THE UNIVERSITY OF CHICAGO

INSIGHTS INTO GENE REGULATION AND DISEASE AT OBESITY GWAS LOCI

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON MOLECULAR METABOLISM AND NUTRITION

BY

AMELIA CHRISTINE JOSLIN

CHICAGO, ILLINOIS

DECEMBER 2020

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

LIST OF TABLES

ACKNOWLEDGEMENTS

My interest in science came from an innate desire to better understand life’s mysteries, and I feel fortunate that I was born during a period of human history where a woman is able to sate this desire through investigative research. Several scientists I met during my undergraduate career were kind enough to take me under their wing and provide me with encouragement to pursue this path. I want to thank Dr. Matthew Lange, Dr. Ted Powers, Dr. Ariadne Vlahakis, Dr. Brad Niles, and Dr. Ralph Green for this gift. It was a unique and humbling experience to work and learn from each of you, and these experiences provided me with a solid knowledge platform which I was able to build upon during my graduate career.

I additionally want to thank my truly amazing committee members. Particularly Dr. Matthew Brady, whose office door is always open. He is an invaluable member of the metabolism community and I will always be grateful for his advice. I also want to thank the other members of my committee, Dr. Xin He and Dr. Anna Di Rienzo, for their kind and helpful guidance during this process.

I also want to thank the friends I had the pleasure to meet here in Chicago, especially Sahar, Ittai, Lindsey, Tiffany, Jessica, Hailey, and Debora who provided a much needed source of light during the hardest periods of graduate school, and for each being unique, strong, independent, intelligent people who I admire.

But, nobody deserves more credit than my family. My mother was critical for fostering early independence and drive, and encouraging me at a young age to make something of myself and to appreciate this opportunity to forge my own path. I want to thank my father for teaching me how to bask in the wonder of life’s minutiae, and I want to thank my sister for being a model of clarity and strength. I want to thank Andrew for his support and love, it would have been much harder without you. Finally, I want to thank Darth Vader and Meatball for the endless squishes.

The work in this thesis, which I am proud to share, evolved from an initial idea from Dr. Marcelo Nobrega and Dr. Debora Sobreira. It really was a dream of mine to work on this exact problem, and I was very lucky to have the opportunity to learn from Dr. Nobrega and the outstanding group of people in his lab. Specifically I want to acknowledge Marcelo, Debora and Grace for their contributions to this project, plus Ivy for her critical help during my early training. I also want to thank Lindsey for our philosophical discussions on life and gene regulation. I am incredibly grateful for the opportunity to be in this lab, and for the investment it took from Dr. Nobrega to get (and keep) this project off the ground.

DISSERTATION ABSTRACT

While genome-wide association studies (GWAS) have identified variants and genes associated with human disease, a comprehensive understanding of the genetic architecture of individual loci and the functional implications of these associations remains incomplete. In this work, we applied an integrated pipeline to chart the regulatory landscapes of obesity-associated loci within two cell types central to obesity etiology. In both adipocytes and hypothalamic neurons, we annotated gene expression, chromatin accessibility, and long-range chromatin interactions across multiple differentiation stages. Additionally, we generated a list of 2,396 variants in high LD with BMI lead SNPs and tested them in a massively parallel reporter assay to identify putatively causal variants modulating enhancer activity. We identified 94 variants within enhancers that displayed enhancer-modulating properties, many of which were active in both cell types. Our data show that individual GWAS loci harbor multiple candidate causal variants within distinct enhancers that display cross-tissue effects. Integrating the identified enhancer modulating variants (EMVars) with chromatin interactions and eQTL information generated a comprehensive list of genes predicted to underlie obesity GWAS associations. Aggregating our data across multiple time points allowed us to assign more candidate causal variants to genes compared to regulatory maps in a single cell type and to prioritize 232 genes with varying degrees of evidence for obesity risk importance. We used these insights during experimental dissection of a complex genomic interval on 16p11.2 where we observed EMVars at two independent GWAS loci exhibiting megabase-range, cross-locus Hi-C chromatin interactions and shared eQTL effects. We provide evidence that EMVars within these two loci converge to regulate a shared gene set. Together, our data chart the genetic architecture of obesity-associated

loci and support a model in which many GWAS loci contain multiple variants that impair the activities of distinct enhancers across tissues, potentially with temporally restricted effects, to impact the expression of multiple genes. This complex network model has broad implications for ongoing variant to function efforts to mechanistically dissect GWAS.

CHAPTER 1: INTRODUCTION

1.1 The allure of the non-coding genome

In 1988 the Human Genome Project (HGP) was brought forth as an attempt to attain a complete map of the base pair sequence encoding human life. Completed in 2003, the HGP was the first large scale endeavor to understand how the nearly 3.2 billion base pair sequence of repeating four nucleotides - A, T, C and G - in the human genome leads to precise control of developmental and organismal functions[1]. What emerged from this effort was both the first human genome sequence, as well as a realization of a critical knowledge gap in our understanding of genome function; e.g., using this code how does a cell know which combination of genes to express, and to what level? This question still remains, and a continued investigation is critical for our understanding of gene expression patterns, which determine not only which proteins are produced, but also the overarching functions of each cell.

The first clue to emerge was the discovery that protein-coding genes comprise only 2% of genomic sequence. The remaining 98% was even colloquially termed “junk DNA”, as its function remained a mystery[2]. But in 2012 the Encyclopedia of DNA Elements (ENCODE) project suggested that up to 80.4% of the genome is functional, and provided evidence through large scale investigation of 147 different cell types that so-called “junk DNA” is riddled with elements capable of regulating the spatiotemporal expression of genes[3]. These gene regulatory elements have thus become a research area of great interest, and deep mechanistic investigation into how these regulatory elements act, and how they are misregulated in disease, will provide great insight into our understanding of gene expression regulation and genome biology.

1.2 Gene expression control via cis-regulatory elements

Defined as enhancers, repressors, and insulators – gene regulatory elements are now known to coordinate amongst each other to orchestrate precise gene expression from development through adulthood. These elements usually contain docking sites (motifs) for transcription factor (TF) proteins, which together allow for cooperative binding of multiple transcription factors to DNA at these non-coding locations (Figure 1.1). In conjunction with defined TF expression within embryonic domains, this combinatorial binding allows for precise control of gene expression[4–6]. These TFs act by recruiting or blocking the actions of RNA polymerase II, the protein responsible for producing mRNA from DNA, at a gene's promoter to modulate its transcription level. On average, a gene's expression can be regulated by up to dozens of enhancers that fine-tune its expression across cell types and developmental stages[3,7].

Transcription factor binding to DNA can be measured directly through chromatin immunoprecipitation (ChIP-seq) or indirectly through chromatin compaction assays. The level of local chromatin compaction, where functional genomic elements can be turned on or off based on loosening (activation) or compaction (inactivation) of the genomic region due to recruitment of activators or repressors, alters regulatory element accessibility and thus availability for TF binding. Accessibility measures have traditionally relied on assays of chromatin compaction such as DNase I hypersensitive sites sequencing (DNase-seq), Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE-seq), or more recently, Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)[8]. ATAC-seq relies on a Tn5 transposase which inserts sequencing adapters into nucleosome free regions of DNA for PCR amplification and quantification of open chromatin via next generation sequencing[8] (Figure 1.1). This measurement allows for identification of accessible genomic regions, thus suggesting the location and activity status of gene regulatory elements under defined cellular states, perturbations, and/or developmental time-points.

A critical biological insight recently uncovered about regulatory elements is their ability to act locally, as well as over great linear distances (typically up to 1Mb), to alter a cognate gene's expression through physical interaction and stabilization mediated by proteins such as CCCTC-binding factor (CTCF) and cohesin[9]. This allows regulatory elements to modulate the expression of local or distal genes, which makes predicting the gene(s) that they regulate challenging. Fortunately the 3D organization of genes and their regulatory elements in the genome can now be measured thanks to a groundbreaking suite of "C" technologies, which provide a snapshot of DNA-DNA interactions within the nucleus and can also be used to suggest the location of a gene's regulatory elements within the vast search space of the non-coding genome[10] (Figure 1.1). A particularly sensitive "C" technology, termed *in situ* promoter capture Hi-C (cHi-C), has become popular in the last several years because it allows for targeted enrichment and sequencing of promoter interacting genomic regions[11–13]. Knowledge of promoter interacting regions makes detection of regulatory elements located far away from their cognate gene feasible, although the proportion of promoter interacting regions that are regulatory, versus structural or artifact, requires further investigation.

1.3 Current methods for enhancer identification

Out of all regulatory elements, enhancers are arguably the best understood. As of yet, enhancers are not easily detectable based on sequence alone, so other methods to identify their location have been implemented. Some of the earliest methods to identify these elements were through estimates of conservation across species to prioritize non-coding regions of the genome under selection, or through ChIP-seq detection of histone modifications such as H3K27ac and H3K4me1 that correlate with enhancer activity (Figure 1.1). Unfortunately, not all enhancers are conserved, and co-localization of enhancers with histone marks is imperfect, making these measures incomplete[14]. These techniques also lack direct assessment of enhancer activity. Historically, the best methods to measure enhancer activity have been mouse transgenic reporter assays for *in-vivo* validation of enhancer activity or luciferase assays for *in-vitro* validation. Luciferase assays and mouse transgenic assays provide reliable measurements of enhancer potential but are low-throughput with limited sensitivity and require individual cloning of tested regions. Recently, a class of high throughput enhancer assays has been developed, which includes Self-transcribing Active Regulatory Region sequencing (STARR-seq)[15], Massively Parallel Reporter Assay (MPRA)[16], High-resolution Dissection of Regulatory Activity (HiDRA)[17], and others, which can test thousands of regions at a time and take advantage of barcoding technology to gain very precise estimations of enhancer activity (Figure 1.1). Although these assays allow for unprecedented numbers of enhancer activity measurements, they suffer from being episomal, meaning the DNA fragments are tested outside their native genomic context, and only short DNA sequences can be tested. Thus, confirmation of predictions with CRISPR-cas9 editing, where you can remove these putative enhancers out of the genome and measure changes in gene expression, is the current gold standard to validate enhancer predictions[18]. All together, multiple data types should be orthogonally integrated and carefully interpreted when evaluating whether or not a region of the genome is an enhancer.

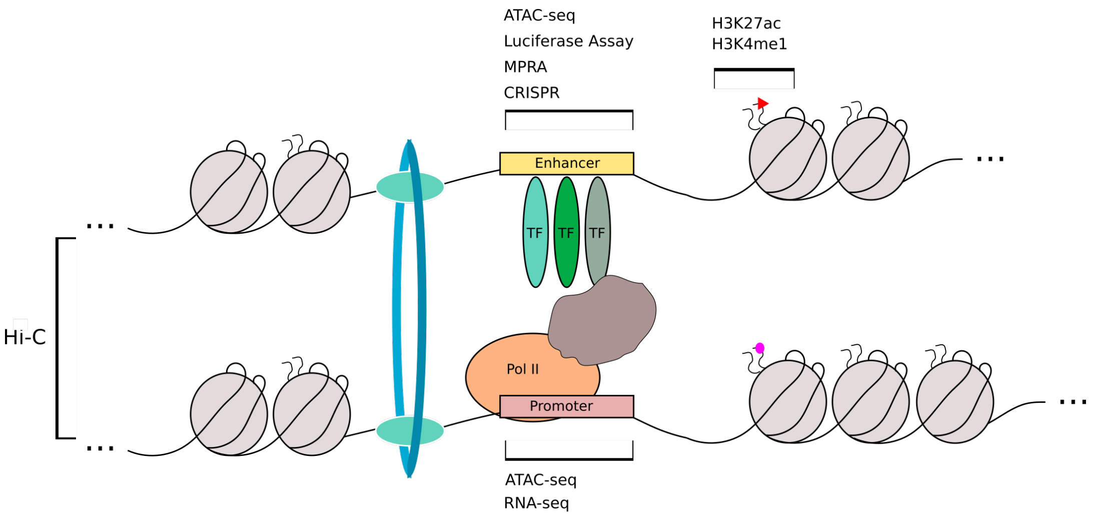


Figure 1.1 Overview of applied methods
The genome is riddled with regulatory elements. Methodologies have arisen in order to identify enhancers and connect them to their cognate promoters within the vast search space of the non-coding genome. To identify enhancers, we used a combination of ATAC-seq, luciferase assays, MPRAs or colocalization with posttranslationally applied marks on histone tails such as H3K27ac or H3K4me1. Gene expression levels were determined via RNA-seq. Promoter accessibility for transcription factor (TF) and Pol II binding, which can also be observed with ATAC-seq. The proximity of enhancers and promoters in 3D nuclear space, sometimes brought together by CTCF/cohesin complexes (blue ring), was measured with a "C" technology termed in situ promoter capture Hi-C.

1.4 Understanding disease risk through Genome Wide Association Studies

For cellular homeostasis to persist it is very important for the orchestration between gene regulatory elements and their cognate genes to be maintained. Gene expression levels naturally differ between individuals, and this diversity is attributed to both lifetime environmental exposures and genetic variants. Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation present between people, as over 100 million SNPs have been identified to date. For over a decade, geneticists have used a method called Genome Wide Association Studies (GWAS) to identify SNPs that alter risk for common diseases. GWAS relies on genotyping a large number of individuals with and without the disease or disease modulating phenotype of interest, and determines whether particular alleles of each SNP are more likely to be present in the individuals with or without disease, thus associating alleles to disease state. The results of these studies have lead to an explosion of interest in the fields of medical genetics and genomics, and have been instrumental in our understanding of the genetic basis of many traits.

Looking back to the first 5 years of GWAS, the results were initially difficult to interpret. Researchers quickly noticed that the vast majority of GWAS disease associations mapped to the non-coding genome far away from nearby protein coding genes. Up until this point, disease had been understood primarily through the lens of Mendelian traits such as Sickle-cell anemia, Hemophilia A, or Phenylketonuria (PKU), where rare mutations within single genes cause disease. It was thought that these findings could be extrapolated to common diseases such as obesity, cardiovascular disease, or type 2 diabetes (T2D) where we would theoretically observe common coding variants with large effect sizes explaining the majority of trait heritability. Examples of a common disease with large effect variants that motivated these assumptions include coding variants for the gene *MTHFR* (Methylenetetrahydrofolate Reductase), where the 677C>T substitution creates a thermolabile and less active form of this enzyme leading to a 10 fold increased risk for homozygous TT individuals to develop hyperhomocysteinemia[19,20]. But upon implementation of GWAS it was discovered that the genetic architecture of complex common disease is very different from the *MTHFR* example or Mendelian traits. The first wave of GWAS identified very few variants that explained only a small portion of the trait heritability predicted by twin studies. This concerning performance of GWAS compared to twin studies even lead researchers to ponder where heritability was hiding in the genome, or why it was "missing"[21].

Today, thanks to larger and more powerfully designed GWAS, that question can now be answered with confidence. Hundreds of regions containing SNPs with very small effect sizes contribute to the heritability of traits such as obesity and T2D, meaning that many genes contribute to risk for these traits, and each explain a small proportion of total disease risk and penetrance. To date, over 90% of all complex trait disease associations map to the non-coding genome[22]. These association regions are highly enriched for regulatory elements such as enhancers, suggesting that genetic variants may act to modulate regulatory element activity directly through altering binding affinity of transcription factors and affecting expression of genes important for disease. It is now more important than ever to focus scientific effort on learning about these noncoding variants and how small single nucleotide polymorphisms in regulatory elements can modulate gene expression and thus contribute to disease risk.

1.5 Limitations of GWAS

Although GWAS have been instrumental for our understanding of the genetic basis of complex traits, to date very few loci have been mechanistically dissected deep enough to provide a confident understanding of how variants in these regions lead to disease. This shortage of mechanistic understanding is primarily due to limitations in GWAS design that create challenging experimental hurdles that must be crossed prior to interpretation.

First, when attempting to interpret the results from any GWAS locus, there are usually many SNPs in each association region that reach genome wide significance. This is because the genome is inherited in blocks, where many SNPs are oftentimes nonrandomly linked together due to lack of recombination in that region during meiosis. This phenomenon is known as linkage disequilibrium or “LD”. When performing a genetic association study where individuals must be genotyped, genomic regions are “tagged” by choosing SNPs within each haplotype block to represent that locus, thus avoiding costly whole-genome sequencing. Because of this, the SNP with the lowest *p* value identified in each GWAS locus is oftentimes not the causal variant, meaning the risk variant driving disease risk. Nearby SNPs in the LD block, due to allele frequency differences or lack of direct genotyping, can instead reach the highest statistical significance in the area[23]. Therefore, SNPs in each GWAS locus that are the most significant (lead SNPs) and nearby SNPs in high LD must all be considered for causality.

Once the causal variant has been predicted you can try to understand how that variant disturbs a gene. To do this for non-coding variation, it is important to first determine the cell type driving the association. To address this, epigenetic enrichment analyses are sometimes performed to determine the cell type(s) where there are active regulatory elements in the region of interest. Although there are many mechanisms by which a non-coding region can contribute to disease risk, modulation of regulatory elements, specifically enhancers, is predicted to be the most common[24–26]. It is then up to the researcher to prove that the SNP affects the regulatory element, potentially by providing evidence that this variant has the capability of affecting enhancer activity in a disease relevant tissue.

Once the likely causal variant has been identified, the target gene affected by the non-coding SNP must be determined. As previously discussed, enhancers do not necessarily regulate the closest gene, so if a causal SNP is in an enhancer, the gene regulated by the enhancer must be identified. Using a set of “gold standard” genes with missense mutations that lead to the same trait, it has been estimated that at least 50% of the causal GWAS target genes are not the closest gene[27,28]. Technologies such as promoter capture Hi-C have become essential tools to help researchers predict target genes within these regions. Integration with expression quantitative trait loci (eQTL) mapping information has also been used to varying degrees of success as well as CRISPR-cas9 editing to validate enhancer-target predictions. Finally, the implicated gene must be linked to a cellular function important for your disease in a causal manner, usually though *in-vivo* mouse models or *in-vitro* cellular studies.

All of these questions create a multi-step process of variant-to-function interpretation that has complicated our understanding of GWAS results. Studies attempting to address these questions have themselves uncovered additional complexities hiding within the genetic architecture of these loci. Because of the regulatory complexity of many genes, where each gene can have multiple enhancers, multiple variants within these regions can then affect the expression of target genes. This has lead to the hypothesis that multiple causal variants may exist within GWAS loci. Additionally, deep dissections of certain loci have taught us that multiple genes may also be affected by these variants where each independently contributes to disease[29]. In the work presented here, we aimed to apply a pipeline to address many of these experimental hurdles and outstanding questions in order to prioritize functional variants and target genes in obesity associated GWAS loci in a high-throughput manner.

1.6 The genetic basis of common obesity

Metabolic disease prevalence is on the rise, and the center for disease control and prevention (CDC) estimates that obesity now affects over 40% of Americans[30]. Defined as having a body mass index (BMI) > 30, obesity is not only a serious condition in isolation, but is also a major risk factor for other diseases such as cancer, heart disease, and type 2 diabetes. The high prevalence and presence of serious comorbidities associated with this disease makes it a significant public health threat that has yet to be addressed. Few successful treatments for obesity actually exist, and a large contributing factor is thought to be a historically inaccurate understanding of obesity etiology. Although environmental influences such as poor nutrition and sedentary lifestyle have long been the sole attributors to obesity risk, genetic factors are now known to play a key role. In support of this, heritability estimates for body mass index (BMI) range from 16-40%[31–34], demonstrating that a sizable proportion of the phenotypic variation within obesity is attributable to genetic variants.

There are several interesting theories on how genetic variants contribute to obesity risk. In 1962 the "thrifty" genotype hypothesis was put forth to suggest that certain genetic variants were positively selected on through human evolution because they confer a biological ability to store energy, and were advantageous during the many thousands of years humans did not have reliable food availability[35]. But after the dawn of the agricultural revolution as food became more widely accessible, it is thought that these variants now predisposed individuals to superfluous fat storage and obesity risk. Although this hypothesis is attractive, it is not well supported in scientific literature, as few obesity loci exhibit signatures of selection[36]. A second, more recent, hypothesis suggests genetic variants may modulate hedonic reward centers in the brain[37]. The brain is responsible for regulating hunger and satiety cues, and circulating hormones act on regions of the hypothalamus to stimulate or inhibit feeding in order to maintain energy balance. In one example, individuals with congenital deficiencies of the hormone leptin were asked to rate how much they liked images of certain foods. Upon treatment with leptin supplementation, these ratings were significantly reduced, demonstrating the capacity of genetic variants to also affect food desirability[38,39]. Thus, a thorough investigation of human genetic variation is important because it can pinpoint genes important for obesity relevant biological processes, which can lead us to identify novel therapeutic targets for this common condition.

In 2015 the genetic investigation of anthropometric traits (GIANT) consortium performed a large-scale effort to identify novel genetic loci associated with obesity. With this effort, the consortium identified 56 novel associations to BMI while confirming 41 from previous studies. This group found that highly expressed genes near BMI associated SNPs are strongly enriched for tissues in the central nervous system, particularly the hippocampus, hypothalamus and limbic systems[40]. They also suggested that common genetic variants (MAF > 5%) explain at least 21% of the heritability for obesity and found that these variants were enriched to be in brain and adipose enhancers. Specifically, hypothesized mechanisms for obesity risk include; cross-talk between adipocytes and regions of the hypothalamus controlling feeding behavior via signaling by the satiety hormone leptin or other metabolism-modulating hormones such as insulin or ghrelin[39,41–43]; regulation of thermogenesis through modulation of beta-adrenergic signaling or beiging/whitening of adipose tissue[29,44]; disruption of reward pathways in the brain leading to imbalances toward stimuli when eating[37,38]; or developmental abnormalities leading to loss of homeostatic mechanisms in either cell type[45]. These hypotheses suggest that processes within brain and adipose drive obesity, where these cell types coordinate to regulate phenotypes related to body weight maintenance.

To date, few genes have been implicated in obesity risk, and the majority of these have been through investigating instances of rare monogenic obesity. Several genes have been implicated in both conditions, including *POMC* (Pro-opiomelanocortin), *BDNF* (Brain-derived neurotropic factor), *PCSK1* (Proprotein convertase subtilisin/kexin Type 1), *LEP* (Leptin), and *MC4R* (Melanocortin 4 receptor) [41,46–49]. Each of these genes are involved in the leptin-melanocortin pathway, a key metabolic signaling pathway that controls food intake. Leptin is a hormone secreted by adipocytes that signals to leptin receptors on neurons in the arcuate nucleus of the hypothalamus. This signal blocks *AGRP* (Agouti-related protein) appetite stimulating signals and instead induces expression of *POMC* to produce the pro-opiomelanocortin pre-hormone. This POMC pre-hormone is then cleaved by PC1/3 [50], the protein produced by *PCSK1,* to make α-MSH. α-MSH can then activate *MC4R* signaling in secondary neurons of the paraventricular nucleus to induce satiety (reviewed in [51]). *BDNF* has found to be regulated downstream of *MC4R* signaling and is thought to play a role in *MC4R*'s regulation of energy balance and neuronal development[52]. *BDNF* also has an important role to play in the hippocampus, where it regulates long-term memory[53]. Humans harboring pathogenic mutations in the coding portion of these genes exhibit hyperphagia and severe, early onset obesity.

It has been proposed that non-coding associations at these regions may tag rare coding variants with large effect sizes at these locations, which is known as the synthetic association hypothesis[54]. But a careful investigation of variants at the *MC4R* locus instead supports the contribution of common non-coding variants that modulate the expression of these genes to a lesser degree, predisposing to the common form of obesity[55]. Secondarily this indicates that identifying common variation with small effect sizes does not speak to the magnitude of the biological relevance of the target gene itself.

Other novel genes such as *NEGR1* (Neuronal growth regulator 1)[56]*, MAP2K5* (Mitogen-activated protein kinase kinase 5)[45]*, ADCY3* (Adenylate cyclase 3)[57,58]*,* and *CADM2* (Cell adhesion molecule 2)[59] have been linked to obesity specifically from GWAS results with varying degrees of supporting evidence. *NEGR1*, *ADCY3,* and *CADM2* are predicted to lead to obesity risk through modulation of homeostatic processes in the hypothalamus. On the other hand, *MAP2K5*'s role in obesity risk has been predicted to be important in adipocytes, where this gene acts as a potent activator of extracellular signal-related kinase 5 (ERK5) signaling, which regulates a wide range of processes, including cell growth, proliferation, and differentiation.

Mice lacking ERK5 specifically in adipocytes had increased adiposity as well as food intake[60]. *MAP2K5* may therefore regulate adiposity via this important pathway[60].

As with other traits, very few obesity associated regions have been mechanistically investigated thoroughly from "variant to function" to identify the genes driving risk. One great exception is for an association that lies within the first intron of the *FTO* (Fat mass and obesity-associated) gene, where these variants consistently represent the strongest genetic variants associated with obesity across studies and human populations. Individuals homozygous for the risk variant in this locus are on average 2.5-3kg heavier than homozygous non-risk individuals[61]. The risk variant in this locus is associated with both increased food intake as well as diminished satiety response in adults as well as children[62–65]. For years gene driving this phenotype was predicted to be *FTO* itself, and mouse models showed that knocking out this gene lead to a body weight phenotype. However, a loss-of-function mutation (R316Q) in *FTO* was then identified in a cohort of human subjects, and neither heterozygous nor homozygous individuals exhibited a high prevalence of obesity, complicating this interpretation [66]. Recent, careful investigation of this locus demonstrated that the first intron of *FTO* contains enhancers active in both brain and adipose[29,44]. Using 3D genome technology as well as CRISPR-cas9 editing, these enhancers were then linked not to *FTO*, but to two distal genes, *IRX3* and *IRX5* that were also found to regulate BMI in mice and humans[29,44]. These enhancers also seem to have temporally restricted effects, so they only modulate the expression of these two genes only during the early stages of adipose and neuronal development. This mechanism has been elucidated through years of work to determine that SNPs within these enhancers likely alter disease risk through disruption of adipose-mediated thermogenesis as well as hypothalamic mediated food intake preferences (unpublished and [29]). Although the *FTO* locus has highlighted novel insights into the complexities of GWAS, questions remain whether these principles are uniform across regions or unique to this locus. Understanding these complexities will allow us to further understand the genetic mechanisms underlying GWAS.

1.7 Overview of thesis research

The thesis research presented here was an attempt to address the gaps discussed above in our knowledge of how to interpret GWAS associations. This work aimed to encompass a pipeline that would allow us to interpret GIANT consortium obesity GWAS associations in a high-throughput manner in order to generate a list of prioritized genes for further investigation into their relevancy for obesity risk. This pipeline is also flexible enough that it can be applied to other heritable complex traits. We first created maps of key genomic features in cell types important for our trait of interest, obesity, to prioritize causal SNPs and target genes important for disease risk (Chapter 2). After integrating these data, we then chose to dig further into an interesting locus on chromosome 16 that emerged from the analysis. An introduction to this locus and our findings are discussed in Chapter 3.

Chapter 2: In order to interpret obesity GWAS associations, we turned to models of white adipocytes and hypothalamic neurons in order to generate useful genomic annotations in these disease relevant cell types. We differentiated Simpson-Golabi-Behmel syndrome (SGBS) preadipocytes to adipocytes and collected ATAC-seq, RNA-seq, and promoter-capture HiC at four differentiation stages. Secondarily, we also differentiated iPSC cells to mature hypothalamic neurons and collected these cells for ATAC-seq, RNA-seq, and promoter-capture HiC at three differentiation stages. These genomic annotations were then integrated with additional data suggesting putatively causal variants that modulate enhancer activity which were identified using a massively parallel reporter assay (MPRA) performed in 3 adipose and 2 neuronal cell lines. Combining these datasets and integrating them with eQTL information from the Genotype-Tissue Expression project (GTEx version 8) we were able to prioritize target genes for importance in obesity risk across 38 GWAS association loci in a cell type specific manner.

Chapter 3: Leveraging data from Chapter 2 of this thesis, we then went on to further investigate an interesting locus on chromosome 16p11.2 containing two independent GWAS loci. Both of these loci exhibited a dense network of enhancers containing variants that were seemingly connected via cHi-C interactions and eQTL evidence. In the *ATP2A1* (ATPase Sarcoplasmic/Endoplasmic Reticulum Ca2+ Transporting 1) locus, we identified a haplotype of 7 common variants that segregate together in European populations. These variants were each found to modulate enhancer activity, were eQTLs for many genes in brain and adipose, and interacted with many promoters in our cHi-C data in both cell types. In the second *SBK1* (SH3 Domain Binding Kinase 1) locus we identified one very interesting candidate causal variant that also exhibited these traits and made long range cHi-C contacts into the *ATP2A1* locus. Using a suite of functional genomics technologies we decided to probe this hypothesis and discovered that there is indeed a physical connection between the two loci, and that at least one of the enhancers is capable of modulating the expression of more than one gene under unique cell-type and temporal situations. These results validate our predictions from Chapter 2 that suggest many GWAS loci will harbor more than one functional variant that participates in either uniform or temporally restricted modulation of gene expression to lead to obesity risk, and may regulate more than one gene.

CHAPTER 2: LEVERAGING GWAS TO IDENTIFY GENES IMPORTANT FOR OBESITY RISK

2.1 Abstract

GWAS interpretation relies on the generation of genomic annotations in cell types relevant to the disease or trait of interest. In order to interpret GWAS association loci for obesity, we generated ATAC-seq, RNA-seq, and promoter capture Hi-C (cHi-C) data across several stages of development for adipose and hypothalamic neurons. We then combine these annotations with additional data generated from a massively parallel reporter assay (MPRA) where we tested enhancer modulating activity for all SNPs in high LD with the 97 independent obesity GWAS loci identified in Locke et al 2015. Using MPRA, we were able to identify 94 variants across 40 of these GWAS loci that exhibited enhancer-modulating activity in either adipose and/or brain cells. Thirty-nine percent of these enhancer-modulating variants (EMVars) were functional in both cell types, suggesting that effects they impart could be shared across tissues. Additionally, two-thirds of the loci harbored more than one EMVar, providing evidence for multiple functional variants across the majority of loci. Using the generated genomic annotations, we prioritized genes in 38 GWAS loci via a classification system into degrees of supporting evidence for obesity risk. Twenty class I genes exhibited the highest levels of support and many had functions with known relevance to metabolic processes. Thirty additional class II genes also had high levels of support. Together, our data support a model in which many GWAS loci contain multiple variants that impair the activities of distinct enhancers across tissues, potentially with temporally restricted effects, to impact the expression of multiple genes. This

complex network model has broad implications for ongoing variant to function efforts to mechanistically dissect GWAS.

2.2 Introduction

While genome wide association studies (GWAS) have been instrumental in associating genetic variation to disease, functional studies delineating specific causal variants or effector genes of these associations have yet to become commonplace. Current evidence indicates that a significant proportion of associated variants impart their phenotypic effect through functional effects on distal gene regulatory elements, such as enhancers. However, a challenge remains to pinpoint the causal variants that modulate these enhancers and to identify their effects on target genes in specific tissues. Recent studies have posited that regulatory variants are common, and may act in a pleiotropic manner to modulate expression across cell types[67,68]. Therefore, questions remain whether genetic associations are driven by single or multiple variants, if the phenotypic impact of causal variants is uniform or cell type specific, and whether this regulation occurs across developmental stages or is confined to specific temporal windows[29,44,69,70]. The ability to characterize the genetic architecture of a disease remains anchored in the need to develop and interpret comprehensive functional genomic maps in disease relevant cell types. Here, we applied a suite of tools to generate genomic annotations that allow for the functional interpretation of GWAS loci associated with obesity.

GWAS meta-analyses for Body Mass Index (BMI) have identified 97 independent loci associated with obesity, where the vast majority of these loci harbor causal variants that are predicted to be noncoding[40]. These noncoding BMI associated variants are strongly enriched to lie within regions containing brain and, to a lesser extent, adipose enhancers identified by

H3K27ac and H3K4me1 ChIP-seq in these cell types. These two cell types are thought to be critical players in BMI maintenance, as they modulate energy intake in the form of hunger and satiety cues and control energy expenditure through central and peripheral circuitry. To functionally interpret obesity-associated loci, we systematically generated key genomic annotations in primary human adipocytes and human iPSC-derived hypothalamic neurons. To capture dynamic features of chromatin accessibility, gene expression, and long-range enhancer-promoter interactions, we assessed these parameters across the differentiation of iPSC derived hypothalamic neurons and during the conversion of pre-adipocytes to mature white adipocytes. Additionally, we identified a set of 2,396 putatively causal variants in high linkage disequilibrium (LD) with the 97 BMI GWAS lead single nucleotide polymorphisms (SNPs) and determined the enhancer activity and allelic effects of each of these variants by performing a massively parallel reporter assay (MPRA) in brain and adipose cell lines.

