evidence of a possible association between one SNP in the Apelin gene and common polygenic obesity in French children.

The p-values reported here do not withstand genome wide correction. Furthermore, permutation analysis using all SNPs tested within this thesis results in non-significance of all tests. This, together with the fact that the association has not been replicated means that the possibility of a false-positive cannot be ruled out. SNPs were analysed using permutations to compensate for the six statistical tests being performed and this resulted in one putative association to obesity in children.

As noted in the introduction to this chapter, a nominal association of the Apelin SNP rs2235306 was reported to be associated with fasting plasma glucose levels in Han Chinese ($p = 0.04$), however this study did not find an association to any obesity related phenotype[472]. In the results presented here, rs2235306 had a low minor allele frequency (0.02) and since the previously reported association was not to obesity it was not analysed. More recently, an association was reported between an *APLN* SNP, rs3115757 and both BMI and WC in Chinese women but not men[474]. Rs3115757 was not genotyped in this study since it was not present in dbSNP at the time these assays were designed. It is not present in HapMap so there is no information on whether it is tagged by any genotyped SNP.

No associations between apelin SNPs and obesity phenotypes have been reported in any GWAS to-date. This is due to the fact that most GWAS exclude X chromosome variants from analysis and highlights a current weakness in genome-wide studies.

Both nominally associated SNPs were found to have a higher frequency of the minor genotype within obese subjects compared to controls suggesting that the minor allele increases obesity risk. The putatively associated SNP, rs2281068 is located within intron 1 of the Apelin gene so is unlikely to be causative unless it affects an enhancer element located within the intron. Therefore, if this is a genuine association, it is probably in linkage disequilibrium with some other mutation responsible for modifying obesity risk.

Sequencing was unable to locate any mutations in LD with the associated SNP that were either in the promoter (1.5kbp upstream of the start codon) or translated regions of the gene. One SNP, rs2281069 was discovered in the 5' UTR. Since this is untranslated and therefore unlikely to be functional it was not genotyped in the case-control subjects. Power to detect variants with MAF >5% and in high LD (r^2>0.8) with the associated SNPs was sufficiently high (>99%), however as MAF and LD values drop (MAF=0.1%, r^2=0.5) this power is reduced (84%) and if the causative SNP is even rarer (<0.1%) then this will drop further. Thus the possibility of a false-negative cannot be ruled out. Furthermore since only 2kb of the promoter region was sequenced, all regulatory regions have not been investigated.

Since no coding SNPs are currently known to exist in the human *APLN* gene and sequencing was not able to locate any variants in LD with the associated SNPs, it is likely that the causative variant is within the promoter region, upstream of the region sequenced in this study or within an enhancer or inhibitor element where it affects transcription of the gene. Using HapMap data the LD block containing *APLN* extends approximately 20kbp upstream of the gene and this region could contain variants that alter expression (see Figure 4.4). Alternatively, the causative SNP could affect splicing of the mRNA and aberrantly spliced transcript may not translate into active apelin peptides. As apelin levels are correlated to adiposity[353] and apelin injections have been shown to stumulate weight loss[343,344,468] an alteration in *APLN* expression or apelin protein synthesis could lead to a malfunction in the control of body-weight thus influencing obesity risk.

Further work is required in order to replicate the association in other cohorts and ethnic groups. If replication is successful then sequencing will be required in order to locate the causative variant. This should be carried out by sequencing more of the promoter region (at least 20kb upstream of the transcription start site) in a larger number of samples. Novel variants will then be typed in the current sample sets and any association will need to be replicated in other populations. After genotyping, any SNPs found to be significantly associated with obesity in the promoter region can then be assessed for their effect on expression of the *APLN* transcript using luciferase assays. An alternative to genotyping any discovered SNPs in all samples would be to impute genotypes from the sub-set of samples that have been sequenced. If this produced a significant result then genotypes could be validated afterwards which means the costs of genoytping are only incurred if a significant result is found.

Figure 4.4 LD plot (D') of *APLN* gene and the surrounding region. Bright red squares indicate D'=1 and LOD>2, shades of pink or red indicate D'<1 and LOD>2, blue indicate D'=1 and LOD<2 and white squares indicate D'<1 and LOD<2.

Chapter 5

Investigation of genetic variants within the Interleukin 11 gene for association to common polygenic obesity

5.1 Introduction

Interleukin 11 (IL-11), also known as adipogenesis inhibitory factor, is a member of the IL-6 family of anti-inflammatory cytokines and is expressed in a wide range of tissues including the CNS, bone, lung, uterus, testis and adipose[475]. IL-11 acts on a number of tissues including central nervous system, testis, bone marrow, intestine and adipose[354,475].

IL-11 is an important regulator of hematopoiesis and in particular megakaryocyte formation[475]. It stimulates the proliferation and differentiation of blood stem cells into mature cells and has been demonstrated to promote platelet recovery in patients undergoing chemotherapy[475]. As well as blood cells it has also been reported to regulate development of bone marrow, epithelial cells and hippocampal neuronal progenitor cells and is thought to be important in the establishment of pregnancy[475,476].

5.1.1 IL11 inhibits adipogenesis

IL11 inhibits the differentiation of pre-adipocytes into adipocytes in human bone marrow culture[358]. It also inhibits expression of lipoprotein lipase[477] and in bone marrow culture it has been shown to inhibit fat accumulation[358]. In fully differentiated adipocyte cell-culture, IL-11 activates a number of signalling pathways[355] and treatment of adipocyte cell culture with IL-11 lowers transcription of leptin which is evidence of an involvement in regulation of adiposity levels[357]. It has also been shown to inhibit TNFa, a pro-inflammatory cytokine that is upregulated in obesity[478]. Given its role in controlling the expression of inflammatory cytokines in adipose tissue it has been hypothesised that IL-11 could regulate insulin resistance associated with obesity[355].

5.1.2 *IL11* variation and obesity

To-date, no associations have been reported between IL-11 gene variants and obesity-related phenotypes. *IL-11* does however, lie within the 1-LOD support interval of a linkage peak found to obesity, which means it is a candidate gene for this linkage[479].

In the French GWAS, which used some of the same sample as this study[253], 16 SNPs within the IL11 region were found to have a p-value <0.05, with the most significant hit, rs8101393 ($p=1.7x10^{-3}$) a SNP 130,000bp downstream of the gene. In the most recent GIANT GWAS investigating associations to BMI[233], a SNP within the IL11 region, rs1042506, was found to be nominally associated however with a p-value of $4.16x10^{-2}$, however this is very unlikely to be a genuine genetic association (accessed online[445], see Figure 5.1).

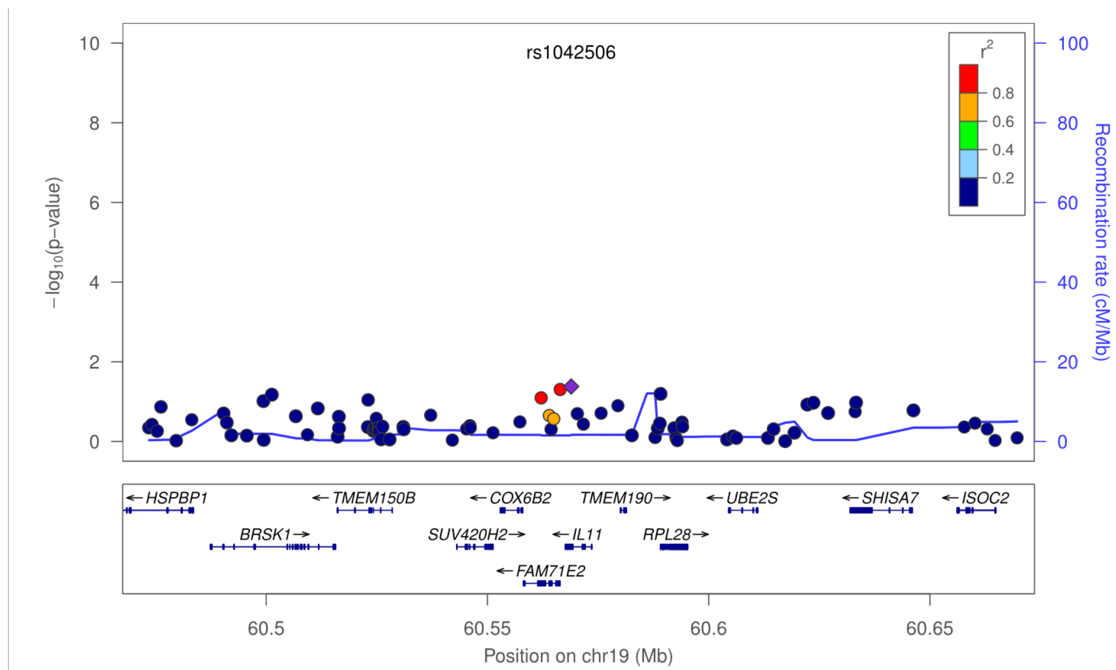

Figure 5.1. SNPs analysed for association to BMI in *IL11* region in the GIANT GWAS[233].

Given the evidence that IL-11 is involved in the development of adipocytes and the obesity linkage peak reported near to the gene, it is a plausible candidate for human polygenic obesity and as such this study was designed to investigate the possibility of genetic associations between variants in the gene, its expression levels and obesity in French obese cases and non-obese controls and Swedish sib-pairs (see Materials and Methods for details).

5.2 Results

5.2.1 SNP selection and case-control genotyping

No tagging SNPs were identified in the IL-11 gene due to a lack of strong linkage disequilibrium (r^2>0.8) in the HapMap genotype data. Seventeen SNPs were selected from the NCBI database (see Figure 5.1). After genotyping, sixteen SNPs survived quality control with a mean call rate of 89%. Of these, ten had low minor allele frequencies (< 5%), leaving six SNPs that were taken forward for analysis (see Table 5.1).

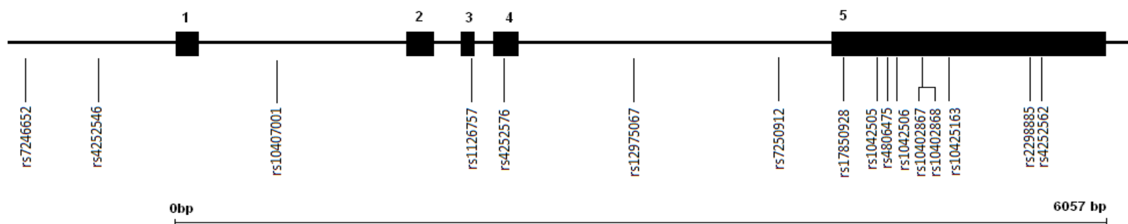

Figure 5.1. Schematic diagram of the *IL-11* gene showing the positions of SNPs successfully genotyped in this study. █ *IL-11* exons.

SNP	Position on Chromosome 19[a]	Position Within IL-11 Gene	Minor Allele Frequency
rs4252562	55876162	3' UTR	<0.01
rs2298885	55876240	3' UTR	0.32
rs10425163	55876773	3' UTR	<0.01
rs10402867	55876945	3' UTR	<0.01
rs10402868	55876946	3' UTR	<0.01
rs1042506	55877111	3' UTR	0.14
rs4806475	55877168	3' UTR	<0.01
rs1042505	55877231	3' UTR	0.32
rs17850928	55877451	Non-synonymous Coding Exon 5	<0.01
rs7250912	55877874	Intron 4	<0.01
rs12975067	55878843	Intron 4	<0.01
rs4252576	55879685	Non-synonymous Coding Exon 4	<0.01
rs1126757	55879872	Synonymous Coding Exon 3	0.44
rs10407001	55881161	Intron 2	0.11
rs4252546	55882345	Upstream	0.48
rs7246652	55882807	Upstream	<0.01

Table 5.1. *IL-11* SNPs successfully genotyped with minor allele frequencies observed. [a] Positions in bp correspond to NCBI build 37.1 measured from the top of the p arm of the chromosome.

Four SNPs were found to be nominally associated with obesity in the French adults (see Table 5.2). These were rs2298885 (p = 0.024, OR = 1.2, CI =1.0-1.4), rs1042506 (p = 0.045, OR = 0.82, CI = 0.68-1.0), rs1042505 (p = 0.003, OR = 1.2, CI = 1.1-1.4) and rs1126757 (p = 0.042, OR = 0.88, CI = 0.77-1.0). Of these, one SNP, rs1042505 was significantly associated with obesity after correction using permutations (p=0.016). None of these SNPs were associated with obesity in the French children (see Table 5.2).

No associations were discovered between SNPs in *IL-11* and the quantitative trait of BMI in the French controls (see Table 5.3).

	Genotype Numbers														
	Controls			Children			Adults			Odds Ratio Children (95% CI)	P-values Children		Odds Ratio Adults (95% CI)	P-values Adults	
SNP	**1,1**[a]	**1,2**	**2,2**	**1,1**	**1,2**	**2,2**	**1,1**	**1,2**	**2,2**		**Empirical**	**Corrected**		**Empirical**	**Corrected**
rs2298885	517	464	112	266	245	58	297	333	84	1.0 (0.87-1.2)	0.89	1.0	1.2 (1.0-1.4)	**0.024**	0.11
rs1042506	838	280	28	461	132	12	620	158	20	0.86 (0.7-1.1)	0.16	0.67	0.82 (0.68-1.0)	**0.045**	0.20
rs1042505	548	485	101	275	256	58	327	373	88	1.1 (0.91-1.2)	0.43	1.0	1.2 (1.1-1.4)	**0.0027**	**0.016**[b]
rs1126757	337	586	229	189	289	130	268	404	139	0.99 (0.86-1.1)	0.93	1.0	0.88 (0.77-1.0)	**0.042**	0.19
rs10407001	888	239	15	483	117	8	638	149	15	0.92 (0.74-1.2)	0.46	1.0	0.94 (0.77-1.2)	0.56	0.98
rs4252546	311	576	253	165	281	135	210	373	202	1 (0.87-1.2)	0.99	1.0	1.1 (0.95-1.2)	0.22	0.66

Table 5.2. Allelic association analysis of *IL-11* SNPs to obesity in children and adult French case-controls. Analysis was performed using PLINK. Empirical p-values result from a single chi-squared test, corrected p-values result from one million permutations. [a] 1 denotes the common allele, 2 denotes the minor allele. [b] Statistically significant ($p < 0.05$) associations are shown in bold.

	Genotype Counts			Mean BMI (95% CI)			P-values	
SNP	**1,1[a]**	**1,2**	**2,2**	**1,1**	**1,2**	**2,2**	**Empirical**	**Corrected**
rs2298885	505	448	112	22.6 (22.5-22.8)	22.3 (22.1-22.4)	22.4 (22.3-22.5)	0.69	1.0
rs1042506	810	275	27	21.6 (21.5-21.8)	22.4 (22.3-22.5)	22.5 (22.3-22.6)	0.52	1.0
rs1042505	535	468	101	22.8 (22.7-22.9)	22.3 (22.2-22.4)	22.4 (22.3-22.5)	0.71	1.0
rs1126757	330	568	220	22.5 (22.4-22.7)	22.4 (22.3-22.5)	22.4 (22.3-22.5)	0.78	1.0
rs10407001	861	234	14	22.3 (22.2-22.4)	22.4 (22.3-22.6)	22.4 (22.3-22.6)	0.83	1.0
rs4252546	303	556	251	22.5 (22.4-22.6)	22.4 (22.2-22.5)	22.4 (22.2-22.5)	0.32	0.96

Table 5.3. Association analysis of *IL-11* SNPs to the quantitative trait of BMI within the French controls using linear regression and age and sex as covariates. Corrected p-values result from one million permutations. [a] 1 denotes the common allele, 2 denotes the minor allele.

Two SNPs were found in high LD, rs2298885 and rs1042505 (r^2=0.95), though LD was otherwise very low (see Figure 5.2). Haplotype analysis was carried out using the four SNPs nominally associated with obesity in adults. All haplotypes were nominally associated with obesity in adults, however only one haplotype survived correction (rs1126757 and rs1042505; p=0.025; OR=1.23) (see Table 5.4).

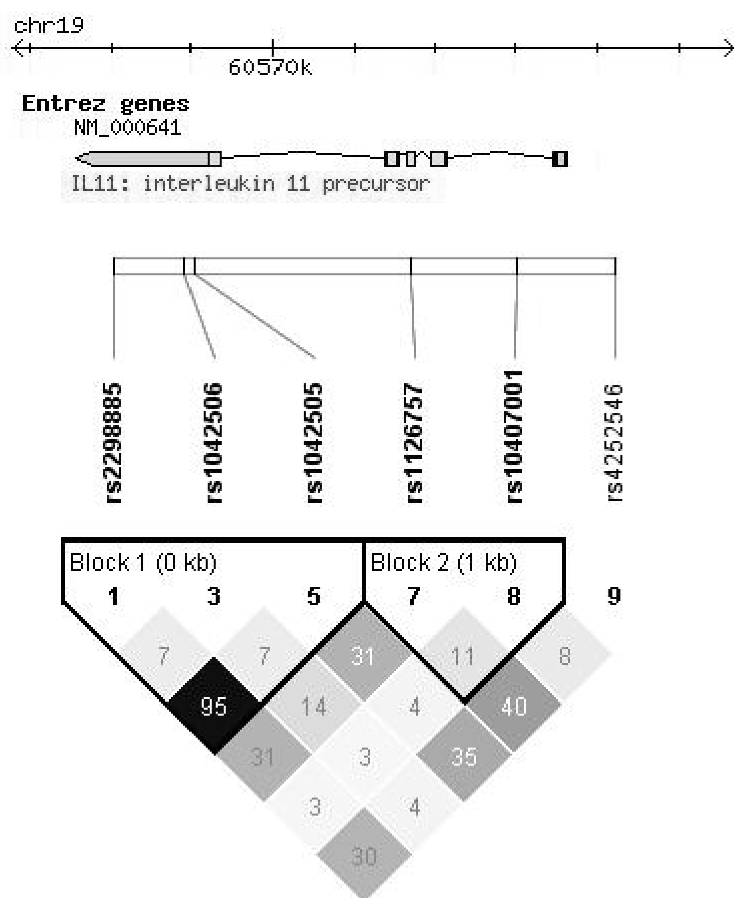

Figure 5.2. Linkage disequilibrium (r^2) plot of *IL-11* SNPs genotyped in the French control samples. Black squares indicate 100% LD and white squares indicate zero LD. The percentage LD level (r^2) within grey squares are displayed.

SNPs				OR	Empirical P-value	Corrected P-value
rs2298885	rs1042506	rs1042505	rs1126757	1.2	**0.011**	0.13
rs2298885	rs1042506	rs1042505		1.2	**0.0081**	0.11
rs2298885	rs1042506			1.2	**0.016**	0.18
rs2298885	rs1042505			1.2	**0.017**	0.18
rs2298885	rs1126757			1.2	**0.028**	0.13
rs1042506	rs1042505			1.2	**0.0057**	0.066
rs1042506	rs1126757			0.81	**0.033**	0.32
rs1042505	rs1126757			1.2	**0.0059**	**0.025**[a]

Table 5.4. Haplotype association analysis with obesity in adults using the four SNPs found to be nominally associated with adult obesity. [a] Statistically significant p-values. Corrected p-values result from one million permutations.

5.2.2 Genotyping within Swedish sib-pair cohort

Genotyping was carried out in the Swedish sib-pair cohort using the same set of assays. Genotyping success rate averaged 93% for the four SNPs nominally associated within the French case-control analysis. These four SNPs were analysed for association to BMI with the QFAM test using the within-families model of association. No associations were found with any *IL-11* SNP (see Table 5.5).

	Genotype Counts			Mean BMI (95% CI)			P-Values	
SNP	**1,1[a]**	**1,2**	**2,2**	**1,1**	**1,2**	**2,2**	**Empirical**	**Corrected**
rs2298885	335	235	42	29.2 (28.4-29.9)	29.2 (28.3-30.1)	26.7 (25.2-28.3)	0.74	1.0
rs1042506	436	180	15	28.8 (28.2-29.4)	29.3 (28.3-30.3)	29.5 (25.7-33.2)	0.81	0.86
rs1042505	348	235	42	29.5 (28.8-30.2)	28.9 (28.1-29.8)	26.5 (25.0-28.0)	0.85	0.86
rs1126757	138	312	143	29.0 (27.8-30.3)	29.4 (28.6-30.1)	28.2 (27.1-29.2)	0.75	0.85

Table 5.5. Association analysis of *IL-11* SNPs with BMI in the Swedish families using the QFAM test and the within-families model of association correcting for age and sex as implemented by PLINK. Corrected p-values result from one million permutations. [a] 1 denotes the common allele and 2 denotes the rare allele.

5.2.3 Analysis of *IL-11* transcription

Microarray data was available that included measurements of *IL-11* transcript levels in subcutaneous adipose tissue for each sibling within the Swedish sib-pair cohort. This was analysed for association with obesity, BMI and *IL-11* genotype.

There was no association between SNPs in the IL-11 gene and its transcript after correcting for age, sex and BMI (see Table 5.6). When siblings were split into two groups of non-obese (BMI<30 kg/m^2) and obese (BMI>30kg/m^2) and the transcription levels corrected for age, sex and relatedness, the two groups were normally distributed as assessed using a normal Q-Q plot. Using an independent samples t-test, a significant difference in the level of IL-11 transcription was found between the two groups ($p=2.8 \times 10^{-9}$, see Figure 5.3 and Table 5.7). When corrected, *IL-11* transcript was plotted against BMI and a weak correlation was observed in the non-obese group ($r^2=0.034$, p=0.005, see Figure 5.4). There was no correlation in the obese group ($r^2= 0.002$, p=0.42, see Figure 5.5).

	Genotype Counts			Mean Transcript Level (95% CI)			P-Values	
SNP	**1,1**[a]	**1,2**	**2,2**	**1,1**	**1,2**	**2,2**	**Empirical**	**Corrected**
rs2298885	174	128	24	5.8 (5.7-5.9)	5.7 (5.6-5.9)	5.9 (5.5-6.3)	0.64	0.63
rs1042506	235	101	6	5.8 (5.7-5.9)	5.8 (5.6-5.9)	5.7 (4.9-6.5)	0.38	0.29
rs1042505	184	131	24	5.8 (5.7-5.9)	5.8 (5.6-5.9)	5.9 (5.5-6.3)	0.36	0.32
rs1126757	74	170	80	5.8 (5.6-6.0)	5.7 (5.6-5.8)	5.9 (5.7-6.1)	0.63	0.62

Table 5.6. Association analysis of *IL-11* SNPs with *IL-11* transcript expression in subcutaneous adipose tissue, corrected for age, sex and BMI in the Swedish families using the QFAM test and the within-families model of association implemented by PLINK. [a] 1 denotes the common allele and 2 denotes the rare allele.

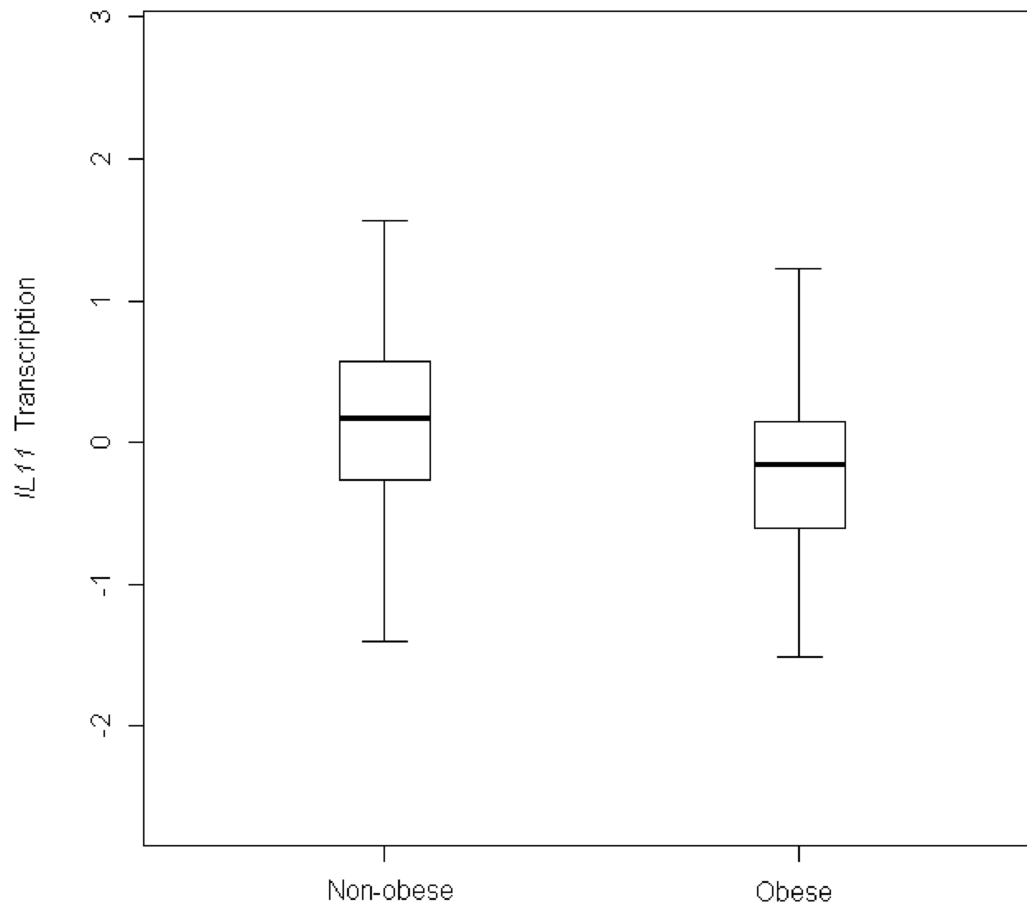

Figure 5.3. Box-plot of *IL-11* transcript levels in subcutaneous adipose tissue with obese and non-obese subjects from Swedish sib-pairs displayed separately. *IL-11* transcription is the microarray signal value corrected for age, sex and relatedness. Mean transcription ± standard deviation were 0.17 ± 0.60 in the non-obese and -0.22 ± 0.60 in the obese group.

Number of Non-obese	Number of Obese	Mean Difference in *IL-11* transcription (95% CI)	P-value
193	138	0.39 (0.27-0.52)	2.8×10^{-9}

Table 5.7. T-test of subcutaneous adipose *IL-11* transcription in obese and non-obese subjects from Swedish sib-pair families. *IL-11* transcription is the microarray signal value corrected for age, sex and relatedness.

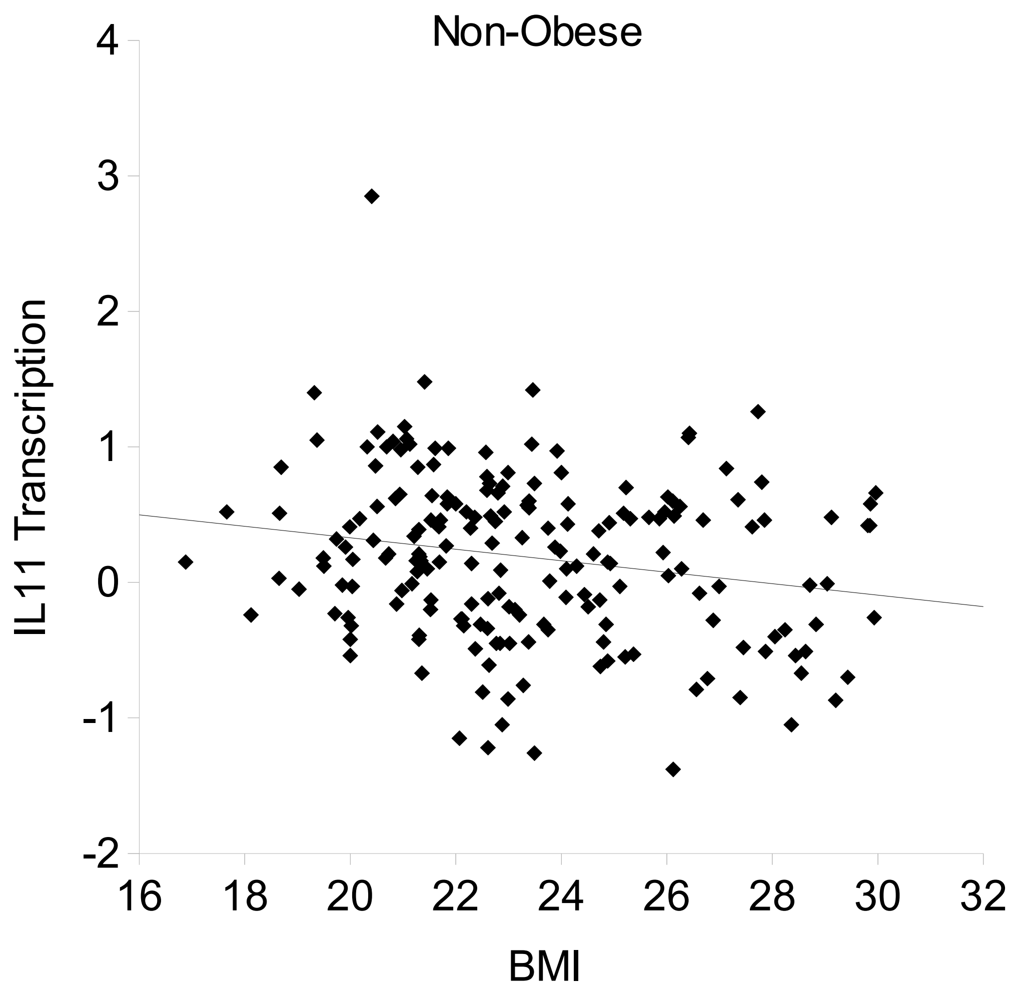

Figure 5.4. *IL-11* transcription levels in subcutaneous adipose tissue plotted against BMI in non-obese siblings (n = 193, r^2 = 0.034, p = 0.005). *IL-11* transcription was the DNA microarray signal value corrected for age, sex and relatedness.

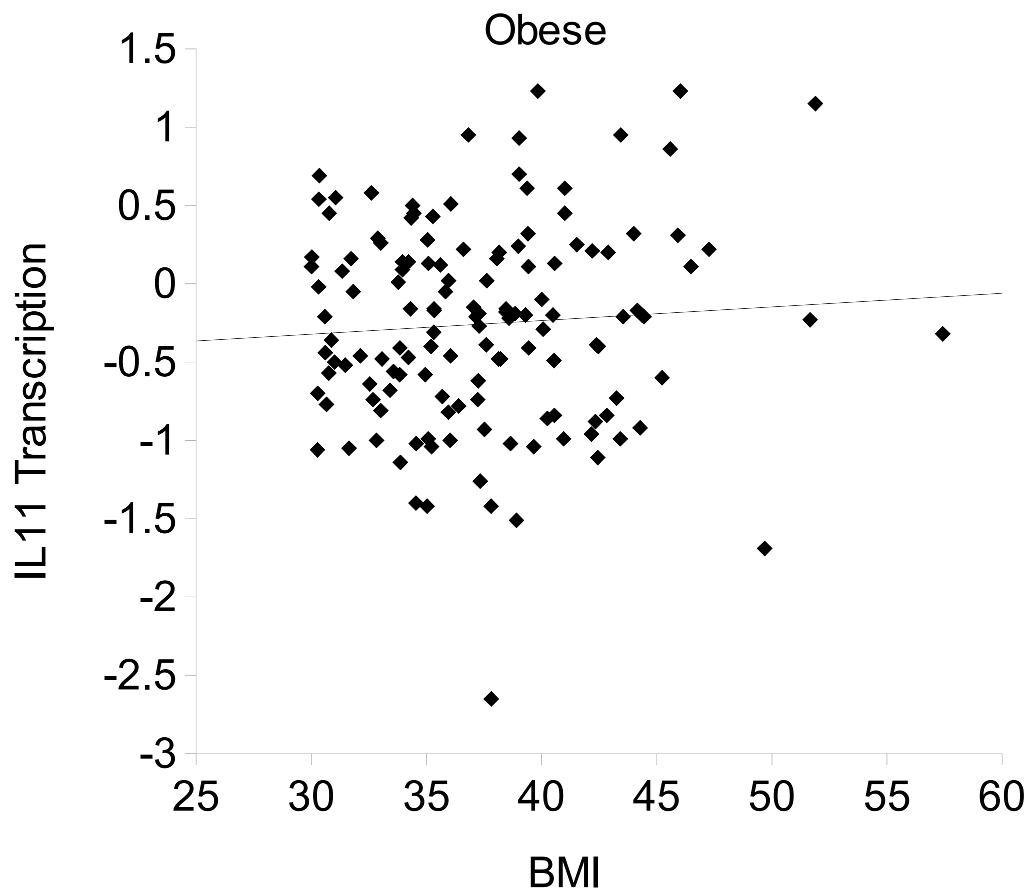

Figure 5.5. *IL-11* transcription levels in subcutaneous adipose tissue plotted against BMI in obese siblings (n = 138 ,r^2 = 0.002, p = 0.42) . *IL-11* transcription was the DNA microarray signal value corrected for age, sex and relatedness.

5.2.4 Thesis-wide multiple testing correction

When *IL11* SNPs were analysed together with all other SNPs investigated in this thesis, no significant associations remain after correction using permutations (see Table 5.8).

SNP	Empirical p-value children	Corrected p-value children	Empirical p-value adults	Corrected p-value adults
rs2298885	0.89	1.0	0.024	0.76
rs1042506	0.16	1.0	0.045	0.82
rs1042505	0.43	1.0	0.0027	0.18
rs1126757	0.93	1.0	0.042	0.95
rs10407001	0.46	1.0	0.56	1.0
rs4252546	0.97	1.0	0.22	1.0

Table 5.8. Thesis-wide correction of p-values for muliple testing using 1 million permutations from association analysis of IL11 SNPs and obesity in French case-controls.

5.2.5 Power calculations

Using an odds ratio of 1.23 (as observed in the most significantly associated SNP, rs1042505), a MAF of 0.30 (as observed in rs1042505 in the controls), and a significance level of 0.0083 (0.05 divided by 6 tests), the power to detect an association was 68% in the adult cohort and 13% in the child cohort (see Table 5.9).

SNP	Significance level	MAF in controls	Odds ratio in children	Odds ratio in adults	Power in children	Power in adults
rs2298885	0.0083	0.31	1.0	1.2	0.084	0.52
rs1042506	0.0083	0.15	0.86	0.82	0.25	0.42
rs1042505	0.0083	0.30	1.1	1.2	0.13	0.68
rs1126757	0.0083	0.45	0.99	0.88	0.085	0.39
rs10407001	0.0083	0.12	0.92	0.94	0.13	0.11
rs4252546	0.0083	0.47	1.0	1.1	0.083	0.23

Table 5.9. Power calculations carried out using MAF and OR observed for each SNP investigated in the IL-11 gene.

To achieve 95% power assuming a significance level of 0.0083, an odds ratio of 1.18 and a MAF of 0.31 would require 12316 samples (6158 cases and 6158 controls). Assuming a significance level of 5×10^{-8} (genome-wide significance) this rises to 33,386 (see figure 5.6). If the real odds ratio is lower then the sample size requirements go up (see Figure 5.7).

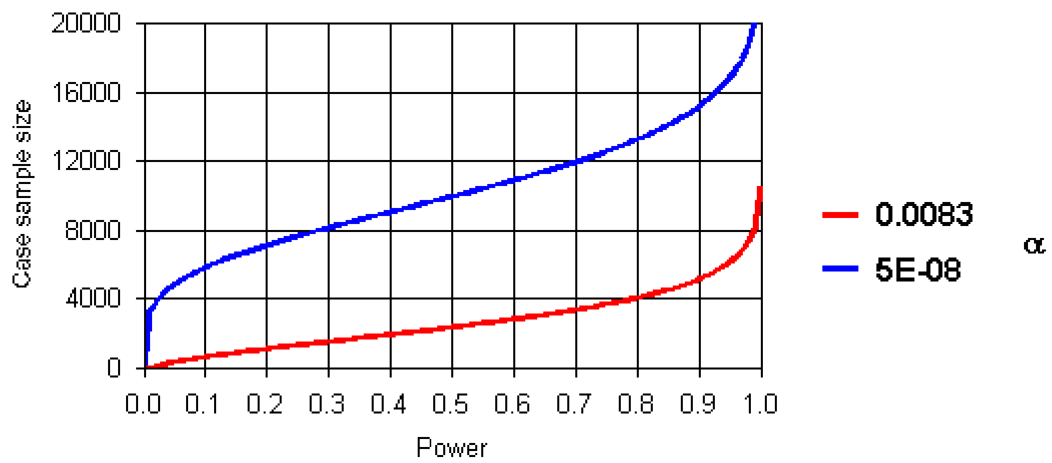

Figure 5.6. Sample size plotted with statistical power for a SNP with MAF= 0.30 (as was observed for rs1042505). α is the significance level: 0.0083 for 6 statistical tests and $5x10^{-8}$ for genome-wide significance.

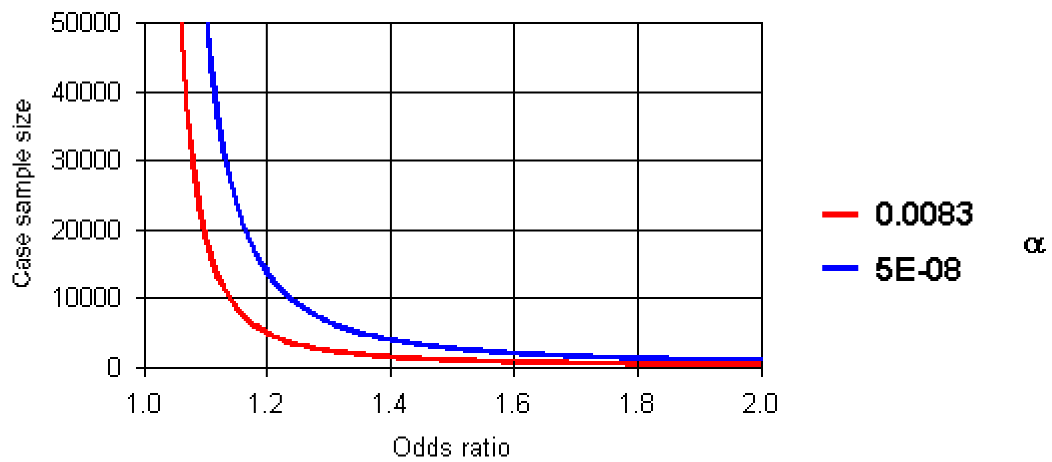

Figure 5.7. Sample size required for varying odds ratios in order to achieve a statistical power of 95% for a SNP with MAF= 0.30 (as was observed for rs1042505). α is the significance level: 0.0083 for 6 statistical tests and $5x10^{-8}$ for genome-wide significance.

5.3 Discussion

These results demonstrate a possible association between one SNP in the IL-11 gene and common polygenic obesity in French adults. Replication in other cohorts is necessary in order to confirm this association. No previous associations of variants in this gene to human obesity or obesity-related phenotypes have been reported to date, which would make this a novel discovery if it were to become validated.

The association does not withstand genome-wide correction for multiple testing, which means that the possibility of a false-positive cannot be ruled out.. For the purposes of this study, genome-wide correction was not considered necessary as IL-11 is a strong biological candidate gene with well-established functional links to adipose tissue and hence obesity. Rs1042505 was found to be significantly associated to obesity after correction using one million permutations. However, future replication will need to reach genome-wide significance in order to confirm the association.

LD between SNPs genotyped in IL-11 was low, which confirms the HapMap data in which LD was too low to generate tag SNPs. Haplotypes were still analysed in order to test whether associated SNPs formed part of a haplotype group that could contain a more significant SNP not typed in this project. All haplotypes constructed from the nominally associated SNPs were found to be nominally associated with obesity in adults, however only one haplotype was significantly associated after correction for multiple tests. This haplotype had a less significant p-value than the single significant SNP. This is not surprising given the low LD found within the gene. Therefore no conclusions can be drawn as to whether there is a more significant SNP in the region causing the association to obesity reported here.

The failure to find an association in the child case-control cohort could be because of the low statistical power. It is possible, however that this is a genetic variant that influences obesity risk in adults but not in children. Alternatively this could be evidence that this is not a real genetic association and merely a statistical artefact. The failure to find an association between SNPs and BMI in the French control groups could be because, if this is a genuine association, it affects disease risk and does not modify BMI in the normal-weight population. The failure to find an association between SNPs and BMI in the Swedish families might be due to the low statistical power in this dataset due to lower sample numbers (n = 661 individuals with BMI data) and the fact that subjects are related which means there is much less genetic variation. Alternatively it might be that this result is a false-positive.

