Identification of Transcription Factors involved in Obesity

A dissertation

submitted by

Jacqueline M Lane

In partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in

Genetics

TUFTS UNIVERSITY

Sackler School of Graduate Biomedical Sciences

May 2011

ADVISER: Jose M Ordovas

List of Figures

List of Tables

Acknowledgements

I'd like to express my appreciation to Dr. Jose Ordovas for all his guidance and support during my Ph.D. I'm deeply grateful to my thesis committee Drs. Donna Slonim, Alan Kopin, Acacia Alcivar Warren, Inga Peter, and Diana Bianchi. In choosing such an interdisciplinary topic I leaned very heavily on all of you for your expertise and I am truly thankful for your input and encouragement. Especially I would like to thank Dr. Alan Kopin for allowing me to perform my wetlab experiments in his lab. Thanks as well to my outside examiner, Dr. Lu Qi.

I would also like to thank all the members of the Nutrition and Genomics Laboratory at JM-HNRCA at Tufts University for great discussions, guidance, friendships, and sharing your expertise with me. I am particularly grateful to Dr. Larry Parnell for helping with day-to-day research life, and a never-ending knowledge of all things random, helpful, and often German. I would also like to give special thanks to my office-mate Dr. Jian Shen for her sharing her knowledge, patience, amazing attitude, kindness, laughter, and great conversations. As well, I'd like to thank Dr. Albert Tai and Allen Parmelle for help with experiments.

I am also incredibly thankful to be a part of the Sackler School. I am grateful to all the professors, students, and support staff of the Genetics Program, Tufts is truly a supportive community and that is thanks to all of you. Specifically I would like to thank Drs. Dean Dawson, Ananda Roy, and Erik Selsing for leading the program during my time as a student. I would also like to thank my fellow incoming Genetics students of 2004, I will always have fond memories of our study sessions together, and your friendship has made my time as a thesis student much brighter. In particular I'd like to thank Drs. Ari Rosenbach, Miriam Enos, and Shauna Hutchinson for traveling the road to "Drs" together with me. Our extended lunch times together kept me sane and made my heart happy; one of the greatest joys of my Ph.D. has been your friendship.

Additional thanks to all the wonderful teachers and role models who have poured into my life. I'd like to give specific thanks to Mr. Howard Young my inspirational high school science teacher who allowed me to conduct my first independent research project. I am sorry that the anaerobic digestion of leaves turned out to be so smelly. I'd also like to thank Rev. Mark and Stella Scafidi for mentoring me during those unruly teenage years; you are great models of the love of God.

I'd like to thank my family for unconditional love and support throughout my life. To my husband Kosta, who loved me back when I was just a little undergrad, thank you for making this journey with me. To my parents, thank you for giving a goofy little middle child dreams and ambition and for making the sacrifices to allow me to seek those. To my siblings Steph and Bud, thank you both for being partners in some of my earliest scientific experiments and for your ever growing and evolving friendship. To my Pop pop and gam-gam, who have always let me know how proud they are and shown me so

much love. To my Aunt Darlene who shares of a love of pranks and humor, thank you for your love and encouragement.

Lastly I want to thank God, for creating such an intriguing universe and giving me the passion and skills to take a small glimpse into it.

Identification of Transcription Factors involved in Obesity

Abstract

Obesity is a leading cause of preventable death in the United States. With the prevalence of obesity >30% in the United States, to effectively treat this complex disorder we must identify the underlying genetic components. To understand complex disorders, understanding complex regulatory networks where many genes are affected is critical. Transcription factors lie at the heart of regulatory networks, small changes in the binding affinity or concentration can perturb entire networks of genes. Therefore we seek to identify transcription factors that are involved in obesity. In order to do this we developed a novel computational methodology to extract information about transcriptional gene networks in genome-scale data, yielding a list of 53 transcription factors potentially involved in obesity. We validated SPI1, our top candidate, by examining the effects of gene knock down in adipogenesis. First we generated and characterized a rapidly differentiating clonal OP9 preadipocyte cell population. We conclude that our cells differentiate in 72 hours, are readily transfectable, and differentiate through gene expression changes comparable to established preadipocyte lines. Gene expression profiles of differentiating OP9 cells express the classical adipogenic marker genes C/EBPbeta, C/EBPalpha, Gata2, and Plin1. Depletion of SPI1 increases lipid accumulation OP9 adipogenesis. In addition, variation in the human SPI1 gene modulates obesity risk. Individuals with the AA genotype of SNP rs4752829 have increased body mass index (BMI) compared to carriers of the G allele. SNP rs3740689 modulates the effect of dietary saturated fatty acid and omega-3 polyunsaturated (N3-PUFA) on BMI. These approaches demonstrate a strong role for SPI1 in obesity, with the potential to target SPI1 with pharmacological or dietary modifications. In addition,

by describing the transcriptome of OP9 preadipocytes, we have validated an adipogenesis model for potential high-throughput screening of the effects of gene depletion on fat cell generation. For the first time the transcriptome of OP9 adipogenesis has been described and a gene-diet interaction influencing BMI has been found in SPI1. Our computational approach directed us to SPI1 and was validated in this study. This study demonstrates that our computational approach can be used to investigate the role of transcription factors in many other disorders.

Chapter 1: Literature Review

Overview of obesity

Obesity is a chronic condition in which individuals accumulate an abnormal amount of body fat. This commonly results from energy imbalance; more calories are taken in than expended. One common method used to diagnose obesity is a surrogate marker for body fat, body mass index (BMI). BMI is calculated by dividing weight (kg) by height (m) squared. The current World Health Organization (WHO) BMI cutoffs for non-Asian populations are as follows: 19-25 normal, 25-30 overweight, >30 obese [1].