Our MPRA data identified putatively causal enhancer-modulating variants (EMVars) with regulatory properties in adipose and/or neuronal cell lines. While we identified a single EMVar in some obesity-associated loci, the majority of loci contained multiple EMVars, demonstrating that the genetic architecture at GWAS loci is often complex. Assaying the regulatory landscapes of human adipocytes and hypothalamic neurons across developmental stages resulted in an increased overlap of functional annotations with EMVars, supporting evidence that a subset of functional variants have temporally restricted phenotypic effects *in vivo*. We synthesized these datasets to provide a ranking system for variant and target gene prioritization across 38 of the 97 GWAS loci to inform functional follow-up in each cell type.

2.3 Results

2.3.1 Charting the regulatory landscape of obesity relevant cell types

To interpret obesity GWAS associations, we aimed to generate comprehensive maps of genome annotations in human hypothalamic neurons and adipocytes. Despite the prominence of hypothalamic neurons in obesity etiology[40,51,71–73], little is known about the regulatory landscape of these cells, which is in part due to the challenge in obtaining them. To overcome this, we differentiated a human induced pluripotent stem cell line (iPSCs) into mature hypothalamic neurons. We modulated sonic hedgehog (SHH), transforming growth factor β (TGFβ), and bone morphogenetic protein (BMP) signaling pathways to induce neuronal differentiation. After neuronal differentiation we introduced BDNF factor to promote maturation of POMC and NPY positive arcuate nucleus type hypothalamic neurons. We collected cells at three time points representing early hypothalamic neuron precursors (D55), early immature (D75) and late (D100) mature hypothalamic neurons. These cells were then processed for cHi-C to elucidate putative enhancer-promoter interactions, ATAC-seq to identify open chromatin, and RNA-seq for global gene expression information. Additionally, we obtained non-immortalized Simpson-Golabi-Behmel syndrome (SGBS)[74] human preadipocytes, the only human preadipocyte cell line available, and differentiated them to mature white adipocytes for collection at four key time points representing preadipocytes, differentiation induction, early mature adipocytes and late mature adipocytes, respectively (Figure 2.1a). For each of the adipocyte time points, we performed ATAC-seq, RNA-seq, and cHi-C (Figure 2.1a)[8,9,13]

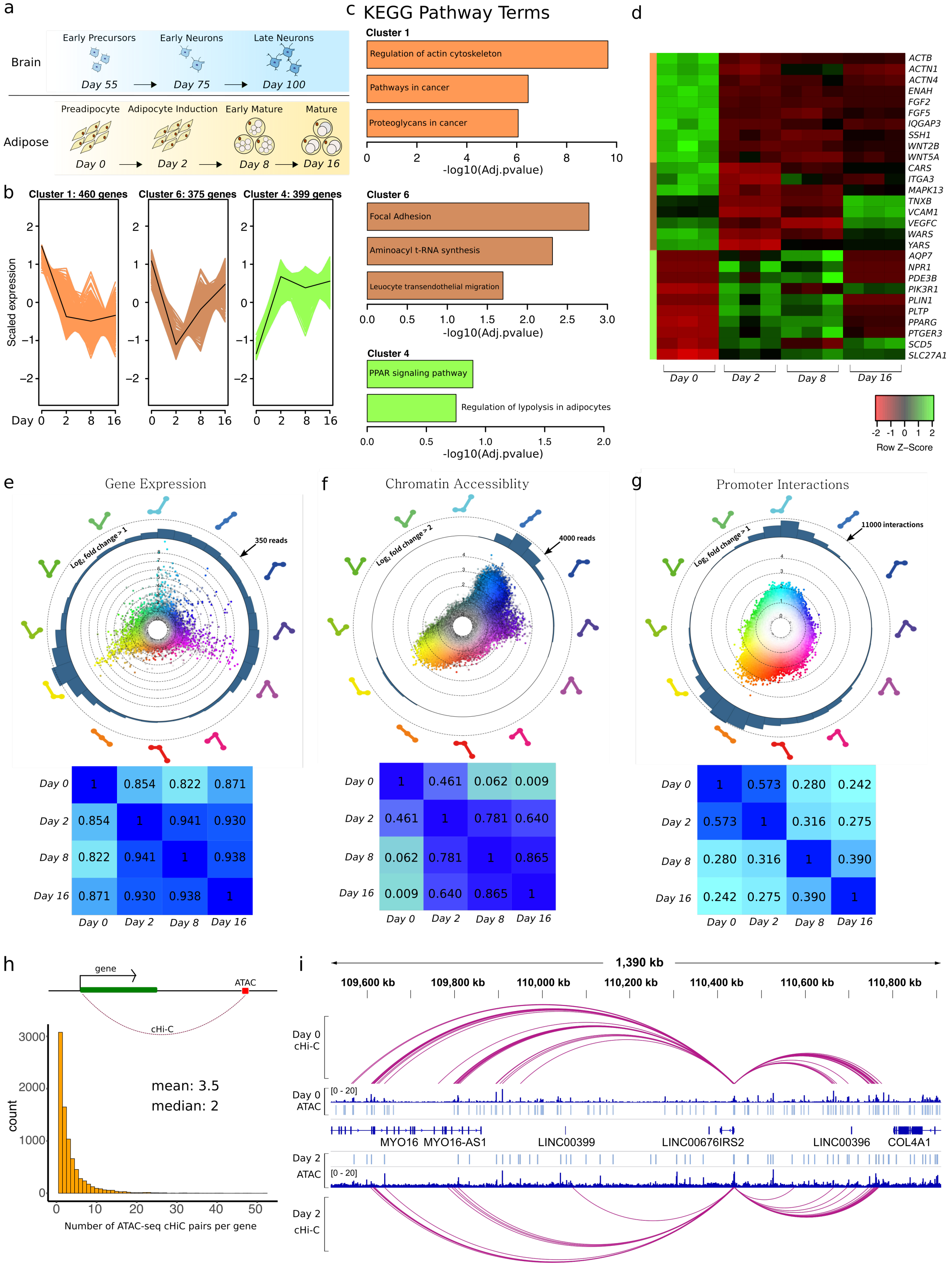
In total, we identified 601,109 - 935,217 significant cHi-C interactions per time point in adipose and 456,653 - 588,929 interactions in neurons, with a median interaction distance between 178-260kb (Supplementary Figure S2.1c). The average fragment size for each interaction was 422 base pairs, allowing us to map putative regulatory regions at very high resolution[75]. In support of these maps use for enhancer identification, we evaluated the promoter distal ends of hypothalamus and adipose interactions and found they were enriched for cell type appropriate enhancer histone marks identified by ENCODE (H3K4me1 and H3K27ac), as well as for open chromatin (Supplementary Figure S2.1e,f).

To characterize dynamic changes of these datasets across differentiation stages, we initially focused on identifying differentially expressed genes (DEGs). For adipose, the largest changes in gene expression occurred upon differentiation induction, with 1,881 DEGs observed between day 0-2 compared to 516 genes between day 2-8 and 611 DEGs between day 8-16. A two-stage comparison between preadipocytes and mature adipocytes would have resulted in 760 DEGs, thus losing a significant portion of temporally restricted transcriptome changes. We used fuzzy-c means clustering to group DEGs for both cell types into predominant patterns (Figure 2.1b & Supplementary Figure S2.1a,b). The three clusters with the highest membership scores are shown (Figure 2.1b & Supplementary Figure S2.2a). The top adipose cluster (Cluster 1) encompassed genes that have the highest expression in preadipocytes and was enriched for genes involved in actin cytoskeleton rearrangement and proliferation, both of which represent important processes for the conversion of preadipocytes to adipocytes[76,77]. The second cluster (Cluster 6) was enriched for genes involved in focal adhesion, a signaling pathway involved in extracellular matrix (ECM) communication[78–81]. The last cluster (Cluster 4) represented genes that were upregulated quickly upon exposure to differentiation stimuli, and included genes from the PPAR signaling pathway as well those involved in lipolysis, both of which are canonical processes that occur during the transition of fibroblast-like preadipocytes to lipid-laden adipocytes[82] (Figure 2.1b-d). For hypothalamic neurons, the three main clusters for neuronal differentiation DEGs were enriched for genes involved in neurogenesis (Cluster 4), cell-cell signaling (Cluster 6), and developmental processes (Cluster 5). The genes involved in neurogenesis and cell-cell signaling were the highest expressed at D100, the most mature state. Alternatively genes involved in development were downregulated by D100 (Supplementary Figure S2.2a-c).

To derive comparisons across time and between datasets, we show all significant data points in adipose and brain using the Hue-Saturation-Value transformation (HSV)[83,84] (Figure 2.1e-g, Supplementary Figure S2.2d-f). With HSV, gene expression, chromatin accessibility and cHi-C promoter interactions can be visualized in a 360° space, where each significant data point is binned into a representative temporal pattern shown on the outside of the plot. We also performed a Pearson's *r* correlation to evaluate relationships between time points. In adipose, gene expression and chromatin accessibility were the least correlated (most changed) between Day 0-2 (*r* =.854; RNA & *r* =.461; ATAC) and reached an equilibrium in later differentiation (*r* =.938-.941; RNA & *r* =.781-.865; ATAC). Conversely, the cHi-C data were the most correlated between Day 0-2 (*r* =.573), but changed dramatically as differentiation continued (*r* =.316-.390) (Figure 2.1e-g). Broadly, gene expression values in adipose exhibited diverse global patterns, where the most common was a strong decrease in expression between the first two time points (yellow pattern). ATAC-seq peaks tended to increase over time (dark blue pattern), but also exhibited some regions of decreased accessibility (yellow pattern). Promoter interactions seemed to be the least dynamic, either decreasing over time (orange pattern) or increasing during the last time point (light blue pattern) (Figure 2.1e-g). These analyses were also performed for neuronal maturation and are presented in Supplementary Figure S2.2. We observed different patterns emerging from these hypothalamic HSV plots compared to adipose, and these differences are likely due to differences in biological processes happening between cell types, or, because we are capturing the late stages of induced maturation rather than the process of stimulated differentiation induction as with the adipocytes.

To evaluate whether changes in chromatin accessibility and/or cHi-C interactions correlate with gene expression changes at gene loci, we obtained a list of genes that were connected to at least one ATAC-seq peak via a significant cHi-C interaction (Figure 2.1h, Supplementary Figure 2.2g). For adipose this list contained 8,288 genes and on average each promoter interacted with 3-4 unique ATAC-seq peaks across this time-course (Figure 2.1h). In the hypothalamic neuron data, 5,129 gene promoters interacted with an average 1-2 ATAC-seq peaks, likely because we captured fewer significant Hi-C interactions in these cells compared to adipose (Supplementary Figure S2.2g). To test the functionality of these interactions, we grouped genes that were differentially expressed at any time point and compared changes in chromatin accessibility and cHi-C interaction strength to static genes. For genes upregulated between two time points, we observed stronger interaction scores and more accessible chromatin compared to genes that were not differentially expressed. Downregulated genes also generally demonstrated stronger suppression of chromatin and interaction scores, supporting the use of these datasets to identify gene regulatory regions (Supplementary Figure S2.1g).

Using these data, we generated a high-resolution map of interactions between promoters and putative regulatory elements across several differentiation time points in both adipose cells and hypothalamic neuronal precursors. This is a critical first step for the overarching goal of obesity GWAS interpretation, and allowed us to wire distant enhancers to promoters in a high-throughput manner (Figure 2.1i, Supplementary Figure S2.2h)



e Gene Expression

	Day 0	Day 2	Day 8	Day 16
Day 0	1	0.854	0.822	0.871
Day 2	0.854	1	0.941	0.930
Day 8	0.822	0.941	1	0.938
Day 16	0.871	0.930	0.938	1

f Chromatin Accessiblity

	Day 0	Day 2	Day 8	Day 16
Day 0	1	0.461	0.062	0.009
Day 2	0.461	1	0.781	0.640
Day 8	0.062	0.781	1	0.865
Day 16	0.009	0.640	0.865	1

g Promoter Interactions

	Day 0	Day 2	Day 8	Day 16
Day 0	1	0.573	0.280	0.242
Day 2	0.573	1	0.316	0.275
Day 8	0.280	0.316	1	0.390
Day 16	0.242	0.275	0.390	1

Figure 2.1: Characterizing Adipocyte Differentiation using Genomic Annotations
a) Time points for data collection

Figure 2.1, continued. b) Adipose DEGs were grouped via fuzzy-c means clustering and the top three clusters with highest membership scores are illustrated. The number of genes in each cluster and scaled expression across the four differentiation time points is depicted. c) Significant KEGG pathway terms identified using Enrichr for the top three clusters. d) Heatmap of gene expression depicting genes from each of the top three clusters that are members of the enriched KEGG pathway terms. The leftmost colored bar indicates cluster membership and each column is an RNA-seq replicate. e-g) HSV transformation of expressed genes, ATAC-seq peaks, and cHi-C interactions across differentiation. The three nodes of each pattern represent day 0, day 2, and day 16 of adipose differentiation. The distance of each point from the center of the circle represents maximum log2 fold change, and color transparency represents the relative number of reads for that data point. Below, heatmaps of Pearson's *r* correlation coefficients estimate overall similarity between time points. h) On average, a promoter interacts with 3-4 ATAC-seq peaks via a cHi-C interaction across time (interactions and ATAC peaks were not required to be significant at the same time point). i) View of cHi-C interactions emanating from the promoter of the *IRS2* gene, which becomes upregulated between differentiation days 0-2. ATAC-seq reads and peaks from day 0 and day 2 are also shown.

2.3.2 Identifying functional variation in obesity GWAS loci

A common mechanism by which noncoding variants lead to disease risk is expression modulation mediated by alterations in enhancer function[67]. In order to identify SNPs capable of affecting enhancer activity at GWAS loci, we employed a massively parallel reporter assay (MPRA) to test variants in high LD with lead SNPs identified in a recent BMI meta-analysis conducted by the GIANT consortium[16,40,85]. Candidate variants were defined as lead SNPs in 97 independent obesity GWAS loci and variants in high LD (MAF >= 5% CEU population, r^2 >.8), for a total of 2,396 variants. We synthesized 175-bp DNA fragments centered on each biallelic SNP, and each allele was placed alongside 18-19 unique 10bp DNA barcodes, allowing for 18-19 measurements of enhancer activity for every allele. This resulted in a pool of 89,964 fragments that was cloned into the pMPRA1 vector[16]. We tested each region containing a SNP for enhancer modulating activity, as well as allele specific differences in activity across three adipose cell types (SGBS preadipocytes, SGBS mature adipocytes, 3T3-L1 preadipocytes) and two neuronal cell types (GT1-7 and HT22 cells) (Figure 2.2a-b). Across the GWAS loci, we identified 807 genomic regions in brain and 543 genomic regions in adipose where at least one of the two alleles acted as an enhancer, and 460 regions were enhancers in both cell types (Figure 2.2c). Of the enhancers, 94 harbored an enhancer-modulating variant (EMVar), which conferred significant differences in enhancer activity between alleles (Figure 2.2b,d). ENCODE ChromHMM provides chromatin state predictions across the genome based on ChIP-seq derived histone marks[3]. Compared to all tested regions, MPRA enhancers were enriched for ENCODE ChromHMM predicted active marks and depleted for inactive marks in adipose and brain tissues (Supplementary Figure S2.3a). They were also more likely to overlap open chromatin and cHi-C interactions compared to other MPRA tested regions without enhancer activity, supporting the potential of enhancer function for these regions in their native chromatin context (Supplementary Figure S2.3b).

Because enhancers are made up of combinatorial transcription factor binding sites that can be up to 1kb in length, we decided to validate enhancer activity for 24 unique ~1,000bp sized regions containing MPRA EMVars using luciferase assays in a brain and/or adipose cell line depending on where they were found to be significant using MPRA. We found that these longer DNA fragments resulted in the same enhancer activity call in luciferase assays for 28/43 (65%) of the conditions tested (Supplementary Figure S2.3c). This indicates that while size of the tested enhancer may affect activity, the calls were relatively consistent across different assay types and fragment sizes.

We next sought to illuminate the network of transcription factors (TFs) putatively bound to these enhancers to understand potential biological processes regulated within these GWAS loci. We performed TF motif enrichment analysis within enhancer sequences identified from each of our MPRA cell types and found that MPRA enhancers at obesity GWAS loci were enriched for motifs for TFs that are involved in critical metabolic processes regulated in both brain and adipose. Multiple members of the AP-1 complex family as well as the ATF family were identified. AP-1 family members are upregulated during early adipogenesis and are critical for proper adipose formation, and also have the ability to regulate whole-body energy expenditure when modulated in the hypothalamus [71,86–88]. ATF factors in particular are important for adipogenesis, and have also been shown to regulate thermogenic programs in the mouse hypothalamus through *Agrp* expression modulation [89–91]. Other TF motifs important for thermogenesis and glucose homeostasis were also enriched, including thyroid hormone receptors, IRF3, ERRα, and USF1/2 [92–95]. Interestingly, TFs important for maintenance of circadian rhythm in central or peripheral clocks such as CLOCK, BHLHE40 (DEC1), and BMAL1 were also enriched (Supplementary Figure S2.4a, Supplementary Table S2.1-S2.2)[96].

In addition to identifying enhancers, the MPRA assay tests for variants capable of modulating enhancer activity (EMVars). Of the 94 EMVars we found 61 brain EMVars and 70 adipose EMVars, and at least one EMVar was identified in 40/97 (41%) of tested GWAS loci. Surprisingly, 2/3 of these contained more than one EMVar, raising the possibility that the genetic architecture of these GWAS loci could be driven by allelic heterogeneity, where multiple variants impart their effects on a phenotype (Figure 2.2e). Additionally, 37/94 (39%) of these variants affected enhancer activity in both cell types, an observation in line with the recent GTEx finding that the majority of expression quantitative trait loci (eQTLs) are not tissue specific (Figure 2.2e)[67]. This suggests that at each of these GWAS loci, multiple variants have the potential to contribute to expression variation. Additionally, the effect of these variants may not be restricted to one obesity relevant tissue.

The two loci that harbor largest number of EMVars identified in our study were the FTO and *ATP2A1* obesity association regions, each representing strong and highly reproducible associations on chromosome 16 (Figure 2.2e)[40]. Overall, we observed the largest number of EMVars mapping to chromosome 16. To investigate this further, we applied stratified LD score regression (s-LDSC)[97] to BMI GWAS summary statistics to estimate heritability across chromosomes and confirmed that chromosome 16 contributes disproportionally to obesity heritability (Figure 2.2f, Supplementary Figure S2.5b,c) [34]. These data suggest that the strong heritability enrichment at chromosome 16 could be driven by an overabundance of functional variants, such as EMVars, that exist within chromosome 16 obesity GWAS loci.

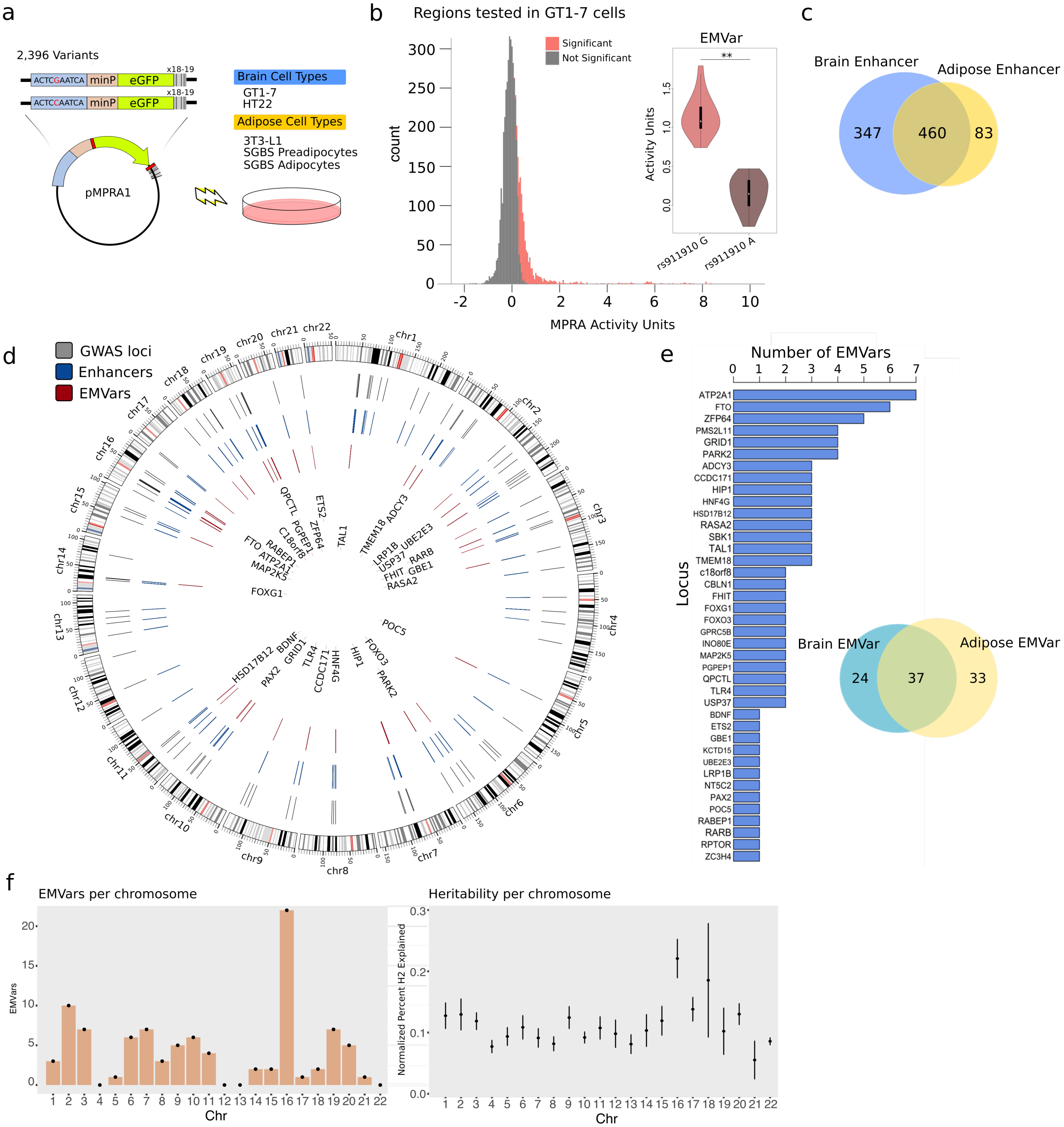


Figure 2.2: MPRA identifies enhancers and functional variants in obesity GWAS loci
a) Variants were synthesized adjacent to 18-19 unique 10bp DNA barcodes and cloned into the pMPRA1 vector. Constructs were transfected into 5 cell lines from the adipose and brain lineages (see Methods). b) Average MPRA activity across replicates is shown for all tested regions in GT1-7 libraries. Significant GT1-7 enhancers ($q < 0.05$; one sided Mann-Whitney U test) are colored red. A SNP was considered an EMVar if the variant significantly affected MPRA enhancer activity levels ($q < 0.05$; two sided Mann-Whitney U test). c) Venn diagram of MPRA enhancers significant in either cell type d) Circos plot of MPRA results. Grey lines within the circle represent the locations of GWAS associations, blue lines represent the locations of MPRA identified enhancers, and the red lines represent identified MPRA EMVars. Locus gene names (closest gene) are shown in the center.

Figure 2.2, continued. e) Bar chart of significant EMVars per locus, along with a Venn diagram of EMVars called in either brain or adipose cell lines. f) (left) Number of significant EMVars identified per chromosome. (right) s-LDSC estimate of the percent of heritability explained per chromosome normalized to the proportion of variants tested; Error Bars = SEM

2.3.3 Assigning functional variants to target genes

Having identified an array of regulatory elements and functional variants in obesity GWAS loci, we next aimed to gain insights into the connections of these regulatory elements with their target genes in 3D genomic space using cHi-C (Figure 2.3a). Understanding the configuration of these functional variants in respect to promoters in the nucleus is important to identify target genes of the EMVar-containing enhancers and generate a list of prioritized genes for future mechanistic studies into their role in obesity etiology.

We intersected the time course cHi-C data generated in adipocytes and neuronal precursors with EMVar locations to identify interactions between EMVars and their target genes in a cell-type specific manner. Having cHi-C data for multiple time points in both lineage differentiations allowed us to assign more EMVars to promoters than using one time point alone (Figure 2.3b). Interestingly, we observed that if an adipose EMVar, enhancer, ATAC-seq peak, or H3K27ac marked region participated in a cHi-C interaction, it contacted a median of 3 different promoters across the adipose cHi-C time-course (data not shown). Similarly, brain EMVars and enhancers contacted a median of 2 or more promoters across the hypothalamic cHi-C time course (Figure 2.3b, data not shown). These data highlight the pervasive opportunity for pleiotropic regulation by these regulatory variants, which could affect gene expression in disparate tissues or in specific developmental stages. Our findings underscore the importance of assaying regulatory landscapes across development or under specific conditions, a finding recently corroborated by reports showing genomic annotations are subject to tissue and temporal specific regulation[29,69,70,98]

To further assess the functionality of these long-range interactions and gain additional evidence for gene targets, we intersected EMVar-promoter interactions with eQTL information in adipose and brain cell types from GTEx[67]. For brain, 26/61 EMVars were eQTLs for a gene in a GTEx(V8) brain cell type and 35/61 interacted with a promoter in at least one cHi-C library. Twenty one EMVars participated in interactions with distant promoters and were eQTLs, while 16/21 were an eQTL for a gene they interacted with (Figure 2.3c). We evaluated the intersection of adipose EMVars with our adipose differentiation cHi-C and GTEx subcutaneous or visceral adipose eQTLs in the same manner (Figure 2.3c). Through the integration of these datasets across time, we were able to assign 31/61 (51%) of brain EMVars and 54/70 (77%) of adipose EMVars to at least one gene in the cell type where the EMVar was found to alter enhancer activity (Figure 2.3c).

Integrating these annotations established gene expression patterns, identified regions of open chromatin, pinpointed enhancers harboring EMVars, and suggested target genes through integration with long-range enhancer-promoter interactions in obesity-associated loci. To summarize this for each GWAS locus, we binned genes into 4 classes of supporting evidence for both adipose and brain based on degrees of supporting evidence for their involvement in obesity etiology (Figure 2.3d). Identified class I genes have functions with known relevance to BMI maintenance, such as cholesterol and steroid metabolism, food intake, fat mass, mitochondrial function, or leptin sensitivity[43,99–104]. There are also several genes with unknown function or functions with unidentified relevance to a BMI phenotype.

As an example, we show that a brain EMVar (rs4776984) in the *MAP2K5* locus interacts with the class I gene *MAP2K5* (Figure 2.3e). This variant is a GTEx eQTL for *MAP2K5* in adipose, brain, and other cell types. This SNP-gene pair was previously tied to obesity risk in a study that identified this locus using eQTLs from the METSIM cohort and cHi-C data from primary human white adipose tissue[45]. We confirmed this SNP's interaction in adipose and show that it also interacts with *MAP2K5* in neuronal precursors (Figure 2.3e). We also identified a second EMVar in this locus, rs2127163, which was an EMVar in both adipose and brain. This variant also interacted with, and is an eQTL for, *MAP2K5* in both adipose and brain, suggesting that the association at this locus may be due to at least two different SNPs in independent enhancers regulating *MAP2K5* across these cell types (Figure 2.3e). Our datasets thus allow us to tease apart loci with multiple potential causal variants and was not influenced by LD in the same manner as eQTL data alone.

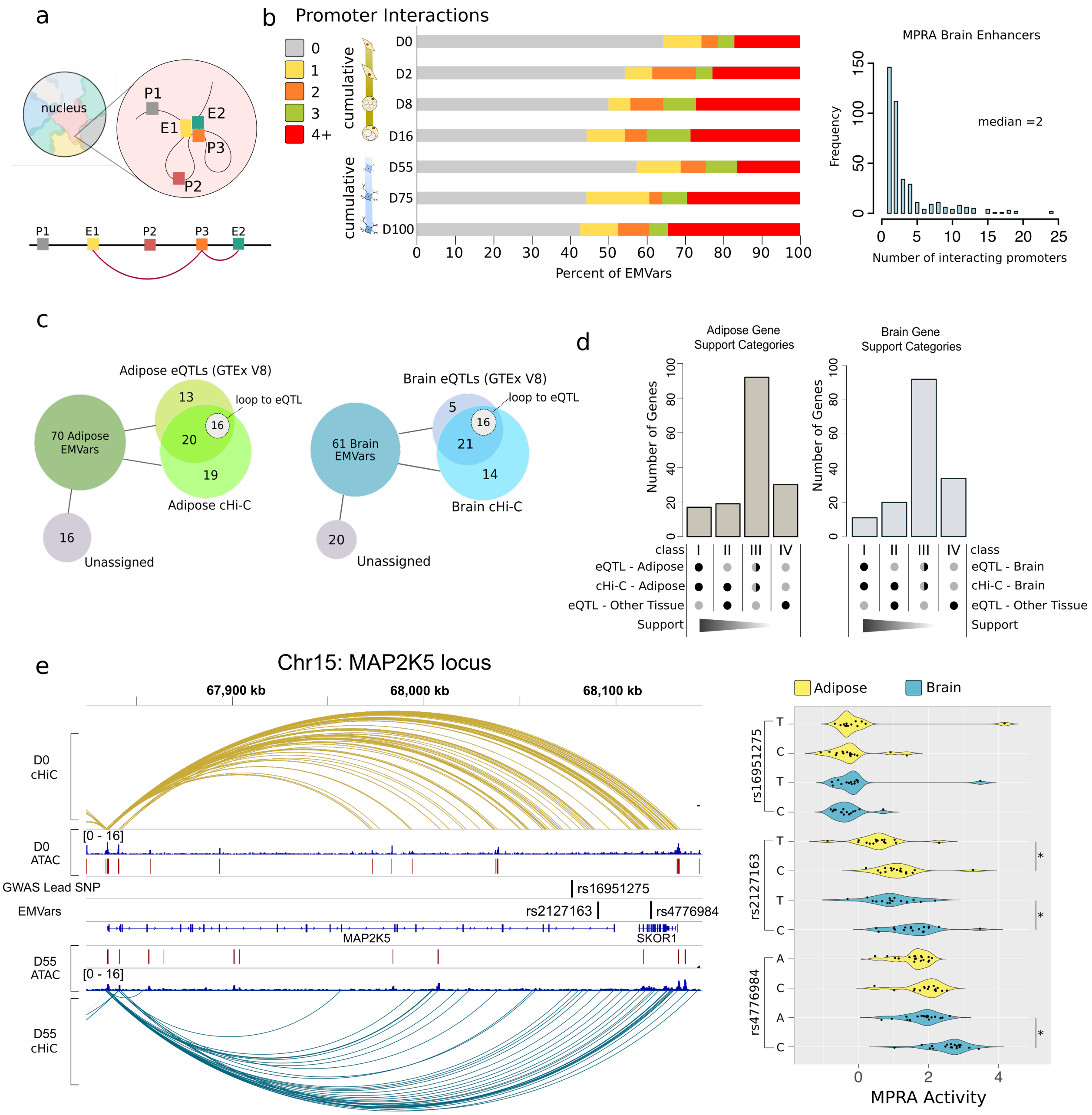


Figure 2.3: Integration of functional variants with genomic annotations prioritizes target genes a) cHi-C allows for identification of physical connections between enhancers (E) and promoters (P) in nuclear space and are shown as arcs on the linear genome (depicted here in pink). b) (left) Cumulative distribution of promoter interactions per EMVar across time in adipose and brain cells. (right) Bar plot showing the number of promoters that each brain MPRA enhancer interacts with across all cHi-C replicates (does not include enhancers that do not interact with a promoter). c) Diagram of EMVars that are either in cHi-C interactions and/or are GTEx eQTLs, and not assigned to a target gene with either method. d) Genes were binned into classes based on strength of evidence supporting them as a GWAS target gene (see Methods). Half shaded circle= eQTL *or* cHi-C support

Figure 2.3, continued. e) *MAP2K5*, a class I gene, is shown here with the local brain and adipose cHi-C interactions emanating from its promoter. Activity units for every barcode for the two EMVars and lead SNP from this locus are shown in a violin plot in HT22 (blue) and 3T3-L1 cells (yellow). **q* < 0.05; two-sided Mann-Whitney U test

2.4 Discussion

In this chapter we generated comprehensive regulatory maps in human adipose and hypothalamic neurons, which lack comprehensive genomic annotations despite their prominence in disease etiology. We profiled these cells across several differentiation stages to catalog chromatin accessibility, expression patterns, and cHi-C enhancer-promoter interactions. We also used a high throughput enhancer assay to identify putatively causal variants within 97 independent obesity GWAS loci identified by the GIANT consortium. Integrating these datasets aided in the interpretation of candidate causal non-coding variants and genes at obesity-associated loci.