The lack of a significant association to IL-11 variants in any GWAS to-date might be evidence that this is not a real association. Another explanation is that coverage of IL-11 SNPs is not complete in genome-wide studies. In the recent GIANT BMI publication[233], three SNPs within *IL-11* were genotyped: rs1042506 (p=0.042), rs4252552 (p=0.37) and rs1126757 (p=0.20). Only two of these SNPs were genotyped in the GIANT WHR paper[262] (rs1042506, p=0.86 and rs4252552, p=0.40). In this thesis, two of these SNPs were nominally associated with obesity in adults (rs1042506, p=0.0045, rs1126757 p=0.042). These were not the most significantly associated SNPs, however. Given the low LD within the gene, this means that coverage of *IL-11* variation in GWAS is not complete. The most significantly associated SNP in the French adult case-controls, rs1042505 is poorly correlated with these two SNPs (r^2=0.07 and 0.31 respectively) which means that if this is a genuine association, it will be missed by current GWAS, highlighting a weakness of current genome-wide investigations.

The significant difference observed in *IL-11* transcript levels between obese and non-obese siblings may be further evidence for a role of IL-11 in the development of obesity. However it is also possible that IL-11 transcription is being modified in some way by an individual's BMI. Since IL-11 has an immunological role its expression could be regulated differently in obesity, which is associated with chronic inflammation and activation of the immune system[480]. The lack of association between *IL-11* SNPs and its expression could again be due to the low number of samples in the Swedish sib-pair cohort (n=359 individuals with IL-11 transcript data). However, since the transcript data is quantitative, statistical power should be sufficient to detect any association so it is more likely that the *IL-11* SNP variation investigated is not associated with transcription of the gene.

The associated SNP, rs1042505 is located within the 3' UTR, therefore it does not alter the amino-acid sequence of the IL-11 protein. If this association is genuine then it is likely that this SNP is within LD with another variant that is causative. This could be a non-synonymous coding SNP, a variant within the promoter region or within an enhancer element that influences transcription. Alternatively, it could be a mutation that alters a splice site. Given the low level of LD found within the region this is less likely, however sequencing of the exons and promoter region should be carried out in order to locate any potentially causative mutations.

In summary, these results demonstrate a possible association between an IL-11 gene variant and obesity in a French Caucasian cohort. This will require further investigation in other populations in order to confirm the associations and sequencing and functional work to locate the causative variant.

Chapter 6

Investigation of genetic variants within the adiponutrin gene for association to common polygenic obesity

6.1 Introduction

Adiponutrin is a cell membrane protein encoded by the patatin-like phospholipase domain-containing protein 3 (PNPLA3) gene in humans, which is located on the long arm of chromosome twenty two and is comprised of eight exons. It belongs to a family of phospholipases that include patatin-like phospholipase 1 (PNPLA1), adipose triglyceride lipase (ATGL/PNPLA2), adiponutrin (ADPN/PNPLA3), gene sequence 2 (PNPLA4) and GS2-like (PNPLA5)[481]. All of these enzymes are highly expressed in adipose tissue[481].

6.1.1 Adiponutrin has a role in lipid metabolism

Adiponutrin is expressed in WAT and BAT and was originally reported to be a triacylglycerol lipase[360]. Lipase activity was subsequently confirmed in a study that also demonstrated acylglycerol transacylase activity (transfers fatty acids to mono- and di-acylglycerol)[482]. Overexpression of adiponutrin *in vitro* does not result in a decrease in intracellular tri-acylglyceride concentration[483].This suggests that adiponutrin may have a role in both lipid hydrolysis and lipogenesis.

6.1.2 Adiponutrin is regulated by metabolic factors

In vitro adiponutrin expression is increased during differentiation of pre-adipocytes and is induced by glucose and insulin treatment[360,484]. Expression in WAT of rodents is downregulated during fasting[360,485] and upregulated after re-feeding[484,485]. Additionally, expression is decreased in insulin-deficient mice and insulin-receptor knockout mice[484]. Expression is normalised in insulin-deficient mice when administered with insulin.

Zucker (*fa/fa*) rats that do not have a functional leptin receptor and over-express leptin have increased adiponutrin expression[360]. Conversely, expression of adiponutrin in *ob/ob* mice, that lack a functional leptin gene, is downregulated[483]. These results suggest that adiponutrin expression is controlled by leptin.

Similar results have been found in humans. Expression of adiponutrin in WAT is significantly increased during a hyperinsulinemic euglycemic clamp[486]. Expression of adiponutrin is increased in obese compared to non-obese subjects probably because of the association between high circulating insulin levels and obesity[369].

6.1.3 Adiponutrin gene variation and non-alcoholic fatty liver disease

A GWAS of nonsynonymous coding SNPs reported an association between rs738409 and hepatic fat levels and hepatic inflammation in a multiethnic population ($p=5.9\times10^{-10}$)[487]. This result was replicated in a Finnish cohort ($p = 0.002$) which also found an association between liver PNPLA3 transcription and BMI ($r^2=0.62$, $p<10^{-4}$)[364] in a European-American cohort[488]. The same SNP was also reported to be associated with various phenotypes that relate to the histological severity of the disease[489]. Another study in children and adolescents with nonalcoholic fatty liver disease reported an association between rs738409 and the same histological markers ($p<10^{-4}$)[490]. As nonalcoholic fatty liver disease is associated with obesity and metabolic syndrome, this supports a possible link between adiponutrin and obesity[491].

6.1.4 Adiponutrin gene variation and obesity

Expression of Adiponutrin in subcutaneous and visceral adipose tissue has been reported to be increased in obese compared to non-obese subjects in a Swedish study[369]. This study also reported an association between two SNPs, rs2072907 and rs1010022, in the Adiponutrin gene and obesity (p-values = 0.015 and 0.021), however these associations become non-significant after correcting for age. This association was subsequently replicated in a second Swedish cohort which survived the correction for age effects[492].

No significant associations between *PNPLA3* variants and obesity phenotypes have been reported in any GWAS to-date. In the French GWAS, which used some of the same sample as this study[253], 44 SNPs within the PNPLA3 region were found to have a p-value <0.05, with the most significant SNP, rs5764455 having a p-value of 3.5×10^{-4}. However, this SNP is located 55,000 bp downstream of the gene and lies within another gene, *PPARVB*. The most recent GIANT publication[233] found a nominally significant p-value of 2.19×10^{-3} with one SNP, rs5764317 within the gene region in which *PNPLA3* is located, however at least three other genes are located closer to this SNP than *PNPLA3* (data acessed online[445], see Figure 6.1). No SNPs located in the *PNPLA3* gene were nominally associated.

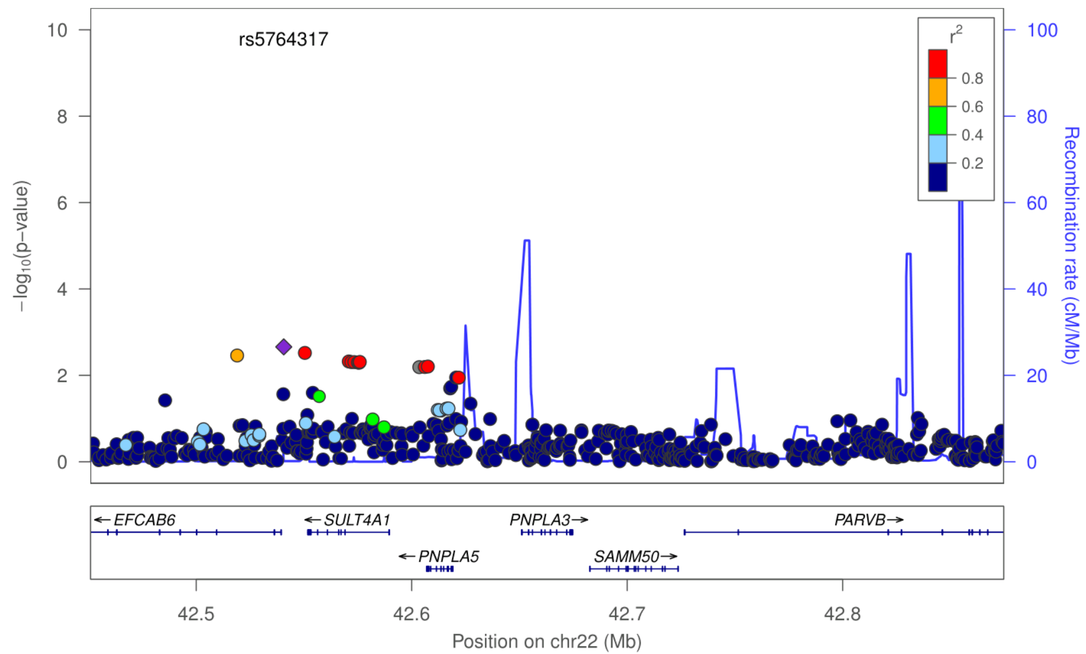

Figure 6.1. SNPs analysed for association to BMI in the PNPLA3 region in the GIANT GWAS[233].

Given its links to lipid metabolism and the genetic association previously reported, adiponutrin is a plausible candidate gene for obesity and so this study was designed to investigate possible genetic associations between variants in the gene and common polygenic obesity in French obese cases and non-obese controls (see Materials and Methods for details).

6.2 Results

6.2.1 SNP Selection and Case-Control Genotyping

Twenty-five SNPs were genotyped within the adiponutrin gene including twenty tag SNPs which covered all thirty-nine common adiponutrin SNPs (MAF>5%) in the HapMap database with a mean r^2 of 0.985 (see Figure 6.1). After genotyping, seventeen SNPs passed quality control with a mean call rate of 94%. Three of these had low minor allele frequencies (<5%) so were not analysed (see Table 6.1). Of the six tag SNPs genotyped, five passed QC. Rs11090617 failed due to low genotyping success rate.

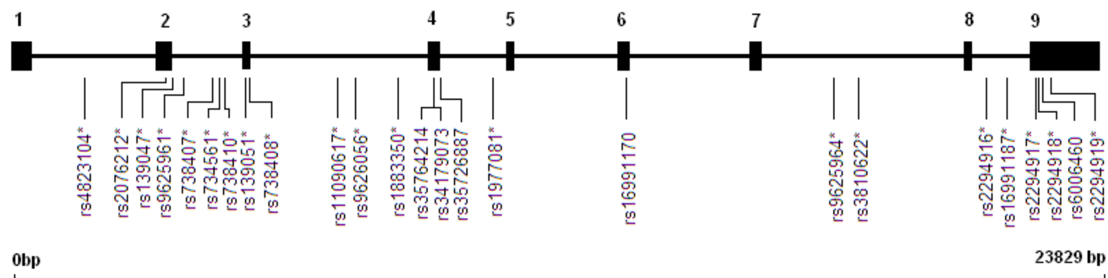

Figure 6.1. Map of adiponutrin gene with positions of SNPs genotyped in this project. * Tag SNPs. █ Adiponutrin exons.

SNP	Position on Chromosome 22[a]	Position Within Adiponutrin Gene	Minor Allele Frequency
rs4823104	44321149	Intron 1	0.21
rs2076212	44322970	Non-Synonymous Coding Exon 2	0.23
rs139047	44323074	Intron 2	0.48
rs734561*	44324104	Intron 2	0.34
rs139051	44324676	Intron 2	0.44
rs738408*	44324730	Synonymous Coding Exon 3	0.35
rs12483959	44325996	Intron 3	0.28
rs9626056	44327075	Intron 3	0.14
rs35764214	44328819	Non-Synonymous Coding Exon 4	<0.01
rs34179073	44328832	Synonymous Coding Exon 4	0.15
rs35726887	44328917	Non-Synonymous Coding Exon 4	<0.01
rs9625964	44337610	Intron 7	0.14
rs2294916*	44340922	Intron 8	0.27
rs2294917	44341986	Intron 8	0.41
rs2294918*	44342116	Non-Synonymous Coding Exon 9	0.48
rs6006460	44342174	Non-Synonymous Coding Exon 9	<0.01
rs2294919*	44342325	3' UTR	0.32

Table 6.1. SNPs successfully genotyped with minor allele frequencies observed. [a]Positions in bp correspond to NCBI build 37.1 measured from the top of the p arm of the chromosome.

One SNP, rs4823104, was nominally associated with obesity in children (p=0.043, OR=0.80, CI=0.64-0.99) but this was not significantly associated after adjusting for false positives using permutations (p=0.391). Another SNP, rs738408, was nominally associated with obesity in adults (p=0.045, OR=1.17, CI=1.01-1.36) but this did not survive correction using permutations (p=0.386) (see Table 6.2).

	Genotype Numbers														
	Controls			Children			Adults			Odds Ratio Children (95% CI)	P-values Children		Odds Ratio Adults (95% CI)	P-values Adults	
SNP	**1,1[a]**	**1,2**	**2,2**	**1,1**	**1,2**	**2,2**	**1,1**	**1,2**	**2,2**		**Empirical**	**Corrected**		**Empirical**	**Corrected**
rs4823104	932	267	21	505	121	4	656	166	8	0.80 (0.64-0.99)	**0.043**	0.39	0.86 (0.71-1.1)	0.14	0.78
rs2076212	890	298	12	476	137	14	658	195	18	0.98 (0.80-1.2)	0.81	1.0	0.99 (0.82-1.2)	0.87	1.0
rs139047	399	589	217	205	302	103	299	392	160	0.98 (0.85-1.1)	0.74	1.0	0.98 (0.87-1.1)	0.81	1.0
rs734561	731	391	52	380	195	22	487	288	33	0.92 (0.78-1.1)	0.36	0.99	1.0 (0.89-1.2)	0.67	1.0
rs139051	517	492	154	254	280	84	364	380	121	1.1 (0.94-1.3)	0.29	0.96	1.1 (0.94-1.2)	0.33	0.98
rs738408	660	328	65	345	201	28	483	329	47	1.0 (0.87-1.2)	0.72	1.0	1.2 (1.0-1.4)	**0.045**	0.39
rs12483959	852	333	30	439	174	16	598	245	29	1.0 (0.84-1.2)	0.91	1.0	1.1 (0.92-1.3)	0.34	0.98
rs9626056	1019	184	9	522	82	3	734	108	14	0.86 (0.67-1.1)	0.28	0.96	0.95 (0.76-1.2)	0.70	1.0
rs34179073	1031	186	14	504	99	2	729	129	6	0.98 (0.76-1.3)	0.84	1.0	0.93 (0.74-1.2)	0.54	1.0
rs9625964	934	148	12	517	90	4	736	126	7	1.0 (0.79-1.3)	0.86	1.0	1.0 (0.81-1.3)	0.82	1.0
rs2294916	792	296	51	421	169	25	586	237	29	1.0 (0.85-1.2)	0.85	1.0	0.99 (0.83-1.2)	0.86	1.0
rs2294917	556	491	116	305	248	64	417	343	88	0.97 (0.84-1.1)	0.71	1.0	0.98 (0.85-1.1)	0.75	1.0
rs2294918	414	581	232	205	293	104	298	409	145	0.96 (0.84-1.1)	0.61	1.0	0.94 (0.83-1.2)	0.35	0.98
rs2294919	762	402	58	383	197	30	552	266	40	1.0 (0.84-1.2)	0.97	1.0	0.94 (0.81-1.1)	0.46	1.0

Table 6.2. Allelic association analysis of Adiponutrin SNPs to obesity in children and adult French case-controls. Analysis was performed using PLINK. Empirical p-values result from a single chi-squared test, corrected p-values result from one million permutations. [a] 1 denotes the common allele, 2 denotes the minor allele.

LD analysed within the controls was low throughout the gene; no pair of SNPs had an r^2 value greater than 0.85 (see Figure 6.2). This reflects the fact that thirteen out of fourteen SNPs analysed were tag SNPs.

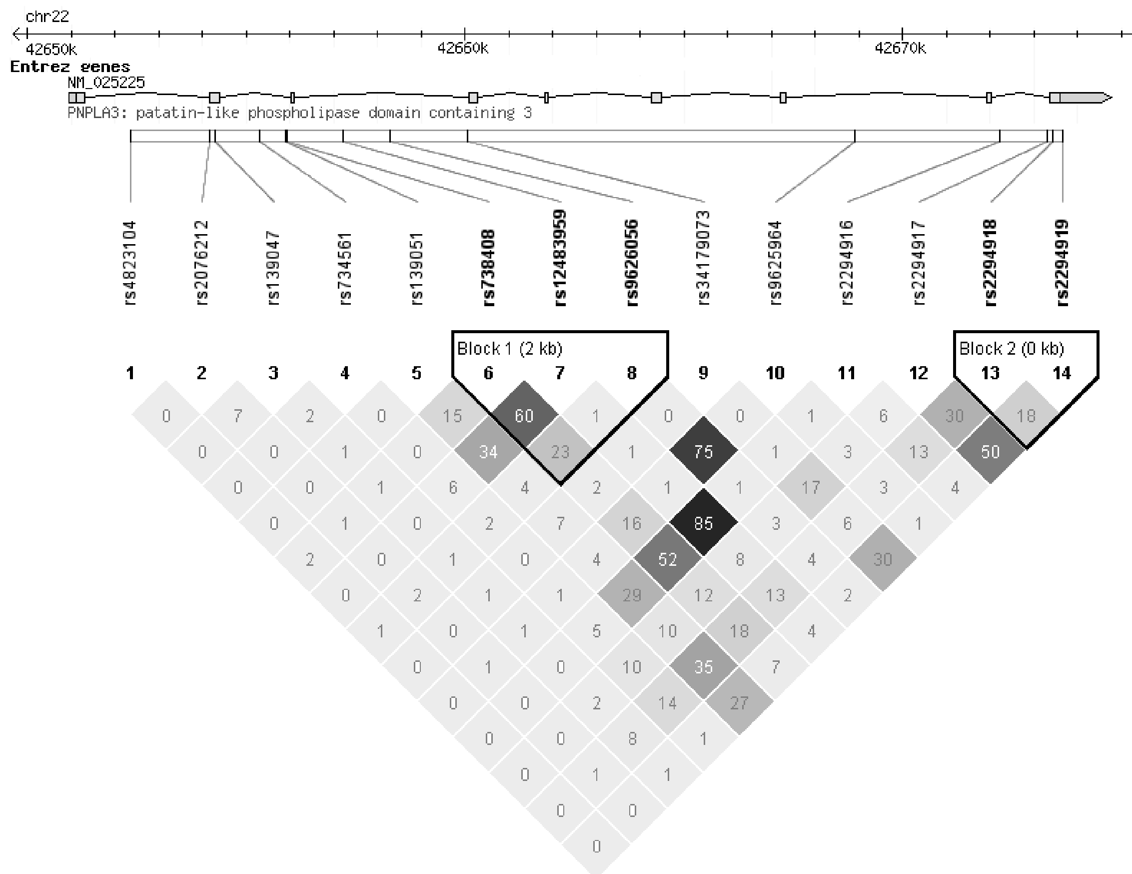

Figure 6.2. Linkage disequilibrium (r^2) plot of Adiponutrin SNPs genotyped in the French control samples. Black squares indicate 100% LD and white squares indicate zero LD. The percentage LD level (r^2) within grey squares are displayed.

One SNP, rs9626056 was found to be associated with BMI in the French controls ($p=0.02$), however this does not survive correction using permutations ($p=0.19$) (see Table 6.3).

	Genotype Numbers			Mean BMI (95% CI)			P-values	
SNP	**1,1[a]**	**1,2**	**2,2**	**1,1**	**1,2**	**2,2**	**Empirical**	**Corrected**
rs4823104	908	256	20	22.4 (22.2-22.5)	22.5 (22.3-22.8)	22.4 (21.5-23.3)	0.28	0.96
rs2076212	864	289	12	22.4 (22.2-22.5)	22.6 (22.3-22.8)	21.7 (20.7-22.7)	0.40	0.99
rs139047	389	573	207	22.5 (22.3-22.7)	22.3 (22.1-22.4)	22.3 (22.0-22.6)	0.22	0.92
rs734561	705	385	51	22.4 (22.2-22.6)	22.2 (22.0-22.4)	23.0 (22.3-23.6)	1.0	1.0
rs139051	502	478	150	22.4 (22.2-22.6)	22.4 (22.2-22.6)	22.2 (21.9-22.5)	0.36	0.99
rs738408	640	321	63	22.4 (22.2-22.5)	22.5 (22.3-22.8)	22.4 (21.9-22.8)	0.37	0.99
rs12483959	825	325	29	22.4 (22.2-22.6)	22.4 (22.2-22.6)	22.6 (21.8-23.3)	0.96	1.0
rs9626056	989	179	8	22.3 (22.2-22.4)	22.7 (22.4-23.0)	22.7 (21.9-23.6)	**0.019**	0.19
rs34179073	1000	180	14	22.4 (22.2-22.5)	22.4 (22.1-22.7)	22.6 (21.7-23.5)	0.57	1.0
rs9625964	908	145	11	22.3 (22.2-22.5)	22.6 (22.2-22.9)	23.3 (22.4-24.2)	0.060	0.48
rs2294916	766	289	49	22.4 (22.2-22.5)	22.4 (22.2-22.6)	22.6 (22.1-23.1)	0.60	1.0
rs2294917	541	474	114	22.5 (22.3-22.6)	22.3 (22.1-22.5)	22.4 (22.0-22.8)	0.29	0.97
rs2294918	405	558	227	22.4 (22.2-22.6)	22.3 (22.2-22.5)	22.5 (22.2-22.7)	0.70	1.0
rs2294919	741	389	56	22.5 (22.3-22.6)	22.3 (22.1-22.5)	22.2 (21.7-22.8)	0.16	0.82

Table 6.3. Association analysis of Adiponutrin SNPs to BMI in the French controls using linear regression with age and sex as covariates. Corrected p-values result from 1 million permutations. [a] 1 denotes the common allele, 2 denotes the minor allele.

6.2.2 Power calculations

Using an odds ratio of 1.17 (as observed in the most significantly associated SNP, rs738408), a MAF of 0.22 (as observed in rs738408 in the controls), and a significance level of 0.0036 (0.05 divided by 14 tests), the power to detect an association was 28% in the adult cohort and 4% in the child cohort (see Table 6.5).

SNP	Significance level	MAF in controls	Odds ratio in children	Odds ratio in adults	Power in children	Power in adults
rs4823104	0.0036	0.13	0.80	0.86	0.25	0.16
rs2076212	0.0036	0.13	0.98	0.99	0.037	0.036
rs139047	0.0036	0.42	0.98	0.98	0.040	0.041
rs734561	0.0036	0.21	0.92	1.0	0.080	0.043
rs139051	0.0036	0.34	1.1	1.1	0.093	0.089
rs738408	0.0036	0.22	1.0	1.2	0.042	0.28
rs12483959	0.0036	0.16	1.0	1.1	0.037	0.079
rs9626056	0.0036	0.083	0.86	0.95	0.097	0.045
rs34179073	0.0036	0.087	0.98	0.93	0.037	0.055
rs9625964	0.0036	0.079	1.0	1.0	0.038	0.039
rs2294916	0.0036	0.17	1.0	0.99	0.039	0.037
rs2294917	0.0036	0.31	0.97	0.98	0.043	0.040
rs2294918	0.0036	0.43	0.96	0.94	0.052	0.083
rs2294919	0.0036	0.21	1.0	0.94	0.036	0.067

Table 6.5. Power calculations carried out using MAF and OR observed for each SNP investigated in the adiponutrin gene.

In order to achieve a statistical power of 95%, assuming a significance level of 0.0036, an odds ratio of 1.2 and a MAF of 0.22, a sample size of 18,822 samples (9,411 cases and 9,411 controls) would be required. Assuming a significance level of $5x10^{-8}$ (genome-wide significance), 45,654 samples would be required.

6.3 Discussion

No evidence was found for an association between variants in the adiponutrin gene and common obesity in either adults or children after correcting for multiple testing. The association that was reported by Johansson *et al.*[369] between three adiponutrin SNPs (rs2072907, rs12483959 and rs139051) and obesity in Swedish Caucasians was not replicated. In this study rs2072907 is tagged by rs2294916 (r^2=1.0 using HapMap data), rs12483959 is tagged by rs11090617 (r^2=1.0 using HapMap data) and rs139051 was genotyped directly. rs11090617 failed QC and rs139051 was not significanlty associated, either before or after correcting for multiple tests. The reason that these data do not agree may be due to the relatively weak associations originally reported. P-values were greater than 0.01 and became non-significant after correcting for age. It is also possible that there was not sufficient statistical power to detect an association using the French case-control cohorts. Using the minor allele frequencies observed, statistical power is estimated at a maximum probability of 70% to detect an association assuming an odds ratio of 0.79 as was reported[369]. Furthermore, tag-SNP coverage of the adiponutrin gene was not complete in this study since rs11090617 did not pass QC. This combined with the limited statistical power means that the possibility of a false-negative cannot be ruled out.

No associations have been reported between *PNPLA3* variants and obesity or BMI in any GWAS to-date. The most recent GIANT BMI publication genotyped 42 SNPs within the adiponutrin gene and found no SNPs to be even nominally associated. This is therefore more evidence that adiponutrin gene variation is not likely to influence obesity risk.

In summary, these results suggest that variants in the adiponutrin gene do not contribute to the development of obesity in French Caucasians.

Chapter 7

Investigation of genetic variants within the Nesfatin gene for association to common polygenic obesity

7.1 Introduction

Nesfatin-1 is an 82 amino acid peptide derived from nucleobindin2 (NUCB2), a protein encoded by the gene NUCB2[372]. In rats, NUCB2 is widely expressed in the CNS as well as liver, adipose and the ghrelin-producing cells of the stomach, which are known to be involved in regulating food intake[372,493]. Cleavage of the NUCB2 preprotein is catalysed by PCSK1 which also produces nesfatin-2 and nesfatin-3 and variants in the *PCSK1* gene have been found to cause a form of monogenic obesity[198].

7.1.1 Central action of nesfatin-1 to reduce food intake

Injection of nesfatin-1 into the brains of both rats and mice has been shown to reduce feeding[372,377]. This effect is not observed for nesfatin-2 or nesfatin-3[372]. Additional studies in rats have consistently demonstrated a suppression of feeding when nesfatin-1 was injected into the lateral[494,495], 3rd[372,377], or 4th brain ventricle[494], cisterna magna[494], and the paraventricular nucleus[377]. Reduction in food intake is also observed in leptin-receptor deficient rats indicating that Nesfatin-1's action is independent of leptin[372]. Furthermore, knockdown of *NUCB2* in the rat hypothalamus using an antisense oligonucleotide increases food intake[372].

In addition to the effects on feeding, continuous central infusion of nesfatin-1 into rats resulted in reduced body weight and administration of a *NUCB2* antisense oligonucleotide increased body weight[372]. Acute injection of nesfatin-1 has also been shown to decrease body weight without altering food intake in a 24 hour cumulative period suggesting a role in controllling energy expenditure[494]. Nesfatin-1 has been shown to directly inhibit NPY neurons in the ARC[496].

7.1.2 *NUCB2* expression is regulated by metabolic factors

Hypothalamic expression of the nesfatin gene has been found to vary under different metabolic conditions, which supports a role as a regulator of energy metabolism. *NUCB2* transcription in the rat hypothalamus is significantly reduced after twenty-four hours fasting and this is restored after re-feeding[372,373]. Re-feeding after a 24 hour fast was found to activate nesfatin-1 immunopositive neurons in the supraoptic nucleus as assessed by double staining for Fos and nesfatin-1 immunoreactivity[494]. In addition, intraperitoneal injections of CCK activated nesfatin-1 immunopositive cells in the PVN and the nucleus of the solitary tract of rats pointing towards a role for central nesfatin-1 in the mediation of gut hormone satiety signalling[494,497].

7.1.3 Peripheral nesfatin-1

The nesfatin-1 protein has been shown to be able to cross the blood-brain barrier in both directions[498],[499]. *NUCB2* is expressed in the stomach and using immunoreactive staining, the nesfatin-1 protein is co-located with ghrelin[500]. These results suggest that nesfatin-1 may act as a gut-brain hormone to influence food intake. Similar to the results of CNS administration of nesfatin-1, intraperitoneal injections have been shown to reduce feeding in mice[501].

In humans, fasting levels of plasma nesfatin-1 have been reported to be significantly reduced in male obese subjects compared to non-obese male controls and were negatively correlated with BMI in the non-obese group[381]. A recent study using female anorexic patients reported a significant increase in circulating nesfatin-1 compared to age-matched healthy controls[382].

7.1.4 Nesfatin-1 as a pharmacological target for the treatment of obesity

Because leptin resistance is normally associated with obesity and peripheral and central injection of nesfatin-1 exerts an anorexigenic effect independent of leptin action[372], targeting the nesfatin-1 pathway may provide a viable approach for pharmacological interventions of obesity.

7.1.5 *NUCB2* variation and obesity

In 2007 when this gene was selected for study in this project, no reported associations between *NUCB2* variation and obesity were found in the literature. In 2011, an association was reported between 3 *NUCB2* SNPs (rs1330, rs214101 and rs757081) and obesity (p=0.016, 0.015 and 0.034) and three SNPs (rs214086, rs214101 and rs757081) and BMI (p=0.023, 0.002 and 0.008) in male but not female Belgian Caucasians[502].

No significant associations between *NUCB2* variation and obesity or BMI have been reported in any GWAS to-date. In the French GWAS[253],18 SNPs within the region were nominally significant with the most significant result, rs7129639, p-value of $1.8x10^{-3}$. The most recent GIANT GWAS[233] found a p-value of $5.4x10^{-4}$ to rs542274, a SNP within the *NUCB2* region (data acessed online[445])(see Figure 7.1).

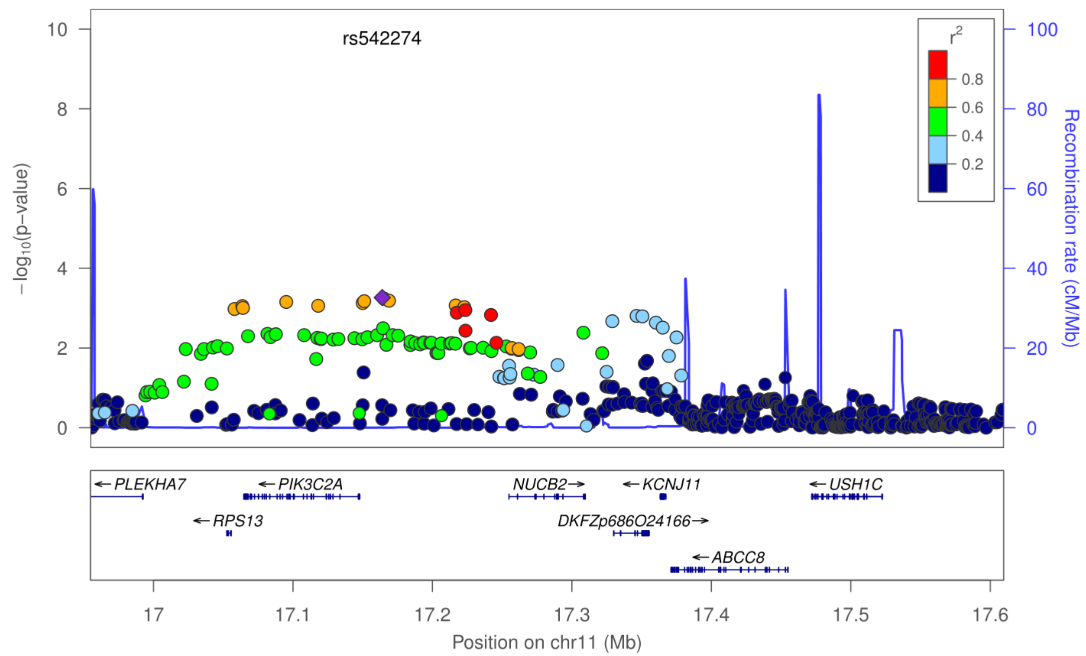

Figure 7.1. SNPs analysed for association to BMI in *NUCB2* region in the GIANT GWAS[233].

Due to its role in regulating appetite, the Nesfatin gene is a plausible candidate for common polygenic obesity in humans. Genetic variants within the NUCB2 gene and the surrounding LD block were genotyped and investigated for associations to common polygenic obesity using French obese cases and non-obese controls (see Materials and Methods for details).

7.2 Results

7.2.1 SNP Selection and Case-Control Genotyping

Thirty-six SNPs were genotyped within the LD block containing the NUCB2 gene, including 33 SNPs tagging all HapMap SNPs within the region (mean r^2=0.965, see Figure 7.2). This included 125kb upstream of the gene and 69kb downstream and the downstream region contains the diabetes gene, KCNJ11. After genotyping quality control, thirty-two SNPs remained with a mean success rate of 94%. Of these, twenty-seven had minor allele frequencies above the required five percent and were taken forward for analysis (see Table 7.1). Of the twenty-one tag SNPs genotyped within the region, seventeen passed QC and were analysed.

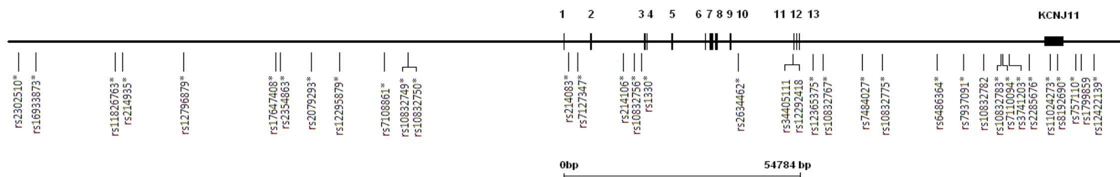

Figure 7.2. Map of *NUCB2* gene and surrounding region with positions of SNPs genotyped in this project. *Tag SNPs. █ *NUCB2* exons.

SNP	Position on Chromosome 11[a]	Position Relative to NUCB2 Gene	Minor Allele Frequency
rs2302510*	17168986	Upstream	<0.01
rs16933873*	17172737	Upstream	0.38
rs11826763	17191534	Upstream	0.05
rs214935*	17193122	Upstream	0.48
rs17647408*	17229466	Upstream	0.24
rs2354863	17230389	Upstream	0.10
rs2079293	17238051	Upstream	0.02
rs12295879	17244473	Upstream	0.03
rs7108861*	17255001	Upstream	0.08
rs10832749*	17260867	Upstream	0.20
rs10832750*	17260918	Upstream	0.48
rs214083*	17298859	Intron 1	0.46
rs7127347	17300844	Intron 1	0.25
rs214106*	17311811	Intron 2	0.50
rs10832756*	17314345	Intron 2	0.09
rs1330	17316029	Intron 2	0.45
rs34405111	17351721	Downstream	<0.01
rs12292418	17351748	Downstream	<0.01
rs12365375	17356826	Downstream	0.08
rs10832767	17359017	Downstream	0.08
rs6486364*	17386257	Downstream	<0.01
rs7937091*	17392300	Downstream	0.32
rs10832782*	17397273	Downstream	0.48
rs10832783	17401134	Downstream	0.13
rs7110094	17401519	Downstream	0.25
rs3741203*	17403163	Downstream	0.37
rs2285676*	17408025	Downstream	0.45
rs11024273*	17412717	Downstream (KCNJ11)	0.39
rs8192690	17414570	Downstream (KCNJ11)	0.13
rs757110*	17418477	Downstream	0.46
rs1799859	17419279	Downstream	0.40
rs12422139	17422397	Downstream	0.14

Table 7.1. *NUCB2* SNPs successfully genotyped with observed minor allele frequencies. [a]Positions in bp correspond to NCBI build 37.1 measured from the top of the p arm of the chromosome.

One SNP, rs10832756 was associated with obesity in children (p=0.02, OR=0.6, CI=0.38-0.93), however this does not survive correction using permutations (p=0.33, see Table 7.2). Analysis was repeated excluding subjects with T2D, impaired fasting glucose or impaired glucose tolerance (any patient with fasting blood glucose >5.6 mmol/l) in order to filter out any association that may be present between SNPs in LD with KCNJ11 and diabetes, which might be masking associations to obesity. After removing these subjects, two SNPs, rs10832756 and rs757110 were nominally associated with obesity in children (p-values: 0.042, 0.039; OR: 0.61, 0.83; CI: 0.38-0.99, 0.70-0.99), however these associations do not withstand correction for multiple testing using permutations (see Table 7.3).

	Genotype Numbers														
	Controls			Children			Adults			Odds Ratio Children (95% CI)	P-values Children		Odds Ratio Adults (95% CI)	P-values Adults	
SNP	1,1[a]	1,2	2,2	1,1	1,2	2,2	1,1	1,2	2,2		Empirical	Corrected		Empirical	Corrected
rs16933873	679	461	65	334	227	42	468	322	55	1.1 (0.91-1.3)	0.41	1.0	1.1 (0.92-1.2)	0.44	1.0
rs11826763	1171	61	0	587	27	0	772	44	0	0.89 (0.56-1.4)	0.62	1.0	1.1 (0.74-1.6)	0.66	1.0
rs214935	377	599	248	195	295	119	269	385	165	0.96 (0.84-1.1)	0.58	1.0	0.96 (0.84-1.1)	0.50	1.0
rs17647408	914	296	18	448	139	18	576	201	22	1.1 (0.89-1.3)	0.44	1.0	1.2 (0.97-1.4)	0.11	0.84
rs2354863	1097	116	8	562	58	4	759	84	2	0.98 (0.72-1.3)	0.89	1.0	0.96 (0.73-1.3)	0.78	1.0
rs12295879	1163	39	0	596	17	0	816	26	0	0.85 (0.48-1.5)	0.57	1.0	0.95 (0.58-1.6)	0.87	1.0
rs7108861	1074	90	5	361	28	0	752	67	0	0.84 (0.55-1.3)	0.42	1.0	0.95 (0.70-1.3)	0.78	1.0
rs10832749	965	261	16	477	123	12	657	151	11	1.0 (0.83-1.3)	0.85	1.0	0.88 (0.72-1.1)	0.22	0.98
rs10832750	414	597	213	218	302	107	295	368	137	0.97 (0.85-1.1)	0.71	1.0	0.93 (0.82-1.1)	0.30	1.0
rs214083	414	507	168	200	257	92	335	355	119	1.1 (0.92-1.2)	0.43	1.0	0.92 (0.80-1.1)	0.20	0.97
rs7127347	871	322	35	465	142	13	608	223	18	0.83 (0.68-1.0)	0.053	0.62	0.95 (0.80-1.1)	0.54	1.0
rs214106	329	612	297	170	307	136	225	413	183	0.94 (0.82-1.1)	0.39	1.0	0.95 (0.84-1.1)	0.43	1.0
rs10832756	1067	117	7	335	22	1	737	83	2	0.60 (0.38-0.93)	0.021	0.33	0.96 (0.73-1.3)	0.78	1.0
rs1330	507	543	152	266	273	74	340	381	108	0.96 (0.83-1.1)	0.60	1.0	1.0 (0.91-1.2)	0.62	1.0
rs12365375	1128	104	6	559	57	0	769	59	1	0.99 (0.71-1.4)	0.94	1.0	0.78 (0.57-1.1)	0.13	0.87
rs10832767	1124	107	6	554	58	0	756	59	1	0.98 (0.71-1.4)	0.93	1.0	0.77 (0.56-1.1)	0.11	0.83
rs7937091	810	380	48	388	196	27	513	272	34	1.1 (0.91-1.3)	0.38	1.0	1.1 (0.94-1.3)	0.23	0.98
rs10832782	432	575	234	200	304	104	280	397	134	1 .0(0.87-1.2)	0.96	1.0	0.96 (0.84-1.1)	0.52	1.0
rs10832783	1065	159	6	545	78	3	735	107	4	0.96 (0.73-1.3)	0.78	1.0	0.98 (0.76-1.3)	0.85	1.0
rs7110094	892	308	28	443	156	15	595	206	19	1.0 (0.85-1.2)	0.80	1.0	1.0 (0.84-1.2)	0.96	1.0
rs3741203	665	469	89	324	234	45	467	283	71	1.0 (0.87-1.2)	0.79	1.0	0.97 (0.84-1.1)	0.69	1.0
rs2285676	509	549	163	254	285	85	360	366	119	1.0 (0.89-1.2)	0.71	1.0	1.0 (0.88-1.1)	0.96	1.0
rs11024273	600	479	98	298	245	53	456	304	74	1.0 (0.89-1.2)	0.63	1.0	0.92 (0.80-1.1)	0.28	0.99
rs8192690	1063	172	6	529	81	4	725	94	5	0.98 (0.75-1.3)	0.86	1.0	0.84 (0.66-1.1)	0.17	0.95
rs757110	480	561	177	260	286	78	328	400	118	0.91 (0.79-1.1)	0.20	0.97	1.0 (0.88-1.1)	0.98	1.0
rs1799859	627	477	74	315	250	43	424	328	72	1.1 (0.91-1.2)	0.47	1.0	1.1 (0.97-1.3)	0.14	0.91
rs12422139	1052	170	6	528	87	3	716	120	8	1.0 (0.78-1.3)	0.90	1.0	1.1 (0.87-1.4)	0.44	1.0

Table 7.2. Allelic association analysis of *NUCB2* SNPs to obesity in children and adult French case-controls. Analysis was performed using PLINK. Empirical p-values result from a single chi-squared test, corrected p-values result from one million permutations. [a] 1 denotes the common allele, 2 denotes the minor allele.