Individuals with obesity have an elevated risk of cardiovascular disease (CVD) [2], type 2 diabetes mellitus (DM2) [3], colon [4], endometrial, and breast cancer, renal, gall bladder [3], and liver diseases, polycystic ovarian syndrome, sleep apnea, osteoarthritis, and many other chronic diseases [5-9]. The presence of obesity predisposes individuals to a cluster of risk factors known as Metabolic Syndrome (atherogenic dyslipidemia, elevated blood pressure, elevated glucose, prothrombotic state, and proinflammatory state) [10-15]. Metabolic Syndrome increases the risk of CVD 2-fold and DM2 5-fold [16]. Patients with DM2 have a 2-3 fold greater risk of coronary heart disease and experience worse outcomes after a coronary event [17]. Approximately 300,000 deaths a year are attributable to obesity, which makes it the second leading cause of preventable death [18].

The impact of the obesity epidemic cannot be measured in medical impact alone; obesity also impairs the ability to live a full and active life. Studies show a correlation

between obesity and restricted activity, depression [19, 20], anxiety [20, 21], vitality, social functioning, and bodily pain [22-25]. Thompson et al. estimates the lifetime healthcare costs associated with a BMI of 27.5 (relative to 22.5) at approximately $4000 per person, a BMI of 32.5 at $10,000 per person, and a BMI of 37.5 at $15,000 [26]. These numbers are staggering when the prevalence of obesity is taken into consideration.

The prevalence of obesity in the United States has been rising since the 1970's. From 1980-1994 the prevalence of obesity rose 8% (National Health and Nutrition Examination Survey, NHANES). The NHANES study tracks the prevalence of obesity and overweight, and shows a continued increase in the prevalence of obesity through 2000. NHANES data also shows the prevalence of obesity leveling off in recent years, although the prevalence is still far above baseline data from 1976 (Figure 1). Recent data indicate the prevalence of obesity is >30% across almost all age and gender groups surveyed [27].

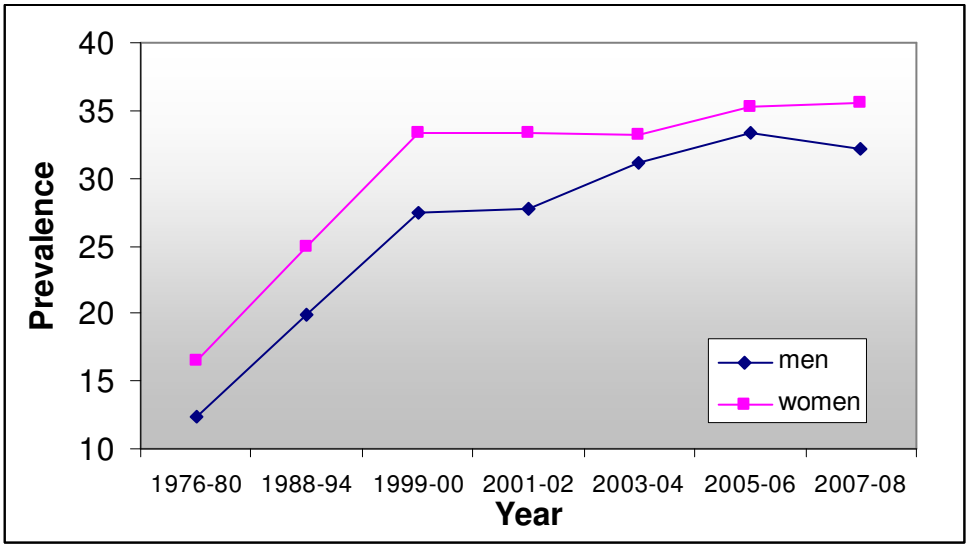


Figure 1 The Prevalence of obesity from 1976-2008. Data collected by the NHANES study [27].

What has caused this surge of obesity over the last 40 years? Obesity is caused by energy imbalance. Energy imbalance can be due to an increase in caloric intake, a decrease in caloric requirements, or a combination of both [28]. Environmental and genetic factors play a role in energy imbalance [29]. Kopelman et al. describes exposure to environmental factors such as television watching or car ownership as "obesogenic" [29].

These environmental factors include; physical activity, food intake, and culture (Figure 2). Several studies have shown an inverse relationship between physical activity and the prevalence of obesity [30-32]. A study conducted in rats showed higher weight gain in groups fed a variety of foods as opposed to one palatable food [33]. In developed nations a wide variety of highly palatable foods is readily available. Food portion sizes have also increased in restaurants and at home in the last 30 years [34]. Rolls et al. demonstrates that increasing portion size leads to an increase in energy intake [35, 36]. Taken together we see a trend towards larger portion sizes concomitant with an increase in energy intake which factor into the growing obesity epidemic.

There is also evidence that societal influences affect the prevalence of obesity. In a study conducted in Sweden, high social class is associated with a lower odds ratio for obesity [37]. A study in England and Wales also reported similar findings; the prevalence of obesity in the highest social class was 10.7% versus 25% for women in the lowest social class [38]. The difference in obesity prevalence across socioeconomic status most likely represents an unequal distribution of resources (access to education, health care providers, healthy foods, opportunities for physical activity etc.) [39].

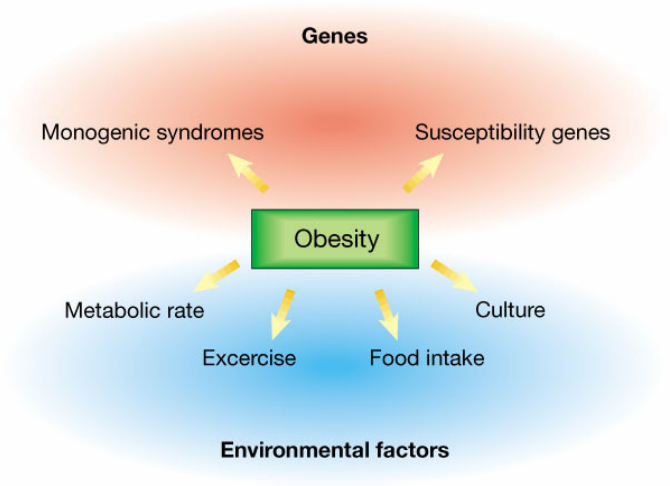


Figure 2 Causes of Obesity. Obesity is caused by environmental and genetic factors.