Using MPRA, we were able to first identify enhancers that were active in brain and/or adipose cells. Using this data we were able to assess the network of transcription factor motifs that were enriched within these regions. The transcription factor enrichments that emerged from HOMER enrichment analysis was a network of TFs that participate in metabolic regulation with roles in energy-expenditure, glucose homeostasis, development, and circadian rhythm. These processes are major players in maintenance of body mass index and modulation within neuronal or adipose cells will have affects on whole body metabolism. Evidence of enrichment for circadian rhythm transcription factors was particularly interesting, as sleep and circadian rhythmicity have recently been identified as modifiable risk factors for obesity[105,106]. Knowledge of these players and their biological importance could illuminate biological processes that are potentially regulated within these regions and thus when misregulated, e.g. in the case of risk loci, leads to disease.

In two thirds of all identified loci, we observed more than one EMVar per locus. This suggests a previously underappreciated degree of allelic heterogeneity present within these loci. Recent reports support this notion, as strongly powered and densely genotyped GWAS have identified independent signals within the same association to suggest the existence of multiple causal variants within single loci[107]. But, few studies to date have investigated how regions containing multiple causal variants affect gene expression across tissues and developmental time points. In thirty three percent of our cases, EMVars were significant in both tested cell types. The GTEx consortium has recently been able to address patterns of tissue specificity in eQTLs with its increasingly larger dataset and found a high level of eQTL effect sharing between tissues[67]. They suggest that only ~20 percent of all eQTLs have their effects restricted to 1-5 cell types. Another ~20 percent have effects in all cell types tested[67]. This indicates functional variants may have pleiotropic effects across tissues as frequently, or more frequently, than tissue specific effects. Therefore, it may be critical to develop a methodology to predict the most likely causal tissue of interest, as well any secondary or tertiary effects in other cell types. This would allow for a better understanding of potential synergistic effects across cell types. Focus on how your SNP or gene of interest modulates biology in one cell type may be providing a limited understanding of how the locus affects disease. Combining the potential for multiple causal variants to exist within the same locus that have cross-tissue effects leads us to infer a complex network model whereby a region with multiple perturbations leads to large effects on cellular phenotypes and disease. The extent to which all perturbed enhancers or genes contribute to your phenotype, or are innocent bystanders, requires further investigation.

During our analysis we also observed that having multi-time point cHi-C data allowed us to assign more EMVars to genes. This could be because sampling these cells multiple times allowed us to pick up more interactions, and thus cell type unique interactions we observed are actually present in all time points but were not measured. But, these interactions could also be due to the presence of temporally restricted interactions between these enhancers and promoters. Our knowledge of the biology of enhancers suggests that regulatory variants associated with human phenotypes may impart their effect during developmental windows, which would be missed in functional assays of a single time point or environmental perturbation[29,69,70,98]. If this were true, we would expect these variants to also have effects on gene expression during some developmental stages and not others. This adds an additional layer of complexity when predicting whether a region of the genome acts as a regulatory element. Plus, it complicates our understanding of the biological importance of assaying a gene's functional effect on an organism when it is modulated at specific developmental time points and not others.

Using all data, we were able to integrate cHi-C data from multiple time points along with GTEx eQTL information to better assign putative gene targets to functional variants and to suggest the importance of 232 genes across classes in obesity risk. Across classes we identified 20 genes that were highly supported class I genes in brain or adipose and thirty that were class II genes. These genes have the highest genomic support for causality in these loci and thus have a high likelihood biological relevance for obesity that will only be realized with further experimental interrogation and understanding.

2.5 Methods

2.5.1: SGBS Culture and Differentiation

SGBS cells were maintained and differentiated as previously described[74,108]. Cells were grown in DMEM/F12 (1:1) media (Life Technologies #11330-032) with 10% fetal bovine serum, 1% penicillin-streptomycin solution (10,000 U/ml; gibco #15140122), 8mg/ml Panthotenic Acid, and 8mg/ml Biotin (Sigma; #B4639). Cells were allowed to grow to 70-80% confluency before splitting 1:3 with 0.25% Trypsin-EDTA (gibco; #25200056). To differentiate, cells were split into 6 well plates, allowed to reach 100% confluency over two days. After maintenance of confluency, day 0 cells were harvested, and the remaining cells were washed twice with 1x PBS and exposed to Quick-Diff media. Quick-Diff media consists of serum free DMEM/F12 media supplemented with 0.01mg/ml human transferrin (Sigma #T2252), 20nM human insulin, 100nM cortisol, and 0.2nM Triiodothyronine. Day 2 cells were harvested two days after addition of quick-diff media. Cells were incubated for a total of 4 days in quick-diff media before 3FC media was added. The 3FC media consists of serum free DMEM/F12 media supplemented with 0.01mg/ml human transferrin (Sigma #T2252), 20nM human insulin, 100nM cortisol, 0.2nM Triiodothyronine, 25nM dexamethasone, 250uM 3-isobutyl-1-methylxanthine (IBMX), and 2uM rosiglitazone. Mature adipocytes are maintained on 3FC media until collection.

2.5.2: Human hypothalamic neuron differentiation

Human induced pluripotent stem cells (hiPSCs) (Findiv 24382) were differentiated into hypothalamic arcuate-like neurons, as previously described by Yao and collaborators in 2017. Briefly, hiPSC were grown as embryoid bodies (EBs) with Neural Induction media, DMEM/F12 (Gibco) without bFGF (Sigma) and supplemented with N2 (Gibco) and

NEAA (Sigma) for 3 days. After induction, EBs were plated in coated dishes (poly-L-ornithine (Sigma) and laminin (Thermo Fisher)) and media was changed daily for 14 days. Primitive neuroepithelial were treated with 50 ng/ml WNT3A (R&D) and neural tube-like rosettes were formed, isolated and transferred onto new coated dishes and grown in media containing WNT3A ((R&D) for 7 days. Cells were supplemented with 1uM cAMP (R&D) and 10ng/ml WNT3A for 3 days. The neurons were then maintained in this medium supplemented with BDNF, GDNF and IGF1 (10ng/ml each) (R&D) to promote neuronal maturation of POMC and NPY neurons. Cells were collected at different time points and processed for *in situ* promoter capture HiC (cHiC), total RNA extraction and ATAC-seq.

2.5.3: Timecourse RNA-seq

Three technical replicates were performed derived from three unique differentiations for both SGBS and hypothalamic neurons. *Adipose:* SGBS cells were grown and differentiated as stated above. At day 0, 2, 8 and 16 of differentiation 1 million cells were collected and frozen per replicate with three technical replicates per time point. When all time points were collected, cells were lysed with a 20g needle and total RNA was extracted using the Qiagen RNeasy kit (#74104). RNA quality was assessed on an agarose gel. RNA-seq libraries were generated from 1ug of total RNA following the Illumina TruSeqRNA Sample Preparation V2 guide and 15 cycles of PCR amplification. *Brain:* iPSC derived hypothalamic neurons were grown and differentiated as stated above. At day 55,75, and 100 cells were collected and frozen per replicate. When all time points were collected, cells were lysed with a 20g needle and total RNA was extracted using the Qiagen RNeasy kit (#74104). RNA quality was assessed on an Agilent Bioanalyzer. RNA-seq libraries were generated from 500ng of total RNA using the NEB Next

Ultra II Directional RNA-seq kit and 10 cycles of PCR amplification. Libraries were sequenced on an Illumina HiSeq 4000 machine. *Analysis:* Gene-level read counts were quantified in each technical replicate at each time point directly using salmon (v0.7.2), correcting for sequence-specific bias and using a gene list derived from GENCODE release grch37.v19. For individual gene expression, read counts per gene were converted into transcripts-per-million (TPM) to account for gene length and library size. For the purposes of HSV visualization, gene counts were normalized converted to TPM to normalize for transcript length and then normalized by library size using trimmed mean of M-values (TMM) normalization and normalized by transcript length. Mean TPM was calculated at each time point, and all genes with mean $\log_2(\text{TPM}) < 1$ at any time point were removed from further analysis**.**

2.5.4: Fuzzy c-means clustering

Gene-level read counts were quantified in each technical replicate at each time point directly using salmon(v0.7.2), correcting for sequence-specific bias and using a gene list derived from GENCODE release grch37.v19. Gene-level read counts were transformed into cpm and any gene with cpm < 1 in more than three samples across all time points was removed from further analysis. The data were normalized to account for library size using TMM normalization. Linear models testing pairwise differential expression between any two time points were then build using limma and tested using a moderated t-test accounting for mean-variance dependence and increased dispersion in limma. All genes with significant differential expression between any two time points were included in the clustering analysis. Raw gene-level counts from salmon were normalized to account for transcript length and scaled to account for differences in gene expression across genes. Fuzzy c-means clustering was performed in R using the e1071 package.

A gene was assigned to the cluster for which it had the highest membership if 1) its membership score was above 0.3 for the averaged replicates and 2) above 0.2 for each individual replicate. The top three clusters were defined by the highest average membership score.

2.5.5: Hue-Saturation-Value (HSV) plots

All Hue-Saturation-Value (HSV) analyses are developed from code originally published in Siersbaek et al[84]. Value (V) indicates the maximum log2(TPM/cpm) for a given gene at any time point, and so is defined as:

$$V = max(C_t)$$

Saturation (S) indicates the maximal fold change between any time points and is defined as:

$$S = 1 - \frac{min_t(C_t)}{V}$$

Hue (H) indicates the pattern of change in gene expression across time, and is defined as:

$$H = 60 * (2 + \frac{C_0 + C_2 - C_{16} - V}{V * S}) * \frac{C_2 - C_0}{|C_2 - C_0|}$$

For visualization purposes, the values of V and S were scaled between 0 and 1 based on rank.

2.5.6: MPRA cell lines culture and transfection

<u>*HT22:*</u> Cells were maintained in DMEM (Gibco # 11995-065) supplemented with 10% FBS and 1% penicillin-streptomycin solution (10,000 U/ml; Gibco #15140122) at 37°C in 5% CO_2. We plated 250K cells into wells of 6 well plates and transfected one day later with Lipofectamine LTX & Plus reagent (Invitrogen; #15338100) when 60-70% confluent. <u>*3T3-L1*</u>: Cells were

maintained in DMEM (Gibco # 11995-065) supplemented with 10% FBS, 1% penicillin-streptomycin solution (10,000 U/ml; Gibco #15140122), 0.8mg/ml Biotin (Sigma; #B4639) and 0.8mg/ml Panthotenic Acid at 37°C in 5% CO_2. We plated 20k cells into wells of 6 well plates and transfected 2.5 days later with Lipofectamine LTX & Plus reagent (Invitrogen; #15338100) when 30-50% confluent. *GT1-7*: Cells were maintained in High Glucose DMEM (Gibco # 10313-021) supplemented with 10% FBS, 1% penicillin-streptomycin solution (10,000 U/ml; gibco #15140122) and 1X Glutamax (Gibco #35050061) at 37°C in 5% CO_2. We plated 750k cells into wells of 6 well plates and transfected the next day with Lipofectamine LTX & Plus reagent (Invitrogen; #15338100) when 60-70% confluent. *SGBS Preadipocyte (D0)*: 30,000 SGBS preadipocyte cells were plated into 24 well plates and transfected with Polyplus jetPEI DNA transfection reagent (Polyplus; #101-10N) when 50% confluent. These cells were collected 48 hours later for RNA processing. *SGBS Adipocyte (D8)*: Cells were plated as described above for differentiation. On differentiation day 8, cells were transfected with Lipofectamine LTX & Plus reagent (Invitrogen; #15338100) and collected on differentiation day 10.

2.5.7: ATAC-seq

Two technical replicates were performed for each time point derived from two unique differentiations for both SGBS and hypothalamic neurons. The two replicate sequencing data was merged and analyzed to produce merged datasets which were used in downstream analyses. We harvested 100,000 fresh SGBS cells per time point, and 100,000 hypothalamic neurons per time point. ATAC-seq libraries for each cell type were generated according to the protocol outlined in Buenrostro et al. 2015[8]. The cells were lysed, centrifuged, and frozen at -80C until all time points were collected. Final processing of all pellets was performed together. Transposed

DNA fragments were PCR amplified using 5- 7 PCR cycles. PCR cycle number was determined using qPCR reactions where the additional cycle numbers were those that corresponded to the inflection point of the qPCR curve. *Peak Calling:* ATAC-seq reads were trimmed to remove Nextera adapters using cutadapt (v8.25) and aligned to the genome using Bowtie2 (v2.3.2). All reads mapping to the mitochondrial genome were removed from further analyses. Peak calling was performed using macs2 (v2.1.1.20160309) using no model and an extension size of 200. Significant peaks were considered those which survived FDR correction ($q<0.05$). *HSV analysis*: The union set of significant peaks across time points was obtained, and peaks of a uniform length of 1kb were obtained by centering around the summit of the highest peak per locus in the union set. The counts per time point mapping to these 1kb union peaks were obtained and transformed to log2cpm format, normalizing by library size(defined as total number of reads in peaks per sample). Hue, saturation, and value were calculated using the same equations as with gene expression, using normalized $\log_2$(cpm) values as input.

2.5.8: in-situ promoter capture Hi-C (cHi-C)

Two technical replicates for each time point were performed derived from two unique differentiations for both SGBS and hypothalamic neurons. Each technical replicate was analyzed alone, and additionally the technical replicate raw sequencing data was merged and analyzed to produce merged datasets. Merged datasets and individual replicate datasets were used in downstream analyses. *In situ* promoter capture HiC was performed and analyzed as previously described[9,13]. 5 million SGBS cells per replicate were harvested, counted and crosslinked using a final 1% (v/v) concentration of formaldehyde for 10 minutes at room temperature while rocking. This reaction was quenched with 0.25M Glycine to a final concentration of 0.2M for 5 minutes and washed with 1X PBS. Cells were frozen in liquid nitrogen and stored at -80C until ready for the next stage of promoter-capture Hi-C processing. Each differentiation time point has two technical replicates and was sequenced on a full lane of an Illumina Hiseq 4000 machine to achieve sufficient read depth for interaction calling. *Data Analysis:* promoter capture Hi-C reads were aligned to the genome using Bowtie2 (v2.3.2) and technical artifacts were removed using HiCUP (v0.5.9). Significant interactions were detected over a background model of null expectation using CHiCAGO (v1.2.0). Only interactions with a CHiCAGO score > 5 at any time point were included in downstream analyses. Trans-chromosomal interactions and interactions between loci greater than 1 megabase apart were filtered from further analysis. Counts were normalized by library size using (TMM) normalization and transformed into counts-per-million (cpm). Hue, saturation, and value were calculated using the same equations as with gene expression, using normalized cpm values as input.

2.5.9: Massively Parallel Reporter Assay (MPRA)

Lead variants for BMI were taken from the 2015 GIANT consortium meta-analysis, which identified 97 independent significant loci. We searched 1,000 genomes phase 3 genotypes (ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502) for the 97 GWAS lead SNPs and obtained all CEU SNPs (Utah residents with Northern and Western European ancestry from the CEPH collection) within 50kb and with r^2 >.8 with a lead SNP. We only retained biallelic SNPs with MAF >= 5% (2,396 in total). Using these variants, MPRA oligo design was performed as previously described with modifications[16]. We synthesized 230bp long DNA fragments as seen below using a 100,000 oligonucleotide Agilent array. 5'-ACTGGCCGCTTCACTG-*enh*-GGTACCTCTAGA-*barcode*-AGATCGGAAGAGCGTCG-3' Each *enh* region was 175 base pairs of endogenous DNA context surrounding one of the biallelic 2,346 SNPs. Each allele of each biallelic variant was synthesized beside 18-19 unique 10bp DNA *barcodes*. Barcodes were randomly generated using a series of A,C,T,or Gs that did not contain three or more of the same base in a row and did not create Kpn1 or Xba1 restriction enzyme sites. We also later determined that barcodes should not end with the sequence "TCT", because it creates a restriction enzyme site with the beginning of the second constant region and they will thus be lost. Upon receipt, this fragment pool was dissolved in 100ul of nuclease free water and PCR amplified using the Micellula DNA emulsion and Purification Kit (EURx #E3600-01) in order to reduce amplification bias of particular oligos over others. This PCR adds homology arms onto the oligos and allowed us to use Gibson Assembly Master Mix (NEB # E2611S) to clone these oligos into a linearized pMPRA1 vector (addgene #49349) that was cut open using the SfiI restriction enzyme. This backbone + oligo insert vector was then linearized using a Kpn1 and Xba1 double digest and a 60bp truncated eGFP containing a minimal promoter and spacer sequence (141bp in size) was ligated in between the *enh* fragments and *barcodes* using T4 DNA ligase (NEB; # B0202S) at a 1:10 ratio of insert to vector and incubated at 16*C overnight. The resulting plasmid library was linearized a final time using Kpn1 and size selected for vectors containing all inserts (oligos + eGFP insert) on a 1% agarose gel. This is then religated using T4 DNA ligase and transformed until enough final plasmid is produced for all transfections. At each cloning and transformation step complexity must be maintained, so we counted colony forming units (CFUs) after each transformation and aimed to attain at least 100 million CFUs. To ensure the best transformation, all reactions were cleaned up with the Minelute

PCR purification kit (Qiagen; # 28004) and then further cleaned on a Millipore drop dialysis membrane (Millipore; #VSWP02500) for an hour before transformation. Cells were transformed into MegaX DH10B T1R electrocompetent bacteria (Invitrogen #C640003) and allowed to grow for only 7-9 hours after recovery to ensure likelihood of getting high CFUs without bias towards particular constructs. Once the final constructs were produced, they were transfected into GT1-7 cells (6 replicates), 3T3-L1 cells (7 replicates), HT22 cells (5 replicates), SGBS Day 0 cells (6 replicates), and SGBS Day 8 cells (5 replicates) as described above. Enough cells were transfected to achieve at minimum 10 million transfected cells per replicate with transfection efficiency estimated using a GFP control plasmid. Cells were collected 48 hours after transfection and flash frozen in liquid nitrogen until all replicates were collected. A replicate was considered "technical" if they were transfected with the same batch of DNA on different days with different cell passages. A replicate was considered "biological" if the input DNA library was separately cloned from the beginning from our Agilent oligonucleotides. MPRA experiments were designed to have 2-4 technical replicates for each of two biological replicates. After transfection, cells were lysed using a 20g needle and RNA was extracted using Qiagen RNeasy mini kit. RNA quality was assessed on an 1% agarose gel. mRNA was isolated from total RNA using Invitrogen Dynabeads (ThermoFisher #61006) and then treated with Promega RQ1 DNAse (Promega; M6101) for 1.5 hours at 37°C with an enzyme boost halfway through the reaction. The isolated and DNA plasmid depleted mRNA is then cleaned up with the Qiagen RNeasy mini kit and quantified using the Promega QuantiFluor RNA system (Promega #E3310). Importantly, enrichment of RNA transcripts emanating from our MPRA plasmid compared to MPRA DNA plasmid contamination is assessed at this point using qPCR primers targeted to the eGFP. To do this, 250ng of mRNA is converted to cDNA while 250ng of mRNA is run through the cDNA reaction without reverse transcriptase (RT). We then perform a qPCR to determine enrichment of eGFP transcripts between RT(+) and RT(-) samples. We set a threshold of a minimum of 8CT enrichment between DNA and RNA as a quality control check. All remaining mRNA is then converted to cDNA using Superscript III Reverse Transcriptase. cDNA is treated with RNAse A (invitrogen #12091-021) and RNAse T1 (ThermoFisher #EN0541) for one hour and then cleaned with the Qiagen Minelute PCR purification kit. 50ng of cDNA is then used as a PCR template for the final Illumina multiplexing primers. All available cDNA should be amplified. Two 50ng reactions of Input DNA (DNA used as transfection material) must also be PCR amplified with Illumina multiplexing primers at this point. Libraries were amplified with 10-11 PCR cycles using Q5 Hot Start High-Fidelity 2x Master Mix (NEB #M0494S) and pooled before clean up using Agencourt AMPure XP beads (0.6x + 1.2x double cleanup; Beckman Coulter; #A63882). Library quality was assessed using the Agilent DNA 1000 bioanalyzer chip (Agilent; #G2938-90014), where a single sharp peak of around 250bp is expected. Samples can then be sent for paired end NGS sequencing. A 25% PhiX genome spike in must be added to each sequencing run due to low sequencing complexity. *Data Analysis:* Barcode counts must be converted to the reverse complement before they can be matched with known barcodes. We required sequenced barcodes to be exact matches with expected barcode sequences. Count data is then analyzed for significance as previously described[85]. In essence, lowly expressed barcodes were removed and enhancer activity was determined from the remaining normalized counts using the following equation: Enhancer activity = log2 (output (CPM) - input (CPM)). Activity was then quantile normalized and enhancer *p* values were calculated using a one-sided Mann-Whitney U Test in R using the wilcox.test function. *P* values were corrected for multiple testing using the p.adjust function, method = “fdr”. All regions where at least one allele was determined to be a significant enhancer were then tested for enhancer modulating effects using a two-sided Mann-Whitney U test in R with *p* values adjusted for multiple testing as previously described. Enhancer modulating variants were retained for downstream analyses if they were significant in half of all technical replicates or both biological MPRA replicates.

Chapter 2.5.10: Luciferase Assays for Enhancer Validation

The luciferase assays used for MPRA validation had between 3-5 technical replicates per construct, where different DNA preps were used and the cells were transfected, collected, and analyzed on different days. Twenty-one regions containing at least one SNP that had an allele in a significant enhancer in either HT22 and/or 3T3-L1 cells were chosen for validation. These regions were PCRed from genomic DNA using Q5 Hot Start 2x Master Mix (NEB #M0494S) and were designed to be ~1kb in size. Each region was cloned into the pGL4.23 luciferase vector containing firefly Luciferase, and were tested for luciferase activity via co-transfection with renilla luiferase at a ratio of (1:50) in both 3T3-L1 cells and HT22 cells. Alleles were determined through Sanger sequencing. Renilla and firefly luciferase fluorescence was measured on a Promega GloMax microplate reader using the Dual-Luciferase Reporter Assay System (Promega #E1910). Firefly luciferase measurements were normalized to renilla measurements and then fold change over a control DNA region was calculated to determine enhancer activity.

2.5.11: s-LDSC Partitioned Heritability Analysis

Heritability per chromosome was calculated via LD score regression analysis using the ldsc package in R (v1.0.0) using Locke et al 2012 BMI GWAS summary statistics downloaded from the GIANT consortium. Briefly, .bim files from 1000 Genomes Phase 1 were downloaded and annotation files were created for each chromosome where chromosome was treated as a binary annotation. LD scores were then computed from these annotation files for input into partitioned heritability analysis. Summary statistics were filtered to contain only HapMap3 variants as advised.

Chapter 2.5.12: Transcription Factor Motif Analysis

All regions identified to be significant enhancers were included in this analysis. Regions were expanded to be 175bp (size of enhancers tested in MPRA) and then if two regions overlapped they were then merged so they would not become overrepresented in the analysis. The program findMotifsGenome.pl from HOMER[109](v4.8.3) was then used in addition to the -size flag to identify motifs that were overrepresented in significant MPRA enhancers from each cell line. These were compared to a size and base composition matched set of background sequences computed by HOMER to determine significance and *p* value. All *p* values from each cell line are included in the supplementary tables.

2.5.13: Calling EMVar Interactions with Promoters

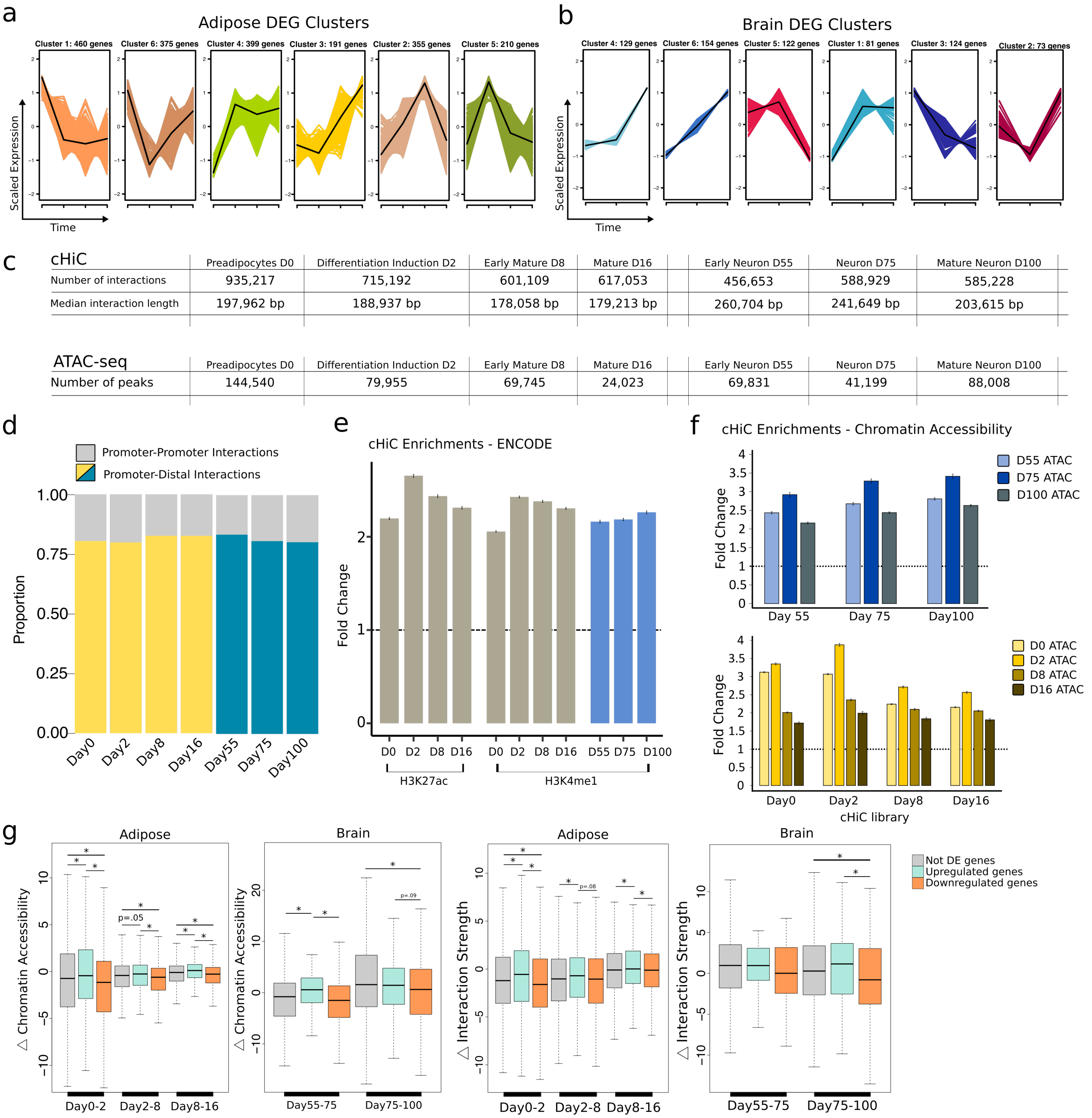
MPRA EMVars were considered to interact with a promoter if the distal end of the promoter interaction came within 1kb of the single base pair SNP location. EMVar SNP location and cHi-C BEDPE files were overlapped using the BEDtools (v2.27.1)[110] pairToBed function.

2.5.14: Gene Support Classes

To develop gene level support for each GWAS locus, we first binned EMVars into their respective loci. For class I genes, we required an EMVar to interact with that gene and the

EMVar must be a GTEx eQTL for that gene in the appropriate cell type (GTEx adipose tissues for adipose EMVars and GTEx Brain tissues for Brain EMVars). Class II genes interact with an EMVar in the appropriate cHi-C dataset and the EMVar is an eQTL for that gene in cell types other than the appropriate one. Class III genes were those that interacted with an EMVar in the appropriate cHi-C libraries or were an eGene for this SNP in the correct cell type. Class IV genes were eGenes for these EMVars in other cell types. Additionally, we only included genes in this analysis that were expressed > 1 TPM in at least 1 time point in their respective cell type from our RNA-seq (All genes with their classes from both cell types are shown in the Supplementary Tables)

2.6: Appendix A, Supplementary Figures

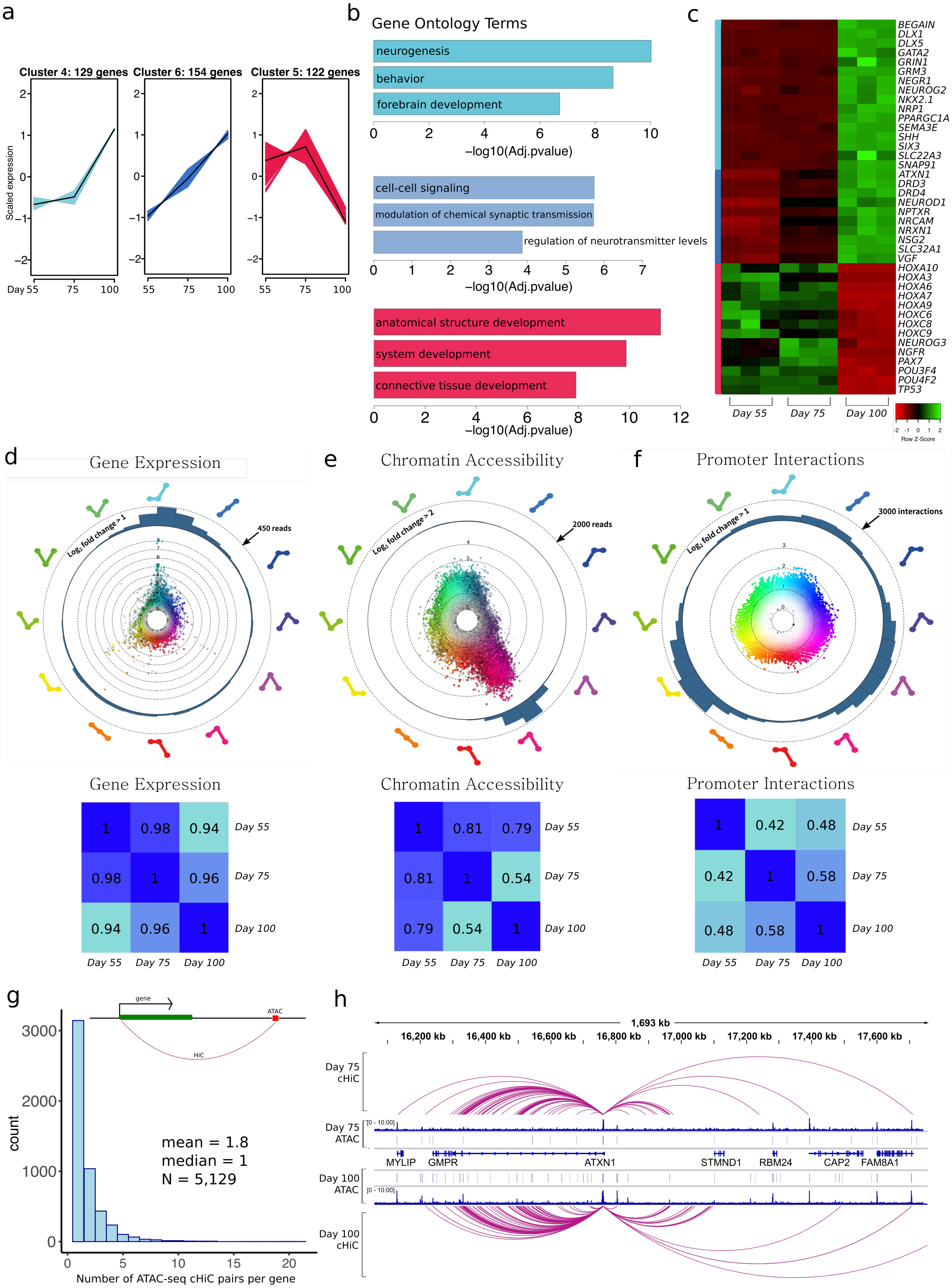


cHiC	Preadipocytes D0	Differentiation Induction D2	Early Mature D8	Mature D16	Early Neuron D55	Neuron D75	Mature Neuron D100
Number of interactions	935,217	715,192	601,109	617,053	456,653	588,929	585,228
Median interaction length	197,962 bp	188,937 bp	178,058 bp	179,213 bp	260,704 bp	241,649 bp	203,615 bp

ATAC-seq	Preadipocytes D0	Differentiation Induction D2	Early Mature D8	Mature D16	Early Neuron D55	Neuron D75	Mature Neuron D100
Number of peaks	144,540	79,955	69,745	24,023	69,831	41,199	88,008

Supplementary Figure S2.1: Features of functional annotations
a) Six fuzzy–c means clusters were identified for adipose and b) brain DEGs from the RNA-seq time course. The number of genes comprising each cluster, along with scaled expression across the time points is shown. c) Overview of median interaction length and number of interactions per time-point in the replicate-merged cHiC datasets. Number of ATAC-seq peaks from the replicate-merged time points. d) Bar plot depicting proportion of promoter-promoter interactions in merged cHiC libraries.