	Controls			Obese Children			Obese Adults			Odds Ratio Children (95% CI)	P-values Children		Odds Ratio Adults (95% CI)	P-values Adults	
	1,1[a]	1,2	2,2	1,1	1,2	2,2	1,1	1,2	2,2		Empirical	Corrected		Empirical	Corrected
rs16933873	304	211	30	304	206	40	205	142	28	1.1 (0.88-1.3)	0.54	1.0	1.1 (0.88-1.3)	0.46	1.0
rs11826763	508	33	0	533	27	0	344	19	0	0.79 (0.47-1.3)	0.36	1.0	0.85 (0.48-1.5)	0.59	1.0
rs214935	181	250	103	179	270	106	130	158	78	1.0 (0.87-1.2)	0.73	1.0	1.0 (0.83-1.2)	0.93	1.0
rs17647408	400	130	6	410	125	17	253	92	12	1.1 (0.86-1.4)	0.44	1.0	1.3 (0.97-1.7)	0.077	0.74
rs2354863	496	48	1	516	51	3	331	43	0	1.1 (0.74-1.6)	0.65	1.0	1.3 (0.83-1.9)	0.26	0.99
rs12295879	515	18	0	545	15	0	361	12	0	0.79 (0.40-1.6)	0.50	1.0	0.95 (0.46-2.0)	0.90	1.0
rs7108861	467	34	2	324	24	0	333	31	0	0.91 (0.54-1.5)	0.72	1.0	1.1 (0.70-1.8)	0.61	1.0
rs10832749	411	128	7	432	115	11	292	67	5	0.94 (0.73-1.2)	0.61	1.0	0.79 (0.59-1.1)	0.12	0.88
rs10832750	170	264	108	194	280	100	132	161	62	0.9 (0.76-1.1)	0.24	1.0	0.84 (0.70-1.0)	0.083	0.76
rs214083	194	215	73	182	233	85	152	158	52	1.1 (0.94-1.4)	0.20	0.97	0.95 (0.78-1.2)	0.60	1.0
rs7127347	386	143	15	422	134	12	264	107	5	0.85 (0.68-1.1)	0.19	0.96	0.97 (0.75-1.3)	0.84	1.0
rs214106	142	256	143	152	284	123	106	173	87	0.90 (0.76-1.1)	0.21	0.98	0.90 (0.74-1.1)	0.26	0.99
rs10832756	454	52	5	294	22	1	321	44	1	0.61 (0.38-0.99)	0.042	0.53	1.0 (0.70-1.5)	0.85	1.0
rs1330	228	227	78	242	254	64	153	167	48	0.92 (0.77-1.1)	0.37	1.0	0.99 (0.82-1.2)	0.93	1.0
rs12365375	509	38	2	512	50	0	345	26	0	1.2 (0.77-1.8)	0.46	1.0	0.91 (0.55-1.5)	0.72	1.0
rs10832767	504	41	2	507	51	0	339	26	0	1.1 (0.74-1.7)	0.60	1.0	0.86 (0.53-1.4)	0.55	1.0
rs7937091	363	156	23	353	181	23	226	124	13	1.1 (0.90-1.4)	0.30	1.0	1.1 (0.90-1.4)	0.29	1.0
rs10832782	189	256	100	180	277	98	130	173	61	1.0 (0.87-1.2)	0.71	1.0	0.95 (0.78-1.2)	0.58	1.0
rs10832783	474	69	3	496	73	3	327	47	1	1.0 (0.73-1.4)	0.97	1.0	0.95 (0.65-1.4)	0.78	1.0
rs7110094	386	137	10	403	144	13	261	96	8	1.0 (0.82-1.3)	0.77	1.0	1.1 (0.81-1.4)	0.72	1.0
rs3741203	293	205	35	293	213	44	218	114	34	1.1 (0.90-1.3)	0.41	1.0	0.95 (0.77-1.2)	0.66	1.0
rs2285676	231	242	65	231	260	79	167	157	51	1.1 (0.92-1.3)	0.30	1.0	1.0 (0.82-1.2)	0.99	1.0
rs11024273	267	219	35	270	223	50	221	118	35	1.1 (0.91-1.3)	0.31	1.0	0.87 (0.71-1.1)	0.22	0.98
rs8192690	463	78	3	481	76	3	324	42	1	0.94 (0.69-1.3)	0.72	1.0	0.76 (0.52-1.1)	0.16	0.94
rs757110	204	245	90	236	267	67	145	169	60	0.83 (0.70-0.99)	0.039	0.50	0.97 (0.8-1.2)	0.73	1.0
rs1799859	293	200	33	287	231	37	188	144	36	1.1 (0.92-1.4)	0.25	0.99	1.2 (0.99-1.5)	0.060	0.63
rs12422139	472	69	4	483	80	1	314	57	4	1.0 (0.75-1.4)	0.85	1.0	1.3 (0.88-1.8)	0.21	0.98

Table 7.3. Allelic association analysis of *NUCB2* SNPs to obesity in non-diabetic children and adult French case-controls. Analysis was performed using PLINK. Empirical p-values result from a single chi-squared test, corrected p-values result from one million permutations. [a] 1 denotes the common allele, 2 denotes the minor allele.

	Genotype Counts			Mean BMI (95% CI)			P-values	
SNP	1,1[a]	1,2	2,2	1,1	1,2	2,2	Empirical	Corrected
rs16933873	665	445	63	22.4 (22.3-22.6)	22.4 (22.2-22.6)	22.4 (21.9-22.9)	0.96	1.0
rs11826763	1139	59	0	22.4 (22.2-22.5)	22.7 (22.1-23.2)	-	0.28	0.99
rs214935	370	583	239	22.4 (22.2-22.6)	22.4 (22.3-22.6)	22.2 (21.9-22.4)	0.21	0.97
rs17647408	892	285	16	22.4 (22.3-22.5)	22.4 (22.1-22.6)	22.3 (21.2-23.4)	0.79	1.0
rs2354863	1067	113	8	22.4 (22.3-22.5)	22.7 (22.3-23.0)	22.3 (21.1-23.5)	0.33	1.0
rs12295879	1130	39	0	22.4 (22.3-22.5)	22.9 (22.2-23.5)	-	0.15	0.92
rs7108861	1047	84	5	22.4 (22.3-22.5)	22.3 (21.9-22.8)	21.2 (19.3-23.2)	0.46	1.0
rs10832749	936	255	16	22.4 (22.2-22.5)	22.5 (22.2-22.7)	21.8 (20.7-22.9)	0.82	1.0
rs10832750	400	580	212	22.3 (22.1-22.5)	22.5 (22.3-22.7)	22.4 (22.1-22.7)	0.31	1.0
rs214083	403	492	166	22.4 (22.2-22.6)	22.5 (22.3-22.6)	22.3 (22.0-22.6)	0.63	1.0
rs7127347	846	314	34	22.4 (22.3-22.6)	22.4 (22.2-22.6)	22.0 (21.3-22.8)	0.57	1.0
rs214106	318	592	293	22.2 (22.0-22.4)	22.5 (22.3-22.6)	22.4 (22.2-22.7)	0.18	0.95
rs10832756	1038	114	7	22.4 (22.2-22.5)	22.4 (22.0-22.8)	22.1 (21.4-22.8)	0.89	1.0
rs1330	492	527	149	22.3 (22.1-22.5)	22.5 (22.3-22.7)	22.4 (22.1-22.8)	0.15	0.91
rs12365375	1099	98	6	22.4 (22.3-22.5)	22.3 (21.9-22.8)	21.9 (19.6-24.2)	0.55	1.0
rs10832767	1096	101	6	22.4 (22.3-22.5)	22.3 (21.9-22.7)	21.9 (19.6-24.2)	0.54	1.0
rs7937091	790	366	47	22.4 (22.3-22.6)	22.3 (22.1-22.5)	22.5 (21.9-23.1)	0.56	1.0
rs10832782	418	562	226	22.4 (22.2-22.6)	22.4 (22.2-22.6)	22.4 (22.1-22.7)	0.86	1.0
rs10832783	1035	154	6	22.4 (22.2-22.5)	22.7 (22.4-23.1)	22.7 (20.4-24.9)	0.042	0.53
rs7110094	868	298	27	22.4 (22.2-22.5)	22.4 (22.1-22.6)	22.2 (21.5-23.0)	0.88	1.0
rs3741203	644	459	86	22.4 (22.2-22.5)	22.4 (22.2-22.6)	22.1 (21.7-22.6)	0.63	1.0
rs2285676	491	535	160	22.4 (22.2-22.6)	22.4 (22.3-22.6)	22.2 (21.9-22.6)	0.38	1.0
rs11024273	582	468	96	22.4 (22.3-22.6)	22.5 (22.3-22.6)	21.9 (21.4-22.3)	0.11	0.85
rs8192690	1033	167	6	22.4 (22.3-22.5)	22.4 (22.1-22.7)	22.4 (21.7-23.0)	0.99	1.0
rs757110	467	545	171	22.3 (22.1-22.5)	22.5 (22.3-22.7)	22.6 (22.2-22.9)	0.076	0.73
rs1799859	613	460	73	22.5 (22.3-22.6)	22.4 (22.2-22.6)	22.4 (21.9-22.9)	0.56	1.0
rs12422139	1020	167	6	22.4 (22.3-22.5)	22.4 (22.1-22.7)	21.9 (20.7-23.2)	0.83	1.0

Table 7.4. Association analysis of *NUCB2* SNPs to BMI using linear regression with age and sex as covariates. Corrected p-values result from 1 million permutations. [a]1 denotes the common allele, 2 denotes the minor allele.

One SNP, rs10832783, was nominally associated with BMI in the French controls ($p=0.042$), however this association was not significant after correction using permutations (see Table 7.4).

Linkage disequilibrium within the region was analysed within the controls (see Figure 7.3). No LD blocks were identified and LD was low within the region, however, NUCB2 SNP, rs214083 was in high LD with KCNJ11 SNP, rs11024273 ($r^2 = 0.71$).

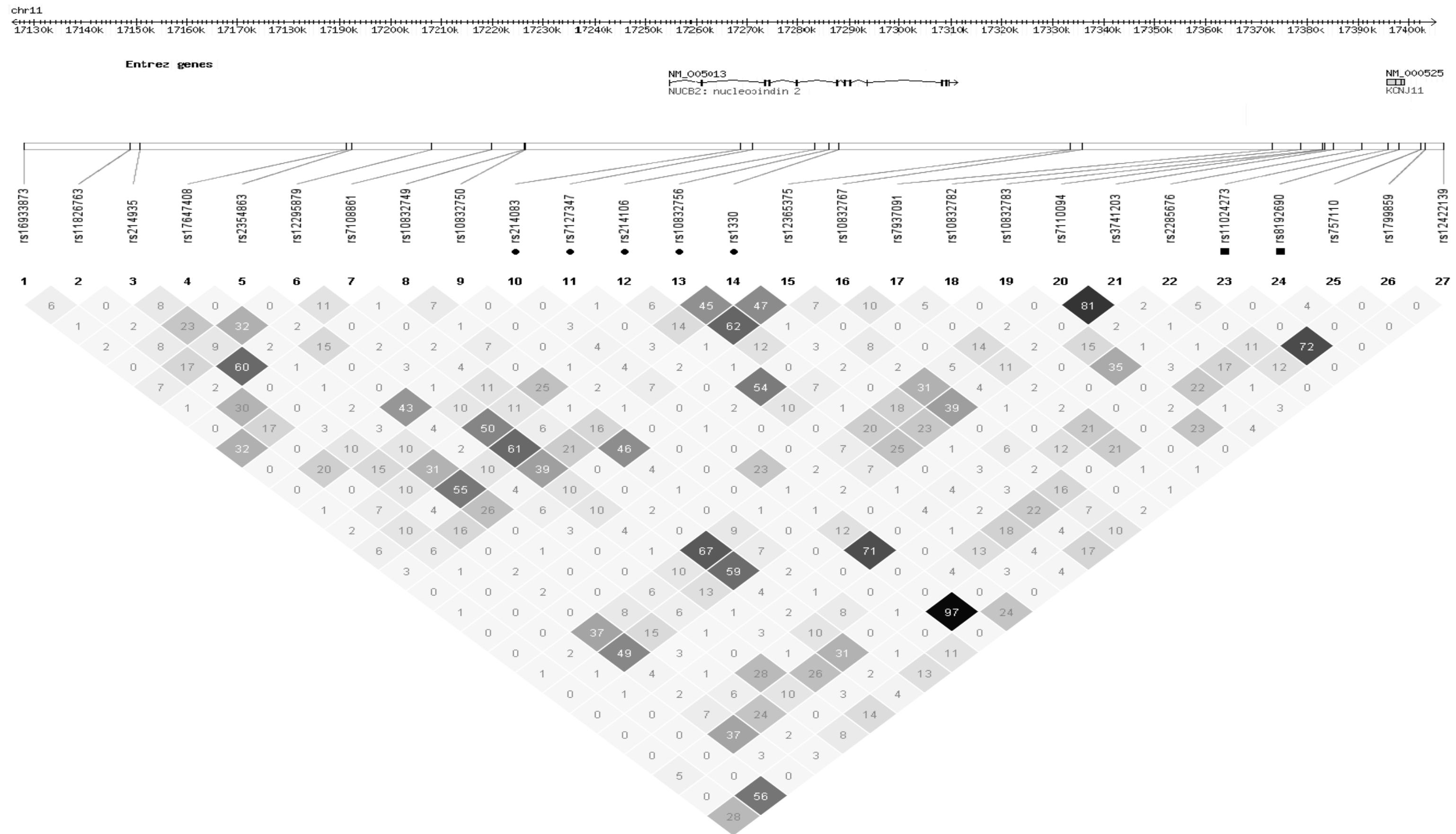

Figure 7.3. Linkage disequilibrium (r^2) plot of SNPs genotyped in *NUCB2* and the surrounding region in the French control samples. ●*NUCB2* SNPs, ■ *KCNJ11* SNPs. Back squares indicate 100% LD and white squares indicate zero LD. The percentage LD level (r^2) within grey squares are displayed.

7.2.2 Power calculations

Using an odds ratio of 0.6 (as observed in the most significantly associated SNP, rs10832756), a MAF of 0.06 (as observed in rs10832756 in the controls), and a significance level of 0.0019 (0.05 divided by 27 tests), the power to detect an association was <1% in the adult cohort and 13% in the child cohort.

SNP	Significance level	MAF in controls	Odds ratio in children	Odds ratio in adults	Power in children	Power in adults
rs16933873	0.0019	0.25	1.1	1.1	<0.01	<0.01
rs11826763	0.0019	0.025	0.89	1.1	<0.01	<0.01
rs214935	0.0019	0.45	0.96	0.96	<0.01	<0.01
rs17647408	0.0019	0.14	1.1	1.2	<0.01	0.029
rs2354863	0.0019	0.054	0.98	0.96	<0.01	<0.01
rs12295879	0.0019	0.016	0.85	0.95	<0.01	<0.01
rs7108861	0.0019	0.043	0.84	0.95	<0.01	<0.01
rs10832749	0.0019	0.12	1.0	0.88	<0.01	0.014
rs10832750	0.0019	0.42	0.97	0.93	<0.01	0.011
rs214083	0.0019	0.39	1.1	0.92	<0.01	0.014
rs7127347	0.0019	0.16	0.83	0.95	0.036	<0.01
rs214106	0.0019	0.49	0.94	0.95	<0.01	<0.01
rs10832756	0.0019	0.055	0.60	0.96	0.13	<0.01
rs1330	0.0019	0.35	0.96	1.0	<0.01	<0.01
rs12365375	0.0019	0.047	0.99	0.78	<0.01	0.022
rs10832767	0.0019	0.048	0.98	0.77	<0.01	0.025
rs7937091	0.0019	0.19	1.1	1.1	<0.01	0.013
rs10832782	0.0019	0.42	1.0	0.96	<0.01	<0.01
rs10832783	0.0019	0.070	0.96	0.98	<0.01	<0.01
rs7110094	0.0019	0.15	1.0	1.0	<0.01	<0.01
rs3741203	0.0019	0.26	1.0	0.97	<0.01	<0.01
rs2285676	0.0019	0.36	1.0	1.0	<0.01	<0.01
rs11024273	0.0019	0.29	1.0	0.92	<0.01	0.012
rs8192690	0.0019	0.074	0.98	0.84	<0.01	0.016
rs757110	0.0019	0.38	0.91	1.0	0.014	<0.01
rs1799859	0.0019	0.27	1.1	1.1	<0.01	0.021
rs12422139	0.0019	0.074	1.0	1.1	<0.01	<0.01

Table 7.6. Power calculations carried out using MAF and OR observed for each SNP investigated in the NUCB2 gene region.

To achieve a statistical power of 95%, assuming a significance level of 0.0019, a MAF of 0.06 and an odds ratio of 0.6, 3,891 cases and 3,891 controls would be required. Assuming a significance level of $5x10^{-8}$, this rises to 8,683 cases and 8,683 controls (see Figure 7.4). If the real odds ratio is greater than 0.6 then this figure will rise (see Figure 7.5).

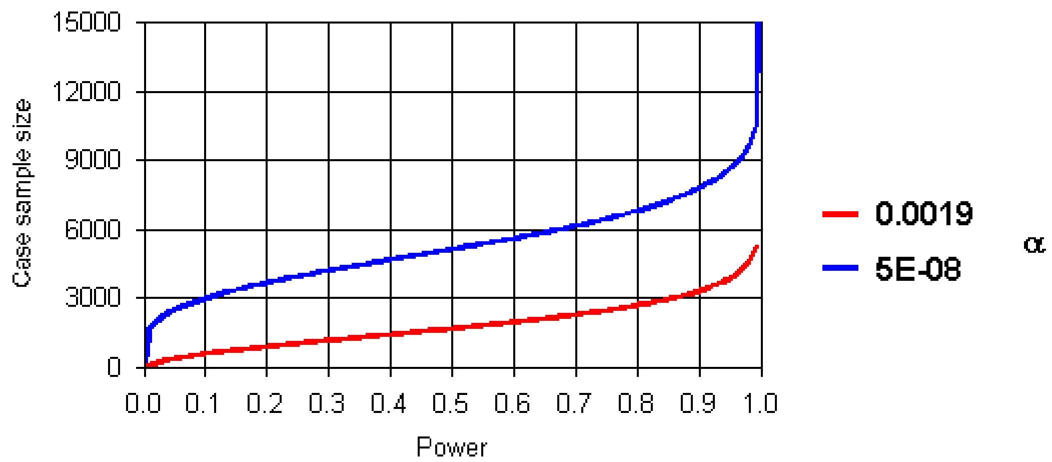

Figure 7.4. Sample size plotted with statistical power for a SNP with MAF=0.0.6 (as was observed for rs10832756). α is the significance level: 0.0019 for 26 statistical tests and $5x10^{-8}$ for genome-wide significance.

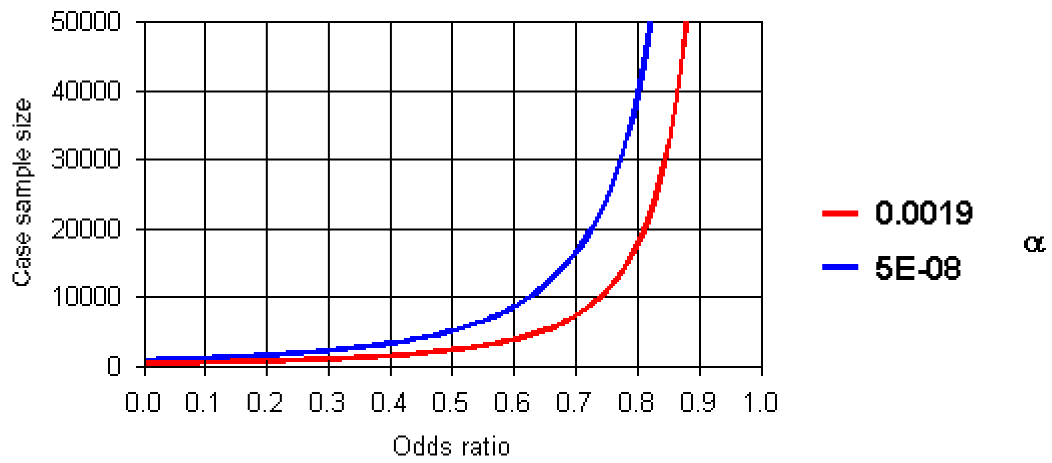

Figure 7.5. Sample size (number of cases, with number of controls=number of cases) required for varying odds ratios in order to achieve a statistical power of 95% for a SNP with MAF=0.06 (as was observed for rs10832756). α is the significance level: 0.0019 for 26 statistical tests and $5x10^{-8}$ for genome-wide significance.

7.3 Discussion

This study was carried out to test the hypothesis that variants in NUCB2 influenced susceptibility to obesity. This was plausible because nesfatin-1, a product of the NUCB2 gene is expressed in appetite controlling regions of the brain and injection of recombinant nesfatin-1 into the brains of rodents suppresses feeding, suggesting a role of the gene in controlling satiety[372]. During the practical stage of this PhD, no studies had been published which investigate such an association.

No significant associations were discovered between SNPs in and near the *NUCB2* gene and obesity in either adults or children. Rs10832756 was nominally associated with obesity in children (p=0.02) however this association became non-significant after correction using permutations, which suggests that this is a false positive rather than a genuine association. When diabetic patients were removed from the analysis the same SNP was nominally significant but at a greater p-value (p=0.042).

No significant associations between *NUCB2* gene variation and obesity phenotypes have been reported in any GWAS to date. The most recent GIANT BMI[233] and WHR[262] studies investigated 25 SNPs within the *NUCB2* region and found no significant associations although nominal associations to BMI were found including to rs542274 ($p=5.4x10^{-4}$). This is tagged by rs10832750 in this study (p=0.30 in adults and 0.71 in children).

Out of thirty-six SNPs genotyped in this study, twenty-seven were analysed for associations to obesity. Nine SNPs failed QC or had low minor allele frequencies. Furthermore, rare variants (<5%) were not investigated and so coverage of the gene was not complete. Statistical power calculations indicate that this study was underpowered to detect associations with the effect sizes observed. These two points mean that the possibility of a false-negative cannot be ruled out.

In summary, these results suggest that variants in the NUCB2 gene are not likely to contribute to the development of obesity in a French Caucasian population.

.

Chapter 8

Investigation of genetic variation within the IRS-1 gene and transcription levels for association to BMI

8.1 Introduction

Insulin receptor substrate 1 (IRS-1) is a signalling molecule encoded by the IRS-1 gene in humans which is located on the long arm of chromosome 2 and contains two exons. IRS-1 is crucial for transmitting signals from insulin and IGF-1 receptors to intracellular pathways that control metabolism[384].

8.1.1 IRS-1 and glucose metabolism

IRS-1 knockout mice have mild diabetic phenotypes as well as a 50% reduction in body weight[384,386]. A number of variants within the IRS-1 gene have been found to influence T2D risk, including a non-synonymous coding SNP, rs1801278 which has been reported by several studies[224,387,504]. A variant within the IRS-1 gene was recently found to be significantly associated with T2D, insulin resistance and hyperinsulinemia in a GWA study carried out by collaborators in France[388]. Genotyping of the associated SNP, rs2943641 in Swedish families was carried out during this study but the results were not published.

8.1.2 IRS-1 and obesity

Adipocytes from IRS-1 knockout mice have reduced differentiation and lipid accumulation[386]. Associations to obesity phenotypes have also been reported including BMI and WHR in a Dutch study[505], BMI in African Americans[506] and blood-leptin concentrations in obese subjects[507] although these associations have not been replicated.

No SNPs within the *IRS1* region were found to be nominally significant in the French GWAS[253] or any of the GIANT publications[263,233]. A more recent GWAS that investigated associations to body fat percentage, however reported a significant association to a SNP near the *IRS1* gene[255]. This finding was published after the completion of the practical work in this thesis.

Given its important role in the insulin signalling pathway, the IRS-1 gene is a plausible candidate gene for obesity and this study was designed to investigate the possibility of genetic associations between rs2943641, its expression levels and obesity in the Swedish sib-pair cohort (see Materials and Methods for details). Analysing transcript expression data is also an opportunity to explore the possibility that the association to T2D reported by Rung *et al.*[388] is mediated via altering levels of transcription.

8.2 Results

Genotyping of the *IRS-1* SNP, rs2943641 in the Swedish families cohort was carried out by collaborators in France[388]. No association was found between rs2943641 and BMI (corrected p=0.5, see Table 8.1).

Genotype Counts			**Mean BMI (95% CI)**			**P-Values**	
1,1[a]	**1,2**	**2,2**	**1,1**	**1,2**	**2,2**	**Empirical**	**Corrected**
214	259	106	29.0 (28.1-29.9)	28.9 (28.1-29.7)	28.4 (27.0-29.7)	0.49	0.50

Table 8.1. Association analysis of *IRS-1* SNP, rs2943641 with BMI corrected for age and sex in the Swedish families using the QFAM test implemented by PLINK. [a]1 denotes the common allele and 2 denotes the rare allele. Corrected p-values result from one million permutations.

Microarray data was available, which included measurements of *IRS-1* transcript levels in subcutaneous adipose tissue for each sibling within the Swedish sib-pair cohort. A significant association was found between rs2943641 and *IRS-1* transcript (corrected $p=1.0 \times 10^{-5}$, see Table 8.2).

Genotype Counts			Mean transcription level (95% CI)			P-values	
1[a],1	**1,2**	**2,2**	**1,1**	**1,2**	**2,2**	**Empirical**	**Corrected**
89	133	42	46.7 (43.2-50.2)	35.9 (34.1-37.7)	33.8 (31.8-35.8)	**$2.8x10^{-9}$**	**1.0×10^{-5}**

Table 8.2. Association analysis of *IRS-1* SNP, rs2943641 with *IRS-1* transcript level corrected for age, sex and BMI in the Swedish families using the QFAM test implemented by PLINK. [a]1 denotes the common allele and 2 denotes the rare allele. Corrected p-values result from one million permutations.

IRS-1 expression was analysed for association with obesity. Between obese and non-obese groups, a significant difference was found in IRS-1 transcript levels after correcting for relatedness, age and sex ($p=1.0x10^{-15}$, see Figure 8.1 and Table 8.3).

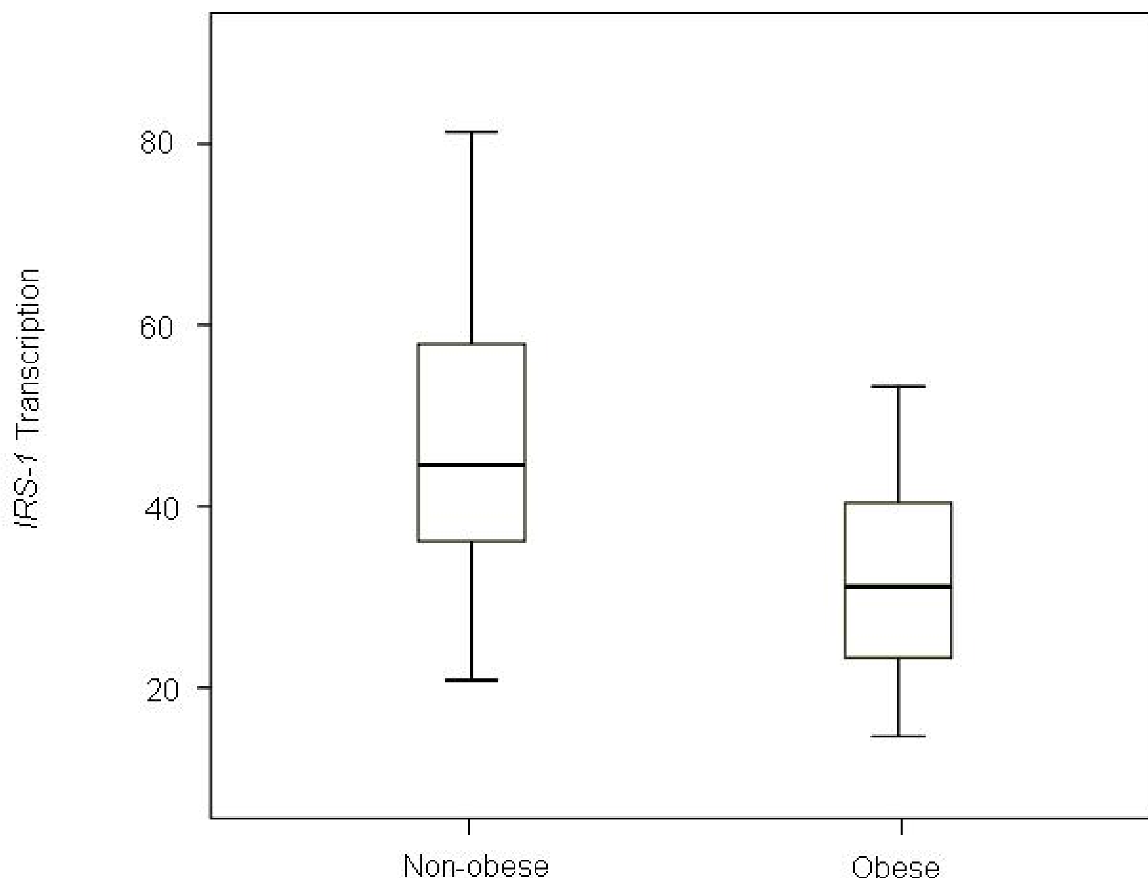

Figure 8.1. Box-plot of *IRS-1* transcript levels in subcutaneous adipose tissue with obese and non-obese subjects from Swedish sib-pairs displayed separately. *IRS-1* transcription is the microarray signal value corrected for age, sex and relatedness.

Number of Non-obese	Number of Obese	Mean Difference in IRS-1 transcription (95% CI)	P-value
190	151	11.1 (8.33-13.9)	**$1.0x10^{-15}$**

Table 8.3. T-test of subcutaneous adipose *IRS-1* transcription in obese and non-obese subjects from Swedish sib-pairs. *IRS-1* transcription is the microarray signal value corrected for age, sex and relatedness.

8.3 Discussion

These results demonstrate a significant association between IRS-1 variant, rs2943641 and IRS-1 transcription levels in the subcutaneous adipose tissue of Swedish sib-pairs. The p-value reported here does remain significant in a genome-wide context, however replication of this result is needed before a genuine association is established. Therefore variation in the *IRS-1* gene is likely to be responsible for modifying its expression in adipose tissue and if this is true in other tissues as well then this may explain the associations reported between the gene and T2D[388]. No significant association was observed between IRS-1 variation and BMI. Rs2943641 is located 500kbp upstream of the transcription start point of the IRS-1 gene. This means it could lie within an enhancer element where it affects binding of a transcription factor. Alternatively it might be in LD with some other variant that is responsible for modifying expression of the gene, either in the promoter or some enhancer region. Further work is required in order to locate the causative variant.

A significant reduction in transcription levels was found in obese compared to non-obese siblings. This difference might be an indication that IRS-1 expression levels influence BMI or it might be the case that the disease state of obesity is responsible for altering expression of the gene. The lack of an association between IRS-1 genetic variation and BMI makes it more likely that an individual's body weight is causing the observed association.

A recent GWAS discovered a significant association between rs2943650, a SNP near IRS-1 and body fat percentage[255]. Rs2943650 is in high LD with rs2943641 (r^2=1.0 in the HapMap CEU population). The failure to replicate this association is likely due to the reduced statistical power in this investigation. Alternatively it may the case that this SNP modifies body fat percentage but not BMI. Since no IRS-1 variants are significantly associated to BMI in any GWAS to-date this is a plausible explanation.

In summary, no evidence was found for an association between *IRS-1* gene variation and BMI in a Swedish family cohort, however a significant association between gene variation and expression in subcutaneous adipose tissue was discovered.

Chapter 9

Investigation of DNA methylation for associations with human polygenic obesity

9.1 Introduction

DNA methylation is an epigenetic modification of DNA in which the 5 position of a cytosine residue within a CpG dinucleotide is methylated. CpG dinucleotides are under-represented in the human genome (observed/expected ratio of 0.2), however certain regions exist that contain a high density of CpG sites (observed/expected ratio of 0.6). These are called CpG islands and are found in the promoter regions of approximately three quarters of all known genes[508]. Elevated methylation in gene promoters is correlated with a reduction in gene expression[509].

There is evidence that DNA methylation patterns can be heritable. In mice CpG methylation at loci affecting expression of the *agouti* gene, which controls fur colour, have been found to be resistant to demethylation during reprogramming[510]. This means that a phenotype of yellow fur, obesity, diabetes and increased susceptibility to cancer can be inherited over multiple generations[278,511]. In humans, inherited methylation of CpG sites within *MLH1* and *MSH2* gene promoters have been found to cause some cases of familial colorectal cancer[512,513]. A recent twin study revealed a significant difference in the DNA methylation patterns between dizygotic twins compared to monozygotic twins which cannot be explained by underlying genetic differences[309]. This demonstrates that DNA methylation is at least partially heritable, which opens the possibility that it may contribute to complex diseases such as obesity. The most established example of altered DNA methylation that results in obesity is Prader-Willi Syndrome which can be caused by deficiencies in the imprinting of the paternal copy of chromosome 15[514,515]. Genomic imprinting has also been reported to influence common obesity at three loci spread across the genome[314]. Overfeeding of rats has been demonstrated to lead to changes in the methylation pattern of the *POMC* gene[291] and methylation of the *TNFa* gene has been shown to predict response to weight loss in humans[315].

The satiety signal leptin is a key protein involved in the control of appetite and mutations in the gene are responsible for a Mendelian form of obesity. Additionally, variants in its gene are associated with both obesity and levels of leptin hormone in circulating blood, although these associations have not been replicated in any GWAS to date[516,517,518]. The leptin CpG island is differentially methylated across tissues, which points to a possible role of methylation in the regulation of this gene[391]. Methylation of one CpG site in the leptin promoter has been shown to be increased in diet-induced obesity in rats[393].

Methylation at a single CpG site can vary within a single tissue type in a single individual and so a suitable experimental method for quantifying methylation levels was required. At the time a method was chosen, methods used for quantification of CpG methylation included cloning-based sequencing[519,520] and pyrosequencing[521,522]. The cloning method is somewhat labour intensive and therefore time consuming when dealing with the large numbers of samples required for association studies. It requires cloning of the PCR product before sequencing and typically at least ten clones (and preferably more) need to be sequenced in order to determine the quantitative level of methylation at each CpG site from a mixture of PCR products. Pyrosequencing can be used to quickly and accurately measure quantitative methylation status at single CpG sites in large numbers of samples but its short read lengths (typically 30bp) mean that it cannot be used to measure more than a few CpG sites at a time.

More recently Sequenom have offered a method that uses their matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry technology to distinguish between methylated and unmethylated residues after PCR and primer extensions carried out on bisulfite treated DNA. This has the advantage of being both high-throughput and accurate for measurements of quantitative methylation at single residues. However it is limited in that it cannot be used to analyse every residue within a CpG island in a single assay[523]. Ilumina have

also developed an array based method that can measure methylation at large numbers of CpG residues across the genome. Typically though, these arrays only investigate a small number of CpG sites per gene promoter and as such are not suitable for investigation of entire CpG islands in candidate genes[524].

Therefore a method that uses direct sequencing was needed in order to provide quick, inexpensive and accurate measurements of quantitative DNA methylation at multiple CpG sites within a CpG island, suitable for investigation of methylation in candidate genes in large sample numbers.

9.2 Results

9.2.1 Developing a Protocol for Measuring CpG Methylation

Firstly, bisulfite sequencing within a stock sample of human genomic DNA (Promega) was carried out to verify that the protocol was effective at detecting methylation states at CpG sites within the POMC promoter region CpG island in genomic DNA derived from blood (see Materials and Methods). This gene was chosen as methylation is well established[525]. Two sets of PCR primers were designed to amplify two regions of 418bp and 369bp within the POMC promoter (see Appendix for primer sequences). PCR of both primer sets was successful but the sequencing reactions were only successfully completed in the first PCR product. This gave a clean read for the first 270bp which provided enough information to determine that nine out of twenty-one CpG sites within the sequenced region were methylated as indicated by unconverted C residues. At each of the nine loci where C residues were observed, smaller T peaks were also present indicating that the DNA samples contained a mixture of methylated and unmethylated DNA. At ten CpG sites that were converted, smaller C peaks were observed, again indicating a mixture of methylated and unmethylated DNA. In two cases a CpG site had both C and T peaks of similar size and these were counted as being methylated (see Figure 1). It is unlikely that the mixed bases observed could be due to incomplete conversion of non-methylated C residues because all non-CpG C residues (that would have appeared in the DNA sequence not treated with bisulfite) had been completely converted to A residues with only background C peaks observed (90.8% and 100% converted in two residues measured).

Bisulfite sequencing was then carried out in a further 9 obese subjects and variation in the methylation status at three loci was observed (see Figure 2.)

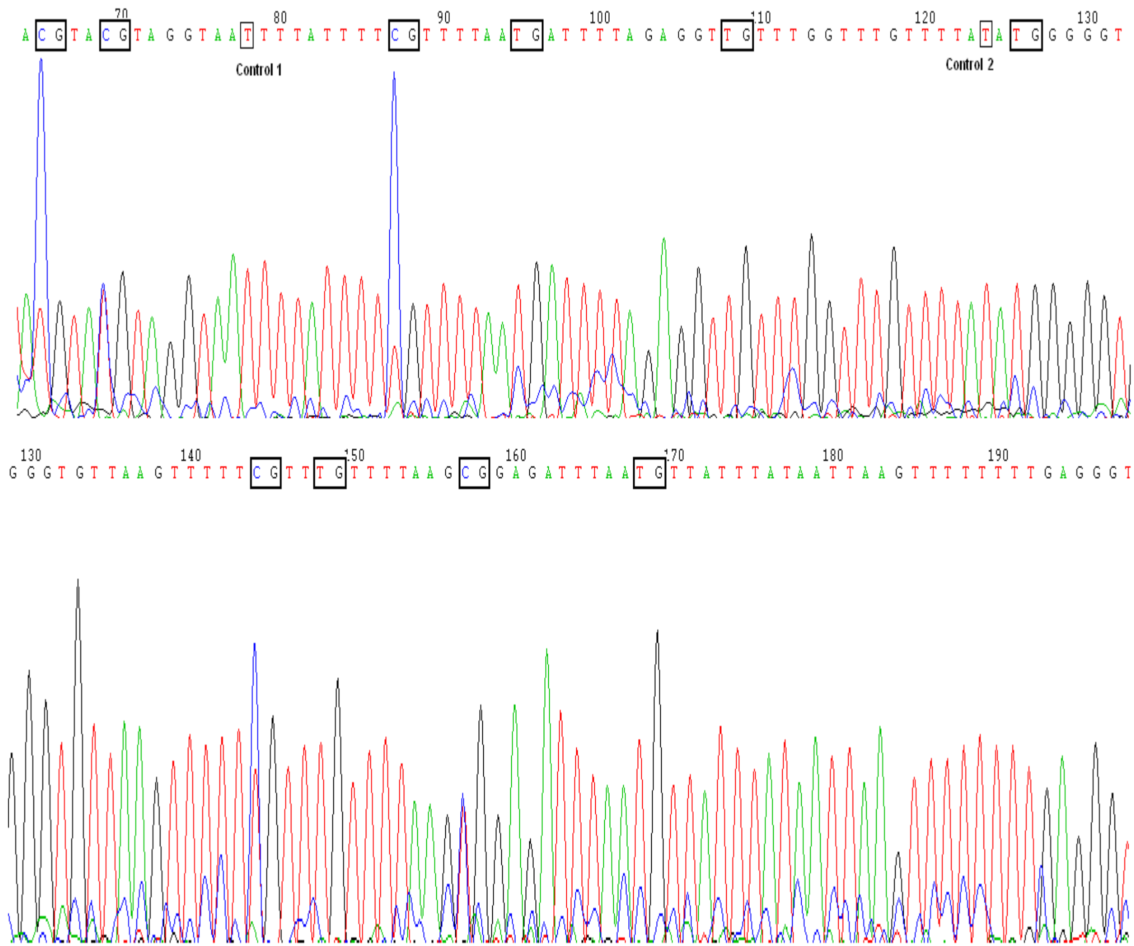

Figure 9.1. Electropherogram showing the sequencing of the CpG island within the POMC promoter. CpG sites are highlighted. Control 1 and Control 2 are non-CpG C-residues that were expected to be fully converted to T residues.