Culture also influences the way people think about the impact, causes, and treatments for obesity. We see evidence of acculturation in ethnic groups that migrate to the U.S. In one study first generation Latino immigrants consumed more fruit and vegetables than White adolescents. With successive generations the Latino diet changed and by the third generation Latinos had poorer nutrition than White adolescents [40]. As the western diet and culture begin to spread it brings with it a global obesity epidemic. However the environment is not the sole cause of obesity. Changes in the environment are acting on genetic susceptibilities.

Obesity and human variation

There are population level prevention recommendations in order to achieve weight loss. Recommendations include increased physical activity, behavioral changes,

and dietary modifications [41-49]. Avenell et al. reviewed the beneficial effects of various weight reducing diets in randomized clinical trials [45]. In the majority of the clinical trials reviewed people lost weight. In a controlled trial, weight loss strategies are effective. However, despite the current recommendations and positive results of controlled, short term lifestyle intervention studies, Americans are not losing weight. The prevalence of obesity in American adults is still >30% [27].

A fraction of the problem may be failing to address individual needs, as opposed to population level recommendations. This implies that subsets of the population might need separate meals tailored to the way their bodies process food. To understand how to address the needs of each subset, we need to understand what is causing the difference in how they process food. Part of this difference traces back to the genetic level.

On average human genomes are 99.9% similar to one another [50, 51]. The remaining 0.1% is variation in the form of copy number variants, copy number polymorphisms, segmental duplications, inversions, translocations, segmental uniparental disomy, and single nucleotide polymorphisms (SNPs) [52], the latter being the most common and widely studied (Figure 3). A single base change is defined as a SNP when it occurs in at least 1% of the population [53]. SNPs occur every 100-300 bases along the 3 billion base pair human genome [53]. Each individual has a unique SNP pattern.

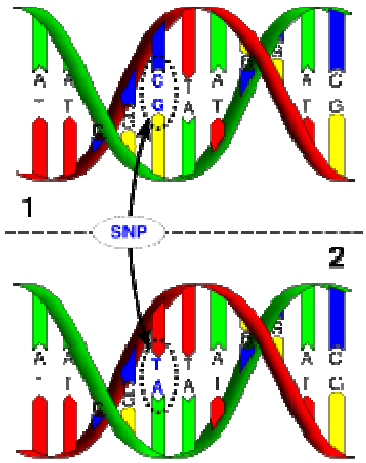


Figure 3 Single nucleotide polymorphisms (SNPs). SNPs are single base pair changes in the genome. SNPs occur with a frequency of at least 1%.

SNPs have been associated with most human traits including predisposition to disease (i.e., obesity [54-57], type II diabetes [58-60], and hypertension [61-63]) and responses to drugs, bacteria, viruses, toxins, chemicals, and, most commonly, to nutrients [53]. Research has established that SNPs can affect the body's response to nutrients. To help individuals maintain a healthy body weight, nutritional recommendations should be based on an understanding of the specific genetic basis that predisposes individuals to obesity.

Genetics of obesity

In order to achieve personalized nutrition recommendations, we must understand the genetic components of obesity. Familial and twin studies demonstrate that obesity has a genetic component [64]. Studies in twins show cognitive restraint, emotional, and

uncontrolled eating are highly heritable [65]. Measures of obesity, including body mass index (BMI), waist circumference, adult weight gain, and waist/hip ratio (W/H) are also highly heritable [66, 67]. In a study of monozygotic twins raised apart the heritability of BMI is .5-.7 [68, 69]. Another twin study estimates heritability of percent body fat and distribution of body fat as .7 and .86 respectively [70]. Rose et al. demonstrated the heritability of W/H and waist circumference as .36-.61 and .72-.82 respectively [71]. Although these studies demonstrate a genetic component to obesity, they do not point to specific genes involved in obesity.

Many specific genes have been linked to obesity by mouse models, through association between genetic markers and obesity phenotypes in humans, and linkage of an area of the human or mouse genome to obesity phenotypes in quantitative trait loci studies (QTLs). As of 2005, approximately 380 genes were listed in the human obesity gene map database [72], and our knowledge of the genetics of obesity has only increased since then. A portion of these genes are involved in monogenic obesity disorders, which are single gene mutations that lead to morbid obesity in childhood [73-78]. For example, several families have been identified with mutations in the gene encoding for Leptin [74, 79, 80]. Individuals with Leptin mutations demonstrate a partitioning defect predisposing individuals to deposition of fat, and suffer from severe hyperphagia that can be corrected with subcutaneous leptin injections [74, 81]. However, these mutations are extremely rare; there are only a handful of families in the world with Leptin mutations. In total 4% of obesity cases are attributable to monogenic disorders. Although monogenic obesity disorders have a large individual risk, because they are so rare, they represent a low population level risk.

On the other hand, common obesity is thought to be the result of many widespread variations. Each individual variation confers a small risk. The human obesity gene map database includes many genes with common mutations. Table 1 gives an overview of various association studies from a review by Ichihara et al [82]. These association studies show a statistical link between obesity related measures and genetic variation. To highlight a few, MCR4 has a well studied variant (SNP) V103I that is protective against obesity [83, 84]. Studies of MCR4 V103I report an odds ratio of obesity between 0.69-0.80 [84]. In TNFA, the -308G -> A SNP is associated with increased BMI, increased W/H, and increased plasma cortisol [85]. SNP 19A -> G in the Leptin gene is associated with obesity in women, and G homozygotes have lower concentrations of leptin [86, 87]. Genes such as MCR4 that include common mutations confer a minor individual risk (1-2% of the variability), but have a large population level risk.

Table 2. Summary of association studies for candidate gene markers and obesity.