Supplementary Figure S2.1, continued. e,f) The promoter-distal ends of interactions are enriched for functional ChIP-seq peaks and ATAC-seq peaks compared to a distribution of randomly chosen, number-matched set of non-promoter MboI fragments within mappable genomic regions (N=100 iterations). The fold change of the observed overlap over our 100 randomized sets is presented. ChIP-seq datasets were obtained from Adipose Nuclei (E063) and Fetal Brain (E081) repositories from the Roadmap Epigenomics project. (All significant with p <0.05; Z-test) Error Bars: SD g) Genes were binned based on upregulation or downregulation across each time point. Plotted are the changes in interaction score or normalized ATAC-seq reads for ATAC peaks connected through these genes via a significant cHiC peak between each time point. *p <0.05; two-sided Mann-Whitney U test



Gene Expression

	Day 55	Day 75	Day 100
Day 55	1	0.98	0.94
Day 75	0.98	1	0.96
Day 100	0.94	0.96	1

Chromatin Accessibility

	Day 55	Day 75	Day 100
Day 55	1	0.81	0.79
Day 75	0.81	1	0.54
Day 100	0.79	0.54	1

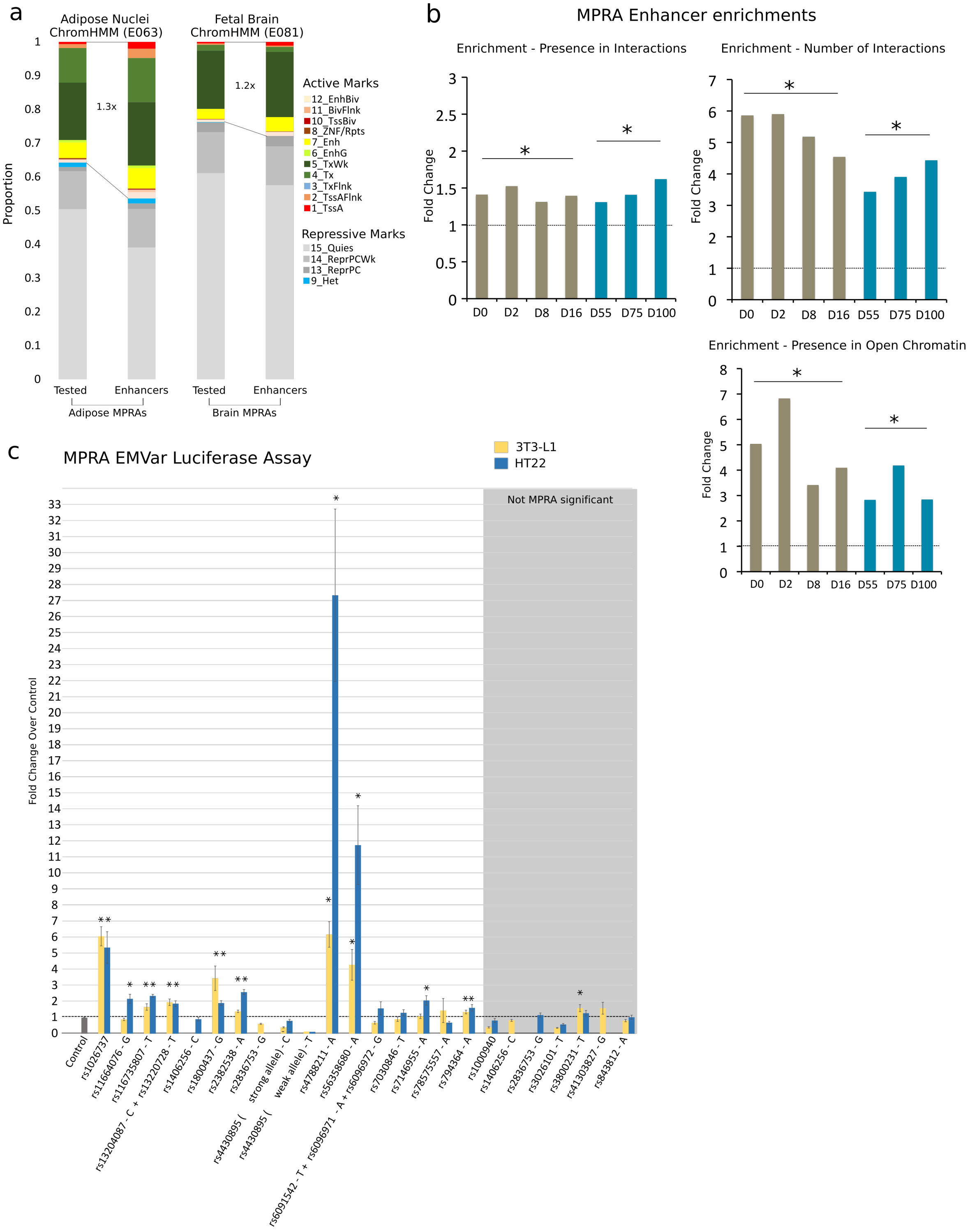
Promoter Interactions

	Day 55	Day 75	Day 100
Day 55	1	0.42	0.48
Day 75	0.42	1	0.58
Day 100	0.48	0.58	1

Supplementary Figure S2.2: Characterizing hypothalamic differentiation using genomic annotations

a) Brain DEGs were grouped via fuzzy-c clustering and the top three clusters with highest membership are illustrated.

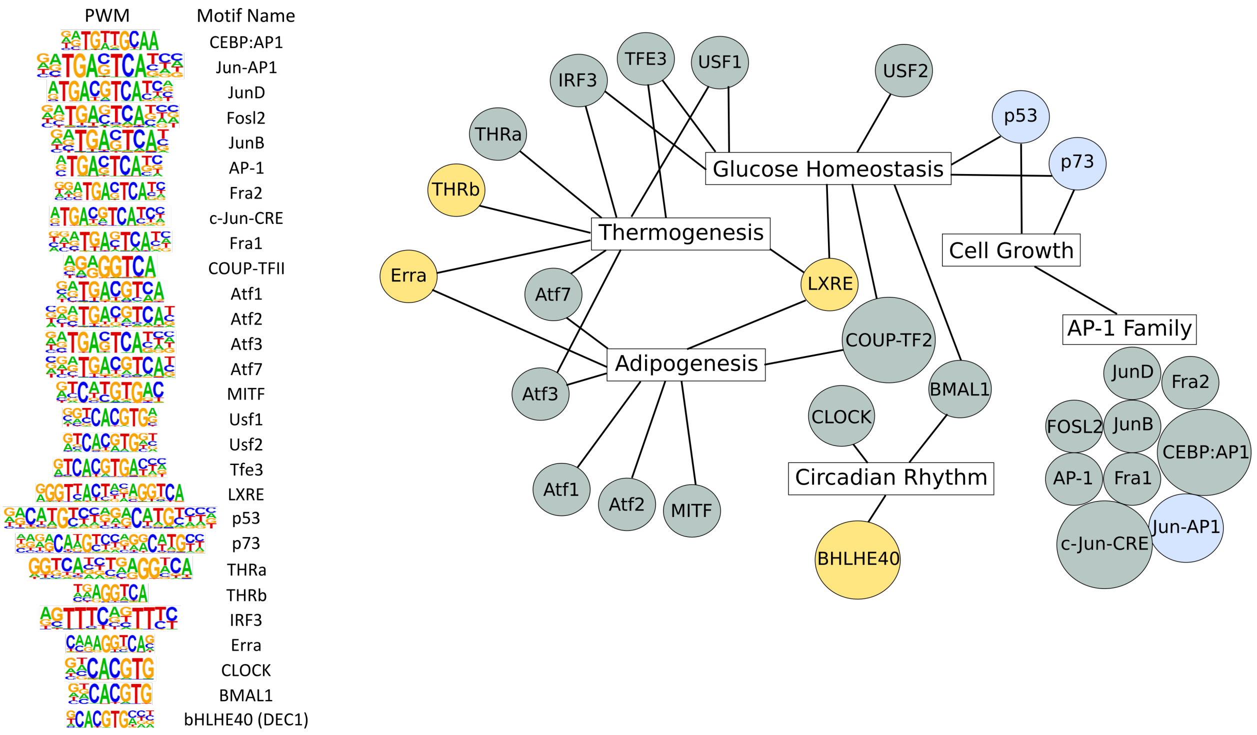
Supplementary Figure S2.2, continued. b) Significant Gene Ontology terms for the top three clusters. c) A heatmap of gene expression depicting genes from each of the top three clusters that are members of the enriched Gene Ontology terms. The leftmost colored bar indicates cluster membership and the columns are RNA-seq replicates. d-f) HSV transformation of gene expression dynamics, ATAC-seq accessibility, and cHiC interactions across differentiation. Each significant data point is categorized and colored based on the temporal pattern it displays shown by the guides on the periphery of each plot. The three nodes of each pattern represent day 55, day 75, and day 100 of neuronal differentiation. The distance of each point from the center of the circle represents maximum log2 fold change, and color transparency represents the relative number of reads for that data point. Below, heatmaps of Pearson's *r* correlation coefficients estimate overall similarity between time points. g) On average, a promoter interacts with 1-2 ATAC-seq peaks via a cHiC interaction across time (interactions and ATAC peaks were not required to be significant at the same time point). h) View of significant cHiC interactions emanating from the promoter of the *ATXN1* gene, which becomes significantly upregulated between differentiation days 75-100. ATAC-seq reads and significant ATAC-seq peaks at day 75 and day100 are also shown.



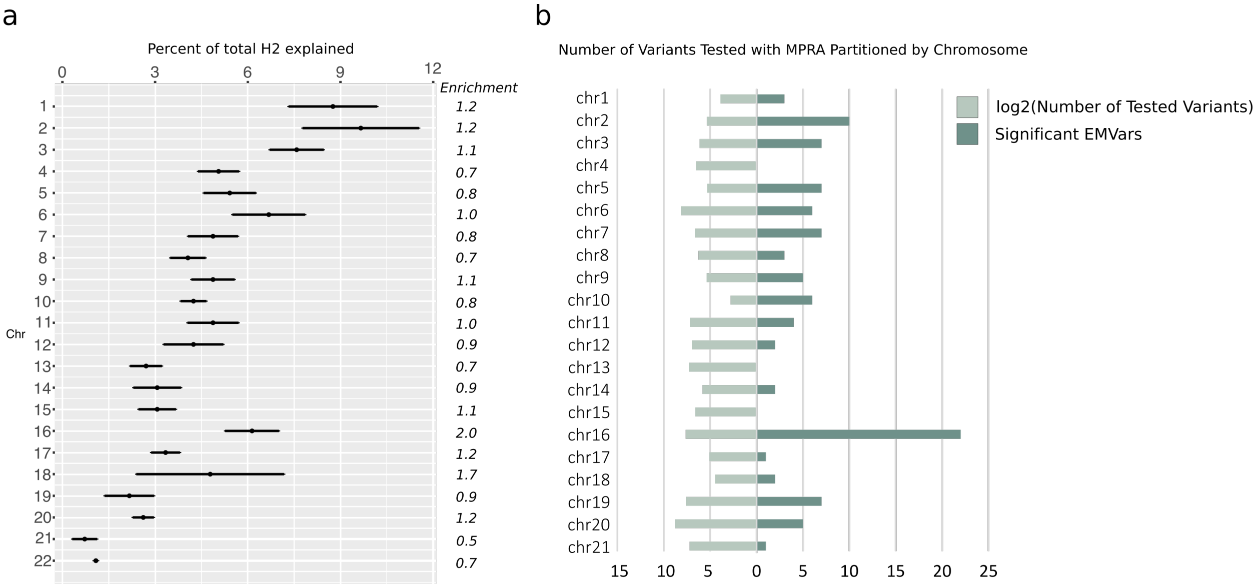
Supplementary Figure S2.3: MPRA enhancers are functional

a) MPRA enhancers are enriched for Epigenome Roadmap's 15 state ChromHMM functional marks in adipose nuclei or fetal brain compared to all tested variants b) MPRA enhancers are enriched for presence in cHiC interactions, number of interactions per enhancer, and open chromatin compared to non-significant regions. (**p* < 0.05; two-sided Student's t-test)

Supplementary Figure S2.3, continued. c) Luciferase assay results for ~1kb sized regions containing an EMVar. Regions were chosen at random, and represent a full spectrum of MPRA enhancer p-values. Because of this, non-EMVar rs1026737 and rs10000940 were included because they had very low and high enhancer *p* values, respectively. If the allele that was captured was not a significant enhancer, the result is colored with a grey background. Interestingly, for the rs4430895 region, we were able to clone both alleles, and although neither allele was an enhancer using the luciferase assay, the allele predicted to be stronger with MPRA had higher Luc2 expression compared to the weak allele. *$p < 0.05$, two-tailed Student's t-test (N=3-4 replicates). Error bars = SEM



Supplementary Figure S2.4: Transcription factors in obesity-associated loci
(left) Position weight matrices for enriched transcription factor motifs from HOMER. Each motif was enriched in either MPRA adipose or brain enhancers (HOMER adjusted *p* value < 0.05). (right) Transcription factors are connected to a BMI relevant phenotype with a line if these factors play a role in that biological process (significant in both brain and adipose = grey circle, significant in adipose = yellow, and significant in brain = blue).



Supplementary Figure S2.5: Chromosome 16 harbors an excess of obesity heritability
a) s-LDSC estimated proportion of total heritability explained per chromosome is depicted along with heritability enrichment values. Error bars = SEM b) Number of EMVars compared to the number of variants tested with MPRA stratified per chromosome.

2.7: Appendix B, Supplementary Tables

Table S2.1: Homer Transcription Factor Enrichment p-values

Motif Name	P-value	Cell Type
Usf2(bHLH)/C2C12-Usf2-ChIP-Seq(GSE36030)/Homer	1.00E-08	3T3-L1
TFE3(bHLH)/MEF-TFE3-ChIP-Seq(GSE75757)/Homer	1.00E-06	3T3-L1
MITF(bHLH)/MastCells-MITF-ChIP-Seq(GSE48085)/Homer	1.00E-06	3T3-L1
CLOCK(bHLH)/Liver-Clock-ChIP-Seq(GSE39860)/Homer	1.00E-05	3T3-L1
USF1(bHLH)/GM12878-Usf1-ChIP-Seq(GSE32465)/Homer	1.00E-04	3T3-L1
JunD(bZIP)/K562-JunD-ChIP-Seq/Homer	1.00E-04	3T3-L1
Atf1(bZIP)/K562-ATF1-ChIP-Seq(GSE31477)/Homer	1.00E-04	3T3-L1
c-Jun-CRE(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	1.00E-03	3T3-L1
AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1.00E-03	3T3-L1
Atf7(bZIP)/3T3L1-Atf7-ChIP-Seq(GSE56872)/Homer	1.00E-03	3T3-L1
Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer	1.00E-03	3T3-L1
BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer	1.00E-02	3T3-L1
c-Myc(bHLH)/LNCAP-cMyc-ChIP-Seq(Unpublished)/Homer	1.00E-02	3T3-L1
Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer	1.00E-02	3T3-L1
Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer	1.00E-02	3T3-L1
Atf2(bZIP)/3T3L1-Atf2-ChIP-Seq(GSE56872)/Homer	1.00E-02	3T3-L1
THRa(NR)/C17.2-THRa-ChIP-Seq(GSE38347)/Homer	1.00E-02	3T3-L1
BMAL1(bHLH)/Liver-Bmal1-ChIP-Seq(GSE39860)/Homer	1.00E-02	3T3-L1
Fosl2(bZIP)/3T3L1-Fosl2-ChIP-Seq(GSE56872)/Homer	1.00E-02	3T3-L1
COUP-TFII(NR)/Artia-Nr2f2-ChIP-Seq(GSE46497)/Homer	1.00E-02	3T3-L1
bHLHE40(bHLH)/HepG2-BHLHE40-ChIP-Seq(GSE31477)/Homer	1.00E-02	3T3-L1
IRF3(IRF)/BMDM-Irf3-ChIP-Seq(GSE67343)/Homer	1.00E-02	3T3-L1
CRX(Homeobox)/Retina-Crx-ChIP-Seq(GSE20012)/Homer	1.00E-02	3T3-L1
JunB(bZIP)/DendriticCells-Junb-ChIP-Seq(GSE36099)/Homer	1.00E-02	3T3-L1
n-Myc(bHLH)/mES-nMyc-ChIP-Seq(GSE11431)/Homer	1.00E-02	3T3-L1
E-box(bHLH)/Promoter/Homer	1.00E-02	3T3-L1
CEBP:AP1(bZIP)/ThioMac-CEBPb-ChIP-Seq(GSE21512)/Homer	1.00E-02	3T3-L1
CLOCK(bHLH)/Liver-Clock-ChIP-Seq(GSE39860)/Homer	1.00E-04	GT1-7
Usf2(bHLH)/C2C12-Usf2-ChIP-Seq(GSE36030)/Homer	1.00E-03	GT1-7
NFAT:AP1(RHD,bZIP)/Jurkat-NFATC1-ChIP-Seq(Jolma_et_al.)/Homer	1.00E-03	GT1-7
ZNF143|STAF(Zf)/CUTLL-ZNF143-ChIP-Seq(GSE29600)/Homer	1.00E-02	GT1-7
Bapx1(Homeobox)/VertebralCol-Bapx1-ChIP-Seq(GSE36672)/Homer	1.00E-02	GT1-7
c-Jun-CRE(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	1.00E-02	GT1-7
ZFX(Zf)/mES-Zfx-ChIP-Seq(GSE11431)/Homer	1.00E-02	GT1-7
MITF(bHLH)/MastCells-MITF-ChIP-Seq(GSE48085)/Homer	1.00E-02	GT1-7
p53(p53)/Saos-p53-ChIP-Seq(GSE15780)/Homer	1.00E-02	GT1-7
p53(p53)/Saos-p53-ChIP-Seq/Homer	1.00E-02	GT1-7
USF1(bHLH)/GM12878-Usf1-ChIP-Seq(GSE32465)/Homer	1.00E-02	GT1-7
IRF3(IRF)/BMDM-Irf3-ChIP-Seq(GSE67343)/Homer	1.00E-02	GT1-7
p73(p53)/Trachea-p73-ChIP-Seq(PRJNA310161)/Homer	1.00E-02	GT1-7

Table S2.1: Homer Transcription Factor Enrichment p-values (continued)

Motif Name	P-value	Cell Type
THRa(NR)/C17.2-THRa-ChIP-Seq(GSE38347)/Homer	1.00E-02	GT1-7
Atf7(bZIP)/3T3L1-Atf7-ChIP-Seq(GSE56872)/Homer	1.00E-02	GT1-7
Atf1(bZIP)/K562-ATF1-ChIP-Seq(GSE31477)/Homer	1.00E-02	GT1-7
TFE3(bHLH)/MEF-TFE3-ChIP-Seq(GSE75757)/Homer	1.00E-02	GT1-7
Nkx3.1(Homeobox)/LNCaP-Nkx3.1-ChIP-Seq(GSE28264)/Homer	1.00E-02	GT1-7
Usf2(bHLH)/C2C12-Usf2-ChIP-Seq(GSE36030)/Homer	1.00E-06	HT22
Atf1(bZIP)/K562-ATF1-ChIP-Seq(GSE31477)/Homer	1.00E-05	HT22
AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1.00E-04	HT22
BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer	1.00E-04	HT22
Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer	1.00E-04	HT22
Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer	1.00E-04	HT22
TFE3(bHLH)/MEF-TFE3-ChIP-Seq(GSE75757)/Homer	1.00E-04	HT22
MITF(bHLH)/MastCells-MITF-ChIP-Seq(GSE48085)/Homer	1.00E-04	HT22
Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer	1.00E-03	HT22
Atf7(bZIP)/3T3L1-Atf7-ChIP-Seq(GSE56872)/Homer	1.00E-03	HT22
c-Jun-CRE(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	1.00E-03	HT22
JunB(bZIP)/DendriticCells-Junb-ChIP-Seq(GSE36099)/Homer	1.00E-03	HT22
CLOCK(bHLH)/Liver-Clock-ChIP-Seq(GSE39860)/Homer	1.00E-03	HT22
Fosl2(bZIP)/3T3L1-Fosl2-ChIP-Seq(GSE56872)/Homer	1.00E-03	HT22
Jun-AP1(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	1.00E-03	HT22
JunD(bZIP)/K562-JunD-ChIP-Seq/Homer	1.00E-03	HT22
Atf2(bZIP)/3T3L1-Atf2-ChIP-Seq(GSE56872)/Homer	1.00E-03	HT22
THRa(NR)/C17.2-THRa-ChIP-Seq(GSE38347)/Homer	1.00E-02	HT22
USF1(bHLH)/GM12878-Usf1-ChIP-Seq(GSE32465)/Homer	1.00E-02	HT22
COUP-TFII(NR)/Artia-Nr2f2-ChIP-Seq(GSE46497)/Homer	1.00E-02	HT22
Cdx2(Homeobox)/mES-Cdx2-ChIP-Seq(GSE14586)/Homer	1.00E-02	HT22
p73(p53)/Trachea-p73-ChIP-Seq(PRJNA310161)/Homer	1.00E-02	HT22
Mef2a(MADS)/HL1-Mef2a.biotin-ChIP-Seq(GSE21529)/Homer	1.00E-02	HT22
Nrf2(bZIP)/Lymphoblast-Nrf2-ChIP-Seq(GSE37589)/Homer	1.00E-02	HT22
BMAL1(bHLH)/Liver-Bmal1-ChIP-Seq(GSE39860)/Homer	1.00E-02	HT22
SpiB(ETS)/OCILY3-SPIB-ChIP-Seq(GSE56857)/Homer	1.00E-02	HT22
GATA3(Zf),DR4/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	1.00E-02	HT22
FOXM1(Forkhead)/MCF7-FOXM1-ChIP-Seq(GSE72977)/Homer	1.00E-02	HT22
CEBP:AP1(bZIP)/ThioMac-CEBPb-ChIP-Seq(GSE21512)/Homer	1.00E-02	HT22
Atf1(bZIP)/K562-ATF1-ChIP-Seq(GSE31477)/Homer	0.000000001	SGBS_Preadipocytes(D0)
Atf7(bZIP)/3T3L1-Atf7-ChIP-Seq(GSE56872)/Homer	0.00000001	SGBS_Preadipocytes(D0)
COUP-TFII(NR)/Artia-Nr2f2-ChIP-Seq(GSE46497)/Homer	0.0000001	SGBS_Preadipocytes(D0)
THRa(NR)/C17.2-THRa-ChIP-Seq(GSE38347)/Homer	0.00001	SGBS_Preadipocytes(D0)
GSC(Homeobox)/FrogEmbryos-GSC-ChIP-Seq(DRA000576)/Homer	0.00001	SGBS_Preadipocytes(D0)
LXRE(NR),DR4/RAW-LXRb.biotin-ChIP-Seq(GSE21512)/Homer	0.00001	SGBS_Preadipocytes(D0)
Mef2a(MADS)/HL1-Mef2a.biotin-ChIP-Seq(GSE21529)/Homer	0.0001	SGBS_Preadipocytes(D0)
Mef2c(MADS)/GM12878-Mef2c-ChIP-Seq(GSE32465)/Homer	0.0001	SGBS_Preadipocytes(D0)

Table S2.1: Homer Transcription Factor Enrichment p-values (continued)

Motif Name	P-value	Cell Type
THRb(NR)/Liver-NR1A2-ChIP-Seq(GSE52613)/Homer	0.001	SGBS_Preadipocytes(D0)
c-Jun-CRE(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	0.001	SGBS_Preadipocytes(D0)
Atf2(bZIP)/3T3L1-Atf2-ChIP-Seq(GSE56872)/Homer	0.001	SGBS_Preadipocytes(D0)
AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	0.001	SGBS_Preadipocytes(D0)
BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer	0.001	SGBS_Preadipocytes(D0)
Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer	0.001	SGBS_Preadipocytes(D0)
Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer	0.001	SGBS_Preadipocytes(D0)
Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer	0.01	SGBS_Preadipocytes(D0)
Mef2b(MADS)/HEK293-Mef2b.V5-ChIP-Seq(GSE67450)/Homer	0.01	SGBS_Preadipocytes(D0)
CRX(Homeobox)/Retina-Crx-ChIP-Seq(GSE20012)/Homer	0.01	SGBS_Preadipocytes(D0)
Bapx1(Homeobox)/VertebralCol-Bapx1-ChIP-Seq(GSE36672)/Homer	0.01	SGBS_Preadipocytes(D0)
Usf2(bHLH)/C2C12-Usf2-ChIP-Seq(GSE36030)/Homer	0.01	SGBS_Preadipocytes(D0)
Erra(NR)/HepG2-Erra-ChIP-Seq(GSE31477)/Homer	0.01	SGBS_Preadipocytes(D0)
JunB(bZIP)/DendriticCells-Junb-ChIP-Seq(GSE36099)/Homer	0.01	SGBS_Preadipocytes(D0)
Fosl2(bZIP)/3T3L1-Fosl2-ChIP-Seq(GSE56872)/Homer	0.01	SGBS_Preadipocytes(D0)
GATA(Zf),IR3/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	0.01	SGBS_Preadipocytes(D0)
Atf1(bZIP)/K562-ATF1-ChIP-Seq(GSE31477)/Homer	1.00E-09	SGBS_Adipocytes(D8)
COUP-TFII(NR)/Artia-Nr2f2-ChIP-Seq(GSE46497)/Homer	1.00E-09	SGBS_Adipocytes(D8)
Mef2c(MADS)/GM12878-Mef2c-ChIP-Seq(GSE32465)/Homer	1.00E-07	SGBS_Adipocytes(D8)
GSC(Homeobox)/FrogEmbryos-GSC-ChIP-Seq(DRA000576)/Homer	1.00E-06	SGBS_Adipocytes(D8)
THRb(NR)/Liver-NR1A2-ChIP-Seq(GSE52613)/Homer	1.00E-06	SGBS_Adipocytes(D8)
Atf7(bZIP)/3T3L1-Atf7-ChIP-Seq(GSE56872)/Homer	1.00E-06	SGBS_Adipocytes(D8)
THRa(NR)/C17.2-THRa-ChIP-Seq(GSE38347)/Homer	1.00E-06	SGBS_Adipocytes(D8)
LXRE(NR),DR4/RAW-LXRb.biotin-ChIP-Seq(GSE21512)/Homer	1.00E-05	SGBS_Adipocytes(D8)
Mef2a(MADS)/HL1-Mef2a.biotin-ChIP-Seq(GSE21529)/Homer	1.00E-04	SGBS_Adipocytes(D8)
Mef2b(MADS)/HEK293-Mef2b.V5-ChIP-Seq(GSE67450)/Homer	1.00E-04	SGBS_Adipocytes(D8)
CRX(Homeobox)/Retina-Crx-ChIP-Seq(GSE20012)/Homer	1.00E-04	SGBS_Adipocytes(D8)
Usf2(bHLH)/C2C12-Usf2-ChIP-Seq(GSE36030)/Homer	1.00E-03	SGBS_Adipocytes(D8)
Erra(NR)/HepG2-Erra-ChIP-Seq(GSE31477)/Homer	1.00E-02	SGBS_Adipocytes(D8)
Pitx1(Homeobox)/Chicken-Pitx1-ChIP-Seq(GSE38910)/Homer	1.00E-02	SGBS_Adipocytes(D8)
Atf2(bZIP)/3T3L1-Atf2-ChIP-Seq(GSE56872)/Homer	1.00E-02	SGBS_Adipocytes(D8)
Nkx3.1(Homeobox)/LNCaP-Nkx3.1-ChIP-Seq(GSE28264)/Homer	1.00E-02	SGBS_Adipocytes(D8)
c-Jun-CRE(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	1.00E-02	SGBS_Adipocytes(D8)
Bapx1(Homeobox)/VertebralCol-Bapx1-ChIP-Seq(GSE36672)/Homer	1.00E-02	SGBS_Adipocytes(D8)

Table S2.1: Homer Transcription Factor Enrichment p-values
Transcription factor motifs identified with HOMER enrichment analysis for each MPRA library. Only significant transcription motifs (HOMER adjusted p-value <0.05) were kept for downstream analysis, which is shown in Supplementary Figure S2.4

Table S2.2: Adipose Genes and Class Rank

Gene	Adipose Class	Gene	Adipose Class	Gene	Adipose Class
ARL3	1	DHX34	3	SEZ6L2	3
ATXN2L	1	DMWD	3	SIX5	3
BCS1L	1	DNAJC27	3	SLC1A5	3
C16orf62(VPS35L)	1	EIF3C	3	SMAD3	3
HSD17B12	1	EML2	3	SSBP4	3
INO80E	1	ERCC1	3	STIL	3
KCTD15	1	ETS2	3	SULT1A1	3
KNOP1	1	FAM178A	3	TAOK2	3
MAP2K5	1	FBXO46	3	TBC1D10B	3
NFATC2IP	1	FKBP8	3	TFDP2	3
NPC1	1	FOSB	3	TMBIM1	3
POC5	1	GATSL2	3	TMEM160	3
SH2B1	1	GDF15	3	TMEM241	3
SULT1A2	1	GIPR	3	TTC39C	3
TUFM	1	GK5	3	TUBA4A	3
USP37	1	GPATCH1	3	UBXN2A	3
ZNF142	1	GPI	3	VASP	3
AS3MT	2	GTF2IRD2	3	YPEL3	3
ATP2A1	2	HIF1AN	3	ZBTB38	3
C10ORF32(BORCS7)	2	HIRIP3	3	ZC3H4	3
CCDC101	2	IRX3	3	ZFP64	3
CMPK1	2	IRX5	3	ZNF181	3
CNNM2	2	ISYNA1	3	ANKRD39	4
COL4A3BP	2	JUND	3	C15ORF61	4
FTO	2	KIF22	3	CNOT9	4
GDPD3	2	KLHL26	3	CYP27A1	4
IFI30	2	LAMA3	3	EIF3CL	4
NT5C2	2	LAT	3	FHIT	4
NUPR1	2	LPCAT2	3	GPRC5B	4
PGPEP1	2	LSM14A	3	IQCH	4
POLK	2	MAPK3	3	IQCK	4
RABEP2	2	MARK4	3	KCTD13	4
SAE1	2	MEIS3	3	LIN7C	4
SFXN2	2	MMP2	3	METTL15	4
SPNS1	2	MVP	3	MFSD13A	4
WBP1L	2	NDUFB8	3	NPIPB12	4
ADCY3	3	NPIPB9	3	NPIPB6	4
ALDOA	3	OPA3	3	NPIPB8	4
AP2S1	3	PAGR1	3	POM121C	4
BBC3	3	PDE4C	3	PPP4C	4
BRWD1	3	PEPD	3	PTRHD1	4
C18orf8(RMC1)	3	PLCD4	3	SNRPD2	4
CALHM2	3	PPM1N	3	STK36	4
CCDC146	3	PPP1R13L	3	SYMPK	4
CCDC171	3	PRKD1	3	SYT15	4
CCDC9	3	PRKD2	3	TMEM219	4
CD2BP2	3	PRRT2	3	TRIM73	4
CD3EAP	3	RASA2	3	TTLL4	4
CDIPT	3	RBL2	3	UBE2E3	4
CDV3	3	RCC1L	3	XPO6	4
CENPO	3	RPGRIP1L	3	ZFHX4	4
CTDSP1	3	RQCD1	3	ZNF771	4
CWC22	3	RTN2	3		

Table S2.2: Adipose Genes and Class Rank
All genes identified using the classification system outlined in Figure 2.3 with their respective ranks in adipose.

Table S2.3: Hypothalamic genes and class rank

Gene	Brain Class	Gene	Brain Class	Gene	Brain Class
ZNF142	1	RTN2	3	CEP89	3
TUFM	1	RNF25	3	CEP57L1	3
SH2B1	1	PXDN	3	CDIPT	3
RABEP1	1	PSMB6	3	CCDC171	3
POC5	1	PRRT2	3	CCDC146	3
NUPR1	1	PRKD2	3	CALML4	3
NUP88	1	PPP4C	3	C2ORF44	3
NFATC2IP	1	PPM1N	3	C19ORF40	3
MAP2K5	1	POMC	3	ATPAF1	3
INO80E	1	POM121C	3	ASPHD1	3
HSD17B12	1	PNKD	3	ARMC2	3
ZNF232	2	PLCD4	3	API5	3
STK36	2	PFN1	3	ALDOA	3
SPNS1	2	PEPD	3	AKTIP	3
SGF29	2	PACRG	3	AGBL4	3
SBK1	2	OPA3	3	ADCY3	3
RPAIN	2	MVP	3	ACP1	3
RABEP2	2	MOB3C	3	ZNF771	4
POLK	2	MAZ	3	ZFHX4	4
PGPEP1	2	LSM14A	3	XPO6	4
KCTD13	2	KNOP1	3	VMO1	4
GDPD3	2	KLHL26	3	UBE2E3	4
DOC2A	2	KIF22	3	TTLL4	4
COL4A3BP	2	KIF1C	3	TMEM219	4
CMPK1	2	KIAA1683	3	SYMPK	4
CAMTA2	2	KDX1	3	SNRPD2	4
C16ORF62	2	KCTD6	3	RPTOR	4
C15ORF61	2	IRX3	3	RCC1L	4
BCS1L	2	INCA1	3	PTRHD1	4
ATXN2L	2	HMGCR	3	PIAS1	4
ATP2A1	2	HIRIP3	3	PARK2	4
YPEL3	3	GSG1L	3	NPTX1	4
VASP	3	GPRC5B	3	NPIPB12	4
USP37	3	GPI	3	NLRP1	4
UPF1	3	GDF15	3	NCOA1	4
TMEM59L	3	FOXO3	3	MIS12	4
TBX6	3	FOXA3	3	MAPK3	4
TAOK2	3	FOSB	3	KCTD15	4
TAL1	3	FEM1B	3	IQCK	4
STIL	3	FBXO46	3	IQCH	4
SSBP4	3	FAM57B	3	GRID1	4
SRCAP	3	ENDOV	3	GCNT4	4
SNX3	3	EML2	3	FTO	4
SMIM7	3	EIF4A3	3	FOXG1	4
SMAD6	3	EIF3C	3	FHIT	4
SMAD3	3	DMWD	3	EFR3B	4
SLC5A5	3	DDX49	3	DNAJC27	4
SKOR1	3	DACT3	3	CYP27A1	4
SIX5	3	CXCL16	3	CENPO	4
SH3YL1	3	COPE	3	C1QBP	4
SEZ6L2	3	CNOT9	3	ALKAL2(FAM150B)	4
SESN1	3	CLN6	3		
S1PR3	3	CHMP6	3		

Table S2.3: Hypothalamic genes and class rank
All genes identified using the classification system outlined in Figure 2.3 with their respective ranks in hypothalamus.