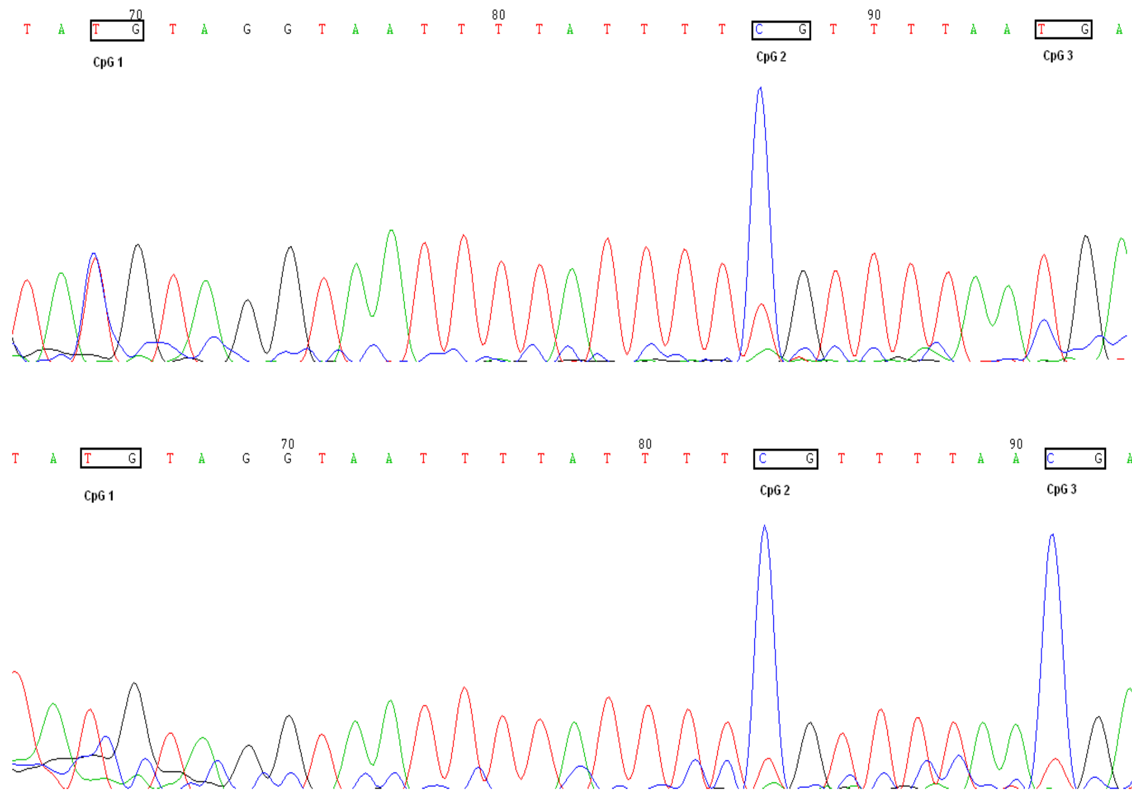

Figure 9.2. Bisulfite sequencing of POMC CpG island in two samples. CpG sites are highlighted and differences in methylation status can be observed at CpG 3.

9.2.2 Locating CpG Islands within Obesity-Associated Genes

Using CpGPlot software[305], putative CpG islands were found in the promoter regions of obesity associated genes: leptin (*LEP*), *FTO*, and peroxisome proliferator–activated receptor gamma (*PPARγ*) (see Figures 9.3-9.5). Putative CpG islands were also found in the obesity candidate gene Adiponutrin (see Figure 9.6).

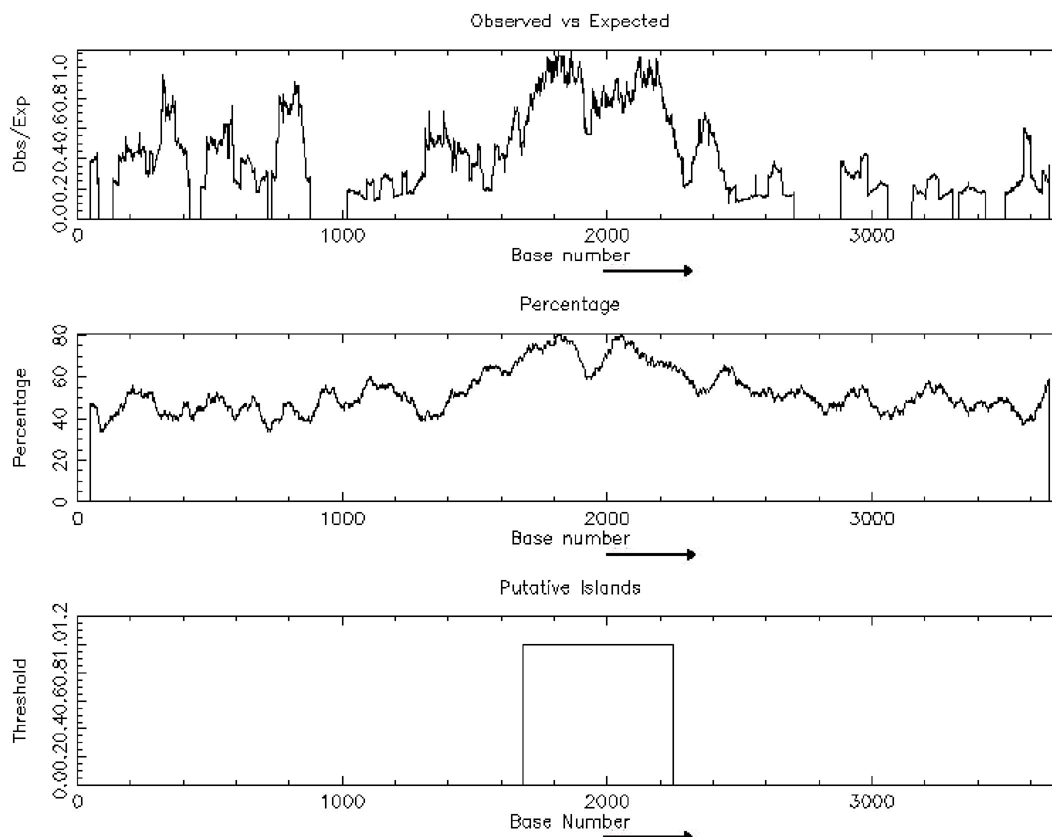

Figure 9.3. CpG plots of the region covering the promoter region and start of the leptin gene. The top graph displays the ratio of observed to expected CpG dinucleotides. The middle graph displays the percentage of C or G nucleotides and the bottom graph displays the position of any putative CpG islands. CpG islands are defined as regions greater than 100bp that contain a ratio of observed to expected CpGs greater than 0.6 and a GC content of greater than 50%. The start of the gene is indicated by an arrow.

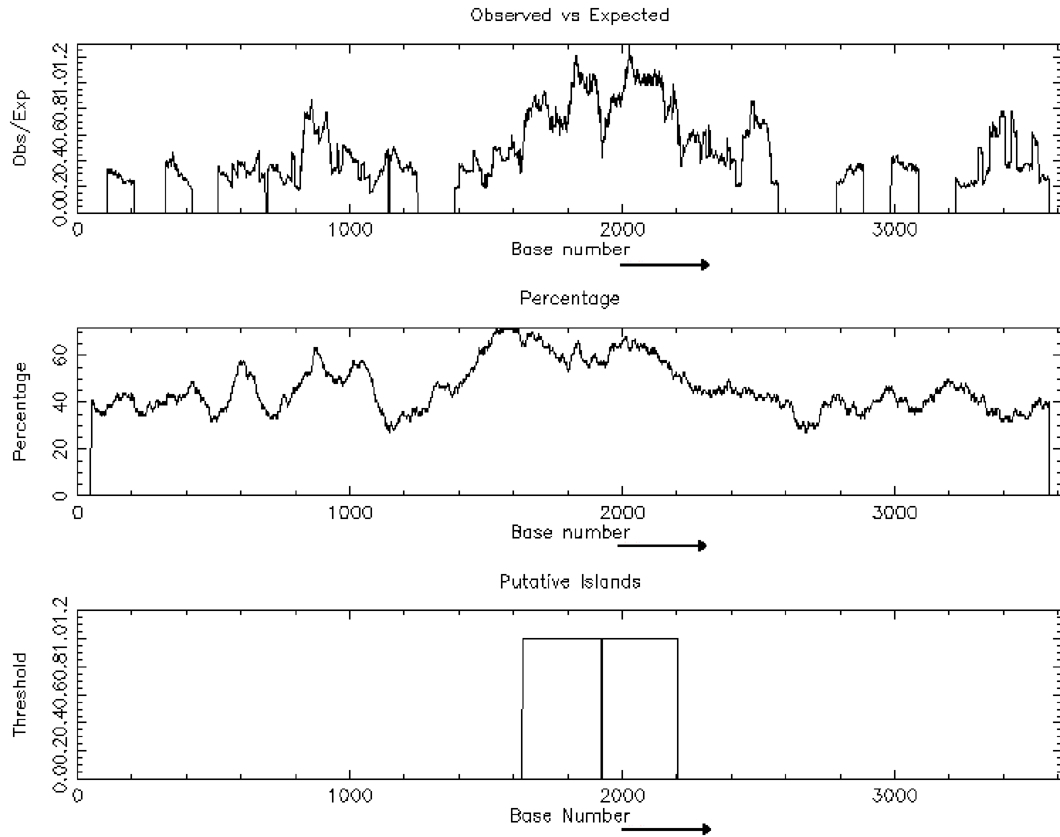

Figure 9.4. CpG plots of the region covering the promoter region and start of the *FTO* gene. The top graph displays the ratio of observed to expected CpG dinucleotides. The middle graph displays the percentage of C or G nucleotides and the bottom graph displays the position of any putative CpG islands. CpG islands are defined as regions greater than 100bp that contain a ratio of observed to expected CpGs greater than 0.6 and a GC content of greater than 50%. The start of the gene is indicated by an arrow.

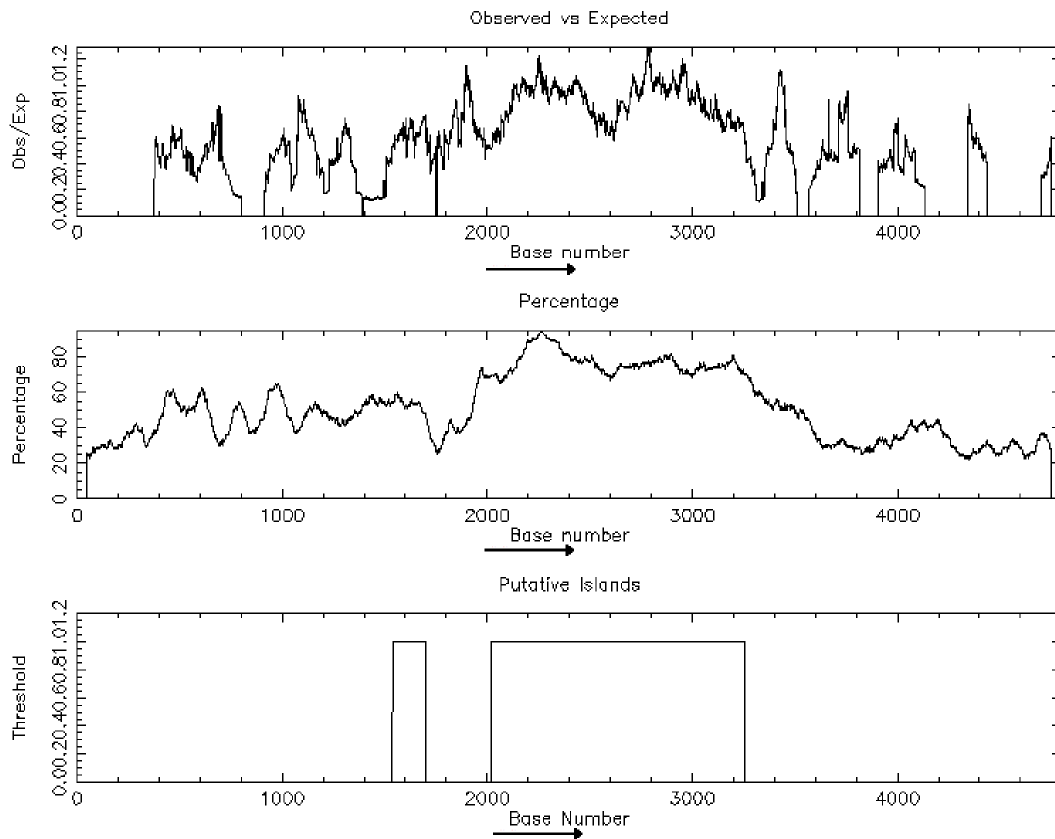

Figure 9.5. CpG plots of the region covering the promoter region and start of the *PPARγ* gene. The top graph displays the ratio of observed to expected CpG dinucleotides. The middle graph displays the percentage of C or G nucleotides and the bottom graph displays the position of any putative CpG islands. CpG islands are defined as regions greater than 100bp that contain a ratio of observed to expected CpGs greater than 0.6 and a GC content of greater than 50%. The start of the gene is indicated by an arrow.

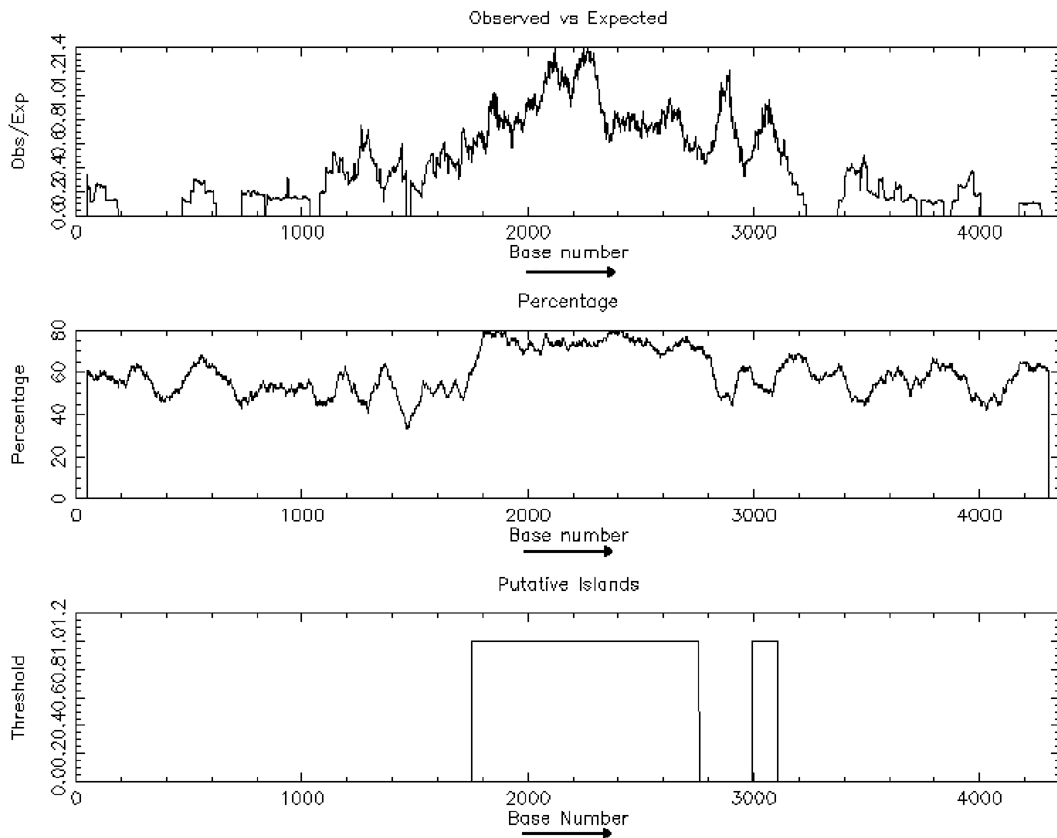

Figure 9.6. CpG plots of the region covering the promoter region and start of the Adiponutrin gene. The top graph displays the ratio of observed to expected CpG dinucleotides. The middle graph displays the percentage of C or G nucleotides and the bottom graph displays the position of any putative CpG islands. CpG islands are defined as regions greater than 100bp that contain a ratio of observed to expected CpGs greater than 0.6 and a GC content of greater than 50%. The start of the gene is indicated by an arrow.

9.2.3 Investigation of DNA methylation in leptin gene

Methylation levels of CpG sites within the leptin gene promoter were investigated for potential associations with human polygenic obesity. Leptin was chosen as its involvement in energy regulation and obesity is well established[224] and it has a clear putative CpG island within its promoter region (see Figure 9.3 above).

PCR primers were designed to amplify a 294bp region within the Leptin CpG island, which was followed by direct sequencing (See Figure 9.7 for PCR product location and Appendix for primer sequences). Using stock gDNA, methylation was observed at all 6 CpG sites within the sequenced region. A small amount of variation between subjects was observed at each CpG site with the highest variation observed at CpG 6 (mean methylation level of 39.2% with a standard deviation of 18.7%, see Table 9.1 and Figure 9.8).

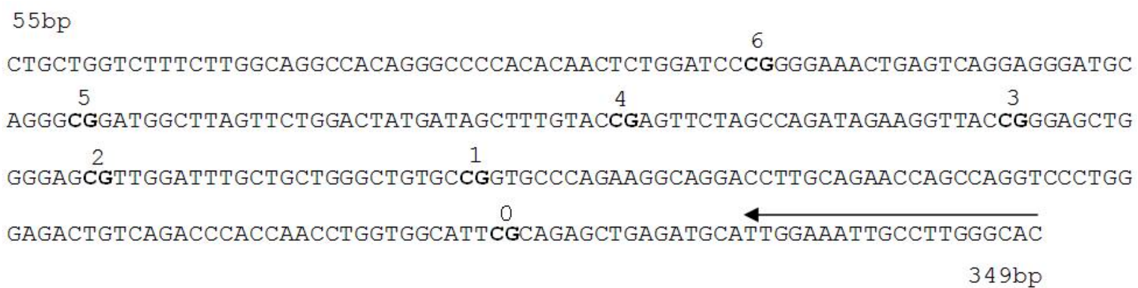

Figure 9.7 The region of the leptin gene analysed by bisulfite sequencing with CpG dinucleotides numbered and shown in bold. Base-pair positions are relative to the transcription start of the Leptin gene. The position of the sequencing primer is also indicated by an arrow. CpG 0 was not within the sequencing read.

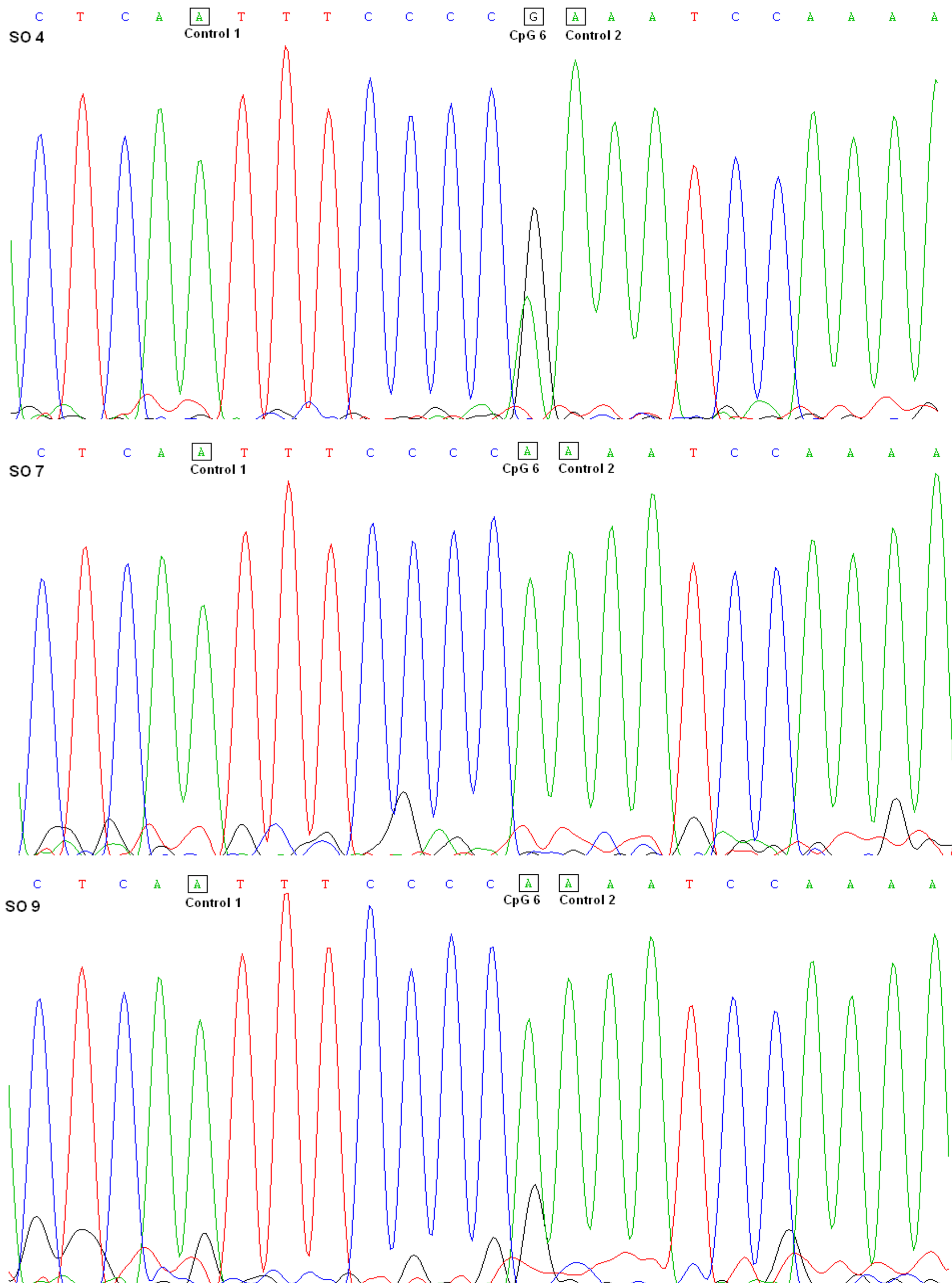

Figure 9.8. Electropherograms showing differing levels of methylation at CpG site 6 from the bisulfite sequencing of Leptin within three obese individuals. Control 1 and Control 2 are non-CpG loci that are expected to be fully converted.

DNA	CpG 1	CpG 2	CpG 3	CpG 4	CpG 5	CpG 6
Control	79.1	76.2	66.3	68.9	84.5	44.8
Obese 1	88.5	80.0	84.4	70.3	87.0	39.1
Obese 2	82.3	81.8	73.6	63.5	84.5	46.4
Obese 3	85.3	85.5	82.4	61.7	60.1	36.0
Obese 4	88.3	82.3	82.7	40.2	63.8	63.3
Obese 5	91.2	89.7	82.5	67.8	81.1	36.6
Obese 6	92.3	85.6	70.4	76.7	72.0	2.6
Obese 7	77.3	91.2	93.5	55.5	76.0	68.5
Obese 8	90.6	91.0	87.6	62.2	46.3	27.7
Obese 9	79.2	83.6	80.2	52.1	83.0	26.9
Mean Methylation Level	85.4	84.7	80.4	61.9	73.8	39.2
Standard Deviation	5.6	4.9	8.2	10.5	13.3	18.7

Table 9.1. Variation in levels of DNA methylation at six CpG sites sequenced in the Leptin promoter CpG island between different DNA samples. Methylation level is measured as peak height of the G residue divided by the sum of the peak heights of the G and C residues at that site as read from DNA sequencing data from that sample.

9.2.3.1 Low volume bisulfite sequencing protocol

The method used above was modified in order to test whether it could be used with smaller quantities of DNA such that a case-control study could be carried out. The protocol was modified from the method presented in the Materials and Methods chapter as follows: 10ng DNA was mixed with 0.7µl Hydroquinone and 12µl Conversion Buffer made up to 14µl with dH_2O. This was placed in a thermocycler for three minutes at 94°C followed by nine hours at 50°C. 50 µl DNA Binding Buffer was mixed with the conversion reaction and pipetted into a DNA purification column. This was span at 10,000 rpm for 30 seconds in a microcentrifuge and the collection tube emptied. 200µl DNA Wash Buffer was added to the column which was then span at 10,000 rpm for 30 seconds. 200µl Desulfonation Buffer was added to the column and incubated at room temperature for 20 minutes before spinning at 10,000 rpm for 30 seconds. 200µl of DNA Wash Buffer was then added and the column was spun at 10,000rpm for 30

seconds. The column was then placed in a new 1.5ml microcentrifuge tube and 50μl of DNA Elution buffer was pipetted onto the surface of the column filter. This was incubated at room temperature for 3 minutes before spinning down at 10,000rpm for 30 seconds. The eluted solution was purified, converted DNA ready for PCR.

This protocol was used to measure leptin CpG methylation in four out of ten samples used above. Peak height ratios were compared and found to differ by an average of 7.4% ±7.0%.

9.2.4 Investigation of leptin gene DNA methylation levels for possible association to human polygenic obesity

DNA methylation in the leptin gene CpG island was measured in 92 morbidly obese French adults and 92 non-obese controls from the D.E.S.I.R cohort. (see Materials and Methods for details of subjects). Out of 184 samples treated with bisulfite and sequenced, eighty-two cases and seventy-five controls met quality control standards (i.e. two non-CpG cytosine residues were >95% converted) and were analysed for association. Because the first three CpG sites were found to be over 90% methylated in the majority (>95%) of samples, they were not analysed. Using the quantitative measure of percent methylation, CpG 5 was found to be nominally associated with both obesity (p=0.013) and age- and sex-corrected BMI (p=0.032) (see Table 9.2). No statistically significant result was observed for any phenotype for the other two variably methylated CpG dinucleotides.

CpG Site	Mean Methylation in Obese (Standard Deviation)	Mean Methylation in Controls (Standard Deviation)	Obesity (P-value)	BMI* (P-value)
CpG 4	73.3% (5.7%)	74.0% (5.1%)	0.86	0.77
CpG 5	79.0% (6.0%)	88.0% (3.6%)	0.013	0.032
CpG 6	46.5% (7.6%)	49.0% (7.3%)	0.65	0.54

Table 9.2. Association of obesity and BMI with the percentage methylation at the three variably-methylated CpG sites in the Leptin gene CpG island. *Age and sex corrected BMI.

Thirty-two SNPs located within a 200kb region containing the Leptin gene were analysed for association with BMI in the 92 control subjects. This genotyping data was made available by collaborators in France[526]. Four SNPs were nominally associated with BMI (rs2167289, rs791600, rs10249476, rs4731429), however none survive correction using permutations (see Table 9.3). These SNPs were then analysed for association to CpG 5 methylation. One of the SNPs nominally associated with BMI, rs4731429 was nominally associated with CpG 5 methylation (p=0.017, see Table 9.4). This does not survive correction using permutations.

	Genotype Counts			Mean BMI (95% CI)			P-values	
SNP	**1[a],1**	**1,2**	**2,2**	**1,1**	**1,2**	**2,2**	**Empirical**	**Corrected**
rs17151739	24	24	2	22.4 (21.8-22.9)	22.5 (21.9-23.1)	21.8 (18.5-25.00)	0.76	1.0
rs6960004	34	14	2	22.5 (22.0-23.0)	22.2 (21.5-23.0)	22.2 (19.5-24.8)	0.44	1.0
rs6953698	32	17	1	22.6 (22.1-23.1)	22.0 (21.3-22.6)	-	0.32	1.0
rs322812	11	31	8	22.6 (21.8-23.3)	22.4 (21.9-23.0)	22.05 (21.24-22.86)	0.56	1.0
rs322760	13	30	7	22.6 (21.9-23.3)	22.3 (21.8-22.9)	22.3 (21.1-23.4)	0.57	1.0
rs4731401	40	10	0	22.4 (21.9-22.8)	22.6 (21.6-23.6)	-	0.55	1.0
rs322737	13	28	9	22.6 (21.7-23.4)	22.3 (21.8-22.8)	22.5 (21.7-23.3)	0.91	1.0
rs1421309	35	14	1	22.5 (22.0-23.0)	22.0 (21.2-22.8)	-	0.61	1.0
rs10447854	26	22	2	22.3 (21.8-22.8)	22.4 (21.8-23.1)	23.8 (23.4-24.2)	0.25	1.0
rs1017607	16	26	8	22.2 (21.5-22.9)	22.2 (21.6-22.7)	23.5 (22.7-24.4)	0.11	0.92
rs12706827	41	8	1	22.5 (22.0-22.9)	21.8 (20.8-22.9)	-	0.41	1.0
rs6976221	29	17	4	22.3 (21.7-22.8)	22.7 (22.1-23.2)	22.4 (20.2-24.5)	0.34	1.0
rs12538332	30	17	3	22.3 (21.8-22.8)	22.5 (21.8-23.2)	23.1 (211-25.2)	0.85	1.0
rs6956123	36	14	0	22.4 (22.0-22.9)	22.3 (21.5-23.1)	-	0.50	1.0
rs6467165	40	9	1	22.5 (22.0-22.9)	22.1 (21.1-23.0)	-	0.64	1.0
rs4731416	40	10	0	22.3 (21.9-22.7)	22.8 (21.7-23.9)	-	0.50	1.0
rs11981584	40	8	2	22.4 (21.9-22.8)	22.4 (21.2-23.6)	23.4 (20.7-26.1)	0.66	1.0
rs2167289	14	25	11	23.2 (22.4-24.0)	22.3 (21.7-22.8)	21.7 (21.0-22.3)	**0.0069**	0.17
rs791595	35	12	3	22.2 (21.7-22.7)	22.8 (21.9-23.7)	23.1 (22.1-24.0)	0.11	0.93
rs791600	14	25	11	23.2 (22.4-24.0)	22.3 (21.7-22.8)	21.6 (21.0-22.3)	**0.0069**	0.17
rs2021808	44	6	0	22.3 (21.9-22.8)	22.8 (21.6-24.1)	-	0.57	1.0
rs791608	44	6	0	22.5 (22.0-22.9)	22.0 (20.8-23.3)	-	0.35	1.0
rs10249476	20	26	4	22.0 (21.4-22.6)	22.6 (22.0-23.1)	23.5 (22.4-24.6)	0.04	0.58
rs10487506	13	29	8	22.8 (22.0-23.6)	22.4 (21.9-23.0)	21.6 (20.8-22.4)	0.13	0.95
rs4731429	17	25	8	21.6 (21.0-22.1)	22.7 (22.1-23.2)	23.3 (22.4-24.2)	**0.0055**	0.13
rs4731437	35	14	1	22.3 (21.8-22.8)	22.7 (22.0-23.3)	-	0.57	1.0
rs1466146	35	14	1	22.4 (21.9-22.9)	22.6 (21.9-23.3)	-	0.72	1.0
rs6979784	31	15	4	22.4 (21.9-22.9)	22.4 (21.7-23.2)	22.18 (20.3-24.0)	0.67	1.0
rs7811892	22	22	6	22.6 (21.9-23.2)	22.3 (21.7-22.9)	22.1 (20.9-23.3)	0.23	1.0
rs1545444	15	29	6	22.3 (21.5-23.0)	22.3 (21.7-22.8)	23.4 (22.3-24.5)	0.33	1.0
rs12532999	13	30	7	22.3 (21.5-23.0)	22.2 (21.7-22.7)	23.5 (22.6-24.4)	0.17	0.98
rs6964936	20	27	3	22.4 (21.7-23.0)	22.5 (21.9-23.0)	21.8 (20.2-23.4)	0.90	1.0

Table 9.3. Association analysis of SNPs in the Leptin gene region with BMI in French controls using linear regression and age and sex as covariates. Corrected p-values result from one million permutations. [a] 1 denotes the common allele, 2 denotes the minor allele.

	Genotype Counts			Mean CpG 5 Methylation (95% CI)			P-values	
SNP	1[a],1	1,2	2,2	1,1	1,2	2,2	Empirical	Corrected
rs17151739	24	24	2	0.85 (0.77-0.94)	0.91 (0.86-0.95)	0.98 (0.94-1.02)	0.18	0.99
rs6960004	34	14	2	0.89 (0.85-0.94)	0.86 (0.73-0.99)	0.95 (0.92-0.97)	0.86	1.0
rs6953698	32	17	1	0.91 (0.88-0.95)	0.84 (0.72-0.96)	-	0.081	0.84
rs322812	11	31	8	0.85 (0.75-0.95)	0.88 (0.82-0.95)	0.92 (0.88-0.96)	0.40	1.0
rs322760	13	30	7	0.84 (0.76-0.93)	0.89 (0.82-0.96)	0.93 (0.89-0.97)	0.28	1.0
rs4731401	40	10	0	0.9 (0.87-0.93)	0.82 (0.61-1.03)	-	0.19	0.99
rs322737	13	28	9	0.87 (0.73-1.02)	0.9 (0.85-0.95)	0.84 (0.77-0.9)	0.70	1.0
rs1421309	35	14	1	0.91 (0.88-0.94)	0.82 (0.67-0.97)	-	0.25	1.0
rs10447854	26	22	2	0.85 (0.77-0.93)	0.93 (0.91-0.96)	0.75 (0.26-1.24)	0.43	1.0
rs1017607	16	26	8	0.92 (0.89-0.94)	0.87 (0.78-0.95)	0.88 (0.75-1)	0.48	1.0
rs12706827	41	8	1	0.87 (0.81-0.93)	0.94 (0.9-0.98)	-	0.27	1.0
rs6976221	29	17	4	0.86 (0.79-0.93)	0.95 (0.91-0.98)	0.78 (0.57-0.99)	0.78	1.0
rs12538332	30	17	3	0.9 (0.86-0.94)	0.85 (0.73-0.97)	0.94 (0.89-1)	0.80	1.0
rs6956123	36	14	0	0.87 (0.81-0.93)	0.92 (0.84-1)	-	0.36	1.0
rs6467165	40	9	1	0.9 (0.84-0.95)	0.82 (0.71-0.93)	-	0.52	1.0
rs4731416	40	10	0	0.86 (0.81-0.92)	0.96 (0.93-0.99)	-	0.11	0.92
rs11981584	40	8	2	0.87 (0.81-0.92)	0.97 (0.95-0.99)	0.9 (0.79-1)	0.24	1.0
rs2167289	14	25	11	0.85 (0.71-0.99)	0.9 (0.85-0.95)	0.89 (0.84-0.94)	0.57	1.0
rs791595	35	12	3	0.88 (0.82-0.94)	0.92 (0.86-0.97)	0.8 (0.51-1.09)	0.85	1.0
rs791600	14	25	11	0.85 (0.71-0.99)	0.9 (0.85-0.95)	0.89 (0.84-0.94)	0.57	1.0
rs2021808	44	6	0	0.87 (0.82-0.93)	0.95 (0.93-0.97)	-	0.31	1.0
rs791608	44	6	0	0.87 (0.82-0.93)	0.96 (0.94-0.99)	-	0.23	1.0
rs10249476	20	26	4	0.87 (0.77-0.96)	0.91 (0.86-0.95)	0.84 (0.62-1.05)	0.81	1.0
rs10487506	13	29	8	0.91 (0.84-0.98)	0.91 (0.86-0.95)	0.75 (0.53-0.97)	0.063	0.76
rs4731429	17	25	8	0.8 (0.68-0.92)	0.92 (0.88-0.96)	0.95 (0.94-0.96)	**0.017**	0.35
rs4731437	35	14	1	0.9 (0.86-0.94)	0.84 (0.7-0.98)	-	0.39	1.0
rs1466146	35	14	1	0.9 (0.86-0.94)	0.84 (0.7-0.97)	-	0.34	1.0
rs6979784	31	15	4	0.89 (0.85-0.94)	0.91 (0.86-0.95)	0.71 (0.24-1.17)	0.18	0.99
rs7811892	22	22	6	0.87 (0.79-0.96)	0.89 (0.82-0.95)	0.91 (0.85-0.98)	0.64	1.0
rs1545444	15	29	6	0.86 (0.74-0.99)	0.88 (0.83-0.94)	0.94 (0.92-0.97)	0.39	1.0
rs12532999	13	30	7	0.86 (0.78-0.95)	0.88 (0.81-0.95)	0.95 (0.92-0.98)	0.35	1.0
rs6964936	20	27	3	0.85 (0.75-0.95)	0.92 (0.87-0.96)	0.83 (0.67-0.99)	0.40	1.0

Table 9.4. Association analysis of SNPs in the Leptin gene region with CpG 5 methylation in French controls using linear regression and age and sex as covariates. Corrected p-values result from one million permutations. [a] 1 denotes the common allele, 2 denotes the minor allele.

9.2.5 Analysis of *LEP* transcript levels

Leptin transcription data from the Swedish sib-pair cohort was analysed in order to ascertain whether or not expression of the gene is altered in subcutaneous adipose tissue. When siblings were split into two groups of non-obese (BMI<30 kg/m^2) and obese (BMI>30kg/m^2) and the transcription levels corrected for age, sex and relatedness, the two groups were normally distributed. Using an independent samples t-test, a significant difference in the level of *LEP* transcription was found between the two groups ($p=2.2x10^{-16}$, see Figure 9.9 and Table 9.5).

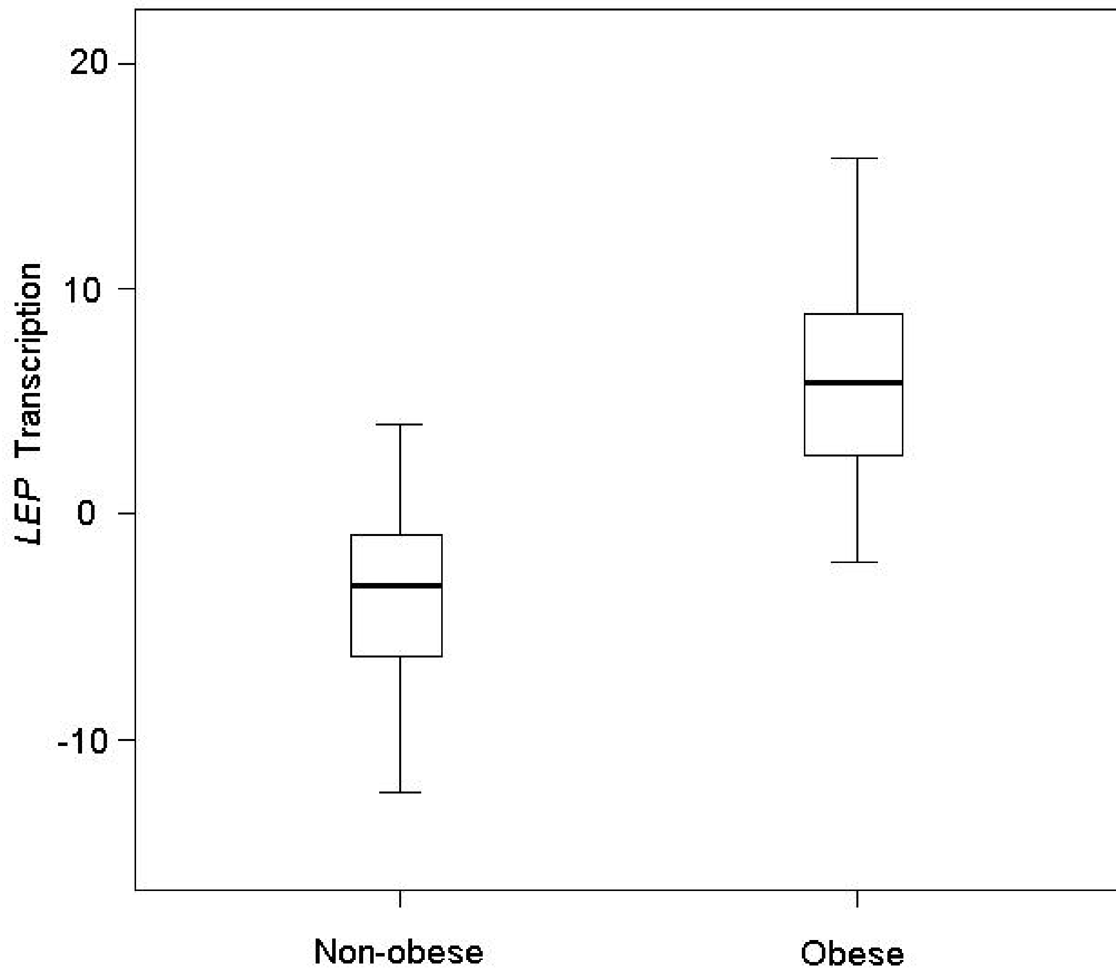

Figure 9.9. Box-plot of LEP transcript levels in obese and non-obese subjects from Swedish sib-pairs. SIRT1 transcription is the microarray signal value corrected for age, sex and relatedness. Mean transcription ± standard deviation were -3.57 ± 3.56 in the non-obese and 6.09 ± 4.19 in the obese group.