Gene	Locus	Polymorphism	Effect of variant allele	dbSNP no.	Reference
Angiotensin-converting enzyme (*ACE*)	17q23	Intron16 (I/D)	Associated with plasma ACE level		[118]
		–240A→T	Associated with plasma ACE level	rs4291	[119]
β2-Adrenergic receptor (*ADRB2*)	5q31-q32	Arg16Gly	Decreased receptor activity	rs1042713	[106, 107]
		Gln27Glu	Decreased receptor activity	rs1042714	[108]
β3-Adrenergic receptor (*ADRB3*)	9p12	Trp64Arg	Decreased receptor activity	rs4994	[109, 110]
G protein β3 subunit (*GNB3*)	12p13.3	825C→T	Modified G protein activation	rs5443	[120–122]
Leptin (*LEP*)	7q31	–2548G→A	Changed leptin concentration	rs7799039	[123]
		19A→G	Changed leptin concentration	rs2167270	[124, 125]
Leptin receptor (*LEPR*)	1p31	Arg223Glu	Changed receptor function	rs1137101	[126, 127]
		Lys656Asn	Changed receptor function	rs8179183	[128]
Peroxisome proliferator–activated receptor γ (*PPARG*)	3p25	Pro115Gln	Reduced transactivation activity	rs1800571	[104]
		Pro12Ala	Reduced transactivation activity	rs1801282	[105]
Tumor necrosis factor-α (*TNFA*)	6p21.3	–308A→G	Changed transcriptional activity	rs1800629	[117]
Uncoupling protein 1 (*UCP1*)	4q28-q31	–3826A→G	No	rs1800592	[113, 114]
Uncoupling protein 2 (*UCP2*)	11q13	–866G→A	Changed UCP2 mRNA abundance	rs659366	[115]
Uncoupling protein 3 (*UCP3*)	11q13	–55C→T	No	rs1800849	[116]

Table 1 Association studies linking variation with obesity.

One common problem with association studies is replication. This is seen in the case of ADRB3 variant Trp64Arg. The variant Arg allele is associated with increased BMI, fat mass, and waist circumference in a population of Mexican-American siblings [88]. Half of the subsequent studies of Trp64Arg fail to replicate these findings [89, 90]. This could be due to several factors. For instance, a study could be underpowered to detect moderate differences [91]. Furthermore, the study designs could be slightly different. Additionally, it could be due to differences in genetic background of the study groups [92]. And lastly, there could be unadjusted factors that influence the outcome, such as diet [93].

Failure to replicate studies due to differences in genetic background or unadjusted factors that influence the outcome highlight the importance of interactions. Several reports have identified genetic variants associated with obesity that interact with dietary factors. These genes fall into several categories such as energy homeostasis, appetite, satiety, lipoprotein metabolism, and signaling peptides. A good example is the peroxisome proliferator activated receptor gamma (PPARγ) variant Pro12Ala. Originally PPARγ Pro12Ala was demonstrated to associate with obesity [94-96]. Conflicting results followed the original positive association studies [97-99]. Ultimately studies revealed an interaction between PPARγ Pro12Ala and the ratio of polyunsaturated to saturated fat [100]. Another example is the 54Thr variant in FABP2. It confers a better response to dietary intervention on triglyceride levels [101]. These and other studies provide initial evidence about the potential benefits of including genetic information when making dietary recommendations to lower BMI.

Genome-Wide Association Studies

To date two major approaches have been used to identify genetic susceptibilities for obesity. The first approach is based on searching for associations between SNPs in candidate genes with phenotypes of interest. Candidate gene studies select genes based on physiological and molecular mechanisms of disease [102]. There are several limitations to candidate gene studies including; small sample size [103], low number of SNPs investigated, difficult to replicate [104, 105], and resolution only for variants of modest effect size.

Lately a second approach, genome-wide association studies (GWAS), has been employed. A genome wide association study investigates associations between variation in the entire genome and a particular disease [106]. The theory behind GWAS is that many common variations, each conferring a small risk, contribute to the overall risk of developing a common disease. GWAS are powered to detect common variation (present in at least 1-5% of the population) with small effect size [107].

GWAS studies address some of the limitations of candidate gene studies. Currently ~1 million SNPs can be investigated per GWA study. GWA studies have been conducted with a population size of up to 40,000 individuals [56, 108-112], and two major loci associated with BMI have been replicated in several populations [56, 111-114].

To date obesity GWA studies have produced two major findings; a SNP located in an intron of the fat mass and obesity associated gene *(FTO) [56, 111, 112, 114]*, and a SNP located 188kb downstream of the melanocortin 4 receptor (*MC4R) [111-114]*. Signals located near *PCSK1, NEGR1, TMEM18, SH2B1, BDNF, AND KCTD15* have also

been replicated in independent studies [111, 112, 114, 115]. The function of *BDNC, MC4R*, and *SH2B1* are validated by mouse studies, as well as data from rare monogenic forms of obesity [116].

All together these findings only account for <2% of the total variation in adult BMI [102, 116]. Bogardus et al. calculates a difference of 1.17 BMI units between compound homozygotes of the *FTO* and *MC4R* risk alleles (1% of population) versus compound homozygotes of non-risk alleles (19% of population) [102]. These finding have led many experts to ask the question, "Where is the missing heritability?"

Disease	Number of loci	Proportion of heritability explained	Heritability measure
Age-related macular degeneration	5	50%	Sibling recurrence risk
Crohn's disease	32	20%	Genetic risk (liability)
Systemic lupus erythematosus	6	15%	Sibling recurrence risk
Type 2 diabetes	18	6%	Sibling recurrence risk
HDL cholesterol	7	5.20%	Residual* phenotypic variance
Height	40	5%	Phenotypic variance
Early onset myocardial infarction	9	2.80%	Phenotypic variance
Fasting glucose	4	1.50%	Phenotypic variance

Table 2 Summary of major GWAS finding as of 2009 [107]

Several theories exist about where the missing heritability of obesity may turn up [102, 107, 117]. The first theory is simply that there are a large number of variants with a small effect size, and they have simply not been discovered yet. It would take 50 genes at a prevalence of 10% with a risk ratio of ~1.2 to explain 50% of the population attributable risk. At a prevalence of 20% it would take 25 genes.