CHAPTER 3: INVESTIGATION OF TWO INDEPENDENT GWAS ASSOCIATIONS THAT EXHIBIT EXTENSIVE PLEIOTROPY

3.1 Abstract

Using the insights gained in Chapter 2, we sought to better understand a region that emerged from our analysis located on chromosome 16. Within chromosome 16p11.2 we observed EMVars within two independent GWAS regions, the *ATP2A1* locus and *SBK1* locus, that participated in cross-locus cHi-C interactions extending over 500kb in distance to land within the reciprocal region. Ten significant EMVars were identified across the two loci, each of them interacting with several promoters. Not only did these variants share interacting genes, but were also eQTLs for many of the shared genes within this region. In order to further investigate this complex network of variants, we used CRISPR-cas9 editing to knock out 0.75-1.2kb regions of the genome containing the two enhancers, one containing the EMVar rs2650492, and the other containing the EMVar rs9972768, and observed changes in local gene expression. With this data, we demonstrate that these enhancers converge to regulate the gene *SBK1,* a serine/threonine protein kinase implicated in the control of brain developmental processes. To better understand the evidence of pleiotropy suggested by cHi-C and eQTL information, we then went on to target the enhancer containing rs2650492 using CRISPRi under a secondary cellular context. We show this enhancer regulates another gene, *NUPR1*, under these conditions. This supports our previous observation that enhancers within these two independent GWAS loci co-localize within the nucleus and coordinate to regulate gene expression. Secondarily, we provide evidence that these functional variants may commonly affect more than one gene depending on the cell type and developmental time point assayed. Thus, when investigating a non-coding element's function, it

may be important to assess gene expression under unique spatio-temporal conditions important for your disease of interest to have a full understanding of a SNP's ability to modulate disease risk.

3.2 Introduction

While the integration of our functional genomics datasets was able to resolve regions such as the *MAP2K5* locus, where a single class I gene emerged as the likely target of the genetic association with obesity, we also uncovered unexpected complexities amongst other loci. As previously mentioned, in spite of its modest length, chromosome 16 had the strongest enrichment for obesity heritability of all human chromosomes and contained the largest number of EMVars in our study (Figure 2.2f). This overrepresentation was partly due to a hotspot of EMVars within a 600kb span on chromosome 16p11.2, which harbored two independent GWAS loci and an overabundance of EMVars (Supplementary Figure S3.1a). These were the *SBK1* and *ATP2A1* loci, which were named for the gene closest to the lead variant.

The *SBK1* and *ATP2A1* association regions have lead variants separated by over 500kb, making them independent GWAS signals. In this interval, local segmental duplications in the great apes lineage have resulted in new genes and transcripts in the human genome[1,111]. These repeat regions leave this region and its surrounding context vulnerable to structural variants, which resulted in large, rare, gene deletions and duplications in the *ATP2A1* locus[112]. Deletions within this locus have been implicated in highly penetrant forms of obesity[113] as well as developmental delay[114].

Although the *ATP2A1* region has been implicated in obesity risk, the gene or genes underlying this GWAS association have not been conclusively determined. The gene with the most evidence for support is *SH2B1*, a ubiquitously expressed adaptor protein located within the *ATP2A1* locus that enhances leptin and insulin signaling pathway potentiation[43,115]. Heterozygous loss-of-function alleles for *SH2B1* have been associated with early-onset obesity as well as insulin resistance and behavioral abnormalities in humans[103], and the highly penetrant forms of obesity arising from deletions in the 220kb *ATP2A1* region frequently encompass this gene. Although this gene seems like the likely causal gene of the association, additional deletions, duplications and inversions associated with obesity within this region do not affect the expression of *SH2B1*[116,117], suggesting the presence of other obesity modulating genes within the 16p11.2 locus.

What immediately became clear after integrating our genome-wide annotations was that functional variants within both the *ATP2A1* and *SBK1* loci participated in very long-range cHi-C interactions that extended over the 500kb distance to land within and between the reciprocal locus in both adipose and brain. Additionally, the EMVars within these two loci were eQTLs for, and physically interacted with, some of the largest number of genes in our dataset, including *SH2B1* but also others. Because of this interesting phenomenon, we wanted to better understand how this region contributes to disease risk, and to determine whether enhancers within these distinct disease associated loci individually or coordinately regulate the expression of genes in hypothalamic neurons. A better understanding of these regions would shed light into a complex network of variants that may regulate multiple obesity-relevant genes, additionally it would serve to test the ability of our functional annotations to uncover gene regulatory insights at obesity GWAS loci.

3.3 Results

3.3.1 Extensive variant and gene level pleiotropy within obesity-associated loci

In total, 10 EMVars were identified within the *SBK1* and *ATP2A1* regions. Using our cHi-C data, we were able to observe that EMVars within these two regions formed an extensive network of long-range interactions with promoters within and between the reciprocal locus (Figure 3.1a, Supplementary Figure S3.2a). This suggested to us that several of the genes in this megabase region may be co-regulated by a set of shared enhancers important for obesity risk.

The first of these two regions harbored rs2650492, a lead SNP emanating from the 3'UTR of *SBK1*. This SNP is not only associated with BMI, but also with other body weight phenotypes in the UK Biobank[118]. We identified three EMVars in this region. Two of these were not present in the GTEx database nor did they participate in cHi-C interactions and were therefore discarded from future investigation (Supplementary Figure S3.1a). But the third, rs2650492, is the lead SNP of this locus and is an eQTL for 18 nearby genes across GTEx cell types, including 5 in adipose and 7 in brain. This EMVar also participated in cHi-C interactions with 15 genes, including those that extended beyond its locus into the neighboring *ATP2A1* locus over 500kb away (Figure 3.1a). In our MPRA datasets the GWAS risk allele rs2650492-A decreased enhancer activity in both adipose and brain cell lines. This variant localized to open chromatin in our data as well as in a DNaseI cluster that encompassed 85/125 ENCODE cell types (Figure 3.1b). Together, this gives us high confidence that this variant may be a causal variant in this locus. In order to better understand the trans-acting factors that are bound to this enhancer we searched ENCODE data for transcription factors binding to this region of commonly open chromatin. In 56 ChIP-seq libraries from different samples, CTCF was bound to this region (Supplementary Figure S3.3a). We ran JASPAR to identify transcription factor binding sites and identified three high confidence CTCF binding sites (score >11.0) within this region of open chromatin (Supplementary Figure S3.2). Although the majority of CTCF binding peaks center on the highest scoring CTCF site (#2), we wanted to perform allele-specific read mapping of CTCF ChIP-seq data to determine whether this variant could be affecting CTCF binding to chromatin in an allele-specific manner to the second highest scoring CTCF site (#3) which contained the rs2650492 variant. There were 7 heterozygous cell lines in the ENCODE database that had appreciable CTCF ChIP-seq read coverage (> 10 reads) over the rs2650492 variant. None of these lines had statistically significant deviation from the expected 50% read coverage over each allele, suggesting that CTCF binding is not affected by this variant within these cell types (Supplementary Figure S3.3c). This suggests that other transcription factors are binding to this SNP, and that CTCF is not bound directly to this enhancer, but is bound near by, likely allowing it to anchor promoters and participate in the long-range interactions we observe.

The second region in this chromosome 16 locus harbored a lead SNP, rs3888190, which mapped closest to the *ATP2A1* gene. SNPs in this locus have been associated with obesity in several studies, and the region harbors rare large copy number variations (CNVs) that lead to early onset obesity[118]. This region contained 7 EMVars, the largest number in our study, and 5 out of the 7 EMVars are in perfect LD with rs3888190 in the CEU population (Figure 3.1c, Supplementary Figure S3.1a). These variants are therefore commonly inherited together on a European haplotype present at 32% frequency (Supplementary Figure S3.1a). Five out of the 7 alleles segregating on this risk haplotype decrease enhancer activity in the MPRAs. Three located between *SH2B1* and *TUFM* decrease enhancer activity, two located within *RABEP2* introns decrease enhancer activity, and two located within introns of *ATXN2L* increase enhancer activity (Supplementary Figure S3.1b). Each SNP was an eQTL for 9 genes in adipose and 7 genes in brain. Although eQTL status can be confounded by linkage, meaning that variants in high LD capture the functional effects of other variants in high LD, each of these EMVars also participated in independent cHi-C interactions with several promoters in both adipose and brain, indicating capacity for independent regulation of multiple genes in the locus across cell types (Supplementary Figure S3.2).

Using luciferase assays where we expanded the tested regions to ~500bp in size, we confirmed EMVar enhancer activity for all *ATP2A1* locus EMVars and rs2650492 in either SGBS preadipocytes and/or HT22 brain cells depending on where they were active in MPRA (6 active in adipose and 6 active in brain). We validated that 3/6 adipose EMVars and 4/6 brain EMVars had allelic effects detectable by the luciferase assay, which has a lower sensitivity to detect allelic effects compared to the MPRA. Both rs2650492 and rs9972768 were confirmed to affect enhancer activity in both cell types (Figure 3.1d). Overall, these data suggested that multiple functional variants within distinct enhancers are present within this obesity-associated locus, each with the potential to regulate multiple genes, thus providing evidence against the canonical GWAS model of a single casual variant affecting a single target gene. The pleiotropic regulatory impact on different genes would result in complex molecular signals emanating from multiple genes that, together, participate in disease etiology. Given that 27/40 (67.5%) of loci were found to have more than one EMVar (Figure 2.2d), our data suggests that this complex connectivity will be a common feature of disease-associated loci, and is not simply an oddity of this region.

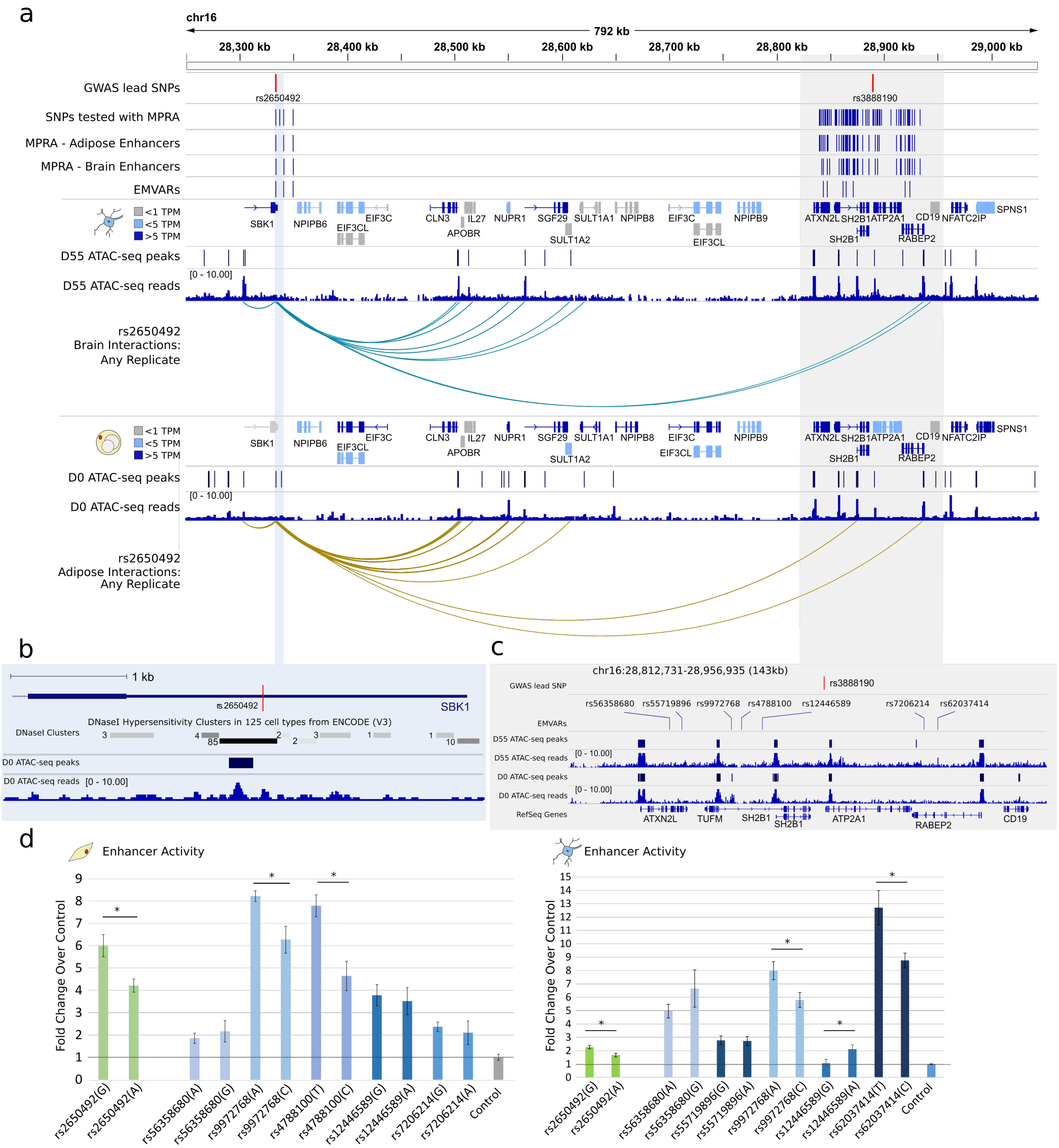


Figure 3.1: Two independent GWAS loci physically converge in nuclear space
a) The locations of two lead GWAS variants separated by >0.5Mb are depicted. cHi-C promoter interactions shown emanate from rs2650492 in brain (blue) and adipose (yellow) and demonstrate cross-locus connections. b) Location of rs2650492 within the 3'UTR of *SBK1*, along with significant DNAseI hypersensitivity clusters in 125 cell types from ENCODE. ATAC-seq peaks and read pileup from day 0 SGBS preadipocytes is also shown c) Location of all EMVars within the *ATP2A1* locus along with ATAC-seq peaks and read pileup from day 0 SGBS preadipocytes and day 55 early neuronal precursors is shown.

Figure 3.1, continued. d) Allele-specific luciferase assay results for EMVars in the HT22 neuronal cell line or SGBS preadipocytes. Fold change is compared to control sequence. *p <0.05, two-tailed Student's t-test, Error Bars = SEM

3.3.2 Evidence of cross-locus connections and implications for gene regulation

To provide additional evidence for a functional connection between the *ATP2A1* and *SBK1* GWAS loci on chromosome 16, we used CRISPR-cas9 editing to delete one EMVar-containing enhancer from each locus to observe effects on gene expression. Out of the 10 EMVars identified across the two loci, rs2650492 and rs9972768 stood out because they were in open chromatin, they participated in many long-range interactions with distal genes, they were eQTLs for several genes within the megabase encompassing these regions, and they were functional EMVars in both adipose and brain. We generated two lines where we deleted the regions harboring rs2650492 or rs9972768 in a human iPSC cell line homozygous for the non-risk haplotypes at both GWAS loci (Figure 3.2a). By deleting these enhancers on the non-risk background we thus recapitulate the effects of the enhancer-lowering risk variants. We used the BrainSpan atlas[119] of gene expression in addition to expression data collected from our hypothalamic differentiation to determine that all genes within this megabase region are expressed uniformly across early development except *NUPR1*, which is lowly expressed until post conception. We therefore chose to assay the effects of these variants during early hypothalamic development. Four homozygous deletion clones from each enhancer deletion line were then differentiated to the hypothalamic lineage. Cells were collected at four time points representing key early developmental stages: iPSCs (TP1), ventralized cells (TP2), neuronal precursors (TP3), and hypothalamic precursors (TP4) (Figure 3.2a). RNA was extracted from each clone at every time point and RNA-seq was performed to identify genes affected by these enhancer deletions. Cells clustered well by developmental stage throughout the differentiation but only separated by genotype at TP4 (Figure 3.3b).

Although these two SNPs map to enhancers separated by over half a megabase, we found that these deletions independently affected the expression of a single gene in the locus during development, *SBK1*, and the effect was not uniform across time (Figure 3.2c). The rs2650492 deletion line significantly decreased expression of *SBK1* during TP1-2, while the rs9972768 variant was trending lower at the first stage but only significantly affected *SBK1* expression during TP2-4 (Figure 3.2c). Both enhancers were critical for proper *SBK1* expression during ventralization (TP2), as *SBK1* expression was reduced in both lines at this stage. The convergent regulation of *SBK1* by variants from two independent GWAS loci supports the cross-locus cHi-C interactions that we observed, as well as the eQTL effect of both variants on *SBK1* expression.

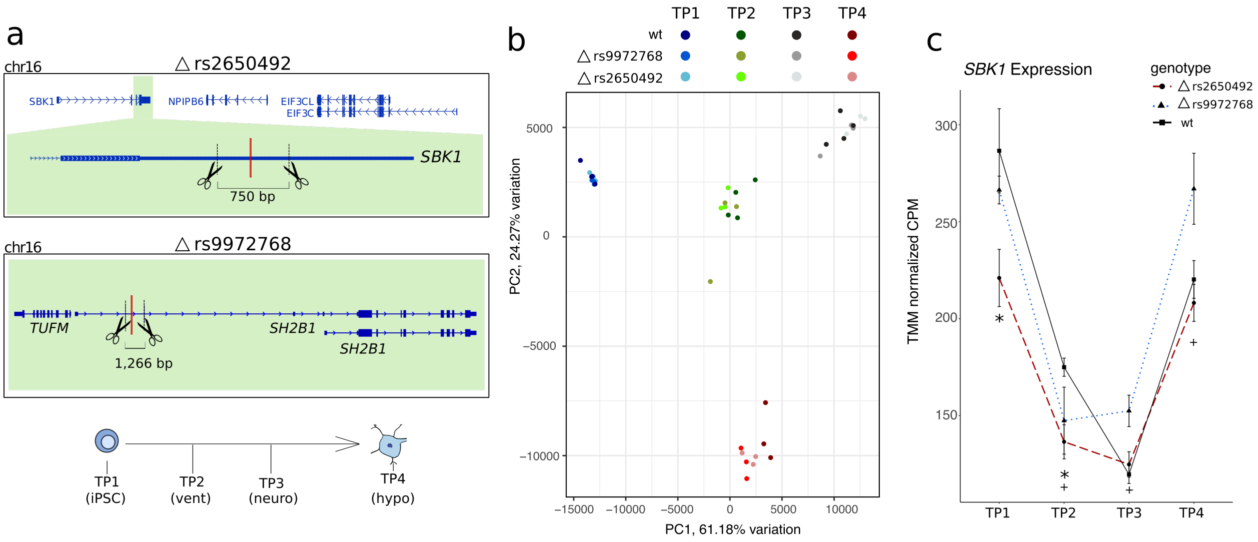


Figure 3.2: Functional variants in two GWAS loci coordinate to regulate *SBK1* in early neuronal development
a) Genomic regions targeted by CRISPR-cas9 editing machinery in iPSCs. (top) A 750bp region within the 3’UTR of *SBK1* containing rs2650492 was targeted for deletion, and (middle) a second 1.3kb region in between *TUFM* and *SH2B1* surrounding rs9972768 was deleted in an independent line. (bottom) iPSCs were differentiated to the hypothalamic lineage and collected at 4 time points for RNA-seq.

Figure 3.2, continued. b) PCA plot showing all genotypes and time points collected for RNA-seq during differentiation to hypothalamic neuronal precursors (N=3-4 clones) c) Plot of TMM normalized counts per million (cpm) for *SBK1* across time points. $*q < 0.05$ rs2650992 deletion lines, $^{+}q < 0.05$ rs9972768 deletion lines; Error Bars = SD

Looking at global gene expression patterns in these two lines, we observed extensive sharing of a large number of differentially expressed genes between the two enhancer deletions (Figure 3.3a). DEG sharing increased dramatically between the ventralization (TP2) and neuronal precursor timepoints (TP3) (Figure 3.3a), suggesting these two enhancers converge to regulate an early driver gene that, when misregulated, leads to a cascade of gene expression effects throughout differentiation that peaks at TP3 (Figure 3.3b).

Although the function of human *SBK1* is unknown, its zebrafish homolog, *Bsk146*, is essential for neuronal development. Upon *Bsk146* knockdown, zebrafish embryos exhibited changes to the midbrain-hindbrain boundary, enlarged hindbrain ventricles, and had small eyes[120]. Mice and rats lacking *SBK1* have also been shown to exhibit an abnormal neurological phenotypes[121]. In both of our enhancer deletion lines, Gene Ontology terms for neural development genes were enriched within differentially expressed genes throughout differentiation, suggesting a conserved role for human *SBK1* in this process (Supplementary Table S3.2).

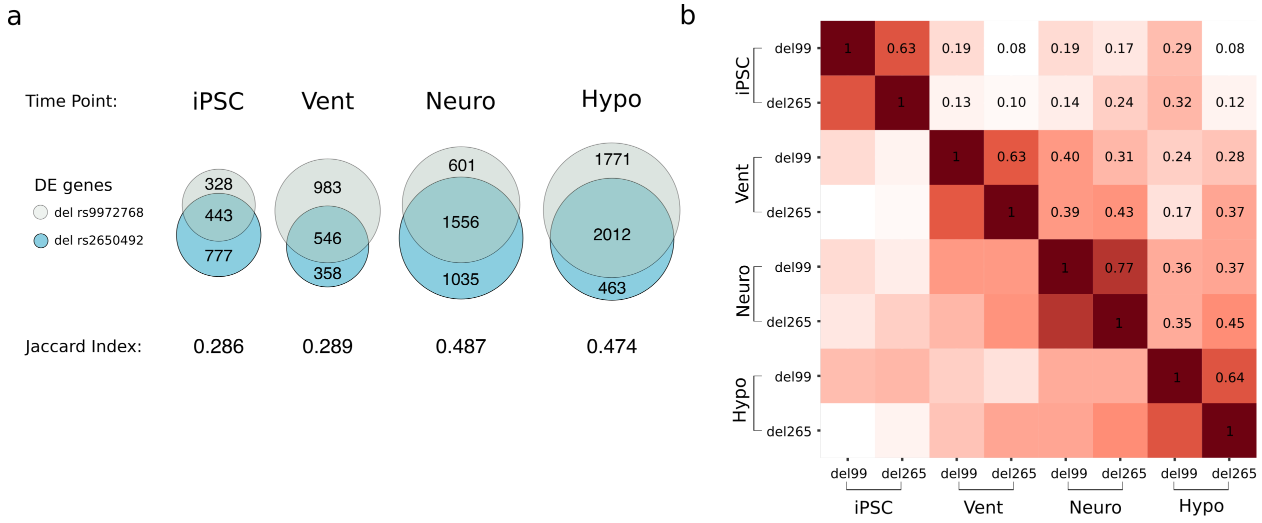


Figure 3.3: A high level of DEG sharing between independent enhancer deletions
a) Venn Diagrams depicting numbers of significantly differentially expressed genes between enhancer deletions and WT cells at each time point. The Jaccard Index is a representation of sharing on a scale of 0-1, where 1 is complete sharing and 0 is no sharing. b) Heatmap showing the Pearson's *r* correlations of log2(fold change over WT) for all expressed autosomal genes in the genome for the two enhancer deletion lines.

3.3.3 Further investigation of the enhancer containing rs2650492

While our data in neural lineage cells demonstrated that enhancers within the two independent GWAS loci regulate *SBK1*, it did not address the evidence of pleiotropy suggested by the many cHi-C interactions and eQTL effects. Therefore we evaluated the ability of the enhancer harboring rs2650492 to regulate additional genes. Using CRISPRi machinery to assess enhancer activity in HEK293t cells, we targeted the rs2650492 EMVar, the promoter of GAPDH as a positive control, and a negative control region within chromosome 16 that was predicted to not have regulatory element activity through looking at ENCODE data across cell types (Supplementary Figure S3.4a). These cells were transfected with plasmids containing either the dCas9-KRAB and/or guide components. We used FACS to sort cells containing the dCas9-KRAB alone, or dCas9-KRAB and CRISPRi guides, to select for cells that were expressing the necessary CRISPRi components (Supplementary Figure S3.4b). We extracted RNA from these cells and performed RNA-seq to look for downregulation of *cis*-genes within the two loci. We observed a significant expression decrease for one gene in the locus, *NUPR1*, which was very lowly expressed throughout our neuronal RNA-seq time course (<5 TPM all stages) but is moderately expressed in HEK293t and in adipose (Figure 3.4). The rs2540492 EMVar was an eQTL for *NUPR1* in several GTEx tissues, and in our data formed frequent long-range cHi-C interactions with *NUPR1* across the adipocyte differentiation and to a much lesser extent in brain. We also observed a significant decrease in *SBK1* expression after targeting rs2650492 in HEK293t cells. However, because rs2650492 maps within the 3'UTR of *SBK1*, we cannot rule out that the change in expression we detect may be due to dCas9-KRAB hindering *SBK1* transcription or reducing mRNA stability from the recruitment to the 3'UTR[122] (Figure 3.4). Together this shows that rs2650492 is capable of regulating at least two genes under unique cellular contexts.

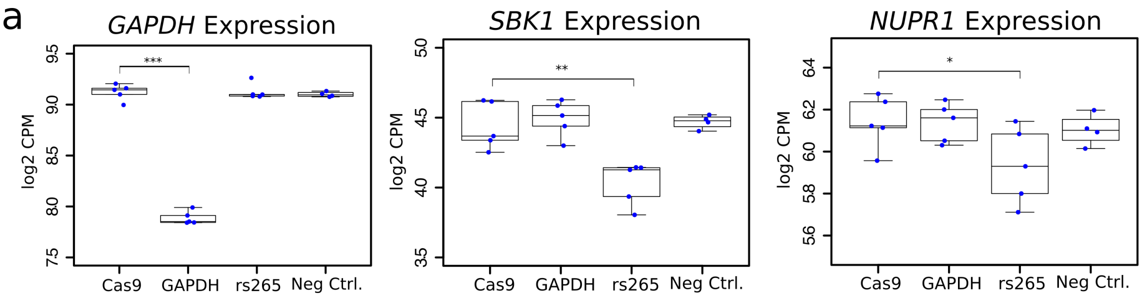


Figure 3.4: The lead GWAS variant, rs2650492, regulates a second gene
a) CRISPRi of enhancer containing rs2650492 in HEK293t cells. Expression after removal of batch effects for significant differentially expressed *cis*-genes and GAPDH identified in RNA-seq analysis across CRISPRi conditions in HEK293t cells (N=4-5 replicates). *$p < 0.0035$; Error Bars = SD

3.4 Discussion

In this Chapter, we aimed to better understand two independent GWAS loci that emerged from our datasets and were seemingly functionally connected. We based our hypothesis on an observation that enhancers within these two loci exhibited long-range cHi-C interactions to promoters within and between the reciprocal locus and exhibited sharing of eQTL effects. These data combined suggested the existence of high levels of regulatory pleiotropy within these regions and colocalization in the nucleus. To assess how these EMVars could lead to phenotypes important for obesity risk, we wanted to specifically understand their gene regulatory capacity in a cell type important for obesity risk. We knocked out two enhancers containing the variants rs2650492 from the *SBK1* locus and rs9972768 from the *ATP2A1* locus, and found that they both regulate *SBK1* expression during early hypothalamic neuronal precursor differentiation. Based on previous research in model organisms, this gene is predicted to be important for brain development in rats and zebrafish[120,123]. Looking at our global differential expression data, we observed enrichment of brain developmental genes, suggesting this gene may play a conserved role in human neuronal development. Although this data alone does not conclusively prove that *SBK1* plays a conserved role in neuronal development, future studies could target *SBK1* in mouse models or human cells to better understand the role this gene may play in neuronal development and potentially obesity. This data strongly supports the existence of a functional link between these two independent GWAS loci, and suggests that as GWAS grow larger and more powered we could observe more instances of this phenomenon due to the presence of more and more weaker GWAS associations occurring in distal enhancers for key disease genes. It also suggests the importance of assaying for gene regulatory effects across several stages of development, as these enhancers display temporally restricted effects.

For the rs2650492 enhancer, which is the lead SNP of the *SBK1* locus association, we wanted to better understand the dense network of promoter interactions emanating from this region. We therefore wanted to test if this enhancer was capable of regulating additional genes beyond *SBK1* under other conditions. We targeted the rs2650492 region using CRISPRi in a secondary cell type and observed that within HEK293t cells this enhancer regulated the expression of *NUPR1*. It was interesting to identify *NUPR1* out of all the other genes, specifically because it is specifically not expressed during early hypothalamic neuronal differentiation. This enhancer interacted with *NUPR1* at one time point in the brain cHi-C timecourse, but in adipose, where *NUPR1* is expressed, the rs2650492 EMVar interacted with the promoter of *NUPR1* at every time point. Altogether, this demonstrates the enhancer containing rs2650492 also regulates *NUPR1*. This also suggests that the cHi-C interactions emanating from this enhancer may predict genes that this enhancer regulates under specific spatio-temporal contexts. Future experiments will need to be performed to determine how many genes seem to be regulated by this enhancer, and which of the genes are capable of contributing to obesity risk.

In summary, the physiological effects stemming from these tissue and temporal regulatory specificities may play a role in the molecular etiology of obesity risk, and highlight complex considerations in the functional experiments that will attempt to better understand the mechanisms underlying GWAS associations.

3.5 Methods

3.5.1: ATP2A1-SBK1 locus EMVar Luciferase Assays

In order to test allele specific enhancer activity, IDT gBlocks™ were ordered containing each allele of each variant. Each SNP tested was centered and surrounded by 470 base pairs of native genomic context and 15bp of homology on each side to the pGL4.23 luciferase vector. These gblocks were then cloned into the pGL4.23 luciferase vector using Gibson assembly. rs4788100 (C) and (T) gblocks were surrounded by 295bp of native genomic context, rs62037414 (C) and (T) gblocks were surrounded by 451bp of genomic context, and rs55719896 (G) and (A) gblocks were surrounded by 469bp of genomic context due to synthesis constraints. Luciferase assays were conducted either in SGBS preadipocytes or HT22 cells as previously described.

3.5.2: iPSC culture conditions

The NA19101 Yoruban iPSC line was grown in complete mTeSR1 media (Stemcell #85850) supplemented with 1% Penicillin-Streptomycin (10,000 U/ml; gibco #15140122) on Matrigel-coated dishes (Corning #354277) at 37°C in 5% CO_2. Cells were passaged 1:10-1:8 every three days or upon reaching 70-80% confluency and ROCK inhibitor was added to the media during each split. Media was changed every day for the duration of culture.