Number of Non-obese	Number of Obese	Mean Difference in *SIRT1* transcription (95% CI)	P-value
190	156	9.66 (8.92-10.41)	2.20×10^{-16}

Table 9.5. T-test of *LEP* transcription in obese and non-obese subjects from Swedish families. *LEP* transcription is the microarray signal value corrected for age, sex and relatedness.

9.3 Discussion

These results demonstrate an effective method for measuring quantitative methylation status at a single CpG site using direct bisulfite-PCR sequencing and low quantities of genomic DNA (~10ng per sample). It should be noted that during the course of this project, the same method was independently developed and published by Jiang *et al.*[527]. The relative speed and simplicity of this method compared to cloning and pyrosequencing techniques make it ideally suited to screening larger numbers of samples for use in case-control association studies.

These results demonstrate a nominal association that does not survive correction for multiple testing between leptin promoter methylation and obesity in humans. This study provides a good example of the utility of analysing DNA methylation as a quantitative trait, as it is both a statistically powerful approach and it provides information at the single base level. At the time the practical work in this thesis was completed, no associations had been reported between DNA methylation measured as a quantitative trait and polygenic human obesity. Since then, a study that reported an association between promoter methylation of serotonin transporter gene in blood leukocytes and obesity in monozygotic twins has been published[528].

Methylation of CpG 5 was found to be decreased in obese subjects compared to non-obese. As increased methylation of a CpG island typically leads to a reduction in expression of a gene[529], this suggests that, in the obese subjects, expression of leptin is upregulated, possibly as part of the feedback mechanism to lower food intake through leptin's action on the brain[530]. This is supported by the evidence that expression of the gene was found to be significantly increased in obese compared to non-obese siblings in the Swedish family cohort, an established phenomenon that might be expected since leptin secretion is increased in obese subjects [531,532].Exploration of leptin CpG methylation levels in this cohort would be a crucial next step in order to acquire evidence that methylation levels in this gene are effecting its expression.

This study examined methylation levels in DNA samples obtained from whole blood and while it may be the case that methylation levels in blood are highly correlated with those in adipose tissue, this has not been proven and so needs to be investigated in the future. Equally, there may be additional environmental signals that are affecting methylation in blood-derived DNA that are driving the changes in leptin methylation or it might be that obesity is modifying methylation patterns as seen in a recent study in rats[393].

The association of rs4731429 with percentage methylation at CpG 5 could indicate a possible link between DNA methylation and the primary DNA sequence, if it were statistically significant. An association between a SNP and percentage methylation at a single CpG could suggest that in this case, epigenetic changes at this site are both heritable and in linkage disequilibrium (LD) with variation in the primary sequence. Such an association has been previously reported in the cancer predisposition gene MGMT[529]. It is not clear what the basis of the association is between the SNP and methylation at the single CpG as there is no evidence whether it is functional or simply in LD with another functional variant. While the significance of the associations of rs4731429 to obesity and BMI do not withstand correction for multiple testing, they underline the possibility that the two types of variation are statistically associated because they are both associated with the same phenotype, either statistically or functionally.

This result will need to be replicated in larger numbers and in other populations to confirm the associations reported here, as it is very clear that this study is underpowered, especially for the SNP association work. Confirming high correlation between methylation levels of these CpG sites in adipose tissue and blood is also desirable. Comparison of percentage methylation and leptin transcript expression levels would be very useful in determining whether differential methylation of the leptin promoter is associated with BMI because of altered leptin transcription. The heritability of the CpG methylation also needs to be investigated in families to confirm the possibility that the methylation and SNP association reported here is highlighting a novel inter-relationship between two types of genomic variation.

In conclusion the results reported here highlight the potential importance of DNA methylation in either the development of obesity or in response to the obese state. It also demonstrates the utility of single-base resolution measurement of methylation within candidate genes.

Chapter 10

Discussion

10.1 Summary of thesis results

The work presented in this thesis used a candidate gene approach to attempt to identify genetic variation contributing to the risk of developing common polygenic obesity. These studies were conducted over a period of rapid development in the field of complex disease genetics. At the start of this project, genome-wide association studies had recently become feasible and numerous reproducibly associated markers to common obesity have been discovered using these methods over the last few years. The first of these were two reports of a novel association to *FTO*[246,247] and the most recent brings the total validated genetic associations to BMI or obesity to 34[233,256]. However, only an estimated 1.5% of the heritability of BMI is currently explained in GWAS association. Due to the hypothesis-free strategy and large numbers of variants being screened, very stringent corrections are required in order to minimize the chances of a false-positive, which unfortunately results in an inflation of the false-negative rate. This means that some variants with small effect sizes will be screened out. .This project was designed in order to investigate the possibility of genetic associations between common variants (MAF>5%) and coding SNPs in candidate genes and human polygenic obesity. Out of 67 successfully genotyped markers, 12 were not present in HapMap and a further six are located on the X chromosome which makes 18 SNPs not likely to be included in GWAS and demonstrates the advantage of this approach.

Chapter 3 described an investigation of the *SIRT1* gene, which was chosen on the basis of its transcriptional regulation of genes involved in lipid metabolism[418] and *SIRT1* knockout mice having increased body weight[421]. Putative associations between variants in the *SIRT1* gene and common obesity were found in both an adult French case-control cohort and a Swedish family cohort. When starting this project, no associations between *SIRT1* SNPs and obesity phenotypes had been published, however three studies have since reported significant associations[442,443,444]. This means that the association between *SIRT1* variants and obesity has

been found across a number of populations which is some evidence that it is genuine and not a false positive. All reported associations found a correlation between the minor allele and a reduced BMI or obesity risk which is evidence that the same association is being detected and therefore more likely that it is genuine. No associations to *SIRT1* have been reported in any GWAS, however nominally significant SNP associations have been detected in the region (e.g. rs471962 with a p-value of 1.25×10^{-4}) indicating the possibility of an association that was screened out by multiple testing correction. The association between adipose *SIRT1* transcript expression and obesity produced more convincing p-values (1.56×10^{-35}) and is further evidence of the influence of SIRT1 on obesity risk. The results of this chapter have been published[533].

In Chapter 4, the apelin gene was selected due to its plasma concentration correlating with BMI[460] and studies that reported reduced food intake in rats injected with apelin peptides[466,467]. One SNP within the gene was found to be nominally associated with obesity in a French child case-control cohort. However, this association does not reach genome-wide significance and has not been replicated so could be a false-positive. No previous associations between *APLN* variation and obesity were found in the literature which means this is a potentially interesting discovery which will require further investigation in order to confirm it's validity. The lack of X chromosome variants reported in GWAS makes this an interesting result.

Chapter 5 detailed the investigation of IL-11, an anti-inflammatory cytokine with inhibitory effects on adipogenesis and lipase activity[354,355,356]. A nominal association between *IL11* variation and obesity was found in the French adult case-control cohort. The lack of genome-wide significance and replication means that the possibility of a false-positive has not been ruled out. No associations between obese phenotypes and IL-11 variation have been reported to-date. The limited coverage of *IL-11* variants in HapMap makes this a particularly interesting result since it is likely that current GWAS do not include any SNPs that tag the most associated SNP.

Chapter 6 reported an investigation into adiponutrin, a lipase that is expressed in adipocytes[483,360]. A study in Swedish subjects reported an association between adiponutrin gene SNPs and obesity[369]. No significant associations were found to any SNP genotyped, including those that had previously been reported. This highlights the importance of replication studies to establish true associations. However, the limited statistical power means that a false-negative cannot be ruled out and this highlights a weakness of using small cohorts in association studies.

The nesfatin gene was chosen in Chapter 7 due to its expression in appetite controlling neurons in the hypothalamus[371,372] and its anorexigenic effect when injected into rodent brains[372]. Unfortunately, there was no evidence of any associations between SNPs in the region and obesity that were statistically significant after permutations. Again, limited statistical power means that the possibility of a false-negative cannot be ruled out.

In Chapter 8, the *IRS1* gene was investigated having recently been reported to be associated with T2D by collaborators in France. Genotype and transcript data from the Swedish sib-pair cohort was analysed and a significant association between *IRS-1* variation and transcription levels of the gene was discovered that could explain the mechanism of the association to T2D. However, *IRS1* variation was not found to be associated with BMI. There was a significant association between *IRS1* transcript levels and obesity but it might be more likely that obesity is altering transcription of the gene rather than the gene causing obesity.

The putative associations described above all survive correction for multiple testing using permutations when each gene is analysed individually, which reduces the chance of false positives. The p-values would not be significant in the context of a genome-wide investigation and do not survive thesis-wide corrections which means that the possibility that these associations are false-positives cannot be ruled out. This, along with replication are neccessary in order to consider a genetic association to be real.

Because there is a large gap between the variation in BMI that can be attributed to published GWAS and the heritability estimates for common obesity, there remain many more associations yet to be discovered. These will most likely come from rare or common variants not investigated in current GWAS or common variants with small effect sizes. This highlights the continued relevance of candidate gene studies which can be used to investigate a much greater proportion of the variation within a single gene region. Candidate gene studies can therefore act as a complementary approach to GWAS by following up on signals that do not survive the genome-wide corrections but are at least nominally significant.

None of the associations reported in this thesis have been found to be significantly associated to BMI or obesity in any GWAS to date. This raises the question of whether or not these are genuine genetic associations and if so why are they not showing up in the GWAS. One possible reason is that they are being screened out by multiple testing corrections. GWAS routinely reject p-values $>5x10^{-8}$ as statistical noise. In order to control the false-positive rate though the false-negative rate is increased and so many real associations may be missed, particularly if effect sizes are small. Another explanation is that the associations are to variants that are not sufficiently tagged by SNPs in current genotyping arrays due to incomplete coverage in the HapMap database or location within the sex chromosomes. Two of the associations reported in this thesis are not likely to be covered in GWAS. Firstly, the apelin SNP, rs2281068, which was putatively associated with obesity in children has not been analysed for association to any obese phenotype in GWAS due to its location on the X chromosome. Secondly, the putatively associated IL-11 SNP, rs1042505, may not be sufficiently tagged in any GWAS since it is not present in HapMap and is not well correlated with any other SNP within the region (maximum r^2=0.31).

Another possibility is that some of this missing heritability is caused by epigenetic variation. This was explored in Chapter 9 where DNA methylation in the leptin gene was measured and analysed for association to common obesity. Initially, a method to measure quantitative methylation in large numbers of samples was developed since CpG methylation can vary within a single tissue type. When using DNA extracted from blood, which contains multiple cell types this was even more important. Direct bisulfite sequencing was used and DNA methylation at the individual base level was analysed by measuring peak height ratios from the electropherogram. Methylation levels were measured at six sites in the leptin promoter CpG island in a subset of French obesity case-controls. A nominal association was discovered between one CpG site and obesity. This is a potentially interesting result since no associations between quantitative methylation in the leptin gene and obesity have been reported to date. However, the p-values obtained are not genome-wide significant and this, together with the fact that the association has not been replicated, means that the possibiity of a false-positive cannot be ruled out so no definitive conclusions can be drawn at present. It is essential that further samples are investigated and with the high-throughput method that was developed here it should be possible to routinely investigate as many as 500 cases and 500 controls.

10.2 Future work

The first objective of any future work will be to replicate the associations discovered here in other populations to establish whether or not they are genuine. Larger cohorts should be used in order to increase the statistical power such that associations that are significant in a genome-wide context can be detected, ideally in line with the numbers suggested by the power calculations. Using the GIANT approach it may be possible to impute the relevant SNP genotypes in sufficient samples to reach the required numbers without further laboratory work being required.

The second objective is discovering the DNA sequences that are directly implicated in the pathophysiology of obesity. SNPs that have been found to be associated with common polygenic obesity may not be causative but simply inherited together with a causative variant that alters the amino acid sequence of the encoded protein or expression levels of a gene. Fine-mapping of associated genes is required in order to discover these causative variants. The 1000 Genomes project may provide the necessary fine mapping information but it is notable that this project is not sequencing subjects chosen on the basis of disease status, in the case of this project, high BMI or morbid obesity. This means that those functional variants underlying obesity may not be being sequenced. Identification of functional variants is important since an understanding of how variation in the gene effects phenotype will lead to a better understanding of the aetiology of the disease which could aid in the development of therapeutics that target the disease. Additionally finding the functional mutations will lead to better estimations of effect sizes and thus how much the currently known associations explain the total heritability of the disease. This is particularly helpful if multiple variants within a gene each affect the phenotype independently. In the future, genetic information could be used to predict disease risk and aid prognosis and so the accuracy of such information will be improved by mapping the functional variants.

The strategy for locating functional mutations will involve fine-mapping of the associated genes by sequencing in a subset of samples followed by association analysis of any SNPs discovered. This is a potentially onerous task since LD blocks can span regions that can be hundreds of kilobases in length and transcription factors can affect gene expression *via* enhancer elements located similarly long distances away from a gene. This difficulty is illustrated by Chapters 3 and 4 where sequencing was carried out in the exons and promoter regions of *SIRT1* and *APLN* but

no significantly associated SNPs were discovered. Next generation sequencing technologies will aid in this process by reducing costs and increasing speed. Follow up genotyping in populations with different LD structures would also be useful to help differentiate functional SNPs from those that are associated due to being in high LD with the functional SNP.

Any SNPs found within exons that are significantly associated with obesity are strong candidates for functional mutations and there are a number of strategies to test this hypothesis. Firstly, software such as Polyphen[534] can be used to evaluate whether a non-synonymous mutation could be damaging to the protein. Other mutations might affect transcription factor binding or splicing and this can be assessed computationally too. However, molecular validation will be necessary. Levels of mRNA expression can be measured using reverse-transcriptase PCR or RNA sequencing in the tissue of interest, to see if the mutant allele alters transcription levels. A mutation may effect phenotype by altering protein expression by altering RNA stability or introducing splice defects. Protein expression can be assessed by Western blotting, assuming that a suitable specific antibody is available. Enzyme activity assays can be used if the protein's has an enzymatic activity that is well understood. For example, sirtuin activity assays are available[449] that could be used to compare mutant versions of SIRT1 with wild type.

Alternatively, if the associated variants are not coding SNPs then the possibility that they affect expression levels of the gene in question can be looked at. This was investigated in Chapter 3 which examined SIRT1 expression in adipose tissue, although no association was found between gene variation and transcript expression levels. In Chapter 8 association was detected between IRS1 gene variation and its transcript expression levels in adipose tissue. Whilst these methods can suggest a mechanism by which gene variation alters mRNA expression, they do not pinpoint the exact mutation that is responsible. Reporter gene assays such as luciferease could be used instead to measure which variants effect transcription.

The association between CpG methylation in the leptin gene and obesity will need replicating in other populations. It is also necessary to confirm the association in DNA extracted from adipose tissue where leptin is expressed as there is no clear mechanism by which altered methylation of the gene in blood cells could influence BMI, assuming the association is genuine. The presumption is that the methylation status in blood simply reflects the methylation status in the adipose tissue. The next step will be to confirm whether this assumption is true and then, ideally, go on to measure transcript levels of the gene to establish whether or not they correlate with CpG methylation. Ultimately, these techniques can then be used to explore the possibility that DNA methylation is associated with obesity in other candidate genes using the same technique or across the genome using techniques such as the Illumina methylation microarrays.

A major challenge in the field of complex genetics is the discovery of the remaining variants that contribute towards the heritability of traits and diseases such as BMI and obesity. There are a number of possibilities for where this missing heritability will be found. It may be a consequence of common variants with small effect sizes that are not currently picked up by GWAS due to the stringent multiple testing corrections. Given that the lower the odds ratio, the larger the study needed to reach statistical significance, studying these variants will require larger and larger cohorts and using current GWAS approaches they will have to be in the millions of samples as the GIANT meta-analysis using approximately 250,000 subjects was unable to account for more than a few percent of the genetic variation.

The missing heritability could also be explained by rare variants that are not adequately tagged in current GWAS. A recent paper reported that rare variants are more likely to be found in functional regions of the genome than common variants suggesting that the majority of phenotypic variation could come from rare alleles[535]. Novel associations to rare variants will require much denser genotyping arrays or, alternatively, as sequencing cost continue to fall, exome or whole genome sequencing. This can be carried out in a subset of samples with the

biggest hits followed up with genotyping in larger sample sets. When sequencing costs drop even further, whole genome sequencing can be carried out in large cohorts. Exome sequencing may not be the best approach when searching for variants that influence complex traits since it is likely that many causative mutations lie outside of exons, however its cost effectiveness will make it an attractive first choice compared to whole genome sequencing.

It is obviously possible that common variants associated with obesity in GWAS may be tagging functional variants that have much lower minor allele frequencies. These rare, functional mutations could be expected to have effect sizes that are larger than those currently reported since some of the statistical association between genetic variant and disease will be lost by the imperfect tagging of the marker variant to the functional variant. Associations to rare variants tend to have larger effect sizes than those to common variants[536] so this is a possibility. Fine mapping of GWAS results will help to pinpoint functional mutations and with the recent technological developments in DNA sequencing, this process will become cheaper and less time-consuming. A simple strategy would involve sequencing in a subset of subjects used in the original GWAS followed by imputation in the remainder.

It might be the case that within each gene that influences disease risk, there are multiple independent variants that each exert effects on phenotype. Current methods might miss this since GWAS are designed to identify the top hit within an associated region and replicate only this single variant. For example, the major histocompatibility complex (MHC) region has been shown to have multiple independent effects in autoimmune diseases, such as type 1 diabetes[537] and rheumatoid arthritis[538]. These effects will be more properly understood after fine-mapping of associations is undertaken.

Another possibility is that the genetic variation that is contributing to the heritability is not related to SNPs at all. CNVs are likely to contribute too. Between any two individuals the genetic base-pair differences due to CNVs is at least 100 times higher than that due to SNPs[539]. The association between BMI and *NEGR1* reported in the first GIANT publication has been hypothesised to be caused by a 45kb deletion upstream of the gene[263]. Rare deletions have been successfully demonstrated to be associated with obesity[271,540] and subsequently a duplication in the same region was associated with low BMI[270]. A better understanding of the effects of CNVs will come from genome sequencing as they are currently relatively difficult to characterise.

One area that has yet to be addressed within the field is epistasis, that is interactions between genetic variants. A recent paper suggests that epistasis could be causing an overestimation of the heritability of complex traits which means that the proportion of heritability currently explained is larger than is thought[541]. Statistical power to detect interactions between two variants are currently very low which means that they are unlikely to be detected in the current GWAS and will either need larger cohorts in order to be explored successfully or novel analytical techniques will need to be developed to avoid the exponential increase in the numbers of statistical tests as gene-gene interactions are examined. The same will be true of gene-environment effects which have the added difficulty of accurately assessing the environmental variables.

Lastly, some portion of heritability may come from epigenetic variation such as DNA methylation and histone modifications. DNA methylation is becoming amenable to examination using genome-wide methylation platforms in the same cohorts used for genetic association studies. Whole methylome sequencing by bisulphite (Bis-seq) or using methylated DNA

immunoprecipitation (MeDIP-seq) can be used initially to discover candidate variants which can then be followed up in larger sample sets using the same method of bisulphite sequencing employed in this thesis. Further technological developments will be needed before the wide range of histone modifications can be screened for routinely.

Another point to consider is phenotyping. There are likely to be several different pathways that can go wrong that result in obesity and therefore there may be several separate phenotypes that are currently grouped together as one disease, with each one having different associated genetics. That is, common diseases are currently classified according to a common effect rather than a common cause. Furthermore, the use of BMI as a phenotype has its limitations. In the non-obese population, from which most GWAS for BMI are sampled (for example the mean BMI of samples in the discovery phase of the first GIANT publication was between 25.1 and 27.5 with less than 20% of subjects >30kg/m^2 [263]), the variation in BMI depends on lean mass as well as fat mass[542,543]. This means that variants affecting fat mass and susceptibility to obesity are not directly assessed. Better phenotyping of samples, to obtain more detailed information on body size and fat distribution, could therefore improve the accuracy of genetic studies.

The next challenge in human disease genetics, after the functional variants that control the heritable portion of disease risk have been identified, will be clinical application of this knowledge. An increased understanding of the pathways involved in controlling BMI and obesity risk has the potential to lead to the development of drugs that target proteins in these pathways. In the future, genome sequencing could be used as a standard tool for diagnosis and prognosis of genetic disease so a more complete understanding of all the variants that effect obesity risk will be important.

In conclusion, this thesis reports the possibility of novel associations between two genes, IL-11 and APLN and common polygenic obesity as well as another gene, SIRT1 for which previous associations have been published. Additionally, the possibility of a novel association was discovered between CpG methylation in the leptin gene and obesity. These findings add to the body of knowledge of obesity genetics, which in the future could aid in developing therapeutics for this widespread and serious disease. It is also clear that there is much still to be discovered in the field of obesity genetics in order to complete our understanding of the heritability of this disease.

References

1. World Health Organization Fact sheet: Obesity and Overweight. at <http://www.who.int/mediacentre/factsheets/fs311/en/index.html>
2. Guh, D. P. *et al.* The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC Public Health* **9**, 88 (2009).
3. Murugan, A. T. & Sharma, G. Obesity and respiratory diseases. *Chron Respir Dis* **5**, 233–42 (2008).
4. Ting, S. M. S., Nair, H., Ching, I., Taheri, S. & Dasgupta, I. Overweight, obesity and chronic kidney disease. *Nephron Clin Pract* **112**, c121–7; discussion c127 (2009).
5. Wearing, S. C., Hennig, E. M., Byrne, N. M., Steele, J. R. & Hills, A. P. Musculoskeletal disorders associated with obesity: a biomechanical perspective. *Obes Rev* **7**, 239–50 (2006).
6. Batty, G. D. *et al.* Obesity and overweight in relation to liver disease mortality in men: 38 year follow-up of the original Whitehall study. *Int J Obes (Lond)* **32**, 1741–4 (2008).
7. Yach, D., Stuckler, D. & Brownell, K. D. Epidemiologic and economic consequences of the global epidemics of obesity and diabetes. *Nat Med* **12**, 62–6 (2006).
8. Müller-Riemenschneider, F., Reinhold, T., Berghöfer, A. & Willich, S. N. Health-economic burden of obesity in Europe. *Eur J Epidemiol* **23**, 499–509 (2008).
9. Leibel, R. L., Rosenbaum, M. & Hirsch, J. Changes in energy expenditure resulting from altered body weight. *N. Engl. J. Med.* **332**, 621–628 (1995).
10. Grande, F., Anderson, J. T. & Keys, A. Changes of Basal Metabolic Rate in Man in Semistarvation and Refeeding. *J Appl Physiol* **12**, 230–238 (1958).
11. Bultman, S. J., Michaud, E. J. & Woychik, R. P. Molecular characterization of the mouse agouti locus. *Cell* **71**, 1195–1204 (1992).
12. INGALLS, A. M., DICKIE, M. M. & SNELL, G. D. Obese, a new mutation in the house mouse. *J. Hered.* **41**, 317–318 (1950).
13. Bahary, N. *et al.* Microdissection of proximal mouse chromosome 6: identification of RFLPs tightly linked to the ob mutation. *Mamm. Genome* **4**, 511–515 (1993).
14. Zhang, Y. *et al.* Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432 (1994).
15. Hummel, K. P., Dickie, M. M. & Coleman, D. L. Diabetes, a new mutation in the mouse. *Science* **153**, 1127–1128 (1966).
16. Chen, H. *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* **84**, 491–495 (1996).
17. Lee, G.-H. *et al.* Abnormal splicing of the leptin receptor in diabetic mice. *, Published online: 15 February 1996; | doi:10.1038/379632a0* **379**, 632–635 (1996).
18. Zucker, L. M. & Zucker, T. F. Fatty, a New Mutation in the Rat. *J Hered* **52**, 275–278 (1961).
19. Ogawa, Y. *et al.* Molecular cloning of rat obese cDNA and augmented gene expression in genetically obese Zucker fatty (fa/fa) rats. *J. Clin. Invest.* **96**, 1647–1652 (1995).
20. Kawano, K. *et al.* Spontaneous long-term hyperglycemic rat with diabetic complications. Otsuka Long-Evans Tokushima Fatty (OLETF) strain. *Diabetes* **41**, 1422–1428 (1992).
21. Takiguchi, S. *et al.* Disrupted cholecystokinin type-A receptor (CCKAR) gene in OLETF rats. *Gene* **197**, 169–175 (1997).
22. Collins, S., Martin, T. L., Surwit, R. S. & Robidoux, J. Genetic vulnerability to diet-induced obesity in the C57BL/6J mouse: physiological and molecular characteristics. *Physiology & behavior* **81**, 243–248 (2004).

23. Hetherington, A. W. & Ranson, S. W. Hypothalamic lesions and adiposity in the rat. *The Anatomical Record* **78**, 149–172 (1940).
24. Aravich, P. F. & Sclafani, A. Paraventricular hypothalamic lesions and medial hypothalamic knife cuts produce similar hyperphagia syndromes. *Behavioral Neuroscience* **97**, 970–983 (1983).
25. Anand, B. K. & Brobeck, J. R. Hypothalamic Control of Food Intake in Rats and Cats. *Yale J Biol Med* **24**, 123–140 (1951).
26. Scott, D. E. & Pepe, G. J. The fetal baboon median eminence as a circumventricular organ: I. Transmission electron microscopy. *Brain Research Bulletin* **19**, 87–94 (1987).
27. Cone, R. D. *et al.* The arcuate nucleus as a conduit for diverse signals relevant to energy homeostasis. *Int. J. Obes. Relat. Metab. Disord.* **25 Suppl 5**, S63–67 (2001).
28. Tatemoto, K. Isolation and characterization of peptide YY (PYY), a candidate gut hormone that inhibits pancreatic exocrine secretion. *Proc Natl Acad Sci U S A* **79**, 2514–2518 (1982).
29. Tatemoto, K., Carlquist, M. & Mutt, V. Neuropeptide Y|[mdash]|a novel brain peptide with structural similarities to peptide YY and pancreatic polypeptide. *, Published online: 15 April 1982; | doi:10.1038/296659a0* **296**, 659–660 (1982).
30. Morris, B. J. Neuronal localisation of neuropeptide Y gene expression in rat brain. *J. Comp. Neurol.* **290**, 358–368 (1989).
31. Levine, A. S. & Morley, J. E. Neuropeptide Y: a potent inducer of consummatory behavior in rats. *Peptides* **5**, 1025–1029 (1984).
32. Yu, Y., Deng, C. & Huang, X.-F. Obese reversal by a chronic energy restricted diet leaves an increased Arc NPY/AgRP, but no alteration in POMC/CART, mRNA expression in diet-induced obese mice. *Behav. Brain Res.* **205**, 50–56 (2009).
33. Shutter, J. R. *et al.* Hypothalamic expression of ART, a novel gene related to agouti, is up-regulated in obese and diabetic mutant mice. *Genes Dev.* **11**, 593–602 (1997).
34. Rossi, M. *et al.* A C-Terminal Fragment of Agouti-Related Protein Increases Feeding and Antagonizes the Effect of Alpha-Melanocyte Stimulating Hormone in Vivo. *Endocrinology* **139**, 4428–4431 (1998).
35. Goto, K. *et al.* Acute intracerebroventricular administration of either carboxyl-terminal or amino-terminal fragments of agouti-related peptide produces a long-term decrease in energy expenditure in rats. *Int. J. Mol. Med.* **12**, 379–383 (2003).
36. Makimura, H., Mizuno, T. M., Mastaitis, J. W., Agami, R. & Mobbs, C. V. Reducing hypothalamic AGRP by RNA interference increases metabolic rate and decreases body weight without influencing food intake. *BMC Neurosci* **3**, 18 (2002).
37. Ollmann, M. M. *et al.* Antagonism of Central Melanocortin Receptors in Vitro and in Vivo by Agouti-Related Protein. *Science* **278**, 135–138 (1997).
38. Fan, W., Boston, B. A., Kesterson, R. A., Hruby, V. J. & Cone, R. D. Role of melanocortinergic neurons in feeding and the agouti obesity syndrome. *, Published online: 09 January 1997; | doi:10.1038/385165a0* **385**, 165–168 (1997).
39. Mizuno, T. M. & Mobbs, C. V. Hypothalamic Agouti-Related Protein Messenger Ribonucleic Acid Is Inhibited by Leptin and Stimulated by Fasting. *Endocrinology* **140**, 814–817 (1999).
40. Hahn, T. M., Breininger, J. F., Baskin, D. G. & Schwartz, M. W. Coexpression of Agrp and NPY in fasting-activated hypothalamic neurons. *Nature Neuroscience* **1**, 271–272 (1998).
41. Ziotopoulou, M., Mantzoros, C. S., Hileman, S. M. & Flier, J. S. Differential Expression of Hypothalamic Neuropeptides in the Early Phase of Diet-Induced Obesity in Mice. *Am J Physiol Endocrinol Metab* **279**, E838–E845 (2000).
42. Qian, S. *et al.* Neither Agouti-Related Protein nor Neuropeptide Y Is Critically Required for the Regulation of Energy Homeostasis in Mice. *Mol. Cell. Biol.* **22**, 5027–5035 (2002).

43. Gropp, E. *et al.* Agouti-related peptide-expressing neurons are mandatory for feeding. *Nat Neurosci* **8**, 1289–91 (2005).
44. Broberger, C., Johansen, J., Johansson, C., Schalling, M. & Hökfelt, T. The neuropeptide Y/agouti gene-related protein (AGRP) brain circuitry in normal, anorectic, and monosodium glutamate-treated mice. *Proc Natl Acad Sci U S A* **95**, 15043–8 (1998).
45. Mercer, J. G. *et al.* Coexpression of Leptin Receptor and Preproneuropeptide Y mRNA in Arcuate Nucleus of Mouse Hypothalamus. *Journal of Neuroendocrinology* **8**, 733–735 (1996).
46. Cowley, M. A. *et al.* Leptin activates anorexigenic POMC neurons through a neural network in the arcuate nucleus. *Nature* **411**, 480–484 (2001).
47. Wu, Q., Boyle, M. P. & Palmiter, R. D. Loss of GABAergic signaling by AgRP neurons to the parabrachial nucleus leads to starvation. *Cell* **137**, 1225–1234 (2009).
48. Qu, D. *et al.* A role for melanin-concentrating hormone in the central regulation of feeding behaviour. *Nature* **380**, 243–247 (1996).
49. Hervé, C. & Fellmann, D. Changes in rat melanin-concentrating hormone and dynorphin messenger ribonucleic acids induced by food deprivation. *Neuropeptides* **31**, 237–242 (1997).
50. Skofitsch, G., Jacobowitz, D. M. & Zamir, N. Immunohistochemical localization of a melanin concentrating hormone-like peptide in the rat brain. *Brain Res. Bull.* **15**, 635–649 (1985).
51. Shimada, M., Tritos, N. A., Lowell, B. B., Flier, J. S. & Maratos-Flier, E. Mice lacking melanin-concentrating hormone are hypophagic and lean. *Nature* **396**, 670–674 (1998).
52. Sakurai, T. *et al.* Orexins and Orexin Receptors: A Family of Hypothalamic Neuropeptides and G Protein-Coupled Receptors that Regulate Feeding Behavior. *Cell* **92**, 573–585 (1998).
53. De Lecea, L. *et al.* The Hypocretins: Hypothalamus-Specific Peptides with Neuroexcitatory Activity. *PNAS* **95**, 322–327 (1998).
54. Cai, X. J. *et al.* Hypothalamic orexin expression: modulation by blood glucose and feeding. *Diabetes* **48**, 2132–2137 (1999).
55. Skofitsch, G. & Jacobowitz, D. M. Immunohistochemical mapping of galanin-like neurons in the rat central nervous system. *Peptides* **6**, 509–546 (1985).
56. Kyrkouli, S. E., Stanley, B. G., Seirafi, R. D. & Leibowitz, S. F. Stimulation of feeding by galanin: Anatomical localization and behavioral specificity of this peptide's effects in the brain. *Peptides* **11**, 995–1001 (1990).
57. Wang, J. & Leibowitz, K. L. Central insulin inhibits hypothalamic galanin and neuropeptide Y gene expression and peptide release in intact rats. *Brain Research* **777**, 231–236 (1997).
58. Sahu, A. Evidence Suggesting That Galanin (GAL), Melanin-Concentrating Hormone (MCH), Neurotensin (NT), Proopiomelanocortin (POMC) and Neuropeptide Y (NPY) Are Targets of Leptin Signaling in the Hypothalamus. *Endocrinology* **139**, 795–798 (1998).
59. Adams, A. C., Clapham, J. C., Wynick, D. & Speakman, J. R. Feeding behaviour in galanin knockout mice supports a role of galanin in fat intake and preference. *J. Neuroendocrinol.* **20**, 199–206 (2008).
60. Juréus, A., Cunningham, M. J., McClain, M. E., Clifton, D. K. & Steiner, R. A. Galanin-like peptide (GALP) is a target for regulation by leptin in the hypothalamus of the rat. *Endocrinology* **141**, 2703–2706 (2000).
61. Matsumoto, Y. *et al.* Galanin-like peptide stimulates food intake in the rat. *Neuroscience Letters* **322**, 67–69 (2002).
62. Krasnow, S. M. *et al.* A role for galanin-like peptide in the integration of feeding, body weight regulation, and reproduction in the mouse. *Endocrinology* **144**, 813–822 (2003).

63. Dubé, D., Lissitzky, J. C., Leclerc, R. & Pelletier, G. Localization of alpha-melanocyte-stimulating hormone in rat brain and pituitary. *Endocrinology* **102**, 1283–1291 (1978).
64. Mains, R. E., Eipper, B. A. & Ling, N. Common precursor to corticotropins and endorphins. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3014–3018 (1977).
65. Grill, H. J., Ginsberg, A. B., Seeley, R. J. & Kaplan, J. M. Brainstem Application of Melanocortin Receptor Ligands Produces Long-Lasting Effects on Feeding and Body Weight. *J. Neurosci.* **18**, 10128–10135 (1998).
66. Yaswen, L., Diehl, N., Brennan, M. B. & Hochgeschwender, U. Obesity in the mouse model of pro-opiomelanocortin deficiency responds to peripheral melanocortin. *Nat. Med.* **5**, 1066–1070 (1999).
67. Hochgeschwender, U., Costa, J. L., Reed, P., Bui, S. & Brennan, M. B. Altered glucose homeostasis in proopiomelanocortin-null mouse mutants lacking central and peripheral melanocortin. *Endocrinology* **144**, 5194–5202 (2003).
68. Mizuno, T. M. *et al.* Hypothalamic pro-opiomelanocortin mRNA is reduced by fasting and [corrected] in ob/ob and db/db mice, but is stimulated by leptin. *Diabetes* **47**, 294–297 (1998).
69. Huszar, D. *et al.* Targeted Disruption of the Melanocortin-4 Receptor Results in Obesity in Mice. (1997).at <http://linkinghub.elsevier.com/retrieve/pii/S0092867400818656>
70. Ste. Marie, L., Miura, G. I., Marsh, D. J., Yagaloff, K. & Palmiter, R. D. A metabolic defect promotes obesity in mice lacking melanocortin-4 receptors. *Proc Natl Acad Sci U S A* **97**, 12339–12344 (2000).
71. Nogueiras, R. *et al.* The central melanocortin system directly controls peripheral lipid metabolism. *J Clin Invest* **117**, 3475–3488 (2007).
72. Mountjoy, K. G., Mortrud, M. T., Low, M. J., Simerly, R. B. & Cone, R. D. Localization of the melanocortin-4 receptor (MC4-R) in neuroendocrine and autonomic control circuits in the brain. *Mol. Endocrinol.* **8**, 1298–1308 (1994).
73. Bagnol, D. *et al.* Anatomy of an endogenous antagonist: relationship between Agouti-related protein and proopiomelanocortin in brain. *J. Neurosci.* **19**, RC26 (1999).
74. Butler, A. A. *et al.* A unique metabolic syndrome causes obesity in the melanocortin-3 receptor-deficient mouse. *Endocrinology* **141**, 3518–3521 (2000).
75. Chen, A. S. *et al.* Inactivation of the mouse melanocortin-3 receptor results in increased fat mass and reduced lean body mass. *Nat. Genet.* **26**, 97–102 (2000).
76. Balthasar, N. *et al.* Divergence of melanocortin pathways in the control of food intake and energy expenditure. *Cell* **123**, 493–505 (2005).
77. Rossi, J. *et al.* Melanocortin-4 receptors expressed by cholinergic neurons regulate energy balance and glucose homeostasis. *Cell Metab.* **13**, 195–204 (2011).
78. Seeley, R. J. *et al.* Behavioral, endocrine, and hypothalamic responses to involuntary overfeeding. *Am. J. Physiol.* **271**, R819–823 (1996).
79. Schwartz, M. W. *et al.* Leptin increases hypothalamic pro-opiomelanocortin mRNA expression in the rostral arcuate nucleus. *Diabetes* **46**, 2119–2123 (1997).
80. Hagan, M. M. *et al.* Role of the CNS melanocortin system in the response to overfeeding. *J. Neurosci.* **19**, 2362–2367 (1999).
81. Douglass, J., McKinzie, A. A. & Couceyro, P. PCR differential display identifies a rat brain mRNA that is transcriptionally regulated by cocaine and amphetamine. *J. Neurosci.* **15**, 2471–2481 (1995).
82. Elias, C. F. *et al.* Leptin Activates Hypothalamic CART Neurons Projecting to the Spinal Cord. *Neuron* **21**, 1375–1385 (1998).
83. Kuhar, M. J. & Dall Vechia, S. E. CART peptides: novel addiction- and feeding-related neuropeptides. *Trends in Neurosciences* **22**, 316–320 (1999).
84. Lambert, P. D. *et al.* CART peptides in the central control of feeding and interactions with neuropeptide Y. *Synapse* **29**, 293–298 (1998).

85. Asnicar, M. A. *et al.* Absence of Cocaine- and Amphetamine-Regulated Transcript Results in Obesity in Mice Fed a High Caloric Diet. *Endocrinology* **142**, 4394–4400 (2001).
86. Kristensen, P. *et al.* Hypothalamic CART is a new anorectic peptide regulated by leptin. *Nature* **393**, 72–76 (1998).
87. Cheung, C. C., Clifton, D. K. & Steiner, R. A. Proopiomelanocortin Neurons Are Direct Targets for Leptin in the Hypothalamus. *Endocrinology* **138**, 4489–4492 (1997).
88. Bagdade, J. D., Bierman, E. L. & Porte, D. The Significance of Basal Insulin Levels in the Evaluation of the Insulin Response to Glucose in Diabetic and Nondiabetic Subjects*. *J Clin Invest* **46**, 1549–1557 (1967).
89. Stephan, F., Reville, P., Thierry, R. & Schlienger, J. Correlations between plasma insulin and body weight in obesity, anorexia nervosa and diabetes mellitus. *Diabetologia* **8**, 196–201 (1972).
90. Polonsky, K. S. *et al.* Quantitative study of insulin secretion and clearance in normal and obese subjects. *J Clin Invest* **81**, 435–441 (1988).
91. MARGOLIS, R. U. & ALTSZULER, N. Insulin in the Cerebrospinal Fluid. *Nature* **215**, 1375–1376 (1967).
92. Schwartz, M. W. *et al.* Kinetics and specificity of insulin uptake from plasma into cerebrospinal fluid. *American Journal of Physiology - Endocrinology And Metabolism* **259**, E378–83 (1990).
93. Woods, S. C., Lotter, E. C., McKay, L. D. & Porte, D. Chronic intracerebroventricular infusion of insulin reduces food intake and body weight of baboons. *Nature* **282**, 503–505 (1979).
94. Brief, D. J. & Davis, J. D. Reduction of food intake and body weight by chronic intraventricular insulin infusion. *Brain Res Bull* **12**, 571–575 (1984).
95. Brown, L. M., Clegg, D. J., Benoit, S. C. & Woods, S. C. Intraventricular insulin and leptin reduce food intake and body weight in C57BL/6J mice. *Physiology & Behavior* **89**, 687–691 (2006).
96. Hallschmid, M. *et al.* Intranasal Insulin Reduces Body Fat in Men but not in Women. *Diabetes* **53**, 3024–3029 (2004).
97. Strubbe, J. & Mein, C. Increased feeding in response to bilateral injection of insulin antibodies in the VMH. *Physiology & Behavior* **19**, 309–313 (1977).
98. McGowan, M. K., Andrews, K. M. & Grossman, S. P. Chronic intrahypothalamic infusions of insulin or insulin antibodies alter body weight and food intake in the rat. *Physiology & Behavior* **51**, 753–766 (1992).
99. Brüning, J. C. *et al.* Role of Brain Insulin Receptor in Control of Body Weight and Reproduction. *Science (80-)* **289**, 2122–2125 (2000).
100. Obici, S., Feng, Z., Karkanias, G., Baskin, D. G. & Rossetti, L. Decreasing hypothalamic insulin receptors causes hyperphagia and insulin resistance in rats. *Nat Neurosci* **5**, 566–572 (2002).
101. Van Houten, M., Posner, B. I., Kopriwa, B. M. & Brawer, J. R. Insulin binding sites localized to nerve terminals in rat median eminence and arcuate nucleus. *Science* **207**, 1081–1083 (1980).
102. Considine, R. V. *et al.* Serum immunoreactive-leptin concentrations in normal-weight and obese humans. *N Engl J Med* **334**, 292–5 (1996).
103. Rosenbaum, M. *et al.* Effects of gender, body composition, and menopause on plasma concentrations of leptin. *Journal of Clinical Endocrinology & Metabolism* **81**, 3424–7 (1996).
104. Halaas, J. L. *et al.* Weight-reducing effects of the plasma protein encoded by the obese gene. *Science (80-)* **269**, 543–6 (1995).