Another theory proposes the presence of rare variation with larger effects. These are not detectable by GWAS, which are powered to find variants present in >5% of the population. To account for 50% of the population attributable risk using rare variants (1

per 5000) with a risk ratio of 1-20 we would need 186-556 genes [102]. Structural variants, including copy number variants (CNV), could also explain the heritability of obesity. However, Mc Carroll et al. demonstrates strong linkage between the majority of common diallelic CNVs and SNPs [118]. Therefore, CNVs would be detected by GWAS. Another theory proffers gene by environment interactions as the key to finding the missing heritability [119].

A major hindrance of GWAS is the sheer size and complexity of the data, which can be difficult to manage [120]. When over 500,000 variants are tested in thousands of individuals with phenotype data the volume of data and statistical tests can be limiting [120, 121]. To account for the multiple testing problem Bonferroni [122] and False Discovery Rate (FDR) [123] adjustments can be used. A proposed threshold for significance in GWAS studies is 10^{-6} [124], although as the number of genotyped SNPs increases a new threshold of 10^{-7} has been suggested [121]. Even GWAS follow-up studies are plagued by multiple testing, often testing hundreds or thousands of SNPs.

Moving forward GWA studies will focus on increasing sample size, combining prior info to restrict the number of SNPs investigated, defining more precise obesity phenotypes, including accurate measures of environmental exposure, expanding to non-European populations, and investigating gene by gene and gene by environment interactions [107, 119]. However, with current GWA studies explaining only 1-2% of the heritability of obesity, further methods of investigation are being developed. Investigators are sequencing the genomes of individuals with extreme obesity phenotypes, as well as implementing whole-genome re-sequencing to confirm GWAS results [107, 125]. Still other approaches use pathway data to reexamine GWAS data

[126, 127]. It is clear that while GWAS have provided insights into complex diseases, such as macular degeneration and type 1 diabetes, significant portions of heritability for obesity remained unexplained.

Tissues involved in obesity

There are twelve genes that have been validated in ten or more genetic association studies with obesity (ADIPOQ, ADRB2, ADRB3, GNB3, HTR2C, NR3C1, LEP, LEPR, PPARG, UCP1, UCP2 and UCP3). These genes fall into a variety of pathways and functions, including energy homeostasis, thermogenesis, adipogenesis, hormonal signaling peptides, taste transduction, and cytokine interactions. Taken together we begin to learn something about the tissues and organs involved in obesity.

The brain regulates energy intake and expenditure [128]. It controls body weight by receiving signals about the body's energy needs and adiposity, and effecting concomitant changes in processing of energy and behavior [6, 129-132]. Genetic evidence shows targeted knockout of LepR in POMC neurons in mice results in loss of body weight homeostasis. In a MC4R knockout mouse, expression of MC4R restores regulation of food intake [133].

Complex feedback loops in the brain regulate food intake and energy expenditure [6, 134-137]. Catabolic neuronal pathways decrease food intake and increase energy expenditure [138]. Melanocortin neurons process signals from leptin and insulin in the basal state to prevent weight gain. On the other hand, anabolic neuronal pathways, involving neurons containing neuropeptide Y, increase food intake and decrease energy

expenditure and are inhibited by leptin and insulin. At the basal state catabolic pathways are activated and anabolic pathways are inhibited. During weight gain, therefore, basally active pathways are only further upregulated or inhibited leading to a mild reaction. However, during weight loss there is a vigorous response with the upregulation of previously inhibited anabolic pathways, and inhibition of previously active catabolic pathways [138]. This vigorous neuronal reaction to weight loss and mild reaction to weight gain is one of the reasons why humans are predisposed towards weight gain.

The gut signals the brain by releasing indicators of energy intake, feeding status, and energy stores. These molecules serve as signals of short term feeding status. The gut peptide ghrelin stimulates the brains appetite response. Ghrelin is released in anticipation of a meal, and signals that the stomach is ready to ingest food. The gut signals satiety to the brain using insulin, cholecystokinin (CCK), and PYY.

Adipose tissue releases signals of long term energy stores. The adipose tissue is made up of adipocytes, vascular tissue, and immune cells [139]. White adipose tissue, the most common form in adults, can be broken down into subcutaneous and intra-abdominal. Adipose tissue functions both as an energy store and an endocrine tissue, releasing signaling molecules called adipokines. Of the genes expressed in adipocytes, 30% produce gene products that are secreted [6, 139-142]. Adipokines are involved in the regulation of energy balance, appetite, glucose homeostasis, immune function, coagulation, and blood pressure [139]. Leptin, RBP4, and adiponectin are adipokines that function as satiety factors released when energy stores are high. Adiponectin is an important adipokine which promotes triglyceride breakdown, inhibits HDL breakdown, increases insulin sensitivity, and signals satiety [139]. Mice with adiponectin deletions

develop severe obesity [139]. Leptin is another critical molecule in signaling long term satiety to the hypothalamus. As described previously, mice and humans lacking leptin are hyperphagic and develop severe obesity.

The central problem in obesity involves the accumulation of excess fat mass. Adipose tissue releases fatty acids when energy is needed, and stores lipids long term when nutrients are in excess, a process which also involves the signaling molecule leptin [6, 143-147] . Leptin binds to cell receptors and triggers the Jak/Stat pathway [6, 148-151]. The Jak/Stat pathway signals to AMP-activated protein kinase (AMPK), which has varied effects depending upon cell type. In liver and muscle cells leptin binding leads to an activation of AMPK. This upregulates fatty acid oxidation, and inhibits triglyceride synthesis in liver and muscle, inhibiting the development of ectopic fat stores. In adipocytes, AMPK activation triggers an increase in lipid synthesis and storage. In the hypothalamus, leptin inhibits AMPK which leads to appetite suppression and increased calorie expenditure [6, 146, 152, 153]. In this way leptin functions across several tissues to insure the body maintains proper fat stores.