3.5.3: CRISPR-cas9 editing of iPSC cells

Fluorescently tagged crRNA-Atto550 and cas9 protein were purchased from IDT. Guides were designed using IDT software in order to maximize cutting efficacy and minimize off target cleavage using their RNP system. Two guides were designed per region in order to delete the enhancers from their endogenous context. IDT crRNA and tracrRNA were complexed according to manufacturer's instructions. The day of transfection, 50uM of each guide and cas9 protein were combined and incubated at room temperature for 20 minutes to form RNPs. Each guide was complexed with cas9 in individual reactions. 900k NA19101 human iPSC cells were harvested and nucleofected using a Lonza nucleofector 2b device with the program A23 and the two RNP complexes. These cells were plated into one 22cm^2 flask. Once recovered, 40k cells were split into a 100cm^2 flask to achieve single cell colonies. Colonies were then picked and transferred into 48 well plates to grow independently. Colonies were screened for the presence of homozygous deletion bands using PCR. Homozygous deletion colonies were grown, transfected, split, and treated identically to WT cells in order to mitigate RNA-seq batch effects for the differentiation. Cells were differentiated as previously described. Cells remained frozen until all timepoints were collected. RNA was then extracted using the Qiagen RNeasy Kit and RNA-quality was assessed via an Agilent Bioanalyzer RNA Chip. 1ug of RNA was used for input into RNA-seq. RNA-seq was performed using the NEB Next Ultra II Directional RNA-seq kit. rs265 guide 1: 5'-GGGAUUGUCCUGACAACUUG-3', rs265 guide 2: 5'-AAAGUGCUCGGAGUUCACUC-3', rs99 guide 1: 5'-UGAGCCAUUCACUAAUACAG-3', rs99 guide 2: 5'-UGUCAACACUGUGGUUCAAU-3', PCR rs265- F: 5'-CCAAGCCCTTGGAAAATGTA-3', PCR rs265-R: 5'-AACTATGGTCCCCTCCCAAC-3', PCR rs99-F: 5'-AGCCGATATCACGCCATTGT-3', PCR rs99-R: 5'-GAACAGAAGCCAGGAGACCC-3'

3.5.4: Early neuronal differentiation time course analysis

Reads were mapped and gene counts were quantified with STAR (v2.5.1a). Counts were filtered to retain autosomal genes and exclude lowly expressed genes (< 1 cpm at all time points). PCA analysis showed that the data clustered well by time-point and genotype so no batch effect correction was necessary. *P* values were identified using glmQLFTest() from edgeR. *P* values were FDR adjusted genome wide using p.adjust to get *q* values.

3.5.5: HEK293t cell culture conditions

HEK293t cells were maintained in DMEM (Gibco # 11995-065) supplemented with 10% FBS and 1% penicillin-streptomycin (10,000 U/ml; gibco #15140122) solution at 37°C in 5% CO_2. Cells were passaged with 0.25% trypsin-EDTA when they reached 70-80% confluency. Media was changed every other day for the duration of culture

3.5.6: HEK293t CRISPRi

Four guide RNAs were designed for each targeted region by CHOPCHOP or MIT's guide design tool based on maximizing cutting efficiency, minimizing off targets, and closest proximity to the region of interest. Guide sequences are provided in the supplementary tables. Guides were then individually cloned using golden gate methodology into the guide vector upstream of an eGFP gene. Cells were transfected into HEK293t cells at 50-70% confluency using Lipofectamine LTX with the 4 guide vectors and dCas9 vector (dCas9-KRAB upstream of BFP fluorophore) at a ratio of three parts Cas9 to one part guide, where each individual guide was added in equal amounts to additional guides for that region. After 48 hours, double positive GFP and BFP fluorescing cells were collected into culture media via FACs sorting, spun down and frozen. In the case of the Cas9 only control population, BFP single positive cells were sorted out of the population via FACS and frozen. FACS gates: Gate 1: FSC-A 50,000-250,000 x SSC-A 1,000-100,000, Gate 2: FSC-W 50,000-125,000 x FSC-W 50,000-125,000, Gate 3: BFP-BV421 1,000+ x GFP-FITC 1,000+ (double positives were collected for conditions and BFP single positives were collected for dCas9 only control). Transfection and sorting was repeated 4-5 times to have technical replicates. RNA was extracted using the Qiagen RNeasy RNA mini extraction kit and 500ng of RNA was used as input for RNA-seq. RNA-seq was performed using the NEB Next Ultra II Directional RNA-seq kit.

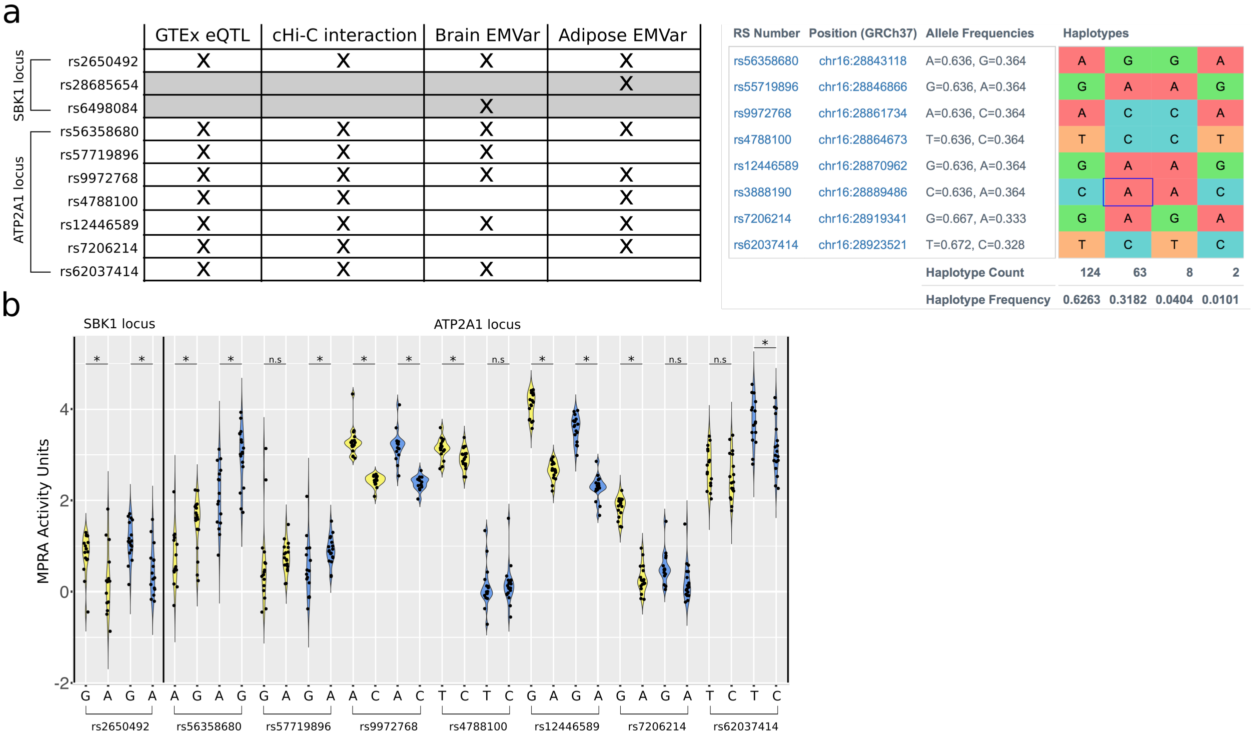
3.5.7 HEK293t CRISPRi Analysis

Reads were mapped and gene counts were quantified with STAR (v2.5.1a). Counts were filtered to retain autosomal genes and exclude lowly expressed genes (< 1 CPM). PC1 clearly associated with sorting batch, so sorting batch was added as a covariate into the final linear model. *P* values were identified using glmQLFTest() from edgeR. Because of the very small effect sizes of CRISPRi in non-coding regions, cis-genes within the *SBK1-ATP2A1* loci were considered for final significance testing. This list included all protein coding genes within these loci passing the expression threshold as well as *GAPDH* (14 genes total). Raw *p* values from these genes were Bonferroni corrected to get adjusted *p* values (P < 0.05/14 genes). Only genes passing the Bonferroni significance threshold were considered significantly affected by the CRISPRi perturbations.

3.5.8: Allele specific mapping of CTCF ChIP-seq Peaks

Encode CTCF ChIP-seq .bam files were found using the region search tool within ENCODE https://www.encodeproject.org/region-search. All files were then sorted using samtools and processed for allele-specific read mapping using WASP (v0.0.3). The location of the rs2650492 variant was provided along with reference and alternate variants and the reads overlapping the reference variant were counted and compared to the reads overlapping the alternate variant.

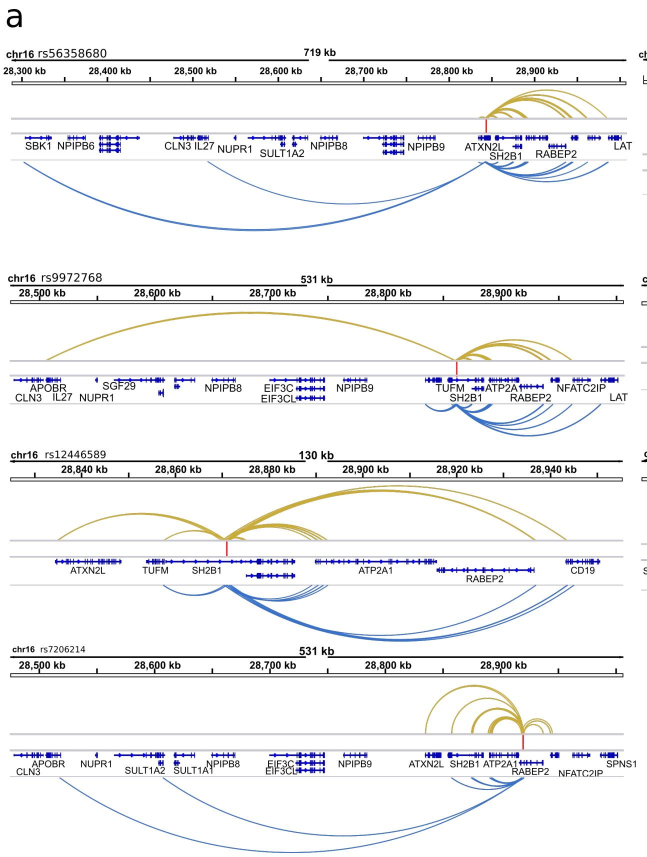
3.6 Appendix C, Supplementary Figures



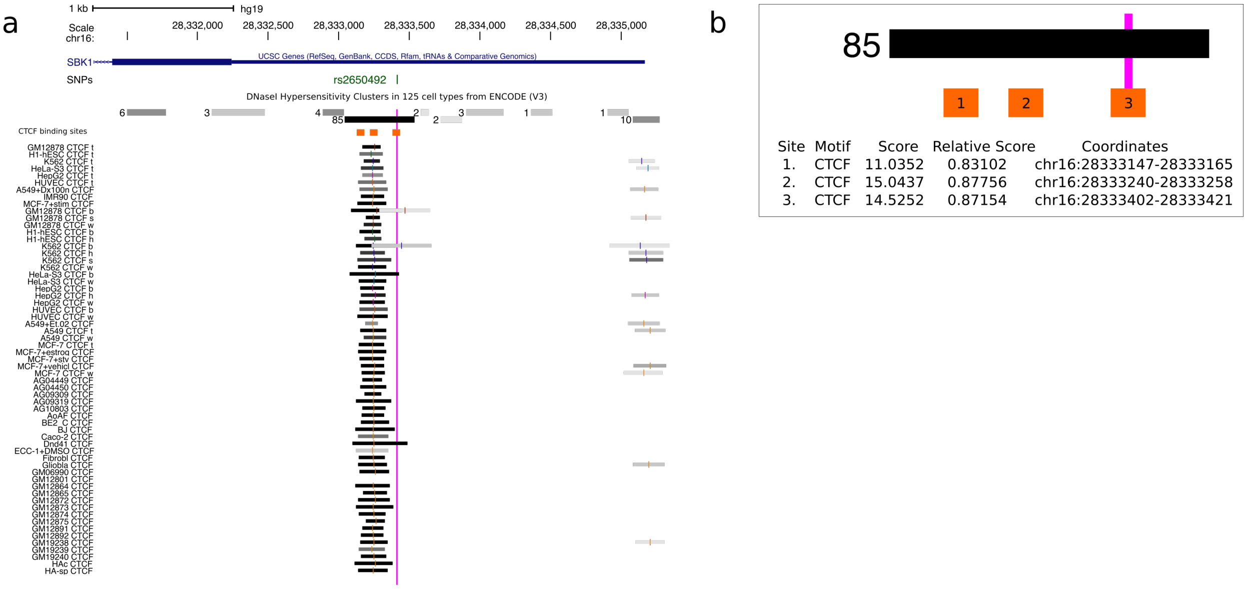
	GTEx eQTL	cHi-C interaction	Brain EMVar	Adipose EMVar
rs2650492	X	X	X	X
rs28685654				X
rs6498084			X	
rs56358680	X	X	X	X
rs57719896	X	X	X	
rs9972768	X	X	X	X
rs4788100	X	X		X
rs12446589	X	X	X	X
rs7206214	X	X		X
rs62037414	X	X	X	

RS Number	Position (GRCh37)	Allele Frequencies	Haplotypes			
rs56358680	chr16:28843118	A=0.636, G=0.364	A	G	G	A
rs55719896	chr16:28846866	G=0.636, A=0.364	G	A	A	G
rs9972768	chr16:28861734	A=0.636, C=0.364	A	C	C	A
rs4788100	chr16:28864673	T=0.636, C=0.364	T	C	C	T
rs12446589	chr16:28870962	G=0.636, A=0.364	G	A	A	G
rs3888190	chr16:28889486	C=0.636, A=0.364	C	A	A	C
rs7206214	chr16:28919341	G=0.667, A=0.333	G	A	G	A
rs62037414	chr16:28923521	T=0.672, C=0.328	T	C	T	C
		Haplotype Count	124	63	8	2
		Haplotype Frequency	0.6263	0.3182	0.0404	0.0101

Supplementary Figure S3.1: Haplotype information for *ATP2A1* locus EMVars
a) (left) Summary information for all 10 EMVars identified in both the *SBK1* and *ATP2A1* loci. Two SNPs in the *SBK1* region were neither eQTLs nor did they participate in cHi-C interactions and were thus removed from future consideration. (right) Allele frequencies and haplotype information in the CEU population for all EMVars in the *ATP2A1* locus (LDhap tool: https://ldlink.nci.nih.gov). The lead risk variant, rs3888190-A, is outlined in blue. b) MPRA allele specific activity levels for EMVars within the *ATP2A1* locus and *SBK1* locus in adipose or brain libraries. The average activity for each barcode across replicates is shown as a dot. *$q < 0.05$; two-sided Mann-Whitney U test. Adipose cHiC data=yellow, Neuronal cHiC data = blue



Supplementary Figure S3.2: SNP specific Hi-C interactions for *ATP2A1* locus EMVars
a) Promoter interactions stemming from each EMVar in the *ATP2A1* locus at any time point in both brain and adipose cells. Location of variant is indicated by a red line. Adipose cHiC data=yellow, Neuronal cHiC data = blue

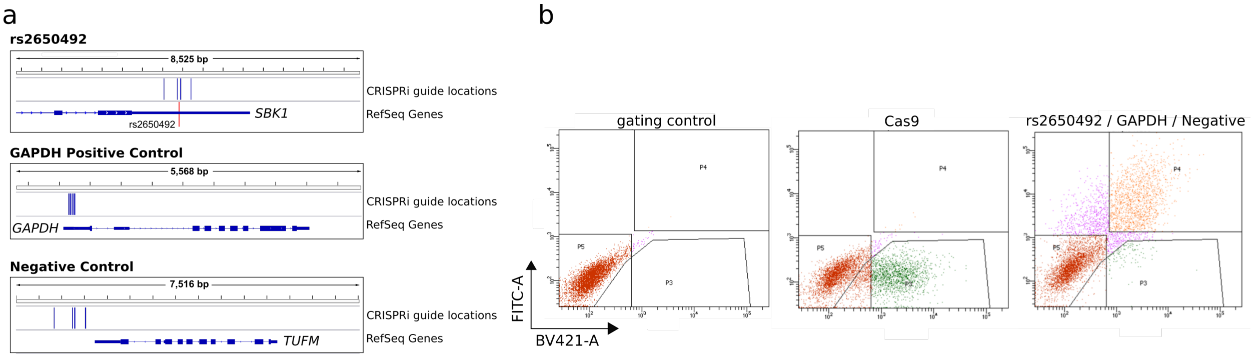


Site	Motif	Score	Relative Score	Coordinates
1.	CTCF	11.0352	0.83102	chr16:28333147-28333165
2.	CTCF	15.0437	0.87756	chr16:28333240-28333258
3.	CTCF	14.5252	0.87154	chr16:28333402-28333421

c

Cell Type	Genotype	Reads Ref	Reads Alt	Total Reads	Percent Alt
BJ Cell Line	G/A	12	14	26	53.84615
Smooth Muscle -female embryo 5 days in vitro differentiated cells	G/A	13	21	34	61.76471
Neural Progenitor -female embryo 5 days in vitro differentiated cells	G/A	28	28	56	50
Prostate Epithelium -male epithelial of prostate primary cell	G/A	111	117	228	51.31579
Brain (SK-N-SH) line -human neuroblastoma cell line	G/A	67	75	142	52.8169
B cell lymphoma (OCI-LY7)	G/A	24	15	39	38.46154
Hela S3 Cells	G/A	5	6	11	54.54545

Supplementary Figure S3.3: CTCF transcription factor binding to rs2650492
a) ENCODE CTCF ChIP-seq for 58 samples showing locations of significant peaks (black bar) and peak point (line within black bar) in relation to rs2650492 within the 3'UTR of *SBK1*. These samples were identified using the region search tool (https://www.encodeproject.org/region-search/). Also depicted are the ENCODE DNaseI hypersensitivity clusters in all 125 ENCODE cell types, and the locations of JASPAR predicted CTCF binding sites (orange). b) Close up of JASPAR predicted binding sites with score > 11 within region of open chromatin. The location of rs2650492 is shown in hot pink, and it is located in the 3rd CTCF peak with the second highest score. c) Allele specific read mapping of seven cell lines with significant read coverage over rs2650492. These lines were heterozygous for this variant and exhibited no significant differences in read mapping to either the non-risk (G) or risk (A) alleles.



Supplementary Figure S3.4: CRISPRi guide locations and selection with flow activated cell sorting (FACS)

a) Locations of CRISPRi guides for each condition. Guides were designed to target the 3’UTR of *SBK1*, the promoter of GAPDH as a positive control, and a region downstream of TUFM as a negative control.
b) Cells were transfected and isolated via FACS based on the presence of either the Cas9 expressing BFP plasmid (BV421-A) and/or the GFP expressing guide plasmid (FITC-A).

3.7 Appendix D, Supplementary Tables

Table S3.1: CRISPRi guide sequences

Guide	Orientation	Sequence	Target
1	5'	GACAATCCCTTGTGGTTAGG	rs2650492
2	5'	GGGCGTAGGACCTGCATGTG	rs2650492
3	5'	TGTGTAGGGTGCAGACGCAT	rs2650492
4	5'	CCCCGCAATAAGCACCACAT	rs2650492
1	5'	AGGAGGAGCAGAGAGCGAAG	GAPDH control
2	5'	CGGGCTCAATTTATAGAAAC	GAPDH control
3	5'	TGGCGACGCAAAAGAAGATG	GAPDH control
4	5'	CGGGCGGAGAGAAACCCGGG	GAPDH control
1	5'	GTATTCTTAAAACTAGAGAG	Negative control
2	5'	GTGTTTGTATGCTATCAGCG	Negative control
3	5'	TAAGAAACGTGAAGACAATG	Negative control
4	5'	TTTCGACGGTCTCTATGGGG	Negative control

Table S3.1: CRISPRi guide sequences
CRISPRi guide sequences used to target either rs2650492, the promoter of GAPDH, or a negative control region on chromosome 16.

Table S3.2: Gene Ontology terms for enhancer deletion DEGs

GO biological Process Complete	Observed	Expected	Fold Enrichment	(+ or -)	raw P value	FDR	Library	Stage
Nervous system development	224	127.15	1.76	+	4.27E-16	5.66E-13	rs2650492 Deletion	iPSC
Nervous system development	145	90.44	1.6	+	2.56E-08	0.0000453	rs2650492 Deletion	Ventralizaton
Nervous system development	410	254.18	1.61	+	7.02E-20	1.86E-16	rs2650492 Deletion	Neuronal Precursors
Nervous system development	477	259.77	1.84	+	1.02E-34	3.25E-31	rs2650492 Deletion	Hypothalamic Precursors
Nervous system development	130	79.36	1.64	+	3.48E-08	0.000111	rs9972768 Deletion	iPSC
Nervous system development	256	150.22	1.7	+	3.37E-16	1.07E-12	rs9972768 Deletion	Ventralization
Nervous system development	356	206.74	1.72	+	3.61E-22	2.87E-18	rs9972768 Deletion	Neuronal Precursors
Nervous system development	648	393.44	1.65	+	3.85E-31	1.53E-27	rs9972768 Deletion	Hypothalamic Precursors

Table S3.2: Gene Ontology terms for enhancer deletion DEGs
Significance of nervous system development GO term in both enhancer deletion differentially expressed genes across all time points tested. Shown are observed number of DE genes within this category, the expected number of DE genes in the category, the fold enrichment for observed/expected, whether it is a + or – enrichment, the raw and FDR adjusted p-values, as well as the library and stage.

CHAPTER 4: CONCLUSIONS, SPECULATIONS AND FUTURE DIRECTIONS

Conclusions

The overarching goal of this dissertation was the development, integration, and interpretation of a functional genomics pipeline in order to better understand the genetic basis of obesity risk. In Chapter 2 we generated comprehensive regulatory maps for human adipose and hypothalamic neurons to profile these cells across differentiation stages. We cataloged data such as chromatin accessibility, expression patterns, and cHi-C enhancer-promoter interactions which together aid in the interpretation of candidate causal non-coding variants at obesity-associated loci. Additionally we applied a massively parallel reporter assay to gain information on enhancer locations within obesity GWAS loci. We identified 94 variants within enhancers in high LD with obesity GWAS lead variants that were capable of modulating enhancer activity in adipose and/or brain cell types. Interestingly, we frequently identified multiple functional SNPs within a locus, many of which were capable of modulating enhancer activity across both tested cell types. After integrating these functional variants with cHi-C, we additionally provided evidence of the capacity for multi-gene regulation, where many of the enhancers and EMVars interacted with 4 or more promoters across the cHi-C time course. We integrated this information with eQTLs and expression data to prioritize 20 high confidence class I genes and 30 class II genes for obesity importance.

In Chapter 3, we focused specifically on a megabase region on chromosome 16. We identified 23% (22/94) of EMVars on chromosome 16 alone. This, and the obesity heritability enrichment of chromosome 16, suggests that this chromosome could harbor a plethora of obesity relevant genes. The locus encompassing the *SBK1* and *ATP2A1* association regions emerged from our datasets due to the high complexity of long-range interactions and patterns of eQTL sharing among many genes across the two regions in adipose and brain. Compared to the *SBK1* locus, where we prioritized one casual variant, the *ATP2A1* locus harbored several variants with evidence of causality, and these were all found to be in very high LD and segregate together on a common haplotype. The gene(s) and mechanisms mediating BMI phenotypes in this region remain a focus of investigation, and the combination of data from this thesis and previously published data suggests that there are likely several obesity relevant genes within this large interval. Using MPRA and CRISPR-cas9 editing we demonstrated that these loci harbor two functional SNPs within enhancers that independently regulate *SBK1* early in neuronal differentiation. Additionally, we demonstrated that rs2650492, the *SBK1* locus GWAS lead variant, also regulates *NUPR1*. Other critical modulators of the obesity phenotype exist within the *ATP2A1* locus, such as *SH2B1*, a gene involved in leptin and insulin signaling[43,103]. This gene and others were eGenes that physically connected to multiple EMVars. Therefore, this data suggests that multiple genes with the potential to regulate the obesity phenotype are regulated by functional variants in this locus. It is likely that investigation of these EMVars under other conditions or developmental stages would uncover additional examples of gene regulation in this locus. It has yet to be elucidated whether many, or only a subset, of genes modulated within this region are capable of leading to an obesity phenotype.

Recent work has also suggested that regulatory variants associated with human phenotypes may impart their effect during temporally restricted windows, which would be missed in functional assays of a single developmental time point or environmental perturbation[29,69,70,98]. We were able to provide some additional support for this hypothesis by assigning more EMVars to promoters using the time-course cHi-C data compared to a single time point, as well as our finding that the rs2650492 and rs9972768 EMVars regulated the expression of *SBK1* during specific stages of early hypothalamic differentiation. This may require assessment of putatively causal variants during several key developmental timepoints of your cell type of interest in order to capture any temporally restricted effects of enhancers and to understand the full range of effects this variant may impart.

All together, this work support a model where the underlying genetic architecture of individual loci associated with obesity will often involve allelic heterogeneity, where multiple variants in distinct regulatory elements impart effects on the expression of gene(s) across tissues during uniform or restricted temporal windows to alter disease risk. Whether the complexities we uncovered here are a rule or exception for loci in variant-to-function studies has yet to be addressed and will thus require careful investigation of causal genetic variation at GWAS loci under multiple cell types and temporal conditions.

How can we improve causal variant prediction?

The method used here to identify causal genetic variation relied on experimental fine-mapping where we tested each SNP in high LD with lead variants identified in GWAS for their ability to modulate enhancer activity, presence in cHi-C interactions, and eQTL status. Although the MPRA allows for a direct measurement of functional outcomes of a SNP, there are limitations to this approach to identify causal variation within a GWAS locus. MPRA technology is currently limited by technical limitations to the size of DNA fragments that can be synthesized. Thus, longer regions that require >175bp for enhancer activity would be missed in this assay. Second, we only test for functional variation that affects enhancer activity. Although enhancer modulation is predicted to be the most common mechanism by which causal variants impart their gene modulatory effects, other mechanisms are also likely at play, such as affecting repressor activities[124], splicing[67,125], RNA modifications[126], alternative polyadenylation[127], and likely others. Thus, we are limiting the true scope of functional variants we observe in obesity GWAS loci. Once more of these functional genomics annotations are generated, it will be interesting to learn more about the combinations of mechanisms potentially at play within these regions. In line with this, it seems unlikely that all functional variation observed is causal, as benign functional variation is likely to exist in close proximity to GWAS lead variants. Until this is conclusively understood, functional studies such as this one would be strengthened with knowledge of high confidence statistical fine-mapping results, where variants with high statistical support are eventually given more weight for causality than those with low statistical support. Some fine-mapping approaches[128–131] currently employ a similar theory, where variants in very high LD that cannot be prioritized over others statistically are weighted by their presence in functional genomics annotations. The best methodology to use is still under debate, but it is likely that a combination of approaches that blends both high confidence statistical fine-mapping and verified functional annotations will be the most powerful approach.

The next phase of GWAS seems to be the integration of information from human populations beyond those with European ancestry. This will provide great insight into which SNPs are causal, since LD changes across human populations and thus different genetic variants exist on population specific haplotypes. This can be used to perform trans-ancestry fine-mapping to narrow down candidate causal variation[132]. For example, African populations are the most genetically diverse in the world. Because of this, LD is lower in Africans, allowing for smaller haplotypes and better causal variation prediction. Another example is the use of founder populations where there has been increased probability for large effect size disease variants to overcome selection due to a historic population bottleneck. In testament to this, a common p.Arg684Ter nonsense variant in the *TBC1D4* gene present at 17% allele frequency in the Greenlandic Inuit population has been found to confer very high risk for type 2 diabetes risk (homozygous OR = 10.3)[133]. Genetic investigations into these populations has been wildly important for genetics research, especially in instances where these genetic data are coupled with detailed phenotypic records, such as what is available via the FinnGen Biobank. It is also important to collect samples from diverse human populations to better treat and predict disease in individuals of these ancestry groups.

In this work, I was able to dig deeper into putatively causal variation in two loci. The *SBK1* locus harbored a single likely causal variant, but an interesting cluster of 7 variants in the *ATP2A1* locus emerged in very high LD that individually participated in cHi-C interactions with multiple genes. Two of the EMVars within the *ATP2A1* locus had alleles falling on the risk haplotype that were predicted to increase enhancer activity, while the others were all predicted to decrease enhancer activity. If all of these variants contribute to causality and regulate a single gene important to obesity risk, this means that these variants that increase enhancer activity might offset the effects of other variants on the haplotype that reduce enhancer activity. In a phenomenon called linkage masking[134], functional variants in high LD that have opposite effects may counterbalance one another and escape negative selection. Alternatively, these variants could affect the expression of distinct genes that individually contribute to obesity risk. In order to have a better understanding of either of these scenarios, an interesting experiment would be individual and combinatorial deletions of each of these enhancers in human iPSCs coupled with differentiation to observe which are capable of participating in gene expression regulation during early to late neuronal development. Ideally these enhancers would also be assessed in the adipocyte lineage, although current methodologies to perform CRISPR-cas9 editing in human adipocytes are challenging and iPSC differentiation to the adipocyte lineage is inefficient, making gene expression estimates after CRISPR perturbations in this model difficult. This type of experiment would confirm our predictions of which variants have tissue specific versus shared effects on gene expression to better understand the generalizable cross-tissue effects observed in this work. Overall it remains to be seen whether all of these functional EMVars contribute to causal variation within this locus, or only a subset of the variants that impact key target genes.

How can we improve causal gene prediction?

Another interesting observation that emerged from our analysis was that that these enhancers may regulate multiple genes under different cell type or developmental conditions, which complicates target prediction for non-coding causal variants. If these enhancers are capable of modulating more than one gene under certain stages of development and/or different cell types, methods to better predict target genes must be developed that take into account these factors.

In this work, we used a combination of promoter capture Hi-C, expression data, and publically available eQTL information as a first pass prioritization method for target genes. We then used CRISPR-cas9 editing to validate some of these predictions in two loci. Other methods have been utilized that employ functional genomics or statistical methodologies to prioritize target genes within GWAS loci. One of the more common statistical methods, termed transcriptome wide association study (TWAS)[135], integrates GWAS summary statistics and gene expression measurements to identify genes whose expression pattern is associated with the trait of interest. A second more recent approach, termed polygenic priority score (PoPS)[27], leverages summary statistics along with biological pathway and protein-protein interaction data to predict target genes. These methodologies lead to gene level predictions but either rely on a priori knowledge of cell type or do not generate specific cell type predictions.

CRISPR editing of non-coding regions in order to wire enhancers to promoters remains the most conclusive measure of target gene identification, but suffers from being technically challenging and very low throughput. In order to make CRISPR technologies more accessible, CRISPR screens have emerged as a popular tool. Using CRISPR screens, you can target all genes genome-wide to test for a phenotype of interest. For example, Hilgendorf et. al performed a genome-wide CRISPR screen for genes important in adipogenesis[136]. These types of screens provide an additional layer of evidence to identify genes important for your trait and prioritization based on function, and can be performed across cell types or potentially at different cell stages to suggest when the gene acts on the trait. The limitation of this approach is that your trait of interest must be able to be selected for within the cell population.

In order to further address the question of relevant cell type, methods are being developed that leverage epigenetic marks assayed amongst a broad range of cell types. For example, *tissue-of-action* (TOA) scores [137] use tissue specific gene expression from GTEx as well as epigenomic annotations for cell type predictions for finemapped genetic variants within GWAS loci. Using this method, authors were able to predict the contribution of various cell types to each locus, and demonstrate whether causal variant is expected to act in one or multiple cell types. The authors were able to provide evidence at 41% of type 2 diabetes GWAS signals of shared regulatory effects across tissues. This type of information is critical for reducing the guesswork involved in designing downstream experiments to test these predictions.

In this thesis, we used cell type enrichment information based on ChIP-seq annotations from several generally metabolically relevant cell types generated in Locke et al 2015. A single tissue or multiple tissues of interest can be implicated but this does not generate individual loci level predictions. In general, for these types of predictions to reach their highest potential, a more comprehensive genomic annotation catalog of all cell types at developmental stages is required. For example, a recent paper used single-cell sequencing data from 727 mouse neuronal cell types and performed enrichment analysis for human obesity GWAS data[138]. They specifically identified 26 brain cell types, including the hypothalamus, cortex, and hippocampus, for BMI heritability enrichment. This, and the fact that obesity seems to trigger reward circuitry, may indicate that several of these loci have primary effects in regions of the brain outside the hypothalamus.

Data from specific cellular subtypes such as this can provide additional fine-grain hypothesis generation tools to predict how these genes lead to obesity risk across closely related cellular subtypes such as in the brain or as those across more distally related cell types. It will not be surprising if more instances of cross-tissue effects leading to unique yet synergistic effects on the trait of interest are uncovered in the near future. These cell type and gene level predictions will allow for focused experimental efforts on the tissue or tissues implicated at each locus to better understand how genes within the region may affect relevant biological processes in certain cell types over others.

In the case of the *SBK1* and *ATP2A1* regions, it would be interesting to test how many genes these enhancers are capable of regulating across cell types. With the iPSC model, the enhancer deletions for each of the identified EMVars could be differentiated into several distinct cell types for gene expression measurement across time as was done for hypothalamic neuronal precursors in Chapter 3 of this thesis. Additionally, since the hypothalamic neuronal differentiation does not lead to pure populations of one specific subtype, it could be interesting to perform the same differentiation in Chapter 3 but analyze the data using single cell sequencing, where differentially expressed genes in each cell type cluster could be identified, thus showing whether the *SBK1* phenotype is a pan-hypothalamic cell phenomenon or restricted to certain cellular subtypes. Data such as this, coupled with a CRISPR screen for function, would be informative for narrowing down the entire landscape of putative target genes within these complex loci for those important for obesity risk versus bystanders. This could also help us understand how predictive many of these promoter capture Hi-C interactions are for functional connections across cell types.