105. Licinio, J. *et al.* Association of a corticotropin-releasing hormone receptor 1 haplotype and antidepressant treatment response in Mexican-Americans. *Mol Psychiatry* **9**, 1075–82 (2004).
106. Farooqi, I. S. *et al.* Effects of recombinant leptin therapy in a child with congenital leptin deficiency. *N Engl J Med* **341**, 879–84 (1999).
107. Heymsfield, S. B. *et al.* Recombinant Leptin for Weight Loss in Obese and Lean Adults A Randomized, Controlled, Dose-Escalation Trial. *JAMA* **282**, 1568–1575 (1999).
108. Cohen, P. *et al.* Selective deletion of leptin receptor in neurons leads to obesity. *J Clin Invest* **108**, 1113–1121 (2001).
109. Morrison, C. D., Morton, G. J., Niswender, K. D., Gelling, R. W. & Schwartz, M. W. Leptin inhibits hypothalamic Npy and Agrp gene expression via a mechanism that requires phosphatidylinositol 3-OH-kinase signaling. *Am. J. Physiol. Endocrinol. Metab.* **289**, E1051–1057 (2005).
110. Baskin, D. G., Breininger, J. F. & Schwartz, M. W. Leptin receptor mRNA identifies a subpopulation of neuropeptide Y neurons activated by fasting in rat hypothalamus. *Diabetes* **48**, 828–833 (1999).
111. Balthasar, N. *et al.* Leptin Receptor Signaling in POMC Neurons Is Required for Normal Body Weight Homeostasis. (2004).at <http://linkinghub.elsevier.com/retrieve/pii/S089662730400354X>
112. Van de Wall, E. *et al.* Collective and Individual Functions of Leptin Receptor Modulated Neurons Controlling Metabolism and Ingestion. *Endocrinology* **149**, 1773–1785 (2008).
113. Enoki, S. *et al.* Plasma islet amyloid polypeptide levels in obesity, impaired glucose tolerance and non-insulin-dependent diabetes mellitus. *Diabetes Res. Clin. Pract.* **15**, 97–102 (1992).
114. Rushing, P. A., Hagan, M. M., Seeley, R. J., Lutz, T. A. & Woods, S. C. Amylin: A Novel Action in the Brain to Reduce Body Weight*. *Endocrinology* **141**, 850 (2000).
115. Lutz, T. A., Geary, N., Szabady, M. M., Del Prete, E. & Scharrer, E. Amylin decreases meal size in rats. *Physiology & Behavior* **58**, 1197–1202 (1995).
116. Mollet, A., Gilg, S., Riediger, T. & Lutz, T. A. Infusion of the amylin antagonist AC 187 into the area postrema increases food intake in rats. *Physiology & Behavior* **81**, 149–155 (2004).
117. Kubota, N. *et al.* Adiponectin stimulates AMP-activated protein kinase in the hypothalamus and increases food intake. *Cell Metab.* **6**, 55–68 (2007).
118. Guillod-Maximin, E. *et al.* Adiponectin receptors are expressed in hypothalamus and colocalized with proopiomelanocortin and neuropeptide Y in rodent arcuate neurons. *J. Endocrinol.* **200**, 93–105 (2009).
119. Powley, T. L. & Phillips, R. J. Gastric satiation is volumetric, intestinal satiation is nutritive. *Physiol. Behav.* **82**, 69–74 (2004).
120. Buffa, R., Solcia, E. & Go, V. L. Immunohistochemical identification of the cholecystokinin cell in the intestinal mucosa. *Gastroenterology* **70**, 528–532 (1976).
121. Gibbs, J., Young, R. C. & Smith, G. P. Cholecystokinin decreases food intake in rats. *Journal of Comparative and Physiological Psychology* **84**, 488–495 (1973).
122. Hopman, W. P., Jansen, J. B. & Lamers, C. B. Comparative study of the effects of equal amounts of fat, protein, and starch on plasma cholecystokinin in man. *Scand. J. Gastroenterol.* **20**, 843–847 (1985).
123. Kissileff, H. R., Carretta, J. C., Geliebter, A. & Pi-Sunyer, F. X. Cholecystokinin and Stomach Distension Combine to Reduce Food Intake in Humans. *Am J Physiol Regul Integr Comp Physiol* **285**, R992–R998 (2003).
124. Pi-Sunyer, X., Kissileff, H. R., Thornton, J. & Smith, G. P. C-terminal octapeptide of cholecystokinin decreases food intake in obese men. *Physiol. Behav.* **29**, 627–630 (1982).

125. Lieverse, R. J., Jansen, J. B., Masclee, A. A. & Lamers, C. B. Satiety effects of a physiological dose of cholecystokinin in humans. *Gut* **36**, 176–179 (1995).
126. Adrian, T. E. *et al.* Human distribution and release of a putative new gut hormone, peptide YY. *Gastroenterology* **89**, 1070–7 (1985).
127. Batterham, R. L. *et al.* Gut hormone PYY(3-36) physiologically inhibits food intake. *Nature* **418**, 650–4 (2002).
128. Challis, B. . *et al.* Acute effects of PYY3–36 on food intake and hypothalamic neuropeptide expression in the mouse. *Biochemical and Biophysical Research Communications* **311**, 915–919 (2003).
129. Batterham, R. L. *et al.* Inhibition of food intake in obese subjects by peptide YY3-36. *N Engl J Med* **349**, 941–8 (2003).
130. Nastech Reports Positive PYY Nasal Spray Clinical Trial Data And Will Advance Obesity Clinical Program In 2007. *Medical News Today* at <http://www.medicalnewstoday.com/releases/59640.php>
131. Batterham, R. L. *et al.* Critical role for peptide YY in protein-mediated satiation and body-weight regulation. *Cell Metabolism* **4**, 223–233 (2006).
132. Wren, A. M. *et al.* Ghrelin enhances appetite and increases food intake in humans. *J. Clin. Endocrinol. Metab.* **86**, 5992 (2001).
133. Cummings, D. E., Frayo, R. S., Marmonier, C., Aubert, R. & Chapelot, D. Plasma ghrelin levels and hunger scores in humans initiating meals voluntarily without time- and food-related cues. *Am. J. Physiol. Endocrinol. Metab.* **287**, E297–304 (2004).
134. Cummings, D. E. *et al.* A preprandial rise in plasma ghrelin levels suggests a role in meal initiation in humans. *Diabetes* **50**, 1714–1719 (2001).
135. Monteleone, P., Bencivenga, R., Longobardi, N., Serritella, C. & Maj, M. Differential Responses of Circulating Ghrelin to High-Fat or High-Carbohydrate Meal in Healthy Women. *JCEM* **88**, 5510–5514 (2003).
136. Wortley, K. E. *et al.* Absence of ghrelin protects against early-onset obesity. *J. Clin. Invest.* **115**, 3573–3578 (2005).
137. Zigman, J. M. *et al.* Mice lacking ghrelin receptors resist the development of diet-induced obesity. *J. Clin. Invest.* **115**, 3564–3572 (2005).
138. Tschöp, M. *et al.* Circulating ghrelin levels are decreased in human obesity. *Diabetes* **50**, 707–709 (2001).
139. Hansen, T. K. *et al.* Weight loss increases circulating levels of ghrelin in human obesity. *Clin. Endocrinol. (Oxf)* **56**, 203–206 (2002).
140. Kamegai, J. *et al.* Central effect of ghrelin, an endogenous growth hormone secretagogue, on hypothalamic peptide gene expression. *Endocrinology* **141**, 4797–4800 (2000).
141. Eberlein, G. A. *et al.* Patterns of Prohormone Processing. Order Revealed by a New Procholecystokinin-Derived Peptide. *J. Biol. Chem.* **267**, 1517–1521 (1992).
142. Schwartz, G. J., McHugh, P. R. & Moran, T. H. Gastric loads and cholecystokinin synergistically stimulate rat gastric vagal afferents. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology* **265**, R872–R876 (1993).
143. Kissileff, H. R., Pi-Sunyer, F. X., Thornton, J. & Smith, G. P. C-Terminal Octapeptide of Cholecystokinin Decreases Food Intake in Man. *Am J Clin Nutr* **34**, 154–160 (1981).
144. Kojima, M. *et al.* Ghrelin is a growth-hormone-releasing acylated peptide from stomach. *Nature* **402**, 656–660 (1999).
145. Eissele, R. *et al.* Glucagon-like peptide-1 cells in the gastrointestinal tract and pancreas of rat, pig and man. *European Journal of Clinical Investigation* **22**, 283–291 (1992).
146. Verdich, C. *et al.* A meta-analysis of the effect of glucagon-like peptide-1 (7-36) amide on ad libitum energy intake in humans. *J. Clin. Endocrinol. Metab.* **86**, 4382–4389 (2001).
147. Kreymann, B., Ghatei, M. A., Williams, G. & Bloom, S. R. GLUCAGON-LIKE PEPTIDE-1 7-36: A PHYSIOLOGICAL INCRETIN IN MAN. *The Lancet* **330**, 1300–1304 (1987).

148. Zhao, C.-M., Furnes, M. W., Stenström, B., Kulseng, B. & Chen, D. Characterization of obestatin- and ghrelin-producing cells in the gastrointestinal tract and pancreas of rats: an immunohistochemical and electron-microscopic study. *Cell and Tissue Research* **331**, 575–587 (2007).
149. Zhang, J. V. *et al.* Obestatin, a peptide encoded by the ghrelin gene, opposes ghrelin's effects on food intake. *Science* **310**, 996–999 (2005).
150. Bataille, D. *et al.* Isolation of glucagon-37 (bioactive enteroglucagon/oxyntomodulin) from porcine jejuno-ileum: Characterization of the peptide. *FEBS Letters* **146**, 79–86 (1982).
151. Wynne, K. *et al.* Oxyntomodulin increases energy expenditure in addition to decreasing energy intake in overweight and obese humans: a randomised controlled trial. *Int J Obes (Lond)* **30**, 1729–1736 (2006).
152. Ekblad, E. & Sundler, F. Distribution of pancreatic polypeptide and peptide YY. *Peptides* **23**, 251–261 (2002).
153. Batterham, R. L. *et al.* Pancreatic Polypeptide Reduces Appetite and Food Intake in Humans. *JCEM* **88**, 3989–3992 (2003).
154. Bartz, R. *et al.* Lipidomics reveals that adiposomes store ether lipids and mediate phospholipid traffic. *J. Lipid Res.* **48**, 837–847 (2007).
155. Bartz, R. *et al.* Dynamic activity of lipid droplets: protein phosphorylation and GTP-mediated protein translocation. *J. Proteome Res.* **6**, 3256–3265 (2007).
156. Greenberg, A. S. *et al.* Perilipin, a major hormonally regulated adipocyte-specific phosphoprotein associated with the periphery of lipid storage droplets. *J. Biol. Chem.* **266**, 11341–11346 (1991).
157. Blanchette-Mackie, E. J. *et al.* Perilipin is located on the surface layer of intracellular lipid droplets in adipocytes. *J. Lipid Res.* **36**, 1211–1226 (1995).
158. Brasaemle, D. L. *et al.* Perilipin A increases triacylglycerol storage by decreasing the rate of triacylglycerol hydrolysis. *J. Biol. Chem.* **275**, 38486–38493 (2000).
159. Miyoshi, H. *et al.* Perilipin promotes hormone-sensitive lipase-mediated adipocyte lipolysis via phosphorylation-dependent and -independent mechanisms. *J. Biol. Chem.* **281**, 15837–15844 (2006).
160. Tansey, J. T. *et al.* Perilipin ablation results in a lean mouse with aberrant adipocyte lipolysis, enhanced leptin production, and resistance to diet-induced obesity. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6494–6499 (2001).
161. Villena, J. A., Roy, S., Sarkadi-Nagy, E., Kim, K.-H. & Sul, H. S. Desnutrin, an adipocyte gene encoding a novel patatin domain-containing protein, is induced by fasting and glucocorticoids: ectopic expression of desnutrin increases triglyceride hydrolysis. *J. Biol. Chem.* **279**, 47066–47075 (2004).
162. Ahmadian, M. *et al.* Adipose overexpression of desnutrin promotes fatty acid use and attenuates diet-induced obesity. *Diabetes* **58**, 855–866 (2009).
163. Duncan, R. E., Sarkadi-Nagy, E., Jaworski, K., Ahmadian, M. & Sul, H. S. Identification and functional characterization of adipose-specific phospholipase A2 (AdPLA). *J. Biol. Chem.* **283**, 25428–25436 (2008).
164. Jaworski, K. *et al.* AdPLA ablation increases lipolysis and prevents obesity induced by high-fat feeding or leptin deficiency. *Nat. Med.* **15**, 159–168 (2009).
165. Kopecky, J., Clarke, G., Enerbäck, S., Spiegelman, B. & Kozak, L. P. Expression of the mitochondrial uncoupling protein gene from the aP2 gene promoter prevents genetic obesity. *J. Clin. Invest.* **96**, 2914–2923 (1995).
166. Zimmermann, R. *et al.* Fat mobilization in adipose tissue is promoted by adipose triglyceride lipase. *Science* **306**, 1383–1386 (2004).
167. French, S. A., Story, M. & Jeffery, R. W. Environmental influences on eating and physical activity. *Annu Rev Public Health* **22**, 309–35 (2001).

168. Pontzer, H. *et al.* Hunter-Gatherer Energetics and Human Obesity. *PLoS ONE* **7**, e40503 (2012).
169. Andersson, J. C. & Walley, A. J. *Obesity Before Birth*. (2011).
170. Stunkard, A. J. *et al.* An adoption study of human obesity. *N. Engl. J. Med.* **314**, 193–198 (1986).
171. Tambs, K. *et al.* Genetic and environmental contributions to the variance of the body mass index in a Norwegian sample of first- and second-degree relatives. *American Journal of Human Biology* **3**, 257–267 (2005).
172. Visscher, P. M. Power of the classical twin design revisited. *Twin Res* **7**, 505–512 (2004).
173. Luke, B., Leurgans, S., Keith, L. & Keith, D. The childhood growth of twin children. *Acta Genet Med Gemellol (Roma)* **44**, 169–178 (1995).
174. Maes, H. H. M., Neale, M. C. & Eaves, L. J. Genetic and Environmental Factors in Relative Body Weight and Human Adiposity. *Behavior Genetics* **27**, 325–351 (1997).
175. Corney, G. *et al.* The effect of zygosity on the birth weight of twins in Aberdeen and northeast Scotland. *Acta Genet Med Gemellol (Roma)* **28**, 353–360 (1979).
176. Feinleib, M. *et al.* The Nhlbi Twin Study of Cardiovascular Disease Risk Factors: Methodology and Summary of Results. *Am. J. Epidemiol.* **106**, 284–295 (1977).
177. Fabsitz, R., Garrison, R., Feinleib, M. & Hjortland, M. A twin analysis of dietary intake: Evidence for a need to control for possible environmental differences in MZ and DZ twins. *Behavior Genetics* **8**, 15–25 (1978).
178. Stunkard, A. J., Foch, T. T. & Hrubec, Z. A Twin Study of Human Obesity. *JAMA* **256**, 51–54 (1986).
179. Bodurtha, J. N. *et al.* Genetic analysis of anthropometric measures in 11-year-old twins: the Medical College of Virginia Twin Study. *Pediatr. Res.* **28**, 1–4 (1990).
180. Brook, C. G., Huntley, R. M. & Slack, J. Influence of heredity and environment in determination of skinfold thickness in children. *Br Med J* **2**, 719–721 (1975).
181. Stunkard, A. J., Harris, J. R., Pedersen, N. L. & McClearn, G. E. THE BODY-MASS INDEX OF TWINS WHO HAVE BEEN REARED APART. *New England Journal of Medicine* **322**, 1483–1487 (1990).
182. Montague, C. T. *et al.* Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature* **387**, 903–8 (1997).
183. Strobel, A., Issad, T., Camoin, L., Ozata, M. & Strosberg, A. D. A leptin missense mutation associated with hypogonadism and morbid obesity. *Nat Genet* **18**, 213–5 (1998).
184. Licinio, J. *et al.* Phenotypic effects of leptin replacement on morbid obesity, diabetes mellitus, hypogonadism, and behavior in leptin-deficient adults. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4531–4536 (2004).
185. Clement, K. *et al.* A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* **392**, 398–401 (1998).
186. Krude, H. *et al.* Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nat Genet* **19**, 155–7 (1998).
187. Krude, H. & Grüters, A. Implications of proopiomelanocortin (POMC) mutations in humans: the POMC deficiency syndrome. *Trends Endocrinol Metab* **11**, 15–22 (2000).
188. Krude, H. *et al.* Obesity due to proopiomelanocortin deficiency: three new cases and treatment trials with thyroid hormone and ACTH4-10. *J Clin Endocrinol Metab* **88**, 4633–40 (2003).
189. Farooqi, I. S. *et al.* Heterozygosity for a POMC-null mutation and increased obesity risk in humans. *Diabetes* **55**, 2549–2553 (2006).
190. Vaisse, C., Clement, K., Guy-Grand, B. & Froguel, P. A frameshift mutation in human MC4R is associated with a dominant form of obesity. *Nat Genet* **20**, 113–114 (1998).
191. Yeo, G. S. *et al.* A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nat. Genet.* **20**, 111–112 (1998).

192. Farooqi, I. S. *et al.* Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N Engl J Med* **348**, 1085–95 (2003).
193. Lubrano-Berthelier, C. *et al.* Intracellular retention is a common characteristic of childhood obesity-associated MC4R mutations. *Hum. Mol. Genet.* **12**, 145–153 (2003).
194. Larsen, L. H. *et al.* Prevalence of Mutations and Functional Analyses of Melanocortin 4 Receptor Variants Identified Among 750 Men with Juvenile-Onset Obesity. *JCEM* **90**, 219–224 (2005).
195. Friedel, S. *et al.* Mutation screen of the brain derived neurotrophic factor gene (BDNF): identification of several genetic variants and association studies in patients with obesity, eating disorders, and attention-deficit/hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* **132B**, 96–9 (2005).
196. Xu, B. *et al.* Brain-derived neurotrophic factor regulates energy balance downstream of melanocortin-4 receptor. *Nat Neurosci* **6**, 736–42 (2003).
197. Yeo, G. S. H. *et al.* A de novo mutation affecting human TrkB associated with severe obesity and developmental delay. *Nat Neurosci* **7**, 1187–9 (2004).
198. Jackson, R. S. *et al.* Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. *Nat Genet* **16**, 303–6 (1997).
199. Jansen, E., Ayoubi, T. A., Meulemans, S. M. & Van de Ven, W. J. Neuroendocrine-specific expression of the human prohormone convertase 1 gene. Hormonal regulation of transcription through distinct cAMP response elements. *J. Biol. Chem.* **270**, 15391–15397 (1995).
200. Ozata, M., Ozdemir, 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. *J. Clin. Endocrinol. Metab.* **84**, 3686–3695 (1999).
201. Farooqi, I. S. & O'Rahilly, S. Monogenic obesity in humans. *Annu Rev Med* **56**, 443–58 (2005).
202. Whittington, J. *et al.* Population prevalence and estimated birth incidence and mortality rate for people with Prader-Willi syndrome in one UK Health Region. *J Med Genet* **38**, 792–798 (2001).
203. Holm, V. A. *et al.* Prader-Willi syndrome: consensus diagnostic criteria. *Pediatrics* **91**, 398–402 (1993).
204. Ledbetter, D. H. *et al.* Deletions of chromosome 15 as a cause of the Prader-Willi syndrome. *N. Engl. J. Med.* **304**, 325–329 (1981).
205. Nicholls, R. D., Knoll, J. H. M., Butler, M. G., Karam, S. & Lalande, M. Genetic imprinting suggested by maternal heterodisomy in non-deletion Prader-Willi syndrome. *, Published online: 16 November 1989; | doi:10.1038/342281a0* **342**, 281–285 (1989).
206. Glenn, C. C. *et al.* Modification of 15q11-q13 DNA methylation imprints in unique Angelman and Prader-Willi patients. *Hum. Mol. Genet.* **2**, 1377–1382 (1993).
207. De Smith, A. J. *et al.* A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Hum Mol Genet* **18**, 3257–65 (2009).
208. Beales, P. L., Elcioglu, N., Woolf, A. S., Parker, D. & Flinter, F. A. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. *J. Med. Genet.* **36**, 437–446 (1999).
209. Mykytyn, K. *et al.* Evaluation of complex inheritance involving the most common Bardet-Biedl syndrome locus (BBS1). *Am. J. Hum. Genet.* **72**, 429–437 (2003).
210. Stoetzel, C. *et al.* BBS10 encodes a vertebrate-specific chaperonin-like protein and is a major BBS locus. *Nat. Genet.* **38**, 521–524 (2006).

211. Beales, P. L. *et al.* Genetic and mutational analyses of a large multiethnic Bardet-Biedl cohort reveal a minor involvement of BBS6 and delineate the critical intervals of other loci. *Am. J. Hum. Genet.* **68**, 606–616 (2001).
212. Katsanis, N. *et al.* Triallelic Inheritance in Bardet-Biedl Syndrome, a Mendelian Recessive Disorder. *Science* **293**, 2256–2259 (2001).
213. Michaud, J. L. *et al.* Sim1 haploinsufficiency causes hyperphagia, obesity and reduction of the paraventricular nucleus of the hypothalamus. *Hum Mol Genet* **10**, 1465–73 (2001).
214. Holder, J. L., Jr *et al.* Sim1 gene dosage modulates the homeostatic feeding response to increased dietary fat in mice. *Am. J. Physiol. Endocrinol. Metab.* **287**, E105–113 (2004).
215. Kublaoui, B. M., Holder, J. L., Jr, Tolson, K. P., Gemelli, T. & Zinn, A. R. SIM1 overexpression partially rescues agouti yellow and diet-induced obesity by normalizing food intake. *Endocrinology* **147**, 4542–4549 (2006).
216. Kolehmainen, J. *et al.* Cohen syndrome is caused by mutations in a novel gene, COH1, encoding a transmembrane protein with a presumed role in vesicle-mediated sorting and intracellular protein transport. *Am J Hum Genet* **72**, 1359–69 (2003).
217. Collin, G. B. *et al.* Mutations in ALMS1 cause obesity, type 2 diabetes and neurosensory degeneration in Alström syndrome. *Nat Genet* **31**, 74–8 (2002).
218. Comuzzie, A. G. *et al.* A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. *Nature Genetics* **15**, 273–276 (1997).
219. Hager, J. *et al.* A genome-wide scan for human obesity genes reveals a major susceptibility locus on chromosome 10. *Nature Genetics* **20**, 304–308 (1998).
220. Palmer, L. J. *et al.* A Whole-Genome Scan for Obstructive Sleep Apnea and Obesity. *Am J Hum Genet* **72**, 340–350 (2003).
221. Mitchell, B. D. *et al.* A quantitative trait locus influencing BMI maps to the region of the beta-3 adrenergic receptor. *Diabetes* **48**, 1863–1867 (1999).
222. Hinney, A. *et al.* Independent confirmation of a major locus for obesity on chromosome 10. *J. Clin. Endocrinol. Metab.* **85**, 2962–2965 (2000).
223. Price, R. A. *et al.* A locus affecting obesity in human chromosome region 10p12. *Diabetologia* **44**, 363–366 (2001).
224. Rankinen, T. *et al.* The human obesity gene map: the 2005 update. *Obesity (Silver Spring)* **14**, 529–644 (2006).
225. Stone, S. *et al.* A major predisposition locus for severe obesity, at 4p15-p14. *Am. J. Hum. Genet.* **70**, 1459–1468 (2002).
226. Saunders, C. L. *et al.* Meta-analysis of genome-wide linkage studies in BMI and obesity. *Obesity (Silver Spring)* **15**, 2263–2275 (2007).
227. The International HapMap Project. *Nature* **426**, 789–96 (2003).
228. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
229. Frazer, K. A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
230. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517 (1996).
231. Redden, D. T. & Allison, D. B. Nonreplication in genetic association studies of obesity and diabetes research. *J. Nutr.* **133**, 3323–3326 (2003).
232. Cardon, L. R. & Palmer, L. J. Population stratification and spurious allelic association. *Lancet* **361**, 598–604 (2003).
233. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
234. Ioannidis, J. P. A. Why most published research findings are false. *PLoS Med.* **2**, e124 (2005).
235. Paracchini, V., Pedotti, P. & Taioli, E. Genetics of leptin and obesity: a HuGE review. *Am. J. Epidemiol.* **162**, 101–114 (2005).

236. Yu, Z. *et al.* Genetic Polymorphisms in Adipokine Genes and the Risk of Obesity: A Systematic Review and Meta-Analysis. *Obesity* **20**, 396–406 (2011).
237. Bender, N. *et al.* Association between variants of the leptin receptor gene (LEPR) and overweight: a systematic review and an analysis of the CoLaus study. *PLoS ONE* **6**, e26157 (2011).
238. Wang, R. *et al.* ENPP1/PC-1 gene K121Q polymorphism is associated with obesity in European adult populations: evidence from a meta-analysis involving 24,324 subjects. *Biomed. Environ. Sci.* **24**, 200–206 (2011).
239. Lyon, H. N. *et al.* Common variants in the ENPP1 gene are not reproducibly associated with diabetes or obesity. *Diabetes* **55**, 3180–3184 (2006).
240. Ferrer-Lorente, R., Cabot, C., Fernández-López, J.-A. & Alemany, M. Combined effects of oleoyl-estrone and a beta3-adrenergic agonist (CL316,243) on lipid stores of diet-induced overweight male Wistar rats. *Life Sci.* **77**, 2051–2058 (2005).
241. Kurokawa, N. *et al.* The ADRB3 Trp64Arg variant and BMI: a meta-analysis of 44 833 individuals. *Int J Obes (Lond)* **32**, 1240–1249 (2008).
242. Young, E. H. *et al.* The V103I polymorphism of the MC4R gene and obesity: population based studies and meta-analysis of 29 563 individuals. *Int J Obes (Lond)* **31**, 1437–1441 (2007).
243. Sookoian, S. C., Gonz|[aacute]|lez, C. & Pirola, C. J. Meta-analysis on the G-308A Tumor Necrosis Factor |[alpha]| Gene Variant and Phenotypes Associated with the Metabolic Syndrome. *Obesity* **13**, 2122–2131 (2005).
244. Kennedy, G. C. *et al.* Large-scale genotyping of complex DNA. *Nat Biotechnol* **21**, 1233–7 (2003).
245. Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G. & Chee, M. S. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* **37**, 549–54 (2005).
246. 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 (80-)* **316**, 889–94 (2007).
247. Dina, C. *et al.* Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet* **39**, 724–726 (2007).
248. Gerken, T. *et al.* The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318**, 1469–1472 (2007).
249. Chambers, J. C. *et al.* Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat Genet* **40**, 716–8 (2008).
250. Loos, R. J. F. *et al.* Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* **40**, 768–775 (2008).
251. Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genetics* **41**, 25–34 (2008).
252. Thorleifsson, G. *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* **41**, 18–24 (2009).
253. 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* **41**, 157–159 (2009).
254. O'Rahilly, S. & Farooqi, I. S. Human Obesity: A Heritable Neurobehavioral Disorder That Is Highly Sensitive to Environmental Conditions. *Diabetes* **57**, 2905–2910 (2008).
255. Kilpeläinen, T. O. *et al.* Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nat Genet* **43**, 753–760 (2011).
256. Bradfield, J. P. *et al.* A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat. Genet.* **44**, 526–531 (2012).
257. Scuteri, A. *et al.* Genome-Wide Association Scan Shows Genetic Variants in the FTO Gene Are Associated with Obesity-Related Traits. *PLoS Genet* **3**, e115 (2007).

258. Liu, Y.-J. *et al.* Genome-wide association scans identified CTNNBL1 as a novel gene for obesity. *Hum Mol Genet* **17**, 1803–13 (2008).
259. Cho, Y. S. *et al.* A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nature Genetics* **41**, 527–534 (2009).
260. Lindgren, C. M. *et al.* Genome-Wide Association Scan Meta-Analysis Identifies Three Loci Influencing Adiposity and Fat Distribution. *PLoS Genet* **5**, e1000508 (2009).
261. Scherag, A. *et al.* Two New Loci for Body-Weight Regulation Identified in a Joint Analysis of Genome-Wide Association Studies for Early-Onset Extreme Obesity in French and German Study Groups. *PLoS Genet* **6**, e1000916 (2010).
262. Heid, I. M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42**, 949–960 (2010).
263. Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genetics* **41**, 25–34 (2008).
264. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–53 (2009).
265. A map of human genome variation from population scale sequencing. *Nature* **467**, 1061–1073 (2010).
266. Iyengar, S. K. & Elston, R. C. The genetic basis of complex traits: rare variants or 'common gene, common disease'? *Methods Mol Biol* **376**, 71–84 (2007).
267. Conrad, D. F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2009).
268. Buchanan, J. A. & Scherer, S. W. Contemplating effects of genomic structural variation. *Genet Med* **10**, 639–47 (2008).
269. Walters, R. G. *et al.* A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* **463**, 671–675 (2010).
270. Jacquemont, S. *et al.* Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* **478**, 97–102 (2011).
271. Bochukova, E. G. *et al.* Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* **463**, 666–670 (2010).
272. Glessner, J. T. *et al.* A genome-wide study reveals copy number variants exclusive to childhood obesity cases. *Am J Hum Genet* **87**, 661–6 (2010).
273. Sha, B.-Y. *et al.* Genome-wide association study suggested copy number variation may be associated with body mass index in the Chinese population. *J Hum Genet* **54**, 199–202 (2009).
274. Russo, P., Lauria, F. & Siani, A. Heritability of body weight: Moving beyond genetics. *Nutrition, Metabolism and Cardiovascular Diseases* **20**, 691–697 (2010).
275. Bird, A. Perceptions of epigenetics. *Nature* **447**, 396–398 (2007).
276. Chong, S. & Whitelaw, E. Epigenetic germline inheritance. *Current Opinion in Genetics & Development* **14**, 692–696 (2004).
277. Reik, W. & Walter, J. Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.* **2**, 21–32 (2001).
278. Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* **23**, 314–8 (1999).
279. Waterland, R. A., Travisano, M., Tahiliani, K. G., Rached, M. T. & Mirza, S. Methyl donor supplementation prevents transgenerational amplification of obesity. *Int J Obes (Lond)* **32**, 1373–1379 (2008).
280. Waterland, R. A. & Jirtle, R. L. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol. Cell. Biol.* **23**, 5293–5300 (2003).

281. Waterland, R. A., Lin, J.-R., Smith, C. A. & Jirtle, R. L. Post-weaning diet affects genomic imprinting at the insulin-like growth factor 2 (Igf2) locus. *Hum. Mol. Genet.* **15**, 705–716 (2006).
282. Weaver, I. C. G. *et al.* Epigenetic programming by maternal behavior. *Nature Neuroscience* **7**, 847–854 (2004).
283. Weaver, I. C. G. *et al.* Reversal of maternal programming of stress responses in adult offspring through methyl supplementation: altering epigenetic marking later in life. *J. Neurosci.* **25**, 11045–11054 (2005).
284. Li, S. *et al.* Neonatal diethylstilbestrol exposure induces persistent elevation of c-fos expression and hypomethylation in its exon-4 in mouse uterus. *Mol. Carcinog.* **38**, 78–84 (2003).
285. Ho, S.-M., Tang, W.-Y., Belmonte de Frausto, J. & Prins, G. S. Developmental exposure to estradiol and bisphenol A increases susceptibility to prostate carcinogenesis and epigenetically regulates phosphodiesterase type 4 variant 4. *Cancer Res.* **66**, 5624–5632 (2006).
286. Niemitz, E. L. & Feinberg, A. P. Epigenetics and assisted reproductive technology: a call for investigation. *Am. J. Hum. Genet.* **74**, 599–609 (2004).
287. Rossignol, S. *et al.* The epigenetic imprinting defect of patients with Beckwith-Wiedemann syndrome born after assisted reproductive technology is not restricted to the 11p15 region. *J. Med. Genet.* **43**, 902–907 (2006).
288. Koturbash, I. *et al.* Epigenetic dysregulation underlies radiation-induced transgenerational genome instability in vivo. *Int. J. Radiat. Oncol. Biol. Phys.* **66**, 327–330 (2006).
289. Rakyan, V. K. *et al.* Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2538–2543 (2003).
290. Anway, M. D., Cupp, A. S., Uzumcu, M. & Skinner, M. K. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* **308**, 1466–1469 (2005).
291. Plagemann, A. *et al.* Hypothalamic POMC promoter methylation becomes altered by early overfeeding: An epigenetic model of obesity and the metabolic syndrome. *J Physiol* (2009).at <http://jp.physoc.org/content/early/2009/08/28/jphysiol.2009.176156.abstract>
292. Dodge, J. E., Ramsahoye, B. H., Wo, Z. G., Okano, M. & Li, E. De novo methylation of MMLV provirus in embryonic stem cells: CpG versus non-CpG methylation. *Gene* **289**, 41–8 (2002).
293. Jones, P. A. & Liang, G. Rethinking how DNA methylation patterns are maintained. *Nat Rev Genet* **10**, 805–11 (2009).
294. Monk, M., Boubelik, M. & Lehnert, S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**, 371–382 (1987).
295. Kafri, T. *et al.* Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev.* **6**, 705–714 (1992).
296. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257 (1999).
297. Frank, D. *et al.* Demethylation of CpG islands in embryonic cells. *Nature* **351**, 239–241 (1991).
298. Ooi, S. K. T. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714–717 (2007).
299. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457–466 (2007).
300. Mohn, F. *et al.* Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30**, 755–766 (2008).

301. Gidekel, S. & Bergman, Y. A unique developmental pattern of Oct-3/4 DNA methylation is controlled by a cis-demodification element. *J. Biol. Chem.* **277**, 34521–34530 (2002).
302. Ehrlich, M. *et al.* Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res* **10**, 2709–2721 (1982).
303. Scarano, E., Iaccarino, M., Grippo, P. & Parisi, E. The heterogeneity of thymine methyl group origin in DNA pyrimidine isostichs of developing sea urchin embryos. *Proc Natl Acad Sci U S A* **57**, 1394–400 (1967).
304. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
305. Larsen, F., Gundersen, G., Lopez, R. & Prydz, H. CpG islands as gene markers in the human genome. *Genomics* **13**, 1095–1107 (1992).
306. Wang, Y. & Leung, F. C. C. An evaluation of new criteria for CpG islands in the human genome as gene markers. *Bioinformatics* **20**, 1170–1177 (2004).
307. Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* **3**, 415–28 (2002).
308. Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A* **102**, 10604–9 (2005).
309. Kaminsky, Z. A. *et al.* DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet* **41**, 240–245 (2009).
310. Buiting, K. Prader-Willi syndrome and Angelman syndrome. *Am J Med Genet C Semin Med Genet* **154C**, 365–76 (2010).
311. Roseboom, T., De Rooij, S. & Painter, R. The Dutch famine and its long-term consequences for adult health. *Early Hum. Dev.* **82**, 485–491 (2006).
312. Ravelli, G. P., Stein, Z. A. & Susser, M. W. Obesity in young men after famine exposure in utero and early infancy. *N. Engl. J. Med.* **295**, 349–353 (1976).
313. Heijmans, B. T. *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17046–17049 (2008).
314. Dong, C. *et al.* Possible genomic imprinting of three human obesity-related genetic loci. *Am J Hum Genet* **76**, 427–37 (2005).
315. Campion, J., Milagro, F. I., Goyenechea, E. & Martinez, J. A. TNF-[alpha] Promoter Methylation as a Predictive Biomarker for Weight-loss Response. *Obesity* **17**, 1293–1297 (2009).
316. Laurent, L. *et al.* Dynamic changes in the human methylome during differentiation. *Genome Res.* **20**, 320–331 (2010).
317. Rauch, T. A., Wu, X., Zhong, X., Riggs, A. D. & Pfeifer, G. P. A human B cell methylome at 100−base pair resolution. *PNAS* **106**, 671–678 (2009).
318. Hellman, A. & Chess, A. Gene body-specific methylation on the active X chromosome. *Science* **315**, 1141–1143 (2007).
319. Bäckdahl, L. *et al.* Gene body methylation of the dimethylarginine dimethylamino-hydrolase 2 (Ddah2) gene is an epigenetic biomarker for neural stem cell differentiation. *Epigenetics* **4**, 248–254 (2009).
320. Jjingo, D., Conley, A. B., Yi, S. V., Lunyak, V. V. & Jordan, I. K. On the presence and role of human gene-body DNA methylation. *Oncotarget* **3**, 462–474 (2012).
321. Freyssenet, D. Energy sensing and regulation of gene expression in skeletal muscle. *J Appl Physiol* **102**, 529–540 (2007).
322. Picard, F. *et al.* Sirt1 promotes fat mobilization in white adipocytes by repressing PPAR-[gamma]. *Nature* **429**, 771–776 (2004).
323. Rodgers, J. T. *et al.* Nutrient control of glucose homeostasis through a complex of PGC-1[alpha] and SIRT1. *Nature* **434**, 113–118 (2005).
324. Banks, A. S. *et al.* SirT1 gain of function increases energy efficiency and prevents diabetes in mice. *Cell Metab* **8**, 333–41 (2008).