Circulating leptin levels increase as stored fat mass increases. This normally results in a decrease in appetite and in increase in energy expenditure. However, if satiety signals from adipocytes are repeatedly ignored and leptin levels continue to rise, leptin resistance can develop and lead to energy imbalance and an increase in fat mass [6, 154-158].

Models of adipogenesis

The pathophysiology of obesity involves amassing excess adipose tissue [156, 159, 160]. Growth of adipose tissue is composed of adipocyte size (hypertrophy) [161] and number (hyperplasia) [156, 162]. In childhood and adolescence our number of fat cells increases, plateaus in late adolescence, and remains constant throughout adulthood. Although the number of adipose cells remains constant throughout adulthood, adipose tissue has 10% turnover each year [163-165]. Therefore the process of adipose cell generation, adipogenesis, has relevance to the development of both childhood and adult onset obesity.

Establishment of immortalized preadipocyte cell lines provides an in vitro system for the study of fat cell development. Mouse cell lines designated 3T3L1, 3T3-F442A, and OP9 readily differentiate into adipocytes [166, 167]. Many features of these cells lines recapitulate the process in vivo, including morphological changes, lipid accumulation, and sensitivity to most or all of the key hormones that affect adipocytes. Through research on these cell lines, mouse studies, and other techniques, three classes of transcription factors have been identified that directly influence adipose cell development: PPARγ [168], C/EBPs [169], and the ADD1/SREBP1 basic helix-loop-helix family [166, 170] (Figure 4).

Although in vitro fat cell differentiation provides insights into the transcriptional network involved, there are limitations to this approach. The majority of current adipogenesis models represent white adipose tissue. White adipose tissue makes up the majority of adipose tissue in adults; however, during the first decade of life there

appears to be widely distributed brown fat [171]. Therefore cell culture models provide information about the role of white adipose tissue only.

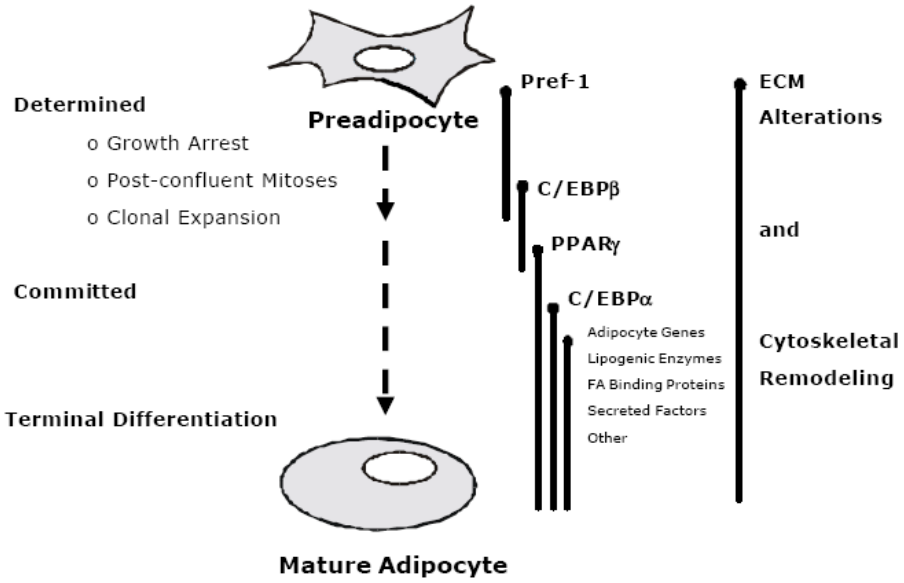


Figure 4 Processes and transcriptional cascades involved in adipogenesis. Adipogenesis goes through three distinct ages (left). Cells also undergo morphological changes (near left). A transcriptional cascade takes places as preadipocytes differentiate into mature adipocytes (near right). In addition cellular processes change as the cell undergoes differentiation (right). From the thesis of Thomas Burkhard.

One major drawback to using a model based on cell culture is the inability to examine fat-depot specific differences in behavior. Preadipocytes isolated from different areas of the body have varied adipogenic potential, the basis for which is completely unknown [172]. This approach will not reveal TFs that are responsible for depot-specific adipogenesis effects.

Despite these limitations this technique is still a useful approach for screening a large number of candidate genes. The method is relatively cost-effective and easy to scale-up. OP9 cells are a line of bone marrow derived mouse stromal cells that rapidly

differentiation into adipocytes [167]. Given stimuli OP9 cells accumulate triacylglycerol, display adipocyte morphology, and express adipocyte late marker proteins such as Glut4 and adiponectin [167]. There are three methods to differentiate OP9 cells: insulin oleate, serum replacement, and adipogenic cocktail. Cells can be stimulated at pre-confluent levels using insulin oleate, and therefore that is the method that will be used in this study. OP9 cells fully differentiate over a three day period, as opposed to the 10 day period needed for differentiation of 3T3-L1 cells.

Transcription Factors

The process of adipogenesis is largely a transcriptional process. In eukaryotes transcription is primarily controlled at the transcription initiation step. This step involves the binding of transcription factors to DNA regulatory sequences. In some cases this binding can involve many TFs binding to various regulatory sequences, allowing complex control [173-176].

Transcription factors generally bind upstream of the mRNA start site, but can also bind downstream, intronically, and occasionally in the coding sequence itself. Often transcription factor binding sites (TFBS) are found within close proximity to the promoter region, but cis-regulatory modules can be found >10,000 base pairs away [177, 178]. One of the first genes to have its control regions mapped was transthyretin (TRR). In hepatocytes, TRR has two control regions, one located between 0-200bp upstream of

the transcription start site, and the other mapped between 1.85-2kb upstream. Interestingly, in choroid plexus cells TRR control regions map elsewhere [175].