Final Remarks

In conclusion, we have provided support for a complex network model within GWAS regions where multiple causal variants affect the expression of multiple genes to lead to disease risk. Regions such as the *ATP2A1* locus are fascinating, and a thorough investigation of this region would lead to additional novel insights into the mechanisms of GWAS associations to disease and further our knowledge of gene regulation. But, regions such as this one on chromosome 16 are seemingly not the best initial targets for therapeutic investigation. Ideally, for the fastest and most efficient use of the data, a focus on less complicated regions, such as the *MAP2K5* locus, would be a more straightforward path to success. Although drug targets with supporting human genetics evidence are twice as likely or more to succeed in clinical trials[139,140], gene predictions outlined in this thesis and related works will not, and should not, replace the mechanistic insights that can be gained through fine-scale single locus efforts where the effects of perturbing these genes is carefully assessed in an in-vivo setting. It simply provides an avenue for gene prioritization before embarking on costly cellular biology based efforts to understand gene function. As we approach the next decade of GWAS interpretation, it will be interesting to learn more about how these complex mechanisms coordinate to regulate disease risk.

BIBLIOGRAPHY

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng J-F, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen H-C, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JGR, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AFA, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang S-P, Yeh R-F, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Patrinos A, Morgan MJ, International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research C for GR, The Sanger Centre:, Washington University Genome Sequencing Center, US DOE Joint Genome Institute:, Baylor College of Medicine Human Genome Sequencing Center:, RIKEN Genomic Sciences Center:, Genoscope and CNRS UMR-8030:, Department of Genome Analysis I of MB, GTC Sequencing Center:, Beijing Genomics Institute/Human Genome Center:, Multimegabase Sequencing Center TI for SB, Stanford Genome Technology Center:, University of Oklahoma's Advanced Center for Genome Technology:, Max Planck Institute for Molecular Genetics:, Cold Spring Harbor Laboratory LAHGC, GBF—German Research Centre for Biotechnology:, *Genome Analysis Group (listed in alphabetical order also includes individuals listed under other headings):, Scientific management: National Human Genome Research Institute UNI of H, Stanford Human Genome Center:, University

of Washington Genome Center:, Department of Molecular Biology KUS of M, University of Texas Southwestern Medical Center at Dallas:, Office of Science UD of E, The Wellcome Trust: Initial sequencing and analysis of the human genome. Nature. Nature Publishing Group; 2001 Feb;409(6822):860–921.

2. Ohno S. So much "junk" DNA in our genome. Brookhaven Symp Biol. 1972;23:366–370. PMID: 5065367

3. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee B-K, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shoresh N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, Dunham I, Ernst J, Furey TS, Gerstein M, Giardine B, Greven M, Hardison RC, Harris RS, Herrero J, Hoffman MM, Iyer S, Kellis M, Khatun J, Kheradpour P, Kundaje A, Lassmann T, Li Q, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SCJ, Reddy TE, Rozowsky J, Schlesinger F, Thurman RE, Wang J, Ward LD, Whitfield TW, Wilder SP, Wu W, Xi HS, Yip KY, Zhuang J, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M, Pazin MJ, Lowdon RF, Dillon LAL, Adams LB, Kelly CJ, Zhang J, Wexler JR, Green ED, Good PJ, Feingold EA, Bernstein BE, Birney E, Crawford GE, Dekker J, Elnitski L, Farnham PJ, Gerstein M, Giddings MC, Gingeras TR, Green ED, Guigó R, Hardison RC, Hubbard TJ, Kellis M, Kent WJ, Lieb JD, Margulies EH, Myers RM, Snyder M, Stamatoyannopoulos JA, Tenenbaum SA, Weng Z, White KP, Wold B, Khatun J, Yu Y, Wrobel J, Risk BA, Gunawardena HP, Kuiper HC, Maier CW, Xie L, Chen X, Giddings MC, Bernstein BE, Epstein CB, Shoresh N, Ernst J, Kheradpour P, Mikkelsen TS, Gillespie S, Goren A, Ram O, Zhang X, Wang L, Issner R, Coyne MJ, Durham T, Ku M, Truong T, Ward LD, Altshuler RC, Eaton ML, Kellis M, Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Batut P, Bell I, Bell K, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena HP, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Li G, Luo OJ, Park E, Preall JB, Presaud K, Ribeca P, Risk BA, Robyr D, Ruan X, Sammeth M, Sandhu KS, Schaeffer L, See L-H, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Hayashizaki Y, Harrow J, Gerstein M, Hubbard TJ, Reymond A, Antonarakis SE, Hannon GJ, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR, Rosenbloom KR, Sloan CA, Learned K, Malladi VS, Wong MC, Barber GP, Cline MS, Dreszer TR, Heitner SG, Karolchik D, Kent WJ, Kirkup VM, Meyer LR, Long JC, Maddren M, Raney BJ, Furey TS, Song L, Grasfeder LL, Giresi PG, Lee B-K, Battenhouse A, Sheffield NC, Simon JM, Showers KA, Safi A, London D, Bhinge AA, Shestak C, Schaner MR, Ki Kim S, Zhang ZZ, Mieczkowski PA, Mieczkowska JO, Liu Z, McDaniell RM, Ni Y, Rashid NU, Kim MJ, Adar S, Zhang Z, Wang T, Winter D, Keefe D, Birney E, Iyer VR, Lieb JD, Crawford GE, Li G, Sandhu KS, Zheng M, Wang P, Luo OJ, Shahab A, Fullwood MJ, Ruan X, Ruan Y, Myers RM, Pauli F, Williams BA, Gertz J, Marinov GK, Reddy TE, Vielmetter J, Partridge E, Trout D, Varley KE, Gasper C, The ENCODE Project Consortium, Overall coordination (data analysis coordination), Data production leads (data production), Lead analysts (data analysis),

Writing group, NHGRI project management (scientific management), Principal investigators (steering committee), Boise State University and University of North Carolina at Chapel Hill Proteomics groups (data production and analysis), Broad Institute Group (data production and analysis), Cold Spring Harbor U of G Center for Genomic Regulation, Barcelona, RIKEN, Sanger Institute, University of Lausanne, Genome Institute of Singapore group (data production and analysis), Data coordination center at UC Santa Cruz (production data coordination), Duke University E University of Texas, Austin, University of North Carolina-Chapel Hill group (data production and analysis), Genome Institute of Singapore group (data production and analysis), HudsonAlpha Institute C UC Irvine, Stanford group (data production and analysis). An integrated encyclopedia of DNA elements in the human genome. Nature. Nature Publishing Group; 2012 Sep;489(7414):57–74.

4. Small S, Blair A, Levine M. Regulation of even-skipped stripe 2 in the Drosophila embryo. EMBO J. John Wiley & Sons, Ltd; 1992 Nov 1;11(11):4047–4057.

5. Halfon MS, Carmena A, Gisselbrecht S, Sackerson CM, Jiménez F, Baylies MK, Michelson AM. Ras Pathway Specificity Is Determined by the Integration of Multiple Signal-Activated and Tissue-Restricted Transcription Factors. Cell. 2000 Sep 29;103(1):63–74.

6. Lettice LA, Williamson I, Wiltshire JH, Peluso S, Devenney PS, Hill AE, Essafi A, Hagman J, Mort R, Grimes G, DeAngelis CL, Hill RE. Opposing Functions of the ETS Factor Family Define Shh Spatial Expression in Limb Buds and Underlie Polydactyly. Dev Cell. 2012 Feb 14;22(2):459–467.

7. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, Ren B. A map of the cis -regulatory sequences in the mouse genome. Nature. Nature Publishing Group; 2012 Aug;488(7409):116–120.

8. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015 Jan 5;109:21.29.1–9. PMCID: PMC4374986

9. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. 2014 Dec 18;159(7):1665–1680. PMID: 25497547

10. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing Chromosome Conformation. Science. American Association for the Advancement of Science; 2002 Feb 15;295(5558):1306–1311. PMID: 11847345

11. Schoenfelder S, Furlan-Magaril M, Mifsud B, Tavares-Cadete F, Sugar R, Javierre B-M, Nagano T, Katsman Y, Sakthidevi M, Wingett SW, Dimitrova E, Dimond A, Edelman LB, Elderkin S, Tabbada K, Darbo E, Andrews S, Herman B, Higgs A, LeProust E, Osborne CS, Mitchell JA, Luscombe NM, Fraser P. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res [Internet]. 2015 Mar 9

[cited 2017 Oct 30]; Available from: http://genome.cshlp.org/content/early/2015/03/07/gr.185272.114 PMID: 25752748

12. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey W, Ewels PA, Herman B, Happe S, Higgs A, LeProust E, Follows GA, Fraser P, Luscombe NM, Osborne CS. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet. 2015 May 4;47(6):ng.3286.

13. Montefiori LE, Sobreira DR, Sakabe NJ, Aneas I, Joslin AC, Hansen GT, Bozek G, Moskowitz IP, McNally EM, Nóbrega MA. A promoter interaction map for cardiovascular disease genetics. Dekker J, McCarthy MI, editors. eLife. 2018 Jul 10;7:e35788.

14. Grzybowski AT, Chen Z, Ruthenburg AJ. Calibrating ChIP-Seq with Nucleosomal Internal Standards to Measure Histone Modification Density Genome Wide. Mol Cell [Internet]. [cited 2015 Jun 3]; Available from: http://www.sciencedirect.com/science/article/pii/S1097276515003044

15. Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. Science. 2013 Mar 1;339(6123):1074–1077. PMID: 23328393

16. Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, Feizi S, Gnirke A, Jr CGC, Kinney JB, Kellis M, Lander ES, Mikkelsen TS. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotechnol. 2012 Mar;30(3):271–277.

17. Wang X, He L, Goggin SM, Saadat A, Wang L, Sinnott-Armstrong N, Claussnitzer M, Kellis M. High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. Nat Commun. Nature Publishing Group; 2018 Dec 19;9(1):5380.

18. Towards a comprehensive catalogue of validated and target-linked human enhancers | Nature Reviews Genetics [Internet]. [cited 2020 Sep 20]. Available from: https://www.nature.com/articles/s41576-019-0209-0

19. Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat Genet. 1995 May;10(1):111–113. PMID: 7647779

20. Yakub M, Moti N, Parveen S, Chaudhry B, Azam I, Iqbal MP. Polymorphisms in MTHFR, MS and CBS Genes and Homocysteine Levels in a Pakistani Population. PLoS ONE [Internet]. 2012 Mar 21 [cited 2020 Jul 21];7(3). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3310006/ PMCID: PMC3310006

21. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L,

Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TFC, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009 Oct 8;461(7265):747–753. PMCID: PMC2831613

22. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science. 2012 Sep 7;337(6099):1190–1195. PMCID: PMC3771521

23. van de Bunt M, Cortes A, Brown MA, Morris AP, McCarthy MI. Evaluating the Performance of Fine-Mapping Strategies at Common Variant GWAS Loci. PLoS Genet [Internet]. 2015 Sep 25 [cited 2020 Sep 20];11(9). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4583479/ PMCID: PMC4583479

24. Nott A, Holtman IR, Coufal NG, Schlachetzki JCM, Yu M, Hu R, Han CZ, Pena M, Xiao J, Wu Y, Keulen Z, Pasillas MP, O'Connor C, Nickl CK, Schafer ST, Shen Z, Rissman RA, Brewer JB, Gosselin D, Gonda DD, Levy ML, Rosenfeld MG, McVicker G, Gage FH, Ren B, Glass CK. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. Science. 2019 29;366(6469):1134–1139. PMCID: PMC7028213

25. Farh KK-H, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shoresh N, Whitton H, Ryan RJH, Shishkin AA, Hatan M, Carrasco-Alfonso MJ, Mayer D, Luckey CJ, Patsopoulos NA, De Jager PL, Kuchroo VK, Epstein CB, Daly MJ, Hafler DA, Bernstein BE. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015 Feb 19;518(7539):337–343.

26. Onengut-Gumuscu S, Chen W-M, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, Farber E, Bonnie JK, Szpak M, Schofield E, Achuthan P, Guo H, Fortune MD, Stevens H, Walker NM, Ward LD, Kundaje A, Kellis M, Daly MJ, Barrett JC, Cooper JD, Deloukas P, Todd JA, Wallace C, Concannon P, Rich SS. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat Genet. Nature Publishing Group; 2015 Apr;47(4):381–386.

27. Weeks EM, Ulirsch JC, Cheng NY, Trippe BL, Fine RS, Miao J, Patwardhan TA, Kanai M, Nasser J, Fulco CP, Tashman KC, Aguet F, Li T, Ordovas-Montanes J, Smillie CS, Biton M, Shalek AK, Ananthakrishnan AN, Xavier RJ, Regev A, Gupta RM, Lage K, Ardlie KG, Hirschhorn JN, Lander ES, Engreitz JM, Finucane HK. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. medRxiv. Cold Spring Harbor Laboratory Press; 2020 Sep 10;2020.09.08.20190561.

28. Consortium SWG of the PG, Ripke S, Walters JT, O'Donovan MC. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. medRxiv. Cold Spring Harbor Laboratory Press; 2020 Sep 13;2020.09.12.20192922.

29. Claussnitzer M, Dankel SN, Kim K-H, Quon G, Meuleman W, Haugen C, Glunk V, Sousa IS, Beaudry JL, Puviindran V, Abdennur NA, Liu J, Svensson P-A, Hsu Y-H, Drucker DJ, Mellgren G, Hui C-C, Hauner H, Kellis M. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N Engl J Med. 2015 Sep 3;373(10):895–907. PMID: 26287746

30. Adult Obesity Causes & Consequences | Overweight & Obesity | CDC [Internet]. [cited 2016 Jun 15]. Available from: http://www.cdc.gov/obesity/adult/causes.html

31. Stunkard AJ, Harris JR, Pedersen NL, McClearn GE. The Body-Mass Index of Twins Who Have Been Reared Apart. N Engl J Med. 1990 May 24;322(21):1483–1487. PMID: 2336075

32. Stunkard AJ, Foch TT, Hrubec Z. A twin study of human obesity. JAMA. 1986 Jul 4;256(1):51–54.

33. Elks CE, Den Hoed M, Zhao JH, Sharp SJ, Wareham NJ, Loos RJF, Ong KK. Variability in the heritability of body mass index: a systematic review and meta-regression. Genomic Endocrinol. 2012;3:29.

34. Shi H, Kichaev G, Pasaniuc B. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. Am J Hum Genet. 2016 07;99(1):139–153. PMCID: PMC5005444

35. Neel JV. Diabetes Mellitus: A “Thrifty” Genotype Rendered Detrimental by “Progress”? Am J Hum Genet. 1962 Dec;14(4):353–362. PMCID: PMC1932342

36. Wang G, Speakman JR. Analysis of Positive Selection at Single Nucleotide Polymorphisms Associated with Body Mass Index Does Not Support the “Thrifty Gene” Hypothesis. Cell Metab. 2016 Oct 11;24(4):531–541.

37. Kenny PJ. Reward Mechanisms in Obesity: New Insights and Future Directions. Neuron. 2011 Feb 24;69(4):664–679. PMCID: PMC3057652

38. Felsted JA, Ren X, Chouinard-Decorte F, Small DM. Genetically Determined Differences in Brain Response to a Primary Food Reward. J Neurosci. 2010 Feb 17;30(7):2428–2432. PMCID: PMC2831082

39. Farooqi IS, Bullmore E, Keogh J, Gillard J, O’Rahilly S, Fletcher PC. Leptin Regulates Striatal Regions and Human Eating Behavior. Science [Internet]. 2007 Sep 7 [cited 2020 Aug 2];317(5843). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3838941/ PMCID: PMC3838941

40. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, Croteau-Chonka DC, Esko T, Fall T, Ferreira T, Gustafsson S, Kutalik Z, Luan J, Mägi R, Randall JC, Winkler TW, Wood AR, Workalemahu T, Faul JD, Smith JA, Hua Zhao J, Zhao W, Chen J, Fehrmann R, Hedman ÅK, Karjalainen J, Schmidt EM, Absher D, Amin N, Anderson D, Beekman M, Bolton JL, Bragg-Gresham JL, Buyske S, Demirkan A, Deng G, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Goel A, Gong J,

Jackson AU, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Mangino M, Mateo Leach I, Medina-Gomez C, Medland SE, Nalls MA, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Shungin D, Stančáková A, Strawbridge RJ, Ju Sung Y, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Isaacs A, Albrecht E, Ärnlöv J, Arscott GM, Attwood AP, Bandinelli S, Barrett A, Bas IN, Bellis C, Bennett AJ, Berne C, Blagieva R, Blüher M, Böhringer S, Bonnycastle LL, Böttcher Y, Boyd HA, Bruinenberg M, Caspersen IH, Ida Chen Y-D, Clarke R, Warwick Daw E, de Craen AJM, Delgado G, Dimitriou M, Doney ASF, Eklund N, Estrada K, Eury E, Folkersen L, Fraser RM, Garcia ME, Geller F, Giedraitis V, Gigante B, Go AS, Golay A, Goodall AH, Gordon SD, Gorski M, Grabe H-J, Grallert H, Grammer TB, Gräßler J, Grönberg H, Groves CJ, Gusto G, Haessler J, Hall P, Haller T, Hallmans G, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hengstenberg C, Holmen O, Hottenga J-J, James AL, Jeff JM, Johansson Å, Jolley J, Juliusdottir T, Kinnunen L, Koenig W, Koskenvuo M, Kratzer W, Laitinen J, Lamina C, Leander K, Lee NR, Lichtner P, Lind L, Lindström J, Sin Lo K, Lobbens S, Lorbeer R, Lu Y, Mach F, Magnusson PKE, Mahajan A, McArdle WL, McLachlan S, Menni C, Merger S, Mihailov E, Milani L, Moayyeri A, Monda KL, Morken MA, Mulas A, Müller G, Müller-Nurasyid M, Musk AW, Nagaraja R, Nöthen MM, Nolte IM, Pilz S, Rayner NW, Renstrom F, Rettig R, Ried JS, Ripke S, Robertson NR, Rose LM, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Scott WR, Seufferlein T, Shi J, Vernon Smith A, Smolonska J, Stanton AV, Steinthorsdottir V, Stirrups K, Stringham HM, Sundström J, Swertz MA, Swift AJ, Syvänen A-C, Tan S-T, Tayo BO, Thorand B, Thorleifsson G, Tyrer JP, Uh H-W, Vandenput L, Verhulst FC, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Warren HR, Waterworth D, Weedon MN, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, The LifeLines Cohort Study, Brennan EP, Choi M, Dastani Z, Drong AW, Eriksson P, Franco-Cereceda A, Gådin JR, Gharavi AG, Goddard ME, Handsaker RE, Huang J, Karpe F, Kathiresan S, Keildson S, Kiryluk K, Kubo M, Lee J-Y, Liang L, Lifton RP, Ma B, McCarroll SA, McKnight AJ, Min JL, Moffatt MF, Montgomery GW, Murabito JM, Nicholson G, Nyholt DR, Okada Y, Perry JRB, Dorajoo R, Reinmaa E, Salem RM, Sandholm N, Scott RA, Stolk L, Takahashi A, Tanaka T, Hooft FM van't, Vinkhuyzen AAE, Westra H-J, Zheng W, Zondervan KT, The ADIPOGen Consortium, The AGEN-BMI Working Group, The CARDIOGRAMplusC4D Consortium, The CKDGen Consortium, The Glgc, The Icbp, The MAGIC Investigators, The MuTHER Consortium, The MIGen Consortium, The PAGE Consortium, The ReproGen Consortium, The GENIE Consortium, The International Endogene Consortium, Heath AC, Arveiler D, Bakker SJL, Beilby J, Bergman RN, Blangero J, Bovet P, Campbell H, Caulfield MJ, Cesana G, Chakravarti A, Chasman DI, Chines PS, Collins FS, Crawford DC, Adrienne Cupples L, Cusi D, Danesh J, de Faire U, den Ruijter HM, Dominiczak AF, Erbel R, Erdmann J, Eriksson JG, Farrall M, Felix SB, Ferrannini E, Ferrières J, Ford I, Forouhi NG, Forrester T, Franco OH, Gansevoort RT, Gejman PV, Gieger C, Gottesman O, Gudnason V, Gyllensten U, Hall AS, Harris TB, Hattersley AT, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Homuth G, Kees Hovingh G, Humphries SE, Hunt SC, Hyppönen E, Illig T, Jacobs KB, Jarvelin M-R, Jöckel K-H, Johansen B, Jousilahti P, Wouter Jukema J, Jula AM, Kaprio J, Kastelein JJP, Keinanen-Kiukaanniemi SM, Kiemeney LA, Knekt P, Kooner JS, Kooperberg C, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimäki T, Lyssenko V, Männistö S, Marette A, Matise TC, McKenzie CA, McKnight B, Moll FL, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Madden PAF, Pasterkamp G, Peden JF, Peters A, Postma DS, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ridker PM, Rioux JD, Ritchie MD, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schunkert H, Schwarz PEH, Sever P, Shuldiner AR, Sinisalo J, Stolk RP, Strauch K, Tönjes A, Trégouët D-A, Tremblay A, Tremoli E, Virtamo J, Vohl M-C, Völker U, Waeber G, Willemsen G, Witteman JC, Carola Zillikens M, Adair LS, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bornstein SR, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PIW, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hui J, Hunter DJ, Hveem K, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, März W, Melbye M, Metspalu A, Moebus S, Munroe PB, Njølstad I, Oostra BA, Palmer CNA, Pedersen NL, Perola M, Pérusse L, Peters U, Power C, Quertermous T, Rauramaa R, Rivadeneira F, Saaristo TE, Saleheen D, Sattar N, Schadt EE, Schlessinger D, Eline Slagboom P, Snieder H, Spector TD, Thorsteinsdottir U, Stumvoll M, Tuomilehto J, Uitterlinden AG, Uusitupa M, van der Harst P, Walker M, Wallaschofski H, Wareham NJ, Watkins H, Weir DR, Wichmann H-E, Wilson JF, Zanen P, Borecki IB, Deloukas P, Fox CS, Heid IM, O'Connell JR, Strachan DP, Stefansson K, van Duijn CM, Abecasis GR, Franke L, Frayling TM, McCarthy MI, Visscher PM, Scherag A, Willer CJ, Boehnke M, Mohlke KL, Lindgren CM, Beckmann JS, Barroso I, North KE, Ingelsson E, Hirschhorn JN, Loos RJF, Speliotes EK. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015 Feb 12;518(7538):197–206.

41. Ramos-Molina B, Martin MG, Lindberg I. PCSK1 Variants and Human Obesity. Prog Mol Biol Transl Sci. 2016;140:47–74. PMCID: PMC6082390

42. Varela L, Horvath TL. Leptin and insulin pathways in POMC and AgRP neurons that modulate energy balance and glucose homeostasis. EMBO Rep. 2012 Dec;13(12):1079–1086. PMCID: PMC3512417

43. Ren D, Zhou Y, Morris D, Li M, Li Z, Rui L. Neuronal SH2B1 is essential for controlling energy and glucose homeostasis. J Clin Invest. 2007 Feb 1;117(2):397–406. PMCID: PMC1765516

44. Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF, Lee JH, Puviindran V, Tam D, Shen M, Son JE, Vakili NA, Sung H-K, Naranjo S, Acemel RD, Manzanares M, Nagy A, Cox NJ, Hui C-C, Gomez-Skarmeta JL, Nóbrega MA. Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature. 2014 Mar 20;507(7492):371–375. PMCID: PMC4113484

45. Pan DZ, Garske KM, Alvarez M, Bhagat YV, Boocock J, Nikkola E, Miao Z, Raulerson CK, Cantor RM, Civelek M, Glastonbury CA, Small KS, Boehnke M, Lusis AJ, Sinsheimer JS, Mohlke KL, Laakso M, Pajukanta P, Ko A. Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. Nat Commun. 2018 Apr 17;9(1):1512.

46. Akıncı A, Türkkahraman D, Tekedereli İ, Özer L, Evren B, Şahin İ, Kalkan T, Çürek Y, Çamtosun E, Döğer E, Bideci A, Güven A, Eren E, Sangün Ö, Çayır A, Bilir P, Törel Ergür A, Ercan O. Novel Mutations in Obesity-related Genes in Turkish Children with Non-syndromic Early Onset Severe Obesity: A Multicentre Study. J Clin Res Pediatr Endocrinol. 2019 22;11(4):341–349. PMCID: PMC6878344

47. Nunziata A, Funcke J-B, Borck G, von Schnurbein J, Brandt S, Lennerz B, Moepps B, Gierschik P, Fischer-Posovszky P, Wabitsch M. Functional and Phenotypic Characteristics of Human Leptin Receptor Mutations. J Endocr Soc. 2019 Jan 1;3(1):27–41. PMCID: PMC6293235

48. Montague CT, Farooqi IS, Whitehead JP, Soos MA, Rau H, Wareham NJ, Sewter CP, Digby JE, Mohammed SN, Hurst JA, Cheetham CH, Earley AR, Barnett AH, Prins JB, O'Rahilly S. Congenital leptin deficiency is associated with severe early-onset obesity in humans. Nature. 1997 Jun 26;387(6636):903–908. PMID: 9202122

49. Nordang GBN, Busk ØL, Tveten K, Hanevik HI, Fell AKM, Hjelmesæth J, Holla ØL, Hertel JK. Next-generation sequencing of the monogenic obesity genes LEP, LEPR, MC4R, PCSK1 and POMC in a Norwegian cohort of patients with morbid obesity and normal weight controls. Mol Genet Metab. 2017;121(1):51–56. PMID: 28377240

50. Wang L, Sui L, Panigrahi SK, Meece K, Xin Y, Kim J, Gromada J, Doege CA, Wardlaw SL, Egli D, Leibel RL. PC1/3 Deficiency Impacts Pro-opiomelanocortin Processing in Human Embryonic Stem Cell-Derived Hypothalamic Neurons. Stem Cell Rep. 2017 Jan 26;8(2):264–277. PMCID: PMC5312251

51. Farooqi IS, O'Rahilly S. Mutations in ligands and receptors of the leptin–melanocortin pathway that lead to obesity. Nat Clin Pract Endocrinol Metab. Nature Publishing Group; 2008 Oct;4(10):569–577.

52. Xu B, Goulding EH, Zang K, Cepoi D, Cone RD, Jones KR, Tecott LH, Reichardt LF. Brain-derived neurotrophic factor regulates energy balance downstream of melanocortin-4 receptor. Nat Neurosci. 2003 Jul;6(7):736–742. PMCID: PMC2710100

53. Bekinschtein P, Cammarota M, Katche C, Slipczuk L, Rossato JI, Goldin A, Izquierdo I, Medina JH. BDNF is essential to promote persistence of long-term memory storage. Proc Natl Acad Sci. National Academy of Sciences; 2008 Feb 19;105(7):2711–2716. PMID: 18263738

54. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. PLoS Biol. 2010 Jan 26;8(1):e1000294. PMCID: PMC2811148

55. Scherag A, Jarick I, Grothe J, Biebermann H, Scherag S, Volckmar A-L, Vogel CIG, Greene B, Hebebrand J, Hinney A. Investigation of a Genome Wide Association Signal for Obesity: Synthetic Association and Haplotype Analyses at the Melanocortin 4 Receptor Gene Locus. PLoS ONE [Internet]. 2010 Nov 15 [cited 2020 Sep 21];5(11). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2981522/ PMCID: PMC2981522

56. Lee AWS, Hengstler H, Schwald K, Berriel-Diaz M, Loreth D, Kirsch M, Kretz O, Haas CA, de Angelis MH, Herzig S, Brümmendorf T, Klingenspor M, Rathjen FG, Rozman J, Nicholson G, Cox RD, Schäfer MKE. Functional inactivation of the genome-wide association study obesity gene neuronal growth regulator 1 in mice causes a body mass phenotype. PloS One. 2012;7(7):e41537. PMCID: PMC3402391

57. Grarup N, Moltke I, Andersen MK, Dalby M, Vitting-Seerup K, Kern T, Mahendran Y, Jørsboe E, Larsen CVL, Dahl-Petersen IK, Gilly A, Suveges D, Dedoussis G, Zeggini E, Pedersen O, Andersson R, Bjerregaard P, Jørgensen ME, Albrechtsen A, Hansen T. Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes. Nat Genet. 2018;50(2):172–174. PMCID: PMC5828106

58. Saeed S, Bonnefond A, Tamanini F, Mirza MU, Manzoor J, Janjua QM, Din SM, Gaitan J, Milochau A, Durand E, Vaillant E, Haseeb A, De Graeve F, Rabearivelo I, Sand O, Queniat G, Boutry R, Schott DA, Ayesha H, Ali M, Khan WI, Butt TA, Rinne T, Stumpel C, Abderrahmani A, Lang J, Arslan M, Froguel P. Loss-of-function mutations in ADCY3 cause monogenic severe obesity. Nat Genet. 2018;50(2):175–179. PMID: 29311637

59. Yan X, Wang Z, Schmidt V, Gauert A, Willnow TE, Heinig M, Poy MN. Cadm2 regulates body weight and energy homeostasis in mice. Mol Metab. 2018 Feb 1;8:180–188.

60. Zhu H, Guariglia S, Li W, Brancho D, Wang ZV, Scherer PE, Chow C-W. Role of Extracellular Signal-regulated Kinase 5 in Adipocyte Signaling. J Biol Chem. 2014 Feb 28;289(9):6311–6322. PMCID: PMC3937697

61. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JRB, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch A-M, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin M-R, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJF, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CNA, Doney ASF, Morris AD, Smith GD, Hattersley AT, McCarthy MI. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007 May 11;316(5826):889–894. PMCID: PMC2646098

62. Cecil JE, Tavendale R, Watt P, Hetherington MM, Palmer CNA. An Obesity-Associated FTO Gene Variant and Increased Energy Intake in Children. N Engl J Med. 2008 Dec 11;359(24):2558–2566. PMID: 19073975

63. Timpson NJ, Emmett PM, Frayling TM, Rogers I, Hattersley AT, McCarthy MI, Davey Smith G. The fat mass–and obesity-associated locus and dietary intake in children. Am J Clin Nutr. Oxford Academic; 2008 Oct 1;88(4):971–978.

64. Speakman JR, Rance KA, Johnstone AM. Polymorphisms of the FTO gene are associated with variation in energy intake, but not energy expenditure. Obes Silver Spring Md. 2008 Aug;16(8):1961–1965. PMID: 18551109

65. Wardle J, Carnell S, Haworth CMA, Farooqi IS, O'Rahilly S, Plomin R. Obesity Associated Genetic Variation in FTO Is Associated with Diminished Satiety. J Clin Endocrinol Metab. Oxford Academic; 2008 Sep 1;93(9):3640–3643.

66. Boissel S, Reish O, Proulx K, Kawagoe-Takaki H, Sedgwick B, Yeo GSH, Meyre D, Golzio C, Molinari F, Kadhom N, Etchevers HC, Saudek V, Farooqi IS, Froguel P, Lindahl T, O'Rahilly S, Munnich A, Colleaux L. Loss-of-Function Mutation in the Dioxygenase-Encoding FTO Gene Causes Severe Growth Retardation and Multiple Malformations. Am J Hum Genet. 2009 Jul 10;85(1):106–111. PMCID: PMC2706958

67. Aguet F, Barbeira AN, Bonazzola R, Brown A, Castel SE, Jo B, Kasela S, Kim-Hellmuth S, Liang Y, Oliva M, Parsana PE, Flynn E, Fresard L, Gaamzon ER, Hamel AR, He Y, Hormozdiari F, Mohammadi P, Muñoz-Aguirre M, Park Y, Saha A, Segrć AV, Strober BJ, Wen X, Wucher V, Das S, Garrido-Martín D, Gay NR, Handsaker RE, Hoffman PJ, Kashin S, Kwong A, Li X, MacArthur D, Rouhana JM, Stephens M, Todres E, Viñuela A, Wang G, Zou Y, The GTEx Consortium, Brown CD, Cox N, Dermitzakis E, Engelhardt BE, Getz G, Guigo R, Montgomery SB, Stranger BE, Im HK, Battle A, Ardlie KG, Lappalainen T. The GTEx Consortium atlas of genetic regulatory effects across human tissues [Internet]. Genetics; 2019 Oct. Available from: http://biorxiv.org/lookup/doi/10.1101/787903

68. Hormozdiari F, Zhu A, Kichaev G, Ju CJ-T, Segrè AV, Joo JWJ, Won H, Sankararaman S, Pasaniuc B, Shifman S, Eskin E. Widespread Allelic Heterogeneity in Complex Traits. Am J Hum Genet. 2017 May 4;100(5):789–802. PMCID: PMC5420356

69. Dynamic genetic regulation of gene expression during cellular differentiation | Science [Internet]. [cited 2020 Aug 5]. Available from: https://science.sciencemag.org/content/364/6447/1287?fbclid=IwAR3fogSPVd-WofbUsz6Sz3Z0u9IEZKQkX-IjODzWWTXshWZYdg5C-Qk772E

70. Beagan JA, Pastuzyn ED, Fernandez LR, Guo MH, Feng K, Titus KR, Chandrashekar H, Shepherd JD, Phillips-Cremins JE. Three-dimensional genome restructuring across timescales of activity-induced neuronal gene expression. Nat Neurosci. Nature Publishing Group; 2020 Jun;23(6):707–717.