325. Moynihan, K. A. *et al.* Increased dosage of mammalian Sir2 in pancreatic [beta] cells enhances glucose-stimulated insulin secretion in mice. *Cell Metab* **2**, 105–117 (2005).
326. Qiao, L. & Shao, J. SIRT1 Regulates Adiponectin Gene Expression through Foxo1-C/Enhancer-binding Protein {alpha} Transcriptional Complex. *J Biol Chem* **281**, 39915–39924 (2006).
327. Qiang, L., Wang, H. & Farmer, S. R. Adiponectin Secretion Is Regulated by SIRT1 and the Endoplasmic Reticulum Oxidoreductase Ero1-L{alpha}. *Mol Cell Biol* **27**, 4698–4707 (2007).
328. Pfluger, P. T., Herranz, D., Velasco-Miguel, S., Serrano, M. & Tschöp, M. H. Sirt1 protects against high-fat diet-induced metabolic damage. *Proc Natl Acad Sci U S A* **105**, 9793–8 (2008).
329. Yeung, F. *et al.* Modulation of NF-kappaB-dependent transcription and cell survival by the SIRT1 deacetylase. *EMBO J* **23**, 2369–80 (2004).
330. Bordone, L. *et al.* SIRT1 transgenic mice show phenotypes resembling calorie restriction. *Aging Cell* **6**, 759–67 (2007).
331. Purushotham, A. *et al.* Hepatocyte-specific deletion of SIRT1 alters fatty acid metabolism and results in hepatic steatosis and inflammation. *Cell Metab* **9**, 327–38 (2009).
332. Schug, T. T. *et al.* Myeloid deletion of SIRT1 induces inflammatory signaling in response to environmental stress. *Mol Cell Biol* **30**, 4712–21 (2010).
333. Lagouge, M. *et al.* Resveratrol Improves Mitochondrial Function and Protects against Metabolic Disease by Activating SIRT1 and PGC-1[alpha]. *Cell* **127**, 1109–1122 (2006).
334. Milne, J. C. *et al.* Small molecule activators of SIRT1 as therapeutics for the treatment of type 2 diabetes. *Nature* **450**, 712–716 (2007).
335. Baur, J. A. *et al.* Resveratrol improves health and survival of mice on a high-calorie diet. *Nature* **444**, 337–342 (2006).
336. Feige, J. N. *et al.* Specific SIRT1 Activation Mimics Low Energy Levels and Protects against Diet-Induced Metabolic Disorders by Enhancing Fat Oxidation. *Cell Metab* **8**, 347–358 (2008).
337. Pedersen, S. B., Ølholm, J., Paulsen, S. K., Bennetzen, M. F. & Richelsen, B. Low Sirt1 expression, which is upregulated by fasting, in human adipose tissue from obese women. *Int J Obes (Lond)* **32**, 1250–5 (2008).
338. Boucher, J. *et al.* Apelin, a Newly Identified Adipokine Up-Regulated by Insulin and Obesity. *Endocrinology* **146**, 1764–1771 (2005).
339. Dray, C. *et al.* Apelin and APJ regulation in adipose tissue and skeletal muscle of type 2 diabetic mice and humans. *Am J Physiol Endocrinol Metab* **298**, E1161–9 (2010).
340. García-Díaz, D., Campión, J., Milagro, F. I. & Martínez, J. A. Adiposity dependent apelin gene expression: relationships with oxidative and inflammation markers. *Mol Cell Biochem* **305**, 87–94 (2007).
341. Dray, C. *et al.* Apelin Stimulates Glucose Utilization in Normal and Obese Insulin-Resistant Mice. **8**, 437–445 (2008).
342. Yue, P. *et al.* Apelin is necessary for the maintenance of insulin sensitivity. *Am J Physiol Endocrinol Metab* **298**, E59–67 (2010).
343. De Mota, N., Lenkei, Z. & Llorens-Cortès, C. Cloning, pharmacological characterization and brain distribution of the rat apelin receptor. *Neuroendocrinology* **72**, 400–7 (2000).
344. O'Carroll, A. M., Selby, T. L., Palkovits, M. & Lolait, S. J. Distribution of mRNA encoding B78/apj, the rat homologue of the human APJ receptor, and its endogenous ligand apelin in brain and peripheral tissues. *Biochim Biophys Acta* **1492**, 72–80 (2000).
345. Sunter, D., Hewson, A. K. & Dickson, S. L. Intracerebroventricular injection of apelin-13 reduces food intake in the rat. *Neurosci Lett* **353**, 1–4 (2003).
346. O'Shea, M., Hansen, M. J., Tatemoto, K. & Morris, M. J. Inhibitory effect of apelin-12 on nocturnal food intake in the rat. *Nutr Neurosci* **6**, 163–7 (2003).

347. Higuchi, K. *et al.* Apelin, an APJ receptor ligand, regulates body adiposity and favors the messenger ribonucleic acid expression of uncoupling proteins in mice. *Endocrinology* **148**, 2690–7 (2007).
348. Rousset, S. *et al.* The biology of mitochondrial uncoupling proteins. *Diabetes* **53 Suppl 1**, S130–5 (2004).
349. Dalgaard, L. T. *et al.* A prevalent polymorphism in the promoter of the UCP3 gene and its relationship to body mass index and long term body weight change in the Danish population. *J Clin Endocrinol Metab* **86**, 1398–402 (2001).
350. Argyropoulos, G. *et al.* Effects of mutations in the human uncoupling protein 3 gene on the respiratory quotient and fat oxidation in severe obesity and type 2 diabetes. *J Clin Invest* **102**, 1345–51 (1998).
351. Heinonen, M. V. *et al.* Apelin, orexin-A and leptin plasma levels in morbid obesity and effect of gastric banding. *Regul Pept* **130**, 7–13 (2005).
352. Li, L. *et al.* Changes and relations of circulating visfatin, apelin, and resistin levels in normal, impaired glucose tolerance, and type 2 diabetic subjects. *Exp Clin Endocrinol Diabetes* **114**, 544–8 (2006).
353. Ziora, K. *et al.* Assessment of serum apelin levels in girls with anorexia nervosa. *J Clin Endocrinol Metab* **95**, 2935–41 (2010).
354. Kawashima, I. *et al.* Molecular cloning of cDNA encoding adipogenesis inhibitory factor and identity with interleukin-11. *FEBS Lett* **283**, 199–202 (1991).
355. Raleigh Tenney, Karrie Stansfield & Phillip H Pekala Interleukin 11 signaling in 3T3-L1 adipocytes. *J Cell Physiol* **202**, 160–166 (2005).
356. Ohsumi, J. *et al.* Adipogenesis inhibitory factor. A novel inhibitory regulator of adipose conversion in bone marrow. *FEBS Lett* **288**, 13–6 (1991).
357. Granowitz, E. V. Transforming Growth Factor-[beta] Enhances and Pro-inflammatory Cytokines Inhibit OB Gene Expression in 3T3-L1 Adipocytes. *Biochem Biophys Res Commun* **240**, 382–385 (1997).
358. Keller, D. C., Du, X. X., Srour, E. F., Hoffman, R. & Williams, D. A. Interleukin-11 inhibits adipogenesis and stimulates myelopoiesis in human long-term marrow cultures. *Blood* **82**, 1428–35 (1993).
359. Okazaki, R. *et al.* Thiazolidinediones inhibit osteoclast-like cell formation and bone resorption in vitro. *Endocrinology* **140**, 5060–5 (1999).
360. Baulande, S., Lasnier, F., Lucas, M. & Pairault, J. Adiponutrin, a Transmembrane Protein Corresponding to a Novel Dietary- and Obesity-linked mRNA Specifically Expressed in the Adipose Lineage. *J Biol Chem* **276**, 33336–33344 (2001).
361. Lake, A. C. *et al.* Expression, regulation, and triglyceride hydrolase activity of Adiponutrin family members. *J Lipid Res* **46**, 2477–87 (2005).
362. He, S. *et al.* A sequence variation (I148M) in PNPLA3 associated with nonalcoholic fatty liver disease disrupts triglyceride hydrolysis. *J Biol Chem* **285**, 6706–15 (2010).
363. Wilson, P. A., Gardner, S. D., Lambie, N. M., Commans, S. A. & Crowther, D. J. Characterization of the human patatin-like phospholipase family. *J Lipid Res* **47**, 1940–9 (2006).
364. Kotronen, A. *et al.* A common variant in PNPLA3, which encodes adiponutrin, is associated with liver fat content in humans. *Diabetologia* **52**, 1056–60 (2009).
365. Kantartzis, K. *et al.* Dissociation between fatty liver and insulin resistance in humans carrying a variant of the patatin-like phospholipase 3 gene. *Diabetes* **58**, 2616–23 (2009).
366. Hoekstra, M. *et al.* The expression level of non-alcoholic fatty liver disease-related gene PNPLA3 in hepatocytes is highly influenced by hepatic lipid status. *J Hepatol* **52**, 244–51 (2010).
367. Liu, Y.-M. *et al.* Adiponutrin: A new gene regulated by energy balance in human adipose tissue. *J Clin Endocrinol Metab* **89**, 2684–9 (2004).

368. Moldes, M. *et al.* Adiponutrin gene is regulated by insulin and glucose in human adipose tissue. *Eur J Endocrinol* **155**, 461–468 (2006).
369. Johansson, L. E. *et al.* Variation in the Adiponutrin Gene Influences Its Expression and Associates With Obesity. *Diabetes* **55**, 826–833 (2006).
370. Brailoiu, G. C. *et al.* Nesfatin-1: distribution and interaction with a G protein-coupled receptor in the rat brain. *Endocrinology* **148**, 5088–94 (2007).
371. Fort, P. *et al.* The satiety molecule nesfatin-1 is co-expressed with melanin concentrating hormone in tuberal hypothalamic neurons of the rat. *Neuroscience* **155**, 174–81 (2008).
372. Oh-I, S. *et al.* Identification of nesfatin-1 as a satiety molecule in the hypothalamus. *Nature* **443**, 709–712 (2006).
373. Kohno, D. *et al.* Nesfatin-1 neurons in paraventricular and supraoptic nuclei of the rat hypothalamus coexpress oxytocin and vasopressin and are activated by refeeding. *Endocrinology* **149**, 1295–301 (2008).
374. Foo, K. S., Brismar, H. & Broberger, C. Distribution and neuropeptide coexistence of nucleobindin-2 mRNA/nesfatin-like immunoreactivity in the rat CNS. *Neuroscience* **156**, 563–79 (2008).
375. Inhoff, T. *et al.* Novel insight in distribution of nesfatin-1 and phospho-mTOR in the arcuate nucleus of the hypothalamus of rats. *Peptides* **31**, 257–62 (2010).
376. Price, C. J., Samson, W. K. & Ferguson, A. V. Nesfatin-1 inhibits NPY neurons in the arcuate nucleus. *Brain Res* **1230**, 99–106 (2008).
377. Maejima, Y. *et al.* Nesfatin-1-regulated oxytocinergic signaling in the paraventricular nucleus causes anorexia through a leptin-independent melanocortin pathway. *Cell Metab* **10**, 355–65 (2009).
378. Pan, W., Hsuchou, H. & Kastin, A. J. Nesfatin-1 crosses the blood-brain barrier without saturation. *Peptides* **28**, 2223–8 (2007).
379. Price, T. O., Samson, W. K., Niehoff, M. L. & Banks, W. A. Permeability of the blood-brain barrier to a novel satiety molecule nesfatin-1. *Peptides* **28**, 2372–81 (2007).
380. Shimizu, H., Ohsaki, A., Oh-I, S., Okada, S. & Mori, M. A new anorexigenic protein, nesfatin-1. *Peptides* **30**, 995–8 (2009).
381. Tsuchiya, T. *et al.* Fasting Concentrations of Nesfatin-1 Are Negatively Correlated with Body Mass Index in Non-Obese Males. *Clin Endocrinol (Oxf)* (2010).at <http://view.ncbi.nlm.nih.gov/pubmed/20550530>
382. Ogiso, K. *et al.* Plasma nesfatin-1 concentrations in restricting-type anorexia nervosa. *Peptides* **32**, 150–3 (2011).
383. Waters, S. B. & Pessin, J. E. Insulin receptor substrate 1 and 2 (IRS1 and IRS2): what a tangled web we weave. *Trends Cell Biol* **6**, 1–4 (1996).
384. Burks, D. J. & White, M. F. IRS proteins and beta-cell function. *Diabetes* **50 Suppl 1**, S140–5 (2001).
385. Saltiel, A. R. & Kahn, C. R. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **414**, 799–806 (2001).
386. Fasshauer, M. *et al.* Essential role of insulin receptor substrate 1 in differentiation of brown adipocytes. *Mol Cell Biol* **21**, 319–29 (2001).
387. Jellema, A., Zeegers, M. P. A., Feskens, E. J. M., Dagnelie, P. C. & Mensink, R. P. Gly972Arg variant in the insulin receptor substrate-1 gene and association with Type 2 diabetes: a meta-analysis of 27 studies. *Diabetologia* **46**, 990–5 (2003).
388. Rung, J. *et al.* Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet* **41**, 1110–5 (2009).
389. Air, E. L. *et al.* Small molecule insulin mimetics reduce food intake and body weight and prevent development of obesity. *Nat Med* **8**, 179–183 (2002).

390. Melzner, I. *et al.* Leptin gene expression in human preadipocytes is switched on by maturation-induced demethylation of distinct CpGs in its proximal promoter. *J Biol Chem* **277**, 45420–7 (2002).
391. Stoger, R. In vivo methylation patterns of the leptin promoter in human and mouse. *Epigenetics* **1**, 155–62 (2006).
392. Noer, A., Sorensen, A. L., Boquest, A. C. & Collas, P. Stable CpG hypomethylation of adipogenic promoters in freshly isolated, cultured, and differentiated mesenchymal stem cells from adipose tissue. *Mol Biol Cell* **17**, 3543–56 (2006).
393. Milagro, F. I. *et al.* High fat diet-induced obesity modifies the methylation pattern of leptin promoter in rats. *J Physiol Biochem* **65**, 1–9 (2009).
394. Jaquet, D., Collin, D., Lévy-Marchal, C. & Czernichow, P. Adult Height Distribution in Subjects Born Small for Gestational Age. *Horm Res* **62**, 92–96 (2004).
395. Jaziri, R. *et al.* The PPARG Pro12Ala Polymorphism Is Associated With a Decreased Risk of Developing Hyperglycemia Over 6 Years and Combines With the Effect of the APM1 G-11391A Single Nucleotide Polymorphism: The Data From an Epidemiological Study on the Insulin Resistance Syndrome (DESIR) Study. *Diabetes* **55**, 1157–1162 (2006).
396. Jernås, M. *et al.* Regulation of carboxylesterase 1 (CES1) in human adipose tissue. *Biochem Biophys Res Commun* **383**, 63–7 (2009).
397. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
398. Sauer, S. & Gut, I. G. Genotyping single-nucleotide polymorphisms by matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry. *Journal of Chromatography B* **782**, 73–87 (2002).
399. Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J Mol Biol* **196**, 261–82 (1987).
400. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).
401. Wigginton, J. E. & Abecasis, G. R. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* **21**, 3445–7 (2005).
402. Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* **30**, 97–101 (2002).
403. Gabriel, S. B. *et al.* The Structure of Haplotype Blocks in the Human Genome. *Science (80-)* **296**, 2225–2229 (2002).
404. Howie, B. N., Donnelly, P. & Marchini, J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet* **5**, e1000529 (2009).
405. Howie, B., Marchini, J. & Stephens, M. Genotype Imputation with Thousands of Genomes. *G3* **1**, 457–470 (2011).
406. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics* **39**, 906–913 (2007).
407. Dupont, W. D. & Plummer Jr, W. D. Power and sample size calculations: A review and computer program. *Control Clin Trials* **11**, 116–128 (1990).
408. Frye, R. A. Phylogenetic classification of prokaryotic and eukaryotic Sir2-like proteins. *Biochem. Biophys. Res. Commun.* **273**, 793–798 (2000).
409. Lin, S. J., Defossez, P. A. & Guarente, L. Requirement of NAD and SIR2 for life-span extension by calorie restriction in Saccharomyces cerevisiae. *Science* **289**, 2126–2128 (2000).
410. Landry, J. *et al.* The silencing protein SIR2 and its homologs are NAD-dependent protein deacetylases. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5807–5811 (2000).

411. Michishita, E., Park, J. Y., Burneskis, J. M., Barrett, J. C. & Horikawa, I. Evolutionarily conserved and nonconserved cellular localizations and functions of human SIRT proteins. *Mol. Biol. Cell* **16**, 4623–4635 (2005).
412. Vaquero, A. *et al.* Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. *Mol. Cell* **16**, 93–105 (2004).
413. Yeung, F. *et al.* Modulation of NF-κB-dependent transcription and cell survival by the SIRT1 deacetylase. *EMBO J* **23**, 2369–2380 (2004).
414. Daitoku, H. *et al.* Silent information regulator 2 potentiates Foxo1-mediated transcription through its deacetylase activity. *PNAS* **101**, 10042–10047 (2004).
415. Motta, M. C. *et al.* Mammalian SIRT1 Represses Forkhead Transcription Factors. *Cell* **116**, 551–563 (2004).
416. Luo, J. *et al.* Negative control of p53 by Sir2alpha promotes cell survival under stress. *Cell* **107**, 137–148 (2001).
417. Vaziri, H. *et al.* hSIR2(SIRT1) functions as an NAD-dependent p53 deacetylase. *Cell* **107**, 149–159 (2001).
418. Picard, F. *et al.* Sirt1 promotes fat mobilization in white adipocytes by repressing PPAR-gamma. *Nature* **429**, 771–776 (2004).
419. Lagouge, M. *et al.* Resveratrol improves mitochondrial function and protects against metabolic disease by activating SIRT1 and PGC-1alpha. *Cell* **127**, 1109–1122 (2006).
420. Rodgers, J. T. *et al.* Nutrient control of glucose homeostasis through a complex of PGC-1alpha and SIRT1. *Nature* **434**, 113–118 (2005).
421. Purushotham, A. *et al.* Hepatocyte-specific deletion of SIRT1 alters fatty acid metabolism and results in hepatic steatosis and inflammation. *Cell Metab.* **9**, 327–338 (2009).
422. Ponugoti, B. *et al.* SIRT1 deacetylates and inhibits SREBP-1C activity in regulation of hepatic lipid metabolism. *J. Biol. Chem.* **285**, 33959–33970 (2010).
423. Schwer, B., Bunkenborg, J., Verdin, R. O., Andersen, J. S. & Verdin, E. Reversible lysine acetylation controls the activity of the mitochondrial enzyme acetyl-CoA synthetase 2. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10224–10229 (2006).
424. Hallows, W. C., Lee, S. & Denu, J. M. Sirtuins deacetylate and activate mammalian acetyl-CoA synthetases. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10230–10235 (2006).
425. Zhang, J. The Direct Involvement of SirT1 in Insulin-induced Insulin Receptor Substrate-2 Tyrosine Phosphorylation. *J. Biol. Chem.* **282**, 34356–34364 (2007).
426. Bordone, L. *et al.* Sirt1 regulates insulin secretion by repressing UCP2 in pancreatic beta cells. *PLoS Biol.* **4**, e31 (2006).
427. Bitterman, K. J., Anderson, R. M., Cohen, H. Y., Latorre-Esteves, M. & Sinclair, D. A. Inhibition of silencing and accelerated aging by nicotinamide, a putative negative regulator of yeast sir2 and human SIRT1. *J. Biol. Chem.* **277**, 45099–45107 (2002).
428. Canto, C. *et al.* AMPK regulates energy expenditure by modulating NAD+ metabolism and SIRT1 activity. *Nature* (2009).at <http://dx.doi.org/10.1038/nature07813http://www.nature.com/nature/journal/vaop/ncurrent/suppinfo/nature07813_S1.html>
429. Cantó, C. *et al.* Interdependence of AMPK and SIRT1 for metabolic adaptation to fasting and exercise in skeletal muscle. *Cell Metab.* **11**, 213–219 (2010).
430. Kim, H.-J. *et al.* Metabolomic analysis of livers and serum from high-fat diet induced obese mice. *J. Proteome Res.* **10**, 722–731 (2011).
431. Cohen, H. Y. *et al.* Calorie restriction promotes mammalian cell survival by inducing the SIRT1 deacetylase. *Science* **305**, 390–392 (2004).
432. Qin, W. *et al.* Neuronal SIRT1 activation as a novel mechanism underlying the prevention of Alzheimer disease amyloid neuropathology by calorie restriction. *J. Biol. Chem.* **281**, 21745–21754 (2006).

433. Chen, D. *et al.* Tissue-specific regulation of SIRT1 by calorie restriction. *Genes Dev* **22**, 1753–1757 (2008).
434. Sun, C. *et al.* SIRT1 improves insulin sensitivity under insulin-resistant conditions by repressing PTP1B. *Cell Metab.* **6**, 307–319 (2007).
435. Nemoto, S., Fergusson, M. M. & Finkel, T. Nutrient availability regulates SIRT1 through a forkhead-dependent pathway. *Science* **306**, 2105–2108 (2004).
436. Han, L. *et al.* SIRT1 is regulated by a PPARγ–SIRT1 negative feedback loop associated with senescence. *Nucleic Acids Res* **38**, 7458–7471 (2010).
437. Moynihan, K. A. *et al.* Increased dosage of mammalian Sir2 in pancreatic beta cells enhances glucose-stimulated insulin secretion in mice. *Cell Metab.* **2**, 105–117 (2005).
438. Howitz, K. T. *et al.* Small molecule activators of sirtuins extend Saccharomyces cerevisiae lifespan. *Nature* **425**, 191–196 (2003).
439. Wood, J. G. *et al.* Sirtuin activators mimic caloric restriction and delay ageing in metazoans. *Nature* **430**, 686–689 (2004).
440. Baur, J. A. *et al.* Resveratrol improves health and survival of mice on a high-calorie diet. *Nature* **444**, 337–342 (2006).
441. Su, H.-C., Hung, L.-M. & Chen, J.-K. Resveratrol, a red wine antioxidant, possesses an insulin-like effect in streptozotocin-induced diabetic rats. *Am. J. Physiol. Endocrinol. Metab.* **290**, E1339–1346 (2006).
442. Peeters, A. V. *et al.* Association of SIRT1 gene variation with visceral obesity. *Hum Genet* **124**, 431–6 (2008).
443. Van den Berg, S. W. *et al.* Genetic variations in regulatory pathways of fatty acid and glucose metabolism are associated with obesity phenotypes: a population-based cohort study. *Int J Obes (Lond)* **33**, 1143–52 (2009).
444. Zillikens, Mc. *et al.* SIRT1 Genetic Variation Is Related to BMI and Risk of Obesity. *Diabetes* **58**, 2828–2834 (2009).
445. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
446. Naslavsky, M. S., Crovella, S., Filho, J. L. L. & Rocha, C. R. C. The sound of silence: Human [beta]-defensin-1 gene untranslated SNPs change the predicted mRNA secondary structure in a length-dependent manner. *Immunol Lett* **129**, 53–55
447. Hayashida, S. *et al.* Fasting promotes the expression of SIRT1, an NAD+-dependent protein deacetylase, via activation of PPARα in mice. *Mol Cell Biochem* **339**, 285–292
448. Satoh, A. *et al.* SIRT1 promotes the central adaptive response to diet restriction through activation of the dorsomedial and lateral nuclei of the hypothalamus. *J Neurosci* **30**, 10220–32 (2010).
449. Liu, Y., Gerber, R., Wu, J., Tsuruda, T. & McCarter, J. D. High-throughput assays for sirtuin enzymes: A microfluidic mobility shift assay and a bioluminescence assay. *Anal Biochem* **378**, 53–59 (2008).
450. Tatemoto, K. *et al.* Isolation and characterization of a novel endogenous peptide ligand for the human APJ receptor. *Biochem. Biophys. Res. Commun.* **251**, 471–476 (1998).
451. Medhurst, A. D. *et al.* Pharmacological and immunohistochemical characterization of the APJ receptor and its endogenous ligand apelin. *J. Neurochem.* **84**, 1162–1172 (2003).
452. Lee, D. K. *et al.* Characterization of apelin, the ligand for the APJ receptor. *J. Neurochem.* **74**, 34–41 (2000).
453. Kawamata, Y. *et al.* Molecular properties of apelin: tissue distribution and receptor binding. *Biochim Biophys Acta* **1538**, 162–71 (2001).
454. Tatemoto, K. *et al.* Isolation and characterization of a novel endogenous peptide ligand for the human APJ receptor. *Biochem Biophys Res Commun* **251**, 471–6 (1998).
455. Reaux, A. *et al.* Physiological role of a novel neuropeptide, apelin, and its receptor in the rat brain. *J Neurochem* **77**, 1085–96 (2001).

456. Lee, D. K. *et al.* Characterization of apelin, the ligand for the APJ receptor. *J Neurochem* **74**, 34–41 (2000).
457. Hosoya, M. *et al.* Molecular and functional characteristics of APJ. Tissue distribution of mRNA and interaction with the endogenous ligand apelin. *J Biol Chem* **275**, 21061–7 (2000).
458. Boucher, J. *et al.* Apelin, a newly identified adipokine up-regulated by insulin and obesity. *Endocrinology* **146**, 1764–1771 (2005).
459. Castan-Laurell, I. *et al.* Effect of hypocaloric diet-induced weight loss in obese women on plasma apelin and adipose tissue expression of apelin and APJ. *Eur. J. Endocrinol.* **158**, 905–910 (2008).
460. Heinonen, M. V. *et al.* Apelin, orexin-A and leptin plasma levels in morbid obesity and effect of gastric banding. *Regul. Pept.* **130**, 7–13 (2005).
461. Dray, C. *et al.* Apelin and APJ regulation in adipose tissue and skeletal muscle of type 2 diabetic mice and humans. *Am J Physiol Endocrinol Metab* **298**, E1161–E1169 (2010).
462. Daviaud, D. *et al.* TNFα up-regulates apelin expression in human and mouse adipose tissue. *FASEB J* **20**, 1528–1530 (2006).
463. Mazzucotelli, A. *et al.* The transcriptional co-activator PGC-1alpha up regulates apelin in human and mouse adipocytes. *Regul. Pept.* **150**, 33–37 (2008).
464. Wei, L., Hou, X. & Tatemoto, K. Regulation of apelin mRNA expression by insulin and glucocorticoids in mouse 3T3-L1 adipocytes. *Regul. Pept.* **132**, 27–32 (2005).
465. Hung, W.-W. *et al.* Blockade of the renin-angiotensin system ameliorates apelin production in 3T3-L1 adipocytes. *Cardiovasc Drugs Ther* **25**, 3–12 (2011).
466. Sunter, D., Hewson, A. K. & Dickson, S. L. Intracerebroventricular injection of apelin-13 reduces food intake in the rat. *Neurosci. Lett.* **353**, 1–4 (2003).
467. O'Shea, M., Hansen, M. J., Tatemoto, K. & Morris, M. J. Inhibitory effect of apelin-12 on nocturnal food intake in the rat. *Nutr Neurosci* **6**, 163–167 (2003).
468. Higuchi, K. *et al.* Apelin, an APJ receptor ligand, regulates body adiposity and favors the messenger ribonucleic acid expression of uncoupling proteins in mice. *Endocrinology* **148**, 2690–2697 (2007).
469. Kunduzova, O. *et al.* Apelin/APJ signaling system: a potential link between adipose tissue and endothelial angiogenic processes. *FASEB J* **22**, 4146–4153 (2008).
470. Yue, P. *et al.* Apelin decreases lipolysis via G(q), G(i), and AMPK-Dependent Mechanisms. *Endocrinology* **152**, 59–68 (2011).
471. Yamamoto, T. *et al.* Apelin-transgenic mice exhibit a resistance against diet-induced obesity by increasing vascular mass and mitochondrial biogenesis in skeletal muscle. *Biochim. Biophys. Acta* **1810**, 853–862 (2011).
472. Zhang, R. *et al.* Association of apelin genetic variants with type 2 diabetes and related clinical features in Chinese Hans. *Chin Med J (Engl)* **122**, 1273–6 (2009).
473. Ohman, M. *et al.* Genome-Wide Scan of Obesity in Finnish Sibpairs Reveals Linkage to Chromosome Xq24. *J Clin Endocrinol Metab* **85**, 3183–3190 (2000).
474. Liao, Y.-C. *et al.* Apelin gene polymorphism influences apelin expression and obesity phenotypes in Chinese women. *Am. J. Clin. Nutr.* **94**, 921–928 (2011).
475. Grosfeld, J. L., Du, X. & Williams, D. A. Interleukin-11: its biology and prospects for clinical use. *JPEN J Parenter Enteral Nutr* **23**, S67–9 (1999).
476. Paiva, P., Menkhorst, E., Salamonsen, L. & Dimitriadis, E. Leukemia inhibitory factor and interleukin-11: critical regulators in the establishment of pregnancy. *Cytokine Growth Factor Rev* **20**, 319–28 (2009).
477. Ohsumi, J. *et al.* Regulation of lipoprotein lipase synthesis in 3T3-L1 adipocytes by interleukin-11/adipogenesis inhibitory factor. *Biochem. Mol. Biol. Int.* **32**, 705–712 (1994).

478. Redlich, C. A., Gao, X., Rockwell, S., Kelley, M. & Elias, J. A. IL-11 enhances survival and decreases TNF production after radiation-induced thoracic injury. *J. Immunol.* **157**, 1705–1710 (1996).
479. Bell, C. G. *et al.* Genome-wide linkage analysis for severe obesity in french caucasians finds significant susceptibility locus on chromosome 19q. *Diabetes* **53**, 1857–65 (2004).
480. Clément, K. *et al.* Weight loss regulates inflammation-related genes in white adipose tissue of obese subjects. *FASEB J* **18**, 1657–69 (2004).
481. Wilson, P. A., Gardner, S. D., Lambie, N. M., Commans, S. A. & Crowther, D. J. Characterization of the human patatin-like phospholipase family. *J. Lipid Res.* **47**, 1940–1949 (2006).
482. Jenkins, C. M. *et al.* Identification, Cloning, Expression, and Purification of Three Novel Human Calcium-independent Phospholipase A2 Family Members Possessing Triacylglycerol Lipase and Acylglycerol Transacylase Activities. *J. Biol. Chem.* **279**, 48968–48975 (2004).
483. Lake, A. C. *et al.* Expression, regulation, and triglyceride hydrolase activity of Adiponutrin family members. *J. Lipid Res.* **46**, 2477–2487 (2005).
484. Kershaw, E. E. *et al.* Adipose Triglyceride Lipase Function, Regulation by Insulin, and Comparison With Adiponutrin. *Diabetes* **55**, 148–157 (2006).
485. Bertile, F. & Raclot, T. Differences in mRNA expression of adipocyte-derived factors in response to fasting, refeeding and leptin. *Biochim. Biophys. Acta* **1683**, 101–109 (2004).
486. Faraj, M. *et al.* Insulin regulation of gene expression and concentrations of white adipose tissue-derived proteins in vivo in healthy men: relation to adiponutrin. *J Endocrinol* **191**, 427–435 (2006).
487. Romeo, S. *et al.* Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* **40**, 1461–5 (2008).
488. Speliotes, E. K., Butler, J. L., Palmer, C. D., Voight, B. F. & Hirschhorn, J. N. PNPLA3 variants specifically confer increased risk for histologic nonalcoholic fatty liver disease but not metabolic disease. *Hepatology* **52**, 904–912 (2010).
489. Rotman, Y., Koh, C., Zmuda, J. M., Kleiner, D. E. & Liang, T. J. The association of genetic variability in patatin-like phospholipase domain-containing protein 3 (PNPLA3) with histological severity of nonalcoholic fatty liver disease. *Hepatology* **52**, 894–903 (2010).
490. Valenti, L. *et al.* I148M patatin-like phospholipase domain-containing 3 gene variant and severity of pediatric nonalcoholic fatty liver disease. *Hepatology* **52**, 1274–80 (2010).
491. Falck-Ytter, Y., Younossi, Z. M., Marchesini, G. & McCullough, A. J. Clinical features and natural history of nonalcoholic steatosis syndromes. *Semin. Liver Dis.* **21**, 17–26 (2001).
492. Johansson, L. E., Lindblad, U., Larsson, C. A., Råstam, L. & Ridderstråle, M. Polymorphisms in the adiponutrin gene are associated with increased insulin secretion and obesity. *Eur J Endocrinol* **159**, 577–583 (2008).
493. Stengel, A. & Taché, Y. Regulation of food intake: the gastric X/A-like endocrine cell in the spotlight. *Curr Gastroenterol Rep* **11**, 448–54 (2009).
494. Stengel, A. *et al.* Central nesfatin-1 reduces dark-phase food intake and gastric emptying in rats: differential role of corticotropin-releasing factor2 receptor. *Endocrinology* **150**, 4911–9 (2009).
495. Yosten, G. L. C. & Samson, W. K. Nesfatin-1 exerts cardiovascular actions in brain: possible interaction with the central melanocortin system. *Am J Physiol Regul Integr Comp Physiol* **297**, R330–6 (2009).
496. Price, C. J., Samson, W. K. & Ferguson, A. V. Nesfatin-1 inhibits NPY neurons in the arcuate nucleus. *Brain Res.* **1230**, 99–106 (2008).
497. Noetzel, S. *et al.* CCK-8S activates c-Fos in a dose-dependent manner in nesfatin-1 immunoreactive neurons in the paraventricular nucleus of the hypothalamus and in the nucleus of the solitary tract of the brainstem. *Regul Pept* **157**, 84–91 (2009).

498. Price, T. O., Samson, W. K., Niehoff, M. L. & Banks, W. A. Permeability of the blood-brain barrier to a novel satiety molecule nesfatin-1. *Peptides* **28**, 2372–2381 (2007).
499. Pan, W., Hsuchou, H. & Kastin, A. J. Nesfatin-1 crosses the blood-brain barrier without saturation. *Peptides* **28**, 2223–2228 (2007).
500. Stengel, A. *et al.* Identification and characterization of nesfatin-1 immunoreactivity in endocrine cell types of the rat gastric oxyntic mucosa. *Endocrinology* **150**, 232–238 (2009).
501. Shimizu, H. *et al.* Peripheral administration of nesfatin-1 reduces food intake in mice: the leptin-independent mechanism. *Endocrinology* **150**, 662–671 (2009).
502. Zegers, D., Beckers, S., Mertens, I. L., Van Gaal, L. F. & Van Hul, W. Association between polymorphisms of the Nesfatin gene, NUCB2, and obesity in men. *Mol. Genet. Metab.* **103**, 282–286 (2011).
503. Gloyn, A. L. *et al.* Large-Scale Association Studies of Variants in Genes Encoding the Pancreatic {beta}-Cell KATP Channel Subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) Confirm That the KCNJ11 E23K Variant Is Associated With Type 2 Diabetes. *Diabetes* **52**, 568–572 (2003).
504. Almind, K. *et al.* Aminoacid polymorphisms of insulin receptor substrate-1 in non-insulin-dependent diabetes mellitus. *Lancet* **342**, 828–32 (1993).
505. Jellema, A., Mensink, R. P., Kromhout, D., Saris, W. H. M. & Feskens, E. J. M. Metabolic risk markers in an overweight and normal weight population with oversampling of carriers of the IRS-1 972Arg-variant. *Atherosclerosis* **171**, 75–81 (2003).
506. Lei, H. H., Coresh, J., Shuldiner, A. R., Boerwinkle, E. & Brancati, F. L. Variants of the insulin receptor substrate-1 and fatty acid binding protein 2 genes and the risk of type 2 diabetes, obesity, and hyperinsulinemia in African-Americans: the Atherosclerosis Risk in Communities Study. *Diabetes* **48**, 1868–1872 (1999).
507. Krempler, F., Hell, E., Winkler, C., Breban, D. & Patsch, W. Plasma leptin levels: interaction of obesity with a common variant of insulin receptor substrate-1. *Arterioscler. Thromb. Vasc. Biol.* **18**, 1686–1690 (1998).
508. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* **103**, 1412–1417 (2006).
509. Van der Ploeg, L. H. T. & Flavell, R. A. DNA methylation in the human γδβ-globin locus in erythroid and nonerythroid tissues. *Cell* **19**, 947–958 (1980).
510. Lane, N. *et al.* Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* **35**, 88–93 (2003).
511. Druker, R., Bruxner, T. J., Lehrbach, N. J. & Whitelaw, E. Complex patterns of transcription at the insertion site of a retrotransposon in the mouse. *Nucleic Acids Res* **32**, 5800–8 (2004).
512. Chan, T. L. *et al.* Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer. *Nat Genet* **38**, 1178–83 (2006).
513. Hitchins, M. P. *et al.* Inheritance of a Cancer-Associated MLH1 Germ-Line Epimutation. *N Engl J Med* **356**, 697–705 (2007).
514. Jiang, Y., Tsai, T. F., Bressler, J. & Beaudet, A. L. Imprinting in Angelman and Prader-Willi syndromes. *Curr Opin Genet Dev* **8**, 334–42 (1998).
515. Weinstein, L. S., Xie, T., Qasem, A., Wang, J. & Chen, M. The role of GNAS and other imprinted genes in the development of obesity. *Int J Obes* **34**, 6–17 (2009).
516. Mammès, O. *et al.* Novel polymorphisms in the 5’ region of the LEP gene: association with leptin levels and response to low-calorie diet in human obesity. *Diabetes* **47**, 487–9 (1998).
517. Hager, J. *et al.* untranslated region of the human ob gene is associated with low leptin levels. *Int J Obes Relat Metab Disord* **22**, 200–5 (1998).

518. LI, W.-D. *et al.* Sequence variants in the 5[prime] flanking region of the leptin gene are associated with obesity in women. *Ann Hum Genet* **63**, 227–234 (1999).
519. Grunau, C., Clark, S. J. & Rosenthal, A. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res* **29**, E65–5 (2001).
520. Paul, C. L. & Clark, S. J. Cytosine methylation: quantitation by automated genomic sequencing and GENESCAN analysis. *Biotechniques* **21**, 126–33 (1996).
521. Tost, J. & Gut, I. G. Analysis of gene-specific DNA methylation patterns by pyrosequencing technology. *Methods Mol Biol* **373**, 89–102 (2007).
522. Tost, J. & Gut, I. G. DNA methylation analysis by pyrosequencing. *Nat Protoc* **2**, 2265–75 (2007).
523. Van den Boom, D. & Ehrich, M. Mass spectrometric analysis of cytosine methylation by base-specific cleavage and primer extension methods. *Methods Mol Biol* **507**, 207–27 (2009).
524. Irizarry, R. A. *et al.* Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res* **18**, 780–90 (2008).
525. Newell-Price, J., King, P. & Clark, A. J. The CpG island promoter of the human proopiomelanocortin gene is methylated in nonexpressing normal tissue and tumors and represses expression. *Mol Endocrinol* **15**, 338–48 (2001).
526. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
527. Jiang, M. H. *et al.* Rapid quantification of DNA methylation by measuring relative peak heights in direct bisulfite-PCR sequencing traces. *Lab Invest* (2009).at <http://dx.doi.org/10.1038/labinvest.2009.132http://www.nature.com/labinvest/journal/vaop/ncurrent/suppinfo/labinvest2009132s1.html>
528. Zhao, J., Goldberg, J. & Vaccarino, V. Promoter methylation of serotonin transporter gene is associated with obesity measures: a monozygotic twin study. *International Journal of Obesity* (2012).doi:10.1038/ijo.2012.8
529. Candiloro, I. L. & Dobrovic, A. Detection of MGMT Promoter Methylation in Normal Individuals Is Strongly Associated with the T Allele of the rs16906252 MGMT Promoter Single Nucleotide Polymorphism. *Cancer Prev Res* **2**, 862–867 (2009).
530. Bogdanović, O. & Veenstra, G. J. C. DNA methylation and methyl-CpG binding proteins: developmental requirements and function. *Chromosoma* **118**, 549–65 (2009).
531. Lönnqvist, F., Arner, P., Nordfors, L. & Schalling, M. Overexpression of the obese (ob) gene in adipose tissue of human obese subjects. *Nat. Med.* **1**, 950–953 (1995).
532. Hamilton, B. S., Paglia, D., Kwan, A. Y. & Deitel, M. Increased obese mRNA expression in omental fat cells from massively obese humans. *Nat. Med.* **1**, 953–956 (1995).
533. Clark, S. J. *et al.* Association of sirtuin 1 (SIRT1) gene SNPs and transcript expression levels with severe obesity. *Obesity (Silver Spring)* **20**, 178–185 (2012).
534. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
535. Zhu, Q. *et al.* A genome-wide comparison of the functional properties of rare and common genetic variants in humans. *Am J Hum Genet* **88**, 458–68 (2011).
536. Bodmer, W. & Bonilla, C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.* **40**, 695–701 (2008).
537. Nejentsev, S. *et al.* Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature* **450**, 887–892 (2007).
538. Plenge, R. M. *et al.* Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* **39**, 1477–1482 (2007).
539. Lupski, J. R. Genomic rearrangements and sporadic disease. *Nat. Genet.* **39**, S43–47 (2007).

540. Walters, R. G. *et al.* A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* **463**, 671–675 (2010).
541. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *PNAS* (2012).doi:10.1073/pnas.1119675109
542. Ode, J. J., Pivarnik, J. M., Reeves, M. J. & Knous, J. L. Body mass index as a predictor of percent fat in college athletes and nonathletes. *Med Sci Sports Exerc* **39**, 403–409 (2007).
543. Kyle, U. G., Schutz, Y., Dupertuis, Y. M. & Pichard, C. Body composition interpretation. Contributions of the fat-free mass index and the body fat mass index. *Nutrition* **19**, 597–604 (2003).