Transcription factors bind to transcriptional control elements and act as activators or repressors of transcription. In most vertebrate systems transcription factors have been discovered by biochemical purification. Transcription control elements, identified by positional cloning and deletion studies, are often used to isolate transcription factors. In the reverse experiment, DNA footprinting, EMSA, and ChIP assays help identify DNA sequences bound by specific TFs [176]. These binding sites are typically 4 to 30 base pairs long (Table 3) [173].

Transcription factor	Consensus binding site
Specificity protein 1 (Sp1)	GGGCGG
CCAAT/Enhancer binding protein (C/EBP)	CCAAT
Activator protein 1 (AP1)	TGACTCA
Octamer binding proteins (OCT-1 & OCT-2)	ATGCAAAT
E-box binding proteins (E12, E47, E2-2)	CANNTGA

Table 3 Transcription factors and their binding sites. Examples transcription factors and the corresponding consensus binding site. N stands for any nucleotide [173].

As binding data are collected for each transcription factor a specific binding motif can be determined. These transcription factor binding sites can be represented as position weight matrices (PWM), where the likelihood of a particular base at each position is represented [173]. Computational programs enable us to identify novel binding sites using these experimentally determined PWMs.

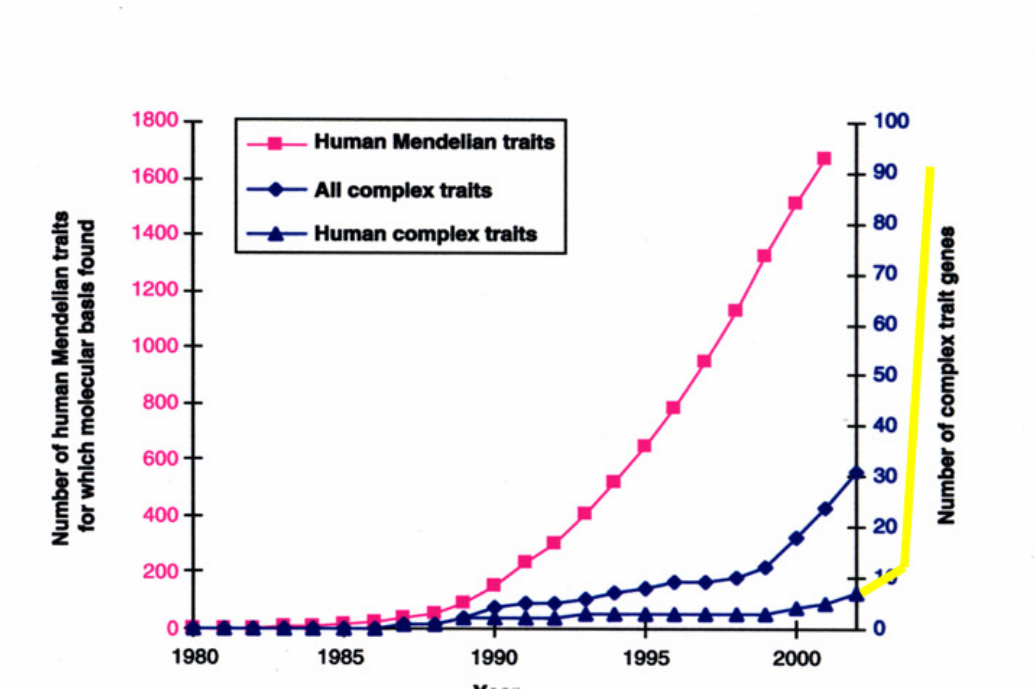
High false positive rates are one drawback to determining the presence of known TFBS. On average human genes are surrounded by 70kb of non-coding DNA [179]. There are several ways to limit the number of false positives. One way is to incorporate conservation data into TFBS searches using phylogenetic footprinting [180, 181]. Phylogenetic footprinting works on the assumption that mutations in functional elements of the genome are more disruptive than in non-functional elements [180]. Therefore, rates of mutation in functional elements will be lower than in other areas of the genome. These conserved regions can be found by searching for conversation across species. In phylogenetic footprinting, these areas are prioritized as search areas for TFBS. In addition to using conservation data, another way to minimize false positives is to use a stringent threshold for positive matrix matches. Lastly, additional information, such as co-expression data, can be included [173, 179]. When these steps are taken Brown et al. states that searching short segments of DNA for known TFBS is a valuable tool [179].

Computational approaches

In order to investigate genetic risk for obesity, either a specific area of the genome or the entire genome must be assessed for linkage with obesity. Traditionally, association and linkage studies have been used to identify genes involved in Mendelian traits. These methods are often not robust enough to detect genes involved in complex traits, such as obesity. Complex traits pose greater research challenges, because genes often interact with one another and/or the environment, complex inheritance patterns cluster in

families, and inheriting a mutation is not a guarantee of developing disease [182]. Progress in identifying genes associated with complex traits has been slow, but has picked up in the last 6 years due to GWAS (Figure 5) [183]. Advances in technology and genomic sequencing efforts now make it possible to study complex traits in simpler systems, and, using conservation across genomes, relate findings back to human disease. One challenge is how to integrate publicly available data with computational methods to discover candidate TFs involved in complex diseases, such as metabolic syndrome.

As mentioned previously, one current method of investigating genes involved in complex diseases is GWAS. GWAS genotype the DNA of individuals at up to 1000K SNP tagging locations, and look for associations between genotype and disease. Currently the cost of genotyping alone in a study with 1000 cases and 1000 controls is $800,000 [184]. Additionally, these studies have yielded a surprising lack of associations with variants responsible for explaining the heritability of many diseases. Therefore, other genome-scale approaches are necessary to identify genes responsible for traits.



A.