71. Idelevich A, Sato K, Nagano K, Rowe G, Gori F, Baron R. Neuronal hypothalamic regulation of body metabolism and bone density is galanin dependent. J Clin Invest. American Society for Clinical Investigation; 2018 Jun 1;128(6):2626–2641. PMID: 0

72. Yao L, Liu Y, Qiu Z, Kumar S, Curran JE, Blangero J, Chen Y, Lehman DM. Molecular Profiling of Human Induced Pluripotent Stem Cell-Derived Hypothalamic Neurones Provides Developmental Insights into Genetic Loci for Body Weight Regulation. J Neuroendocrinol. 2017 Feb;29(2). PMCID: PMC5328859

73. van der Klaauw AA, Farooqi IS. The hunger genes: pathways to obesity. Cell. 2015 Mar 26;161(1):119–132. PMID: 25815990

74. Fischer-Posovszky P, Newell FS, Wabitsch M, Tornqvist HE. Human SGBS Cells – a Unique Tool for Studies of Human Fat Cell Biology. Obes Facts. 2008;1(4):184–189. PMID: 20054179

75. Cairns J, Freire-Pritchett P, Wingett SW, Várnai C, Dimond A, Plagnol V, Zerbino D, Schoenfelder S, Javierre B-M, Osborne C, Fraser P, Spivakov M. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. Genome Biol [Internet]. 2016 Jun 15 [cited 2016 Oct 31];17. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4908757/ PMCID: PMC4908757

76. Chen L, Hu H, Qiu W, Shi K, Kassem M. Actin depolymerization enhances adipogenic differentiation in human stromal stem cells. Stem Cell Res. 2018 May 1;29:76–83.

77. Yang W, Guo X, Thein S, Xu F, Sugii S, Baas PW, Radda GK, Han W. Regulation of adipogenesis by cytoskeleton remodelling is facilitated by acetyltransferase MEC-17-dependent acetylation of α-tubulin. Biochem J. 2013 Feb 1;449(3):605–612. PMCID: PMC5573127

78. Aratani Y, Kitagawa Y. Enhanced synthesis and secretion of type IV collagen and entactin during adipose conversion of 3T3-L1 cells and production of unorthodox laminin complex. J Biol Chem. 1988 Nov 5;263(31):16163–16169. PMID: 2460444

79. Huang G, Greenspan DS. ECM roles in the function of metabolic tissues. Trends Endocrinol Metab. Elsevier; 2012 Jan 1;23(1):16–22. PMID: 22070921

80. Nakajima I, Yamaguchi T, Ozutsumi K, Aso H. Adipose tissue extracellular matrix: newly organized by adipocytes during differentiation. Differ Res Biol Divers. 1998 Aug;63(4):193–200. PMID: 9745710

81. Vaicik MK, Blagajcevic A, Ye H, Morse MC, Yang F, Goddi A, Brey EM, Cohen RN. The Absence of Laminin α4 in Male Mice Results in Enhanced Energy Expenditure and Increased Beige Subcutaneous Adipose Tissue. Endocrinology. 2018 01;159(1):356–367. PMCID: PMC5761598

82. Rosen ED, Sarraf P, Troy AE, Bradwin G, Moore K, Milstone DS, Spiegelman BM, Mortensen RM. PPAR gamma is required for the differentiation of adipose tissue in vivo and in vitro. Mol Cell. 1999 Oct;4(4):611–617. PMID: 10549292

83. Stavreva DA, Coulon A, Baek S, Sung M-H, John S, Stixova L, Tesikova M, Hakim O, Miranda T, Hawkins M, Stamatoyannopoulos JA, Chow CC, Hager GL. Dynamics of chromatin accessibility and long-range interactions in response to glucocorticoid pulsing. Genome Res. 2015 Feb 12;gr.184168.114. PMID: 25677181

84. Siersbæk R, Madsen JGS, Javierre BM, Nielsen R, Bagge EK, Cairns J, Wingett SW, Traynor S, Spivakov M, Fraser P, Mandrup S. Dynamic Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte Differentiation. Mol Cell. 2017 May 4;66(3):420-435.e5.

85. Ulirsch JC, Nandakumar SK, Wang L, Giani FC, Zhang X, Rogov P, Melnikov A, McDonel P, Do R, Mikkelsen TS, Sankaran VG. Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. Cell. 2016 Jun 2;165(6):1530–1545.

86. Stephens JM, Butts MD, Pekala PH. Regulation of transcription factor mRNA accumulation during 3T3-L1 preadipocyte differentiation by tumour necrosis factor-alpha. J Mol Endocrinol. 1992 Aug;9(1):61–72. PMID: 1515026

87. Distel RJ, Ro HS, Rosen BS, Groves DL, Spiegelman BM. Nucleoprotein complexes that regulate gene expression in adipocyte differentiation: direct participation of c-fos. Cell. 1987 Jun 19;49(6):835–844. PMID: 3555845

88. White UA, Stephens JM. Transcriptional factors that promote formation of white adipose tissue. Mol Cell Endocrinol. 2010 Apr 29;318(1–2):10–14. PMCID: PMC3079373

89. Adipocyte browning and resistance to obesity in mice is induced by expression of ATF3 | Communications Biology [Internet]. [cited 2020 Apr 7]. Available from: https://www.nature.com/articles/s42003-019-0624-y

90. Liu Y, Maekawa T, Yoshida K, Muratani M, Chatton B, Ishii S. The Transcription Factor ATF7 Controls Adipocyte Differentiation and Thermogenic Gene Programming. iScience. 2019 Feb 18;13:98–112. PMCID: PMC6402263

91. Lee Y-S, Sasaki T, Kobayashi M, Kikuchi O, Kim H-J, Yokota-Hashimoto H, Shimpuku M, Susanti V-Y, Ido-Kitamura Y, Kimura K, Inoue H, Tanaka-Okamoto M, Ishizaki H, Miyoshi J, Ohya S, Tanaka Y, Kitajima S, Kitamura T. Hypothalamic ATF3 is involved in regulating glucose and energy metabolism in mice. Diabetologia. 2013 Jun;56(6):1383–1393. PMCID: PMC3648686

92. Pelletier P, Gauthier K, Sideleva O, Samarut J, Silva JE. Mice lacking the thyroid hormone receptor-alpha gene spend more energy in thermogenesis, burn more fat, and are less sensitive to high-fat diet-induced obesity. Endocrinology. 2008 Dec;149(12):6471–6486. PMID: 18719022

93. Dahle MK, Taskén K, Taskén KA. USF2 inhibits C/EBP-mediated transcriptional regulation of the RIIβ subunit of cAMP-dependent protein kinase. BMC Mol Biol. 2002 Jun 21;3(1):10.

94. Laurila P-P, Soronen J, Kooijman S, Forsström S, Boon MR, Surakka I, Kaiharju E, Coomans CP, Berg SAAVD, Autio A, Sarin A-P, Kettunen J, Tikkanen E, Manninen T, Metso J, Silvennoinen R, Merikanto K, Ruuth M, Perttilä J, Mäkelä A, Isomi A, Tuomainen AM, Tikka A, Ramadan UA, Seppälä I, Lehtimäki T, Eriksson J, Havulinna A, Jula A, Karhunen PJ, Salomaa V, Perola M, Ehnholm C, Lee-Rueckert M, Eck MV, Roivainen A, Taskinen M-R, Peltonen L, Mervaala E, Jalanko A, Hohtola E, Olkkonen VM, Ripatti S, Kovanen PT, Rensen PCN, Suomalainen A, Jauhiainen M. USF1 deficiency activates brown adipose tissue and improves cardiometabolic health. Sci Transl Med. American Association for the Advancement of Science; 2016 Jan 27;8(323):323ra13-323ra13. PMID: 26819196

95. Shimomura K, Kumar V, Koike N, Kim T-K, Chong J, Buhr ED, Whiteley AR, Low SS, Omura C, Fenner D, Owens JR, Richards M, Yoo S-H, Hong H-K, Vitaterna MH, Bass J, Pletcher MT, Wiltshire T, Hogenesch J, Lowrey PL, Takahashi JS. Usf1, a suppressor of the circadian Clock mutant, reveals the nature of the DNA-binding of the CLOCK:BMAL1 complex in mice. eLife [Internet]. 2013 Apr 9 [cited 2020 Apr 7];2. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3622178/ PMCID: PMC3622178

96. Honma S, Kawamoto T, Takagi Y, Fujimoto K, Sato F, Noshiro M, Kato Y, Honma K. Dec1 and Dec2 are regulators of the mammalian molecular clock. Nature. 2002 Oct 24;419(6909):841–844. PMID: 12397359

97. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, Anttila V, Xu H, Zang C, Farh K, Ripke S, Day FR, Consortium R, Purcell S, Stahl E, Lindstrom S, Perry JRB, Okada Y, Raychaudhuri S, Daly M, Patterson N, Neale BM, Price AL. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015 Nov;47(11):1228–1235. PMCID: PMC4626285

98. Calderon D, Nguyen MLT, Mezger A, Kathiria A, Müller F, Nguyen V, Lescano N, Wu B, Trombetta J, Ribado JV, Knowles DA, Gao Z, Blaeschke F, Parent AV, Burt TD, Anderson MS, Criswell LA, Greenleaf WJ, Marson A, Pritchard JK. Landscape of stimulation-responsive chromatin across diverse human immune cells. Nat Genet. Nature Publishing Group; 2019 Oct;51(10):1494–1505.

99. Praggastis M, Tortelli B, Zhang J, Fujiwara H, Sidhu R, Chacko A, Chen Z, Chung C, Lieberman AP, Sikora J, Davidson C, Walkley SU, Pipalia NH, Maxfield FR, Schaffer JE, Ory DS. A Murine Niemann-Pick C1 I1061T Knock-In Model Recapitulates the Pathological Features of the Most Prevalent Human Disease Allele. J Neurosci. 2015 May 27;35(21):8091–8106. PMCID: PMC4444535

100. Rantakari P, Lagerbohm H, Kaimainen M, Suomela J-P, Strauss L, Sainio K, Pakarinen P, Poutanen M. Hydroxysteroid (17β) Dehydrogenase 12 Is Essential for Mouse Organogenesis and Embryonic Survival. Endocrinology. Oxford Academic; 2010 Apr 1;151(4):1893–1901.

101. Gamero-Villarroel C, González LM, Rodríguez-López R, Albuquerque D, Carrillo JA, García-Herráiz A, Flores I, Gervasini G. Influence of TFAP2B and KCTD15 genetic variability on personality dimensions in anorexia and bulimia nervosa. Brain Behav. 2017;7(9):e00784. PMCID: PMC5607548

102. Williams MJ, Goergen P, Rajendran J, Zheleznyakova G, Hägglund MG, Perland E, Bagchi S, Kalogeropoulou A, Khan Z, Fredriksson R, Schiöth HB. Obesity-Linked Homologues TfAP-2 and Twz Establish Meal Frequency in Drosophila melanogaster. PLoS Genet [Internet]. 2014 Sep 4 [cited 2020 Apr 9];10(9). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4154645/ PMCID: PMC4154645

103. Doche ME, Bochukova EG, Su H-W, Pearce LR, Keogh JM, Henning E, Cline JM, Dale A, Cheetham T, Barroso I, Argetsinger LS, O'Rahilly S, Rui L, Carter-Su C, Farooqi IS.

Human SH2B1 mutations are associated with maladaptive behaviors and obesity. J Clin Invest. 2012 Dec 3;122(12):4732–4736. PMCID: PMC3533535

104. Hershkovitz T, Kurolap A, Gonzaga-Jauregui C, Paperna T, Mory A, Wolf SE, Overton JD, Shuldiner AR, Saada A, Mandel H, Baris Feldman H. A novel TUFM homozygous variant in a child with mitochondrial cardiomyopathy expands the phenotype of combined oxidative phosphorylation deficiency 4. J Hum Genet. Nature Publishing Group; 2019 Jun;64(6):589–595.

105. Engin A. Circadian Rhythms in Diet-Induced Obesity. Adv Exp Med Biol. 2017;960:19–52. PMID: 28585194

106. Chaix A, Lin T, Le HD, Chang MW, Panda S. Time-Restricted Feeding Prevents Obesity and Metabolic Syndrome in Mice Lacking a Circadian Clock. Cell Metab. 2019 05;29(2):303-319.e4. PMID: 30174302

107. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Steinthorsdottir V, Scott RA, Grarup N, Cook JP, Schmidt EM, Wuttke M, Sarnowski C, Mägi R, Nano J, Gieger C, Trompet S, Lecoeur C, Preuss M, Prins BP, Guo X, Bielak LF, Bennett AJ, Bork-Jensen J, Brummett CM, Canouil M, Eckardt K-U, Fischer K, Kardia SL, Kronenberg F, Läll K, Liu C-T, Locke AE, Luan J, Ntalla I, Nylander V, Schönherr S, Schurmann C, Yengo L, Bottinger EP, Brandslund I, Christensen C, Dedoussis G, Florez JC, ford I, Franco OH, Frayling TM, Giedraitis V, Hackinger S, Hattersley AT, Herder C, Ikram MA, Ingelsson M, Jørgensen ME, Jørgensen T, Kriebel J, Kuusisto J, Ligthart S, Lindgren CM, Linneberg A, Lyssenko V, Mamakou V, Meitinger T, Mohlke KL, Morris AD, Nadkarni G, Pankow JS, Peters A, Sattar N, Stančáková A, Strauch K, Taylor KD, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tuomilehto J, Witte DR, Dupuis J, Peyser PA, Zeggini E, Loos RJF, Froguel P, Ingelsson E, Lind L, Groop L, Laakso M, Collins FS, Jukema JW, Palmer CNA, Grallert H, Metspalu A, Dehghan A, Köttgen A, Abecasis G, Meigs JB, Rotter JI, Marchini J, Pedersen O, Hansen T, Langenberg C, Wareham NJ, Stefansson K, Gloyn AL, Morris AP, Boehnke M, McCarthy MI. Fine-mapping of an expanded set of type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet. 2018 Nov;50(11):1505–1513. PMCID: PMC6287706

108. Wabitsch M, Brenner RE, Melzner I, Braun M, Möller P, Heinze E, Debatin KM, Hauner H. Characterization of a human preadipocyte cell strain with high capacity for adipose differentiation. Int J Obes Relat Metab Disord J Int Assoc Study Obes. 2001 Jan;25(1):8–15. PMID: 11244452

109. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell. 2010 May;38(4):576–589.

110. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. Oxford Academic; 2010 Mar 15;26(6):841–842.

111. Recent Segmental Duplications in the Human Genome | Science [Internet]. [cited 2020 Oct 2]. Available from: https://science.sciencemag.org/content/297/5583/1003

112. Antonacci F, Kidd JM, Marques-Bonet T, Teague B, Ventura M, Girirajan S, Alkan C, Campbell CD, Vives L, Malig M, Rosenfeld JA, Ballif BC, Shaffer LG, Graves TA, Wilson RK, Schwartz DC, Eichler EE. A large and complex structural polymorphism at 16p12.1 underlies microdeletion disease risk. Nat Genet. 2010 Sep;42(9):745–750. PMCID: PMC2930074

113. Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K, Saeed S, Hamilton-Shield J, Clayton-Smith J, O'Rahilly S, Hurles ME, Farooqi IS. Large, rare chromosomal deletions associated with severe early-onset obesity. Nature. 2010 Feb 4;463(7281):666–670. PMCID: PMC3108883

114. Bachmann-Gagescu R, Mefford HC, Cowan C, Glew GM, Hing AV, Wallace S, Bader PI, Hamati A, Reitnauer PJ, Smith R, Stockton DW, Muhle H, Helbig I, Eichler EE, Ballif BC, Rosenfeld J, Tsuchiya KD. Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity. Genet Med. Nature Publishing Group; 2010 Oct;12(10):641–647.

115. Maures TJ, Kurzer JH, Carter-Su C. SH2B1 (SH2-B) and JAK2: a multifunctional adaptor protein and kinase made for each other. Trends Endocrinol Metab. 2007 Jan 1;18(1):38–45.

116. Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, Martinet D, Shen Y, Valsesia A, Beckmann ND, Thorleifsson G, Belfiore M, Bouquillon S, Campion D, de Leeuw N, de Vries BBA, Esko T, Fernandez BA, Fernández-Aranda F, Fernández-Real JM, Gratacòs M, Guilmatre A, Hoyer J, Jarvelin M-R, Frank Kooy R, Kurg A, Le Caignec C, Männik K, Platt OS, Sanlaville D, Van Haelst MM, Villatoro Gomez S, Walha F, Wu B, Yu Y, Aboura A, Addor M-C, Alembik Y, Antonarakis SE, Arveiler B, Barth M, Bednarek N, Béna F, Bergmann S, Beri M, Bernardini L, Blaumeiser B, Bonneau D, Bottani A, Boute O, Brunner HG, Cailley D, Callier P, Chiesa J, Chrast J, Coin L, Coutton C, Cuisset J-M, Cuvellier J-C, David A, de Freminville B, Delobel B, Delrue M-A, Demeer B, Descamps D, Didelot G, Dieterich K, Disciglio V, Doco-Fenzy M, Drunat S, Duban-Bedu B, Dubourg C, El-Sayed Moustafa JS, Elliott P, Faas BHW, Faivre L, Faudet A, Fellmann F, Ferrarini A, Fisher R, Flori E, Forer L, Gaillard D, Gerard M, Gieger C, Gimelli S, Gimelli G, Grabe HJ, Guichet A, Guillin O, Hartikainen A-L, Heron D, Hippolyte L, Holder M, Homuth G, Isidor B, Jaillard S, Jaros Z, Jiménez-Murcia S, Joly Helas G, Jonveaux P, Kaksonen S, Keren B, Kloss-Brandstätter A, Knoers NVAM, Koolen DA, Kroisel PM, Kronenberg F, Labalme A, Landais E, Lapi E, Layet V, Legallic S, Leheup B, Leube B, Lewis S, Lucas J, MacDermot KD, Magnusson P, Marshall C, Mathieu-Dramard M, McCarthy MI, Meitinger T, Antonietta Mencarelli M, Merla G, Moerman A, Mooser V, Morice-Picard F, Mucciolo M, Nauck M, Coumba Ndiaye N, Nordgren A, Pasquier L, Petit

F, Pfundt R, Plessis G, Rajcan-Separovic E, Paolo Ramelli G, Rauch A, Ravazzolo R, Reis A, Renieri A, Richart C, Ried JS, Rieubland C, Roberts W, Roetzer KM, Rooryck C, Rossi M, Saemundsen E, Satre V, Schurmann C, Sigurdsson E, Stavropoulos DJ, Stefansson H, Tengström C, Thorsteinsdóttir U, Tinahones FJ, Touraine R, Vallée L, van Binsbergen E, Van der Aa N, Vincent-Delorme C, Visvikis-Siest S, Vollenweider P, Völzke H, Vulto-van Silfhout AT, Waeber G, Wallgren-Pettersson C, Witwicki RM, Zwolinksi S, Andrieux J, Estivill X, Gusella JF, Gustafsson O, Metspalu A, Scherer SW, Stefansson K, Blakemore AIF, Beckmann JS, Froguel P. Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. Nature. Nature Publishing Group; 2011 Oct;478(7367):97–102.

117. González JR, Cáceres A, Esko T, Cuscó I, Puig M, Esnaola M, Reina J, Siroux V, Bouzigon E, Nadif R, Reinmaa E, Milani L, Bustamante M, Jarvis D, Antó JM, Sunyer J, Demenais F, Kogevinas M, Metspalu A, Cáceres M, Pérez-Jurado LA. A Common 16p11.2 Inversion Underlies the Joint Susceptibility to Asthma and Obesity. Am J Hum Genet. 2014 Mar 6;94(3):361–372. PMCID: PMC3951940

118. UK Biobank [Internet]. Neale lab. [cited 2020 Apr 21]. Available from: http://www.nealelab.is/uk-biobank

119. Miller JA, Ding S-L, Sunkin SM, Smith KA, Ng L, Szafer A, Ebbert A, Riley ZL, Royall JJ, Aiona K, Arnold JM, Bennet C, Bertagnolli D, Brouner K, Butler S, Caldejon S, Carey A, Cuhaciyan C, Dalley RA, Dee N, Dolbeare TA, Facer BAC, Feng D, Fliss TP, Gee G, Goldy J, Gourley L, Gregor BW, Gu G, Howard RE, Jochim JM, Kuan CL, Lau C, Lee C-K, Lee F, Lemon TA, Lesnar P, McMurray B, Mastan N, Mosqueda N, Naluai-Cecchini T, Ngo N-K, Nyhus J, Oldre A, Olson E, Parente J, Parker PD, Parry SE, Stevens A, Pletikos M, Reding M, Roll K, Sandman D, Sarreal M, Shapouri S, Shapovalova NV, Shen EH, Sjoquist N, Slaughterbeck CR, Smith M, Sodt AJ, Williams D, Zöllei L, Fischl B, Gerstein MB, Geschwind DH, Glass IA, Hawrylycz MJ, Hevner RF, Huang H, Jones AR, Knowles JA, Levitt P, Phillips JW, Sestan N, Wohnoutka P, Dang C, Bernard A, Hohmann JG, Lein ES. Transcriptional landscape of the prenatal human brain. Nature. 2014 Apr 10;508(7495):199–206. PMCID: PMC4105188

120. Chou C-M, Chen Y-C, Lee M-T, Chen G-D, Lu I-C, Chen S-T, Huang C-J. Expression and characterization of a brain-specific protein kinase BSK146 from zebrafish. Biochem Biophys Res Commun. 2006 Feb;340(3):767–775.

121. Blake JA, Eppig JT, Kadin JA, Richardson JE, Smith CL, Bult CJ, the Mouse Genome Database Group. Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. Nucleic Acids Res. 2017 04;45(D1):D723–D729. PMCID: PMC5210536

122. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell. 2013 Feb 28;152(5):1173–1183. PMCID: PMC3664290

123. Nara K, Akasako Y, Matsuda Y, Fukazawa Y, Iwashita S, Kataoka M, Nagai Y. Cloning and characterization of a novel serine/threonine protein kinase gene expressed predominantly in developing brain. Eur J Biochem. 2001;268(9):2642–2651.

124. Aneas I, Decker DC, Howard CL, Sobreira DR, Sakabe NJ, Blaine KM, Stein MM, Hrusch CL, Montefiori LE, Tena J, Magnaye KM, Clay SM, Gern JE, Jackson DJ, Altman MC, Naureckas ET, Hogarth DK, White SR, Gomez-Skarmeta JL, Schoetler N, Ober C, Sperling AI, Nobrega MA. Asthma-associated variants induce IL33 differential expression through a novel regulatory region. bioRxiv. Cold Spring Harbor Laboratory; 2020 Sep 11;2020.09.09.290098.

125. Li YI, Geijn B van de, Raj A, Knowles DA, Petti AA, Golan D, Gilad Y, Pritchard JK. RNA splicing is a primary link between genetic variation and disease. Science. American Association for the Advancement of Science; 2016 Apr 29;352(6285):600–604. PMID: 27126046

126. Zhang Z, Luo K, Zou Z, Qiu M, Tian J, Sieh L, Shi H, Zou Y, Wang G, Morrison J, Zhu AC, Qiao M, Li Z, Stephens M, He X, He C. Genetic analyses support the contribution of mRNA N 6 -methyladenosine (m 6 A) modification to human disease heritability. Nat Genet. Nature Publishing Group; 2020 Sep;52(9):939–949.

127. Mittleman BE, Pott S, Warland S, Zeng T, Mu Z, Kaur M, Gilad Y, Li Y. Alternative polyadenylation mediates genetic regulation of gene expression [Internet]. eLife. eLife Sciences Publications Limited; 2020 [cited 2020 Sep 15]. Available from: https://elifesciences.org/articles/57492

128. Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E. Identifying Causal Variants at Loci with Multiple Signals of Association. Genetics. Genetics; 2014 Oct 1;198(2):497–508. PMID: 25104515

129. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinforma Oxf Engl. 2016 15;32(10):1493–1501. PMCID: PMC4866522

130. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine-mapping. bioRxiv. 2019 Jul 29;501114.

131. Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, Berulava T, Lee H, Oskolkov N, Fadista J, Ehlers K, Wahl S, Hoffmann C, Qian K, Rönn T, Riess H, Müller-Nurasyid M, Bretschneider N, Schroeder T, Skurk T, Horsthemke B, DIAGRAM+Consortium, Spieler D, Klingenspor M, Seifert M, Kern MJ, Mejhert N, Dahlman I, Hansson O, Hauck SM, Blüher M, Arner P, Groop L, Illig T, Suhre K, Hsu Y-H, Mellgren G, Hauner H, Laumen H. Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. Cell. 2014 Jan 16;156(1–2):343–358. PMID: 24439387

132. Mahajan A, Spracklen CN, Zhang W, Ng MC, Petty LE, Kitajima H, Yu GZ, Rueger S, Speidel L, Kim YJ, Horikoshi M, Mercader JM, Taliun D, Moon S, Kwak S-H, Robertson NR, Rayner NW, Loh M, Kim B-J, Chiou J, Miguel-Escalada I, Parolo P della B, Lin K, Bragg F, Preuss MH, Takeuchi F, Nano J, Guo X, Lamri A, Nakatochi M, Scott RA, Lee J-J, Huerta-Chagoya A, Graff M, Chai J-F, Parra EJ, Yao J, Bielak LF, Tabara Y, Hai Y, Steinthorsdottir V, Cook JP, Kals M, Grarup N, Schmidt EM, Pan I, Sofer T, Wuttke M, Sarnowski C, Gieger C, Nousome D, Trompet S, Long J, Sun M, Tong L, Chen W-M, Ahmad M, Noordam R, Lim VJ, Tam CH, Joo YY, Chen C-H, Raffield LM, Lecoeur C, Maruthur NM, Prins BP, Nicolas A, Yanek LR, Chen G, Jensen RA, Tajuddin S, Kabagambe E, An P, Xiang AH, Choi HS, Cade BE, Tan J, Abaitua F, Adair LS, Adeyemo A, Aguilar-Salinas CA, Akiyama M, Anand SS, Bertoni A, Bian Z, Bork-Jensen J, Brandslund I, Brody JA, Brummett CM, Buchanan TA, Canouil M, Chan JC, Chang L-C, Chee M-L, Chen J, Chen S-H, Chen Y-T, Chen Z, Chuang L-M, Cushman M, Das SK, Silva HJ de, Dedoussis G, Dimitrov L, Doumatey AP, Du S, Duan Q, Eckardt K-U, Emery LS, Evans DS, Evans MK, Fischer K, Floyd JS, Ford I, Fornage M, Franco OH, Frayling TM, Freedman BI, Fuchsberger C, Genter P, Gerstein HC, Giedraitis V, Gonzalez-Villalpando C, Gonzalez-Villalpando ME, Goodarzi MO, Gordon-Larsen P, Gorkin D, Gross M, Guo Y, Hackinger S, Han S, Hattersley AT, Herder C, Howard A-G, Hsueh W, Huang M, Huang W, Hung Y-J, Hwang MY, Hwu C-M, Ichihara S, Ikram MA, Ingelsson M, Islam MT, Isono M, Jang H-M, Jasmine F, Jiang G, Jonas JB, Jorgensen ME, Jorgensen T, Kamatani Y, Kandeel FR, Kasturiratne A, Katsuya T, Kaur V, Kawaguchi T, Keaton JM, Kho AN, Khor C-C, Kibriya MG, Kim D-H, Kohara K, Kriebel J, Kronenberg F, Kuusisto J, Lall K, Lange LA, Lee M-S, Lee NR, Leong A, Li L, Li Y, Li-Gao R, Ligthart S, Lindgren CM, Linneberg A, Liu C-T, Liu J, Locke AE, Louie T, Luan J, Luk AO, Luo X, Lv J, Lyssenko V, Mamakou V, Mani KR, Meitinger T, Metspalu A, Morris AD, Nadkarni GN, Nadler JL, Nalls MA, Nayak U, Ntalla I, Okada Y, Orozco L, Patel SR, Pereira MA, Peters A, Pirie FJ, Porneala B, Prasad G, Preissl S, Rasmussen-Torvik LJ, Reiner AP, Roden M, Rohde R, Roll K, Sabanayagam C, Sander M, Sandow K, Sattar N, Schonherr S, Schurmann C, Shahriar M, Shi J, Shin DM, Shriner D, Smith JA, So WY, Stancakova A, Stilp AM, Strauch K, Suzuki K, Takahashi A, Taylor KD, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tomlinson B, Torres JM, Tsai F-J, Tuomilehto J, Tusie-Luna T, Udler MS, Valladares-Salgado A, Dam RM van, Klinken JB van, Varma R, Vujkovic M, Wacher-Rodarte N, Wheeler E, Whitsel EA, Wickremasinghe AR, Dijk KW van, Witte DR, Xiang Y-B, Yajnik CS, Yamamoto K, Yamauchi T, Yengo L, Yoon K, Yu C, Yuan J-M, Yusuf S, Zhang L, Zheng W, Raffel LJ, Igase M, Ipp E, Redline S, Cho YS, Lind L, Province MA, Hanis CL, Peyser PA, Ingelsson E, Zonderman AB, Psaty BM, Wang Y-X, Rotimi CN, Becker DM, Matsuda F, Liu Y, Zeggini E, Yokota M, Rich SS, Kooperberg C, Pankow JS, Engert JC, Chen Y-DI, Froguel P, Wilson JG, Sheu WH, Kardia SL, Wu J-Y, Hayes MG, Ma RC, Wong T-Y, Groop L, Mook-Kanamori DO, Chandak GR, Collins FS, Bharadwaj D, Pare G, Sale MM, Ahsan H, Motala AA, Shu X-O, Park K-S, Jukema JW, Cruz M, McKean-Cowdin R, Grallert H, Cheng C-Y, Bottinger EP, Dehghan A, Tai E-S, Dupuis J, Kato N, Laakso M, Kottgen A, Koh W-P, Palmer CN, Liu S, Abecasis G, Kooner JS, Loos RJ, North KE, Haiman CA, Florez JC, Saleheen D, Hansen T, Pedersen O, Magi R, Langenberg C, Wareham NJ, Maeda S, Kadowaki T, Lee J, Millwood IY, Walters RG, Stefansson K, Myers SR, Ferrer J, Gaulton KJ, Meigs JB, Mohlke KL, Gloyn AL, Bowden DW, Below JE, Chambers JC, Sim X,

Boehnke M, Rotter JI, McCarthy MI, Morris AP. Trans-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. medRxiv. Cold Spring Harbor Laboratory Press; 2020 Sep 23;2020.09.22.20198937.

133. Moltke I, Grarup N, Jørgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, Korneliussen TS, Andersen MA, Nielsen TS, Krarup NT, Gjesing AP, Zierath JR, Linneberg A, Wu X, Sun G, Jin X, Al-Aama J, Wang J, Borch-Johnsen K, Pedersen O, Nielsen R, Albrechtsen A, Hansen T. A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. Nature. 2014 Aug 14;512(7513):190–193. PMID: 25043022

134. Brown BC, Price AL, Patsopoulos NA, Zaitlen N. Local Joint Testing Improves Power and Identifies Hidden Heritability in Association Studies. Genetics. Genetics; 2016 Jul 1;203(3):1105–1116. PMID: 27182951

135. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, Jansen R, de Geus EJC, Boomsma DI, Wright FA, Sullivan PF, Nikkola E, Alvarez M, Civelek M, Lusis AJ, Lehtimäki T, Raitoharju E, Kähönen M, Seppälä I, Raitakari OT, Kuusisto J, Laakso M, Price AL, Pajukanta P, Pasaniuc B. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. Nature Publishing Group; 2016 Mar;48(3):245–252.

136. A CRISPR-based genome-wide screen for adipogenesis reveals new insights into mitotic expansion and lipogenesis | bioRxiv [Internet]. [cited 2020 Sep 16]. Available from: https://www.biorxiv.org/content/10.1101/2020.07.13.201038v1.full

137. A multi-omic integrative scheme characterizes tissues of action at loci associated with type 2 diabetes | bioRxiv [Internet]. [cited 2020 Sep 8]. Available from: https://www.biorxiv.org/content/10.1101/2020.06.25.169706v3

138. Timshel PN, Thompson JJ, Pers TH. Genetic mapping of etiologic brain cell types for obesity. Loos R, Barkai N, editors. eLife. eLife Sciences Publications, Ltd; 2020 Sep 21;9:e55851.

139. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, Floratos A, Sham PC, Li MJ, Wang J, Cardon LR, Whittaker JC, Sanseau P. The support of human genetic evidence for approved drug indications. Nat Genet. 2015 Aug;47(8):856–860. PMID: 26121088

140. King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. PLOS Genet. Public Library of Science; 2019 Dec 12;15(12):e1008489.