Appendix

A1. Tagging SNPs genotyped and the corresponding captured SNPs

Gene	Tagging SNP	SNPs Captured
SIRT1	rs11596401	rs10997868, rs1467568, rs2224573, rs11596401, rs10997866, rs10997875, rs7069102, rs10823111, rs10997870, rs1885472, rs3758391, rs7091896, rs10997860
SIRT1	rs3818292	rs10823108, rs10823112, rs3740051, rs3818292, rs2236319, rs7096385
SIRT1	rs12413112	rs12778366, rs12413112
Nesfatin	rs214935	rs4356203,rs618331,rs214910,rs10832733,rs214935,rs214940,rs7123301,rs7125607,rs214901,rs7478986,rs620241,rs7342262,rs11024158,rs7107283,rs7946010,rs603618,rs10766381,rs214921,rs11024204,rs17472886,rs12280930,rs11024159,rs7936310,rs11024208,rs12286505,rs214925,rs214900,rs7925692,rs214914,rs214939,rs512852
Nesfatin	rs3741203	rs11024251,rs9645621,rs10766388,rs3741203,rs10832757,rs10832769,rs10832768,rs10766393,rs10766384,rs10766391,rs9633836,rs10832770,rs10466382,rs10832779,rs10766383,rs1002227
Nesfatin	rs1987694	rs17561348,rs586785,rs17473243,rs12577418,rs214933,rs3950680,rs17560341,rs12577525,rs1987694
Nesfatin	rs10832750	rs10832752,rs542274,rs7108315,rs214076,rs1488933, rs214071,rs10832750,rs214101
Nesfatin	rs16933873	rs11605718,rs1974527,rs16933827,rs16933873,rs16924899,rs16933823,rs11603765,rs2040859
Nesfatin	rs10832775	rs10832781,rs11024262,rs2354867,rs10832763,rs10832775,rs11024263,rs10832772
Nesfatin	rs7108861	rs7108861,rs10832738,rs10832741,rs12285874,rs11024165,rs10832742
Nesfatin	rs7937091	rs7121987,rs12791318,rs7110037,rs7937091,rs16933984,rs4451707
Nesfatin	rs757110	rs2051772,rs1557765,rs5215,rs757110,rs7928810,rs10832778
Nesfatin	rs2051773	rs214105,rs2051773,rs214070,rs757081,rs214091,rs2521999
Nesfatin	rs17647408	rs11604470,rs11604561,rs17647408,rs1979602,rs11607974
Nesfatin	rs16933837	rs16933829,rs11024218,rs16924900,rs16933837
Nesfatin	rs214083	rs214083,rs214097,rs214092,rs214093
Nesfatin	rs10832749	rs10832749,rs10832753,rs7111505,rs12226898
Nesfatin	rs11024273	rs4148638,rs2074312,rs11024273

Nesfatin	rs10832782	rs7395484,rs7395484,rs11024272
Nesfatin	rs6486364	rs2214285,rs8192691,rs6486364
Nesfatin	rs10832756	rs10832756,rs16933952,rs12419530
Nesfatin	rs214106	rs10741725,rs214106
Nesfatin	rs2302510	rs3802962,rs2302510
Nesfatin	rs2285676	rs2285676,rs886288
Adiponutrin	rs2294916	rs4823179,rs1883349,rs2072905,rs1010023,rs2294916,rs2073081,rs2072907,rs926633,rs1010022,rs2896019,rs2281135
Adiponutrin	rs2294918	rs4823174,rs2294918,rs2076210,rs5764034
Adiponutrin	rs11090617	rs11090617,rs2076211,rs4823173,rs12483959
Adiponutrin	rs734561	rs734561,rs2006943
Adiponutrin	rs738408	rs738409,rs738408
Adiponutrin	rs2294919	rs2076208,rs2294919

A2. Primer Sequences Used in Sequenom Genotyping Assays

SNP	Gene	PCR Primer 1	PCR Primer 2	iPlex Extension Primer
rs34122272	SIRT1	ACGTTGGATGACAATGTATAGCCA GGCACG	ACGTTGGATGAACTCCTGACCTTG TGATCC	ATCCTGGCACTTTGG
rs35689145	SIRT1	ACGTTGGATGTAGTCCCAGCTGTT CGGAAG	ACGTTGGATGGCAGCATATTATTGG CTCAC	GGAGAATCGCTTGAACTC
rs33957861	SIRT1	ACGTTGGATGCTTTAGGAATTCTG CTCACTC	ACGTTGGATGGCTCCCGGCTATCT TTTCTG	TTCTGCTCACTCAGTTTCA
rs10997871	SIRT1	ACGTTGGATGAGGGTAGCTTTATG TAGTTG	ACGTTGGATGCTTCCTTCTCCATAT CATGT	CAGTTCGTAATCAGAAAGG
rs41274092	SIRT1	ACGTTGGATGACTACTTCGCAACT ATACCC	ACGTTGGATGACCATGACACTGAAT TATCC	CCTCACATAGACACGCTGGA
rs7073231	SIRT1	ACGTTGGATGCTCTGCCTTCTAAA GTGCTG	ACGTTGGATGACACAACAAGACCT CATCTC	CTTCTAAAGTGCTGGGATTA
rs33955981	SIRT1	ACGTTGGATGAGGGTCTCTGTCAT ATGTTC	ACGTTGGATGCCTCTTTATTATGCT ACAGAC	CGAACAAATTAAAAAACCCAGTC
rs11594238	SIRT1	ACGTTGGATGCCATGCCAGGCTAA TTTTTG	ACGTTGGATGGGTCGATCATTTGA GGTCAG	CCAGGCTAATTTTTGTATTTTTA
rs7904945	SIRT1	ACGTTGGATGTAAGTGCCGGGATA CATGTG	ACGTTGGATGGATAGATGCAGCAA ACCACC	GGAGTGATACATGTGCAGAATGC
rs10997865	SIRT1	ACGTTGGATGTGAGTCAGTGTACT AGGCAG	ACGTTGGATGCCAATATTTCCTTTT CATGGC	TGTCCTAGGCAGAAAGTTCACTAA
rs11596401	SIRT1	ACGTTGGATGGAGAACTTGACTCT ACAAAG	ACGTTGGATGCTCCTGAGTAGCTG AGACCA	AACATTTAGCTAGGTATATGGTAC
rs34414573	SIRT1	ACGTTGGATGGCTGTTAATGAAGA AATGTG	ACGTTGGATGCATTTCTGAATCTCT GAATCC	GGACTAAAAGAACTGGATAGTTGA
rs2894057	SIRT1	ACGTTGGATGCCAGGTTCAAGTGA TTCTGC	ACGTTGGATGAAAAATAGCCAGAC GTGGTG	CCCCCTCCCGAGTAGCTGGAATTA C
rs2236318	SIRT1	ACGTTGGATGGGGAAAAGGCTTA AAGTCAAC	ACGTTGGATGCAAAAAACCCTCAC AGAATGC	TCAAAATCTCATTTATTTTCTGAAGT

rs3818292	SIRT1	ACGTTGGATGAGAAGTATTGGCCTTGACAG	ACGTTGGATGATGCATGCAACTGCAGCATC	GAAGTTAATTATAGAAAACCTCAGAT
rs10823103	SIRT1	ACGTTGGATGCCTTATCTGTGTTAAAGATGG	ACGTTGGATGCAACCCTACCCTTATTTCAG	TGTGTTAAAGATGGAATATGAGGTAT
rs2394443	SIRT1	ACGTTGGATGTTAAACCCCATCACGTGACC	ACGTTGGATGACCCGTAGTGTTGTGGTCTG	GGCCGCCCCCGCCCTCT
rs11599176	SIRT1	ACGTTGGATGCCGGAGTTCAAGAGATCTTC	ACGTTGGATGGATTTTTTGGATGCAGTAGC	CCAAAGCACTGGGATTA
rs35224060	SIRT1	ACGTTGGATGGTTCATGTCTGTTACTTCCTG	ACGTTGGATGTGGATTTGGGACTGATGGAG	TACTTCCTGTTTCACAGATA
rs2273773	SIRT1	ACGTTGGATGATTCCAGCCATCTCTCTGTC	ACGTTGGATGTCCAGCGTGTCTATGTTCTG	CTCTGTCACAAATTCATAGCC
rs35592342	SIRT1	ACGTTGGATGACTTTCTGACTTCTGCCACG	ACGTTGGATGAAGAGCAGACGTGTTCACAG	GGACTGCCACGATTTGCAAAC
rs12413112	SIRT1	ACGTTGGATGATGTGGCTGTTGAGTACTTG	ACGTTGGATGCCATACTGGATCTCTAAAATC	AGTACTTGAAATGTATTGATATGT
rs737477	SIRT1	ACGTTGGATGAAAAGGGGACAAGTGAGGTG	ACGTTGGATGTCTTGAACTCCCGACCTCAG	GGGGCACATGCCTGTAA
rs36067477	SIRT1	ACGTTGGATGACCATGACACTGAATTATCC	ACGTTGGATGACTACTTCGCAACTATACCC	TGAATTATCCTTTGGATTCC
rs2234975	SIRT1	ACGTTGGATGGACTTTAAAACAGTGTACAAG	ACGTTGGATGGTAGACTGTTTAATGACTGG	AGTGTACAAGTAAAAAACACT
rs1126757	IL11	ACGTTGGATGTCCCTTGCCCTTACCTGTAG	ACGTTGGATGACCACAACCTGGATTCCCTG	AGCTCCCAGTGCCCC
rs10407001	IL11	ACGTTGGATGGGAGGAGAGAGACTGGGCG	ACGTTGGATGCCGCCAGCCGTCGGTCTGT	AGATGCGGCCGACGGA
rs12975067	IL11	ACGTTGGATGCTGGGCAACAAGAGCAAAAC	ACGTTGGATGGGTCCATGGTTTTCTTTTCC	GCAAAACTCCCTCTCTC
rs17850928	IL11	ACGTTGGATGCAGTCAAGTGTCAGGTGCAG	ACGTTGGATGTGGCGCCCCCCTCCTCAG	GCCCCCCCAGGATGGCG
rs1042506	IL11	ACGTTGGATGTATGTTCCTGCCCAGGCCTA	ACGTTGGATGTCCCGGATTCTTGGGTCTC	GAGGGCAGAAGTCTGTGG
rs10425163	IL11	ACGTTGGATGTCGGTCACCCAGACAGCTCT	ACGTTGGATGACAAACAAAAGGCCCTATGC	GGGGACAGCTCTGTGGAGGC

rs2298885	IL11	ACGTTGGATGAAGCAAGCCTCTCTCCTTAG	ACGTTGGATGAGCAACATGGTGCATCTGTG	GGGTACGGGCGGCTGGGTGG
rs8104023	IL11	ACGTTGGATGAGGGACGCAGGGACTGGTG	ACGTTGGATGAATCCCAGGGAGTCTCCCG	CGCGGGAGGTGACCCCCATCG
rs10402868	IL11	ACGTTGGATGAAGGGAAGCCTGGGTTTTTG	ACGTTGGATGATCGCCCTCAAGTGGATATG	GGTGAGACAGAGAACAGGGAA
rs10402867	IL11	ACGTTGGATGATCGCCCTCAAGTGGATATG	ACGTTGGATGAAGGGAAGCCTGGGTTTTTG	AGTGGATATGTATGACACATTT
rs4252557	IL11	ACGTTGGATGCGGTGAAACCCTGTGTCTAC	ACGTTGGATGTCCTGAGTAGCTGGGACTTC	TGTGTCTACTAAAATACAAAAAATTAT
rs4252576	IL11	ACGTTGGATGAGCGGACCTACTGTCCTAC	ACGTTGGATGCAGGGTCTTCAGGGAAGAG	TCCTACCTGCGGCAC
rs4806475	IL11	ACGTTGGATGTATTTCAGGAGCAGGGGTGG	ACGTTGGATGCTTGGAGACCCAAGAATCCG	GGCAGGTGGACTCCT
rs7246652	IL11	ACGTTGGATGTAATCCCAGCTACTTGGGAG	ACGTTGGATGAATGGCGCGATCAATGGCAC	GGAGGTTGCAGTGAG
rs4252546	IL11	ACGTTGGATGTTCTCTGGTGTGTCTCTCTG	ACGTTGGATGCACACAGAGTGAAGGACAG	CATCTCTGTGGATCTCC
rs7250912	IL11	ACGTTGGATGCAAAGCGCTGGGATAACAGG	ACGTTGGATGATCTCAGAGCTGGGTTTTTG	AATTGTTCTAACAGCAGC
rs4252553	IL11	ACGTTGGATGAGCCTGGGCAGCATAGCAA	ACGTTGGATGTCTAGCCTGGGCAACAGAG	GGCAGCATAGCAAGACTCT
rs1042505	IL11	ACGTTGGATGCCACCCCTGCTCCTGAAATA	ACGTTGGATGCATTATCTCCCCCTAGTTAG	AGTATAAATAAGGCACAGATG
rs4252562	IL11	ACGTTGGATGGAGGCTTGCTTGGGATATAG	ACGTTGGATGCCAAAGTGCCAGGATTACAG	CTGCTTGGGATATAGAAAGATAT
rs10832783	Nesfatin	ACGTTGGATGTCTAATGGTGCTGGGAGAAG	ACGTTGGATGGAGCAACAAGGATGTGGTAG	GCCAGCATCTCTCAT
rs12422139	Nesfatin	ACGTTGGATGTGTGGCGATTCGTACCTATG	ACGTTGGATGACTTACTGTATGCCAGGCAC	GGCACCGTGCTAAGT
rs757110	Nesfatin	ACGTTGGATGTGTCCTGCAGCATTGGGTTG	ACGTTGGATGGGTGCTGAAGCACGTCAATG	CGTCAATGCCCTCATC
rs11024273	Nesfatin	ACGTTGGATGTTTGCTCCATCCTGCACTTG	ACGTTGGATGGGAGAAGGGTTTTGAGCAAG	TTGAGCAAGAAAGCGA

rs2285676	Nesfatin	ACGTTGGATGACACCCTCTCTCATCAACTG	ACGTTGGATGGCTCTACTTGGTCCCTGAAA	GTCCCTGAAAAAGCACC
rs2354863	Nesfatin	ACGTTGGATGCCCCTTCTCCTACTGCTTAA	ACGTTGGATGGGGCACAAAGATGTTAAGGG	AAGGGGAGAGGAACGAT
rs7484027	Nesfatin	ACGTTGGATGGCAAATTTTGCCAGTTATCC	ACGTTGGATGAATGCAATGTGTGATCCCAG	GATCTTAGCGCAGAAAAA
rs16933837	Nesfatin	ACGTTGGATGGAGTAGCTGGGATTACAGGC	ACGTTGGATGCTGTTATCTTCCAGAGAGTC	AGAGAGTCTAGAAAAGCAA
rs2079293	Nesfatin	ACGTTGGATGATGATGAGGGGATGGGAAAG	ACGTTGGATGCTGTCTACCTCATTTGGATG	AGAACTCTCAATTTATCACC
rs10832750	Nesfatin	ACGTTGGATGACCCCTTCTCAATGACACTG	ACGTTGGATGGGAGACTAAAGGTTTCCCAG	CACTATGGATTGCATTAAGGT
rs7127347	Nesfatin	ACGTTGGATGCAAGGAAAAACTTCAATCAC	ACGTTGGATGGTACTGGAATCTGGCCATGA	TAAGTATACTTTTCACCACATC
rs16933873	Nesfatin	ACGTTGGATGACAGTGCTACAAGGCCTTTC	ACGTTGGATGCCTTCACAGGCAGTTCTTTC	TCAATTCCCTAAGTTTTAGAAA
rs34405111	Nesfatin	ACGTTGGATGGTTTCTGAAGCTCATCTGCC	ACGTTGGATGGCAACAGTTCTTCACAGAGG	CAGAGGAAGAACTAAAAGAATA
rs2051773	Nesfatin	ACGTTGGATGTCTGTTCTTGGTGTGACAAC	ACGTTGGATGAAGGACATAGTAAATATGC	TGCTTTAAAAAGGAAATCATCTA
rs2521998	Nesfatin	ACGTTGGATGAAGGCTTAGAATGAGGGAAC	ACGTTGGATGCCATCTTAAAGATTTATTGC	CTTAAACTTTTAAAACATACCTCA
rs1330	Nesfatin	ACGTTGGATGTTACAGAGCATTGTCTTGCC	ACGTTGGATGTGGGTGAATTTTATGACGTG	TCAATAAAGATGCTAAAAAATTGG
rs12295879	Nesfatin	ACGTTGGATGCCTTCAAGTAATCCCCAATG	ACGTTGGATGACTTCAAACTTTCTCTGCAC	TTATCAAAAGCTCAAGTGAAAAAAA
rs1799859	Nesfatin	ACGTTGGATGACCATTAGGGCGTAGGTAAG	ACGTTGGATGGTGACCTCCATCTCCAACTC	GACCTCCATCTCCAACTCCCTGCACAG
rs11024249	Nesfatin	ACGTTGGATGGGGTGATTAACTTTAGGGTG	ACGTTGGATGGATGAAGGAACATGAAAGGAG	ATGAAGGAACATGAAAGGAGAGAATA
rs12796879	Nesfatin	ACGTTGGATGGGGCCAAATATTAGTGTGC	ACGTTGGATGAACTTGTCACTCCTCTGCAC	CAGAGAAAAGGTACAATCTTTACAATG
rs214083	Nesfatin	ACGTTGGATGAACCAAAAGGGTTCATCTCG	ACGTTGGATGCAGCAACCTCTGACACTTAG	GGAGAAAGTTAGTGTAAGTCGTGATAT

rs10832775	Nesfatin	ACGTTGGATGTGCGACCTCCTGAAGCAAAC	ACGTTGGATGTTGGGTTCTGAACTTTGCGG	TTGCGGTTAAATTGATTTAAAAACTGGT
rs1987694	Nesfatin	ACGTTGGATGTGTGAGCCCTTAAAAGGGAC	ACGTTGGATGATTGGCAAGACTCCTGTCTC	TGTCTCGAGCCGAGC
rs17647408	Nesfatin	ACGTTGGATGTCTACCACAACCTGAGCCG	ACGTTGGATGTGGAGTCGGCCGAAGCTTTC	TGAGCCAGCTGAGGG
rs7110094	Nesfatin	ACGTTGGATGTACAGATCTCAGGGATCTGC	ACGTTGGATGTGCTCAGTCAGTGGCAGAAG	CTGGCACCAGCTGGAA
rs8192690	Nesfatin	ACGTTGGATGCCTTGAGTTCGATAAGCCAG	ACGTTGGATGACTTGTCTGCACGGACGAAG	GCGAAGGAGGCGAAGA
rs11826763	Nesfatin	ACGTTGGATGCAGTTGTGTCAGGAAATGGG	ACGTTGGATGGTGAAGGAACACTTAACACC	ACCTTATACACTGTTGGG
rs2634462	Nesfatin	ACGTTGGATGCTCTTCTCCAATCTCTAGAC	ACGTTGGATGTGATTGTTTCAGATGGAAGG	TCAGATGGAAGGAGACTA
rs10832767	Nesfatin	ACGTTGGATGAACCTGCTAGAGAAGTCCAC	ACGTTGGATGCACCTTAGTATCCCAATTCC	GGAGTTGAGGCACCAGTGG
rs214106	Nesfatin	ACGTTGGATGTTGCATTGTTGGTGCTTGTC	ACGTTGGATGCAGCTGCTTTGTTTCTGTGC	ATGCCAAATCATAGCTTAAT
rs10832749	Nesfatin	ACGTTGGATGCCCCAACCTGAGAAAATGAG	ACGTTGGATGGTCAGCATATTACCAACCTG	GGGCAACCTGGGAAACCTTT
rs10832782	Nesfatin	ACGTTGGATGCTGCTGACTTAGCTCTTTGG	ACGTTGGATGATTGCAGAGGGCATCAGAAG	GAAGCGGGTAGATGAAATAGC
rs12365375	Nesfatin	ACGTTGGATGTGATGACTACTATCTGTGAC	ACGTTGGATGCGATTTAACCACAGTTTACC	TAACCACAGTTTACCATCTAAA
rs10832756	Nesfatin	ACGTTGGATGTTGACTTCACAGAGTCATTG	ACGTTGGATGGTTTACTATGTACCAAGCAG	CCAAGCAGTATGCTAAACTTTT
rs214935	Nesfatin	ACGTTGGATGTGTAGTCTACCACAGAAGGG	ACGTTGGATGGGAGGTCATGAAGAAAGAGG	AAGGTCCTTGGAATGTTCTTAG
rs7108861	Nesfatin	ACGTTGGATGGAGGTTACCTTTGGTTATGAG	ACGTTGGATGACAGGAAAATACACAGTAGG	GTTGGAAAAATACATATGGAAAG
rs7937091	Nesfatin	ACGTTGGATGTCAGTGTTACCTGAAGGTCC	ACGTTGGATGAAAGTCTGGAAGTCCTGGTG	TCAGTCCTGGTGCCAACATTCTGA
rs2302510	Nesfatin	ACGTTGGATGCCGAACAGTAAGTGTGTACC	ACGTTGGATGGTTCATCACCCCATTCTAAC	TCTAACTATGAAATCATGAAAAATG

rs3741203	Nesfatin	ACGTTGGATGTTTGGGTAGTCTTTCCTCTG	ACGTTGGATGTTGAAGTTACTCCCCTTCCC	GCAATAATAATGGCATTATGAAAAAA
rs6486364	Nesfatin	ACGTTGGATGAGTGATTATCCTGCCTCAGC	ACGTTGGATGTGGAGAAACCCCATTTCTAC	CCCATTTCTACTAAAAATACAAAATTA
rs12292418	Nesfatin	ACGTTGGATGGTTTCTGAAGCTCATCTGCC	ACGTTGGATGGCAACAGTTCTTCACAGAGG	ATGAAAATATTATTGCTTTACAAGAAAA
rs3115758	Apelin	ACGTTGGATGGTCTCCTTAGGACTGAAGGG	ACGTTGGATGTCCAGAGAAGCAGACCAATC	GAGGAGACATAACCGC
rs3761581	Apelin	ACGTTGGATGATCAGCTCCTCCCCACTGTT	ACGTTGGATGTAGAATGAGGACAGCTCCAG	AGGGAACAAGAAAGGG
rs5975126	Apelin	ACGTTGGATGCAAGCAAAAGGGGAGAAAGC	ACGTTGGATGAGCAGGAATAGCACCCTCTC	AGTAGGGGCCATCACCAG
rs2281068	Apelin	ACGTTGGATGTGGTTCTCTTACATCCTGGC	ACGTTGGATGTAGCGGCAGGGAGACTAAAG	GAGACTAAAGTGAAGCATG
rs3131264	Apelin	ACGTTGGATGCCCAATAGTATCCAGTTGAC	ACGTTGGATGAAAGGATGATGGGCAGAGAC	GCCACCAGGGTCTATGGAAC
rs3115759	Apelin	ACGTTGGATGTTCATTTGGGCCTTCCTTCC	ACGTTGGATGGAAAACAAGGGATCTGCTGG	GGGAGCCCACAGAAGGGAGCA
rs2235307	Apelin	ACGTTGGATGACAAGGCTCACCCTGATCAC	ACGTTGGATGCTTGGTTCACTTCACAAGGC	CTTATCAGCATCATTCAGATCC
rs2235310	Apelin	ACGTTGGATGAGGAGTGTACGTGGTGGTAG	ACGTTGGATGGTCTGTCCTGACTTGAAATG	TGTCCTGACTTGAAATGTCCACCCCCG
rs909657	Apelin	ACGTTGGATGGACATGCTGCTGAAGGTGAC	ACGTTGGATGTGAGACCAGATTTAGAGACC	CATGAATATTTGAGAATGAATAAATGA
rs2235308	Apelin	ACGTTGGATGCTAGGTCTGCAGAGAAGAAC	ACGTTGGATGACGCAATGGGAGGGATGCAA	GGGATGCAAGAAGCC
rs2235306	Apelin	ACGTTGGATGTTCCTGGGTAGATGGCAATG	ACGTTGGATGAATAACAGTCTCTCTCCCCC	CTGCACACCATCTGAT
rs2235312	Apelin	ACGTTGGATGTTTGCTCCCTCTGTTACCTG	ACGTTGGATGAGACCTGAACACGAAGTGAG	ACCACAGTAAGAAGTGGG
rs41334247	Apelin	ACGTTGGATGGTTTGATGGATGTGTCTGGC	ACGTTGGATGCTCTGTTCAGTGGACAGTGC	TGGACAGTGCTAAAGCGGAG
rs3810622	Adiponutrin	ACGTTGGATGCTCTGCTCCAGAAAAACCAG	ACGTTGGATGAGCTACATCCTGCCATCTTC	CCCTCCTGGAGCCAC

rs4823104	Adiponutrin	ACGTTGGATGCCGTACATTTGAATCTTGTCC	ACGTTGGATGAGAAACAAACCCTCCGTCTG	ACACAGGTACCCACA
rs2076212	Adiponutrin	ACGTTGGATGCAGACACCAGAACGTTTTCC	ACGTTGGATGAATGTCCACCAGCTCATCTC	TCATCTCCGGCAAAATA
rs12483959	Adiponutrin	ACGTTGGATGGCAATCCTGTATGTAGCACC	ACGTTGGATGGAGCACGTAGTCTGTAATGG	ATGGGAGGCTCCTAATG
rs35764214	Adiponutrin	ACGTTGGATGTGTCGTACTCCCCATAGAAG	ACGTTGGATGATGGAGGAGTGAGTGACAAC	AATGCCAAAACAACCATCA
rs139051	Adiponutrin	ACGTTGGATGTTTGTTGCCCTGCTCACTTG	ACGTTGGATGCCTGTGAAAGCAAAGGAGAG	GCAAAGGAGAGAGAAGTTA
rs2294917	Adiponutrin	ACGTTGGATGCCTGAAACACATGGAGGAAG	ACGTTGGATGAATGTGGGTCCACCGTAGC	CCCGTAGCTCAGACTGCACAC
rs738410	Adiponutrin	ACGTTGGATGCTAAACAGCCAGCACTTGTC	ACGTTGGATGTGCTACAGAGCAGTCAGCAG	AGAGCAGTCAGCAGGCACCAC
rs738408	Adiponutrin	ACGTTGGATGTCTCCTTTGCTTTCACAGGC	ACGTTGGATGCTGAAGGAAGGAGGGATAAG	TAGGGATAAGGCCACTGTAGAA
rs9625964	Adiponutrin	ACGTTGGATGTGTGACCTCAGGCAAGTTAC	ACGTTGGATGTGAGGTCGGTTTTCCTATCC	TCGGTTTTCCTATCCATATATTG
rs2294916	Adiponutrin	ACGTTGGATGACCTTTCCCTGCTTATCTGG	ACGTTGGATGAAGACTTCCCGTTGGATGAG	CCCGTTGGATGAGTCTCTTTGAG
rs738407	Adiponutrin	ACGTTGGATGGGAGCGCTTATGAAAGCATC	ACGTTGGATGATTTGCCTGTATCCTCAGGG	GGGTCTAGCAGAGAGAAGATAAT
rs9625961	Adiponutrin	ACGTTGGATGCCTAGTAGCTGGGACTATGG	ACGTTGGATGTGATGAAACCCAGTCTCTAC	ACACAGTCTCTACAAAAATTACAGA
rs16991187	Adiponutrin	ACGTTGGATGATAGCAAGAGAGGCCATACC	ACGTTGGATGTTTCCGTGAGCTCCTGAGAG	GCCGTGAGCTCCTGAGAGCCCGTAG
rs1977081	Adiponutrin	ACGTTGGATGAGAGTCTTGCTCTGTTGCCC	ACGTTGGATGGCAGGAAAATGGCTTGAACC	AGGAAAATGGCTTGAACCTCGGAAG
rs34179073	Adiponutrin	ACGTTGGATGTTAGGGCAGATGTCGTACTC	ACGTTGGATGTGAGTGACAACGTACCCTTC	CACCGTGTCCCCCTT
rs16991170	Adiponutrin	ACGTTGGATGGCAAACATGAGTCTGGATTC	ACGTTGGATGAGGTGGTCTAGCAGCTCATC	CCAAGGCAGCCGACTC
rs35726887	Adiponutrin	ACGTTGGATGATCACCAAGCTCAGTCTACG	ACGTTGGATGACCAACTCACCTTGAGATCC	AGGTAGAGGTTCCCTG

rs2294918	Adiponutrin	ACGTTGGATGCTCTGCTTTGGTCTCTGCTG	ACGTTGGATGATGCACACCTGAGCAGGAC	TGGACTCCCTGCTCCCCC
rs2294919	Adiponutrin	ACGTTGGATGCACTACACAGCAATGCGGAG	ACGTTGGATGTCTGTGAGTCACTTGAGGAG	AAGTTTCCCATCTTTGTG
rs734561	Adiponutrin	ACGTTGGATGATGTGGGAACAGACAAGTGC	ACGTTGGATGAACCTGCAGCCAGTTGCATC	CCTCTGTTTAACCTTGTTTG
rs139047	Adiponutrin	ACGTTGGATGCTTTCGGTCCAAAGACGAAG	ACGTTGGATGCCATCCCAAAACAGCTTTTC	AAACAGCTTTTCTAACTTGA
rs1883350	Adiponutrin	ACGTTGGATGAGTTTCACAGATGCAGGCTC	ACGTTGGATGACCACCTTATTGCCCATAGC	GGGCAGGCAGCTTTTGCTACA
rs6006460	Adiponutrin	ACGTTGGATGCAGCAGGTACTTTATTGCCC	ACGTTGGATGAGCAGAGACCAAAGCAGAGG	CCGCGGTCCATCCTCAGGTCCA
rs9626056	Adiponutrin	ACGTTGGATGTGCCACTGGCATCTGATGTA	ACGTTGGATGGTCAGGAAGATGCCTATGAG	AACGGGTGATGACACAGCATCTC
rs2394443	SIRT1 promoter	ACGTTGGATGTTAAACCCCATCACGTGACC	ACGTTGGATGACCCGTAGTGTTGTGGTCTG	CCGCCCCCGCCCTCT
rs932658	SIRT1 promoter	ACGTTGGATGTGTTGCGTCTACCGCTCCG	ACGTTGGATGGAATTTGGCTGCACTACACG	GCGGCCGGCGGCCCT
rs12778366	SIRT1 promoter	ACGTTGGATGCTAAGGTCCTATCTACATCC	ACGTTGGATGTAAGGCTTCTAGGACTGGAG	TCATCTGGTCACCACT
rs3758391	SIRT1 promoter	ACGTTGGATGGCCATAACAAACACTGGCTC	ACGTTGGATGGCACACTGTGACTCCATATC	CTGGCTCTAGATCTACCA
rs12250285	SIRT1 promoter	ACGTTGGATGCATCATACTGGTCAGGCTGG	ACGTTGGATGAAATCCCAGCACTTTGGGAG	TCCTGACATCAGGTGATC
rs10740280	SIRT1 promoter	ACGTTGGATGAGCAAGAGCGAAACTCCGTC	ACGTTGGATGGGTTTGCTGCACCAATCAAC	GAAACTCCGTCTCAAAAAAA

A3. Sequencing PCR primers

Gene	Region	PCR primer 1	PCR primer 2	Annealing temperature (ºC)	Magnesium Concentration (mM)	Q solution
SIRT1	Exon 1a	CGGAGAGATGGTCCCGGCCT	TAGGAGCCCGGGGAGAGGGA	50	1.5	Yes
SIRT1	Exon 1b	AGTGTTGTGGTCTGGCCCGC	AGGCCGGGACCATCTCTCCG	50	3	Yes
SIRT1	Exon 2	CCCCATAGGCATGCTTTACA	AGTGAGGTGGGGCAAGGT	55	3	No
SIRT1	Exon 3	TACAGCTGAAGGGGGTCTTC	TGCACTCAGTTTTTGGCAGT	65	3	No
SIRT1	Exon 4	TTTCTGAAATAAGGGTAGGGTTG	CTTACTCCGCCACAGTAGCA	60	3	No
SIRT1	Exon 5	TAGGTGTGTGTCGCATCCAT	TGTCAAGGCCAATACTTCTGTTT	60	3	No
SIRT1	Exon 6	TCTGCAGTGTGTTCTGAGGTTT	TTTTACCAAAATTCCCAAATGC	60	3	No
SIRT1	Exon 7	TCCATAGCTTTTCTGTTTTGTTTTT	GGGCAACAAGAGGGAAACTC	60	3	No
SIRT1	Exon 8	AAGTATTTAGTGCATGGGTCTTTTT	GATTTACCTTTCTGATTACGAACTG	60	3	No
SIRT1	Exon 9a	CATGCCTCCCAGTTGTTTTT	CCGATGGCTTTTTGAAAACT	60	3	No
SIRT1	Exon 9b	TTTACTTGTGAACTCGATAGAGCA	CCAAAGCATTTATTTGTTTTTCA	60	3	No
SIRT1	Exon 9c	GGAGAGCACTCGGTTGTCTT	TGGTCCTAGCTGGGTGTTTT	60	3	No
SIRT1	Exon 9d	GCCCTAGTATTATGGAGATGAACA	GATCAGCACCATCAGGGTTT	60	3	No
SIRT1	Promoter 1	CGAATTTGGCTGCACTACAC	CCATCTTCCAACTGCCTCTC	50	1.5	Yes

SIRT1	Promoter 2	GGTGTGAGGAGAGTGGGAAA	CGTGTAGTGCAGCCAAATTC	55	1.5	Yes
SIRT1	Promoter 3	GAAACGGCTAGATAGCTCACG	AAATCACTACCGCCGGAAC	60	3	Yes
SIRT1	Promoter 4	CCTTAGCGATATATTTCCAGCTTT	CAGCTTGGCTTCAAACGTG	60	3	No
SIRT1	Promoter 5	AGGCTGGTGTGCAATGGT	GAGCACAGCGTTTCTATCTGTTT	60	3	No
SIRT1	Promoter 6	TTGGGATACTGACTCTCAACATT	GGACAGATCACCTGAAGTCG	60	3	No
SIRT1	Promoter 7	GCCACCACTCTGGGCTAA	CACTCCTCCCAATTCCTGAC	60	3	No
SIRT1	Promoter 8	TCCTGGGTTCACGCCATT	TGCAGTATGCTTTTCAGGTAGG	60	3	No
SIRT1	Promoter 9	ACAGGAATGTGCCACCACAC	ATAAATTGGCCTGCCAAGCA	60	3	No
SIRT1	Promoter 10	CCTTCCACCTCAGCTTCCTA	CCCTACTTCACACAATATACGAAA	60	3	No
SIRT1	Promoter 11	ACTACAGGCGCCCACCAC	GGATCAAGAAAGAACTTTTTAGCAA	55	3	No
Apelin	Exon 1	GGAGCGGTTGTAGGTTGTC	GATGCCAAAGGCCGAGTTGA	60	1	Yes
Apelin	Exon 2	GTAGGTTCCGTTTGTCCCAG	CAGACGAGGAGATCGTCTGT	60	3	No
Apelin	Promoter 1	TGCCTCCTGGGGGATGCAGG	CCGCGAGCCCACTTGGTCAT	60	3	No
Apelin	Promoter 2	GAGGGAAGCAGCCCCATGCC	CCTGGTCCTTGCCCTTGCCG	60	3	No
Apelin	Promoter 3	TGGCACACACGCACCCTGTC	TTGTTCCCTGGAGCTGTCCTCA	60	3	No
Apelin	Promoter 4	AGGGGTGCCATGCTTCACATGC	TCAGCAAGCCAGGCAGTTTGC	60	3	No

A4. Primer Sequences Used in DNA Methylation Study

Gene	PCR primer 1	PCR Primer 2	Sequencing Primer
POMC	TTTAAAGTGGAATAGAGAGAATATGA	CACAAAAAACAACTCCCC	TTATAGTTTTTAAATAATGGGGAAA
POMC	GTAGTTGAGTTGGAGGGTTTA	CAACCTAACCAACATAAAAAAA	TCACCCCTTTACTCTCCAACA
Leptin	AAATTTTTGGGAGGTATTTAAG	AAAAAAACCAACAAAAAAAAA	TTTTTGGGAGGTATTTAAGG
Leptin	TTGTTGGTTTTTTTTGGTAGG	ATACCCAAAACAATTTCCAATAC	ATACCCAAAACAATTTCCAA

A 6. R scripts

Data was read from a tab-delimited text file with column titles indicating the names of the variables. The phrase “file name” indicates that the full path and file name should be inserted e.g “C:/Documents and Settings/sjclark/Desktop/FILENAME.txt"”.

A 6.1. Correction of transcript data for covariate effects using linear regression

```
Table<-  read.table("file name", header=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)

linearRegr<- lm(Transcript~age+bmi+sex, data=Table)

residuals<-residuals(linearRegr)

write.table(residuals, file="file name")
```

A 6.2 Correction of BMI or transcript data for effects of relatedness using clustering

```
Table<-  read.table("file name", header=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)

library(Ecdat)
 LSDV <- lm(BMI ~ Family, data=Table)
 gcenter <- function(df1,group) {
variables <- paste(
```

```
 rep("C", ncol(df1)), colnames(df1), sep=".")
 copydf <- df1
 for (i in 1:ncol(df1)) {
 copydf[,i] <- df1[,i] - ave(df1[,i], group,FUN=mean)}
 colnames(copydf) <- variables
 return(cbind(df1,copydf))}
centerTable <- gcenter(Table[,1:65], Table$Family)
fmlm <- lm(C.BMI ~ C.Family, data=centerTable)
library(sandwich)
 M <- length(unique(Table$Family))
 dfcw <- fmlm$df / (fmlm$df - (M -1))
 library(lmtest)
 coeftest(fmlm, dfcw*vcov(fmlm))
clx <-
function(fm, dfcw, cluster){
 library(sandwich)
 library(lmtest)
 M <- length(unique(cluster))
 N <- length(cluster)
 dfc <- (M/(M-1))*((N-1)/(N-fm$rank))
 u <- apply(estfun(fm),2,
 function(x) tapply(x, cluster, sum))
 vcovCL <- dfc*sandwich(fm, meat=crossprod(u)/N)*dfcw
 coeftest(fm, vcovCL) }
 clx(fmlm, dfcw, Table$Family)
residuals<-residuals(clx)

write.table(residuals, file="file name")
```

A 7. Statistical power calculations for locating SNPs in LD with associated variants

The following calculates the probability of finding a SNP (assuming such a SNP exists) with certain MAF that is in LD with an associated SNP (with MAF =0.13, as was observed in the controls for rs11599176, the most significantly associated SNP in the *SIRT1* gene) when sequencing 10 samples that are homozygous for the minor variant of the associated SNP.

$$r^2 = \frac{H_q^2}{A_p . A_q . B_p . B_q}$$

Where r^2 is the correlation coefficient of LD between two markers A and B (A being the associated SNP and B the marker that is being discovered), H_q is the frequency of the haplotype, H containing the minor alleles of markers A and B, A_q and B_q are the minor allele frequencies of markers A and B and A_p and B_p are the common allele frequencies of markers A and B.

By inputting an r^2 value of 0.8, an A_q of 0.13 and a B_q of 0.05, the value of H_q is determined:

$$0.8 = \frac{H_q^2}{0.87 . 0.13 . 0.95 . 0.05}$$

$$0.00425 = H_q^2 - 0.0130\, H_q$$

$$0.00425 = (H_q - 0.00650)^2 - 0.00650^2$$

$$\sqrt{0.00425 + 0.0065^2} = H_q - 0.00650$$

$$0.072 = H_q$$

Thus the probability of a sample having the haplotye, H, given that the sample has the minor allele of the associated SNP is:

$$\frac{0.072}{0.13}=0.55$$

And the probability of at least one in 20 samples (10 diploid samples) containing the haplotype given that they each contain the minor allele of the associated SNP:

$$1-(1-0.55)^{20}=0.99999$$

Thus there is a probability of >99% of the minor allele of marker B being present in at least one sample out of ten samples that are homozygous for the minor allele of marker A if the MAF of marker A = 0.13, the MAF of marker B =0.05 and LD (r^2) between the two SNPs is 0.8..

Calculating the probability using r^2 = 0.5 and B_q = 0.01:

$$0.5=\frac{H_q^2}{0.87.0.13.0.99.0.01}$$

$$0.0224=H_q$$

$$\frac{0.0224}{0.13}=0.172$$

$$1-(1-0.172)^{20}=0.977$$

Thus there is a probability of 97% of the minor allele of marker B being present in at least one sample out of ten samples that are homozygous for the minor allele of marker A if the MAF of marker A = 0.13, the MAF of marker B =0.01 and LD (r^2) between the two SNPs is 0.5..

Calculating the probability using r^2 = 0.5 and B_q = 0.001:

$$0.5=\frac{H_q^2}{0.87\,.\,0.13\,.\,0.999\,.\,0.001}$$

$$0.0076=H_q$$

$$\frac{0.0076}{0.13}=0.0588$$

$$1-(1-0.0588)^{20}=0.702$$

Thus there is a probability of 70% of the minor allele of marker B being present in at least one sample out of ten samples that are homozygous for the minor allele of marker A if the MAF of marker A = 0.13, the MAF of marker B =0.001 and LD (r^2) between the two SNPs is 0.5.