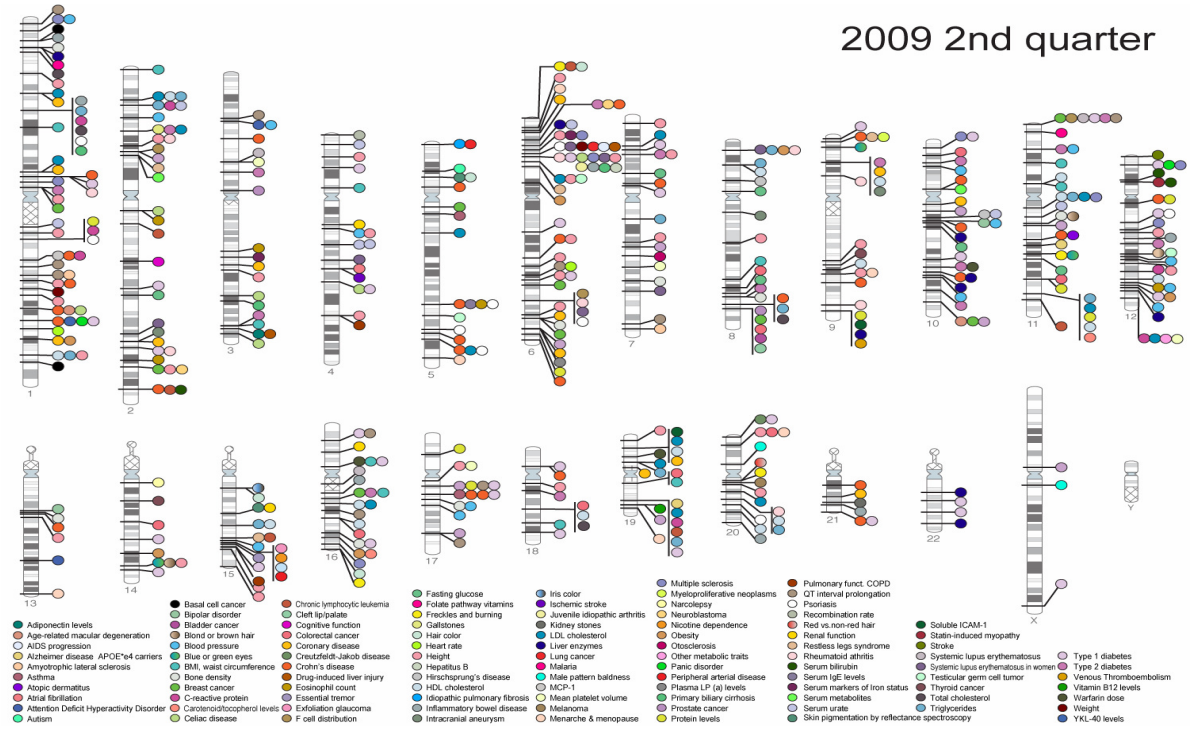


Figure 5 Gene discovery in complex traits. A) Mendelian traits with a discovered molecular basis according to year. The yellow line represents new findings through GWAS added by A. Guttmacher. B) GWAS results shown on the human genome. [183, 185]

Current studies show that using bioinformatics to sort through publicly available data yields strong candidates [186]. Xiong et al.[187] and Bischof et al.[188] outline methods which capitalize on the vast array of publicly available data, and harness bioinformatics to screen for candidate genes. Both propose and implement, using different approaches, the integration of DNA variation, gene expression, protein-protein interaction, and phenotypic variation to increase the reliability of discovering causative genes for complex diseases. As a proof of concept for these approaches, Xiong et al. show the identification of new regulatory relationships between SPARC and collagen type I, SPARC and CTGF, and collagen type I and collagen type III. Likewise, Bischof et al. identified and characterized a set of 11 genes not previously known to be differentially expressed in the murine hypothalamus.

Perhaps the most conclusive proof that obesity related candidate TF genes can be identified from publicly available data comes from Mootha et al.[189] Mootha et al. demonstrated using microarray data they could group functionally related genes and identify a TF, peroxisome proliferator-activated receptor gamma, coactivator 1 alpha (PPARGC1A), which impacts coordinate regulation of genes in DM2. This suggests that a more generalized methodology can be established where genes from a dataset are grouped according to function, and then traced back to common transcriptional control.

Research overview

In this project we use three approaches to understand novel roles of TFs in obesity. In our first approach, we create a new computational method to identify networks of co-regulated genes in genomic data. The computational approach yields a list of TFs that are potentially involved in obesity, and are candidates for functional validation. Our next approach to identify TFs involved in obesity is to characterize the transcriptome of OP9 cells undergoing adipogenesis, in order to validate OP9 cells as a cell model of adipogenesis. Once we have validated the OP9 cell model, we will use it to screen for the effects of TF knockdown on adipogenesis. Using this approach we can identify additional TFs that play a critical role in adipose tissue development, and can serve as targets of obesity therapeutics. Our last approach focuses on variation in a specific TF, SPI1. With this approach we will assess the impact of variation in SPI1 on BMI in a population of European descent. Identifying variation in SPI1 with an impact on BMI will allow us to identify groups of individuals at greater risk of developing obesity, and create specific lifestyle interventions for them. Through our computational approach we identify TF candidates with a potential role in obesity. We then use our adipogenesis model to functionally validate the role of these candidate TFs in adipocytes. After we functionally validate the role of one of our computational TF candidates, SPI1, in adipogenesis, we then investigate specifically how variation in SPI1 can increase susceptibility to obesity in subsets of the population. In this way our three approaches complement one another, as well as provide distinct insights into the role of TFs in obesity.

CHAPTER 2: Computation identification of TFs in obesity

Introduction

Obesity is a complex disorder. To understand complex disorders, we must understand complex regulatory networks where many genes are affected [72]. At the heart of gene regulation is TFs that bind to promoter regions and repress or initiate transcription. Single nucleotide changes in a transcription factor (TF) can affect the concentration or target binding affinity, which in turn perturbs an entire network of targets [189-192].

Obesity is a disorder in which many tissues and systems are affected. Therefore we want to identify novel TFs involved in obesity, because of the impact they can have on an entire network of genes. By focusing our approach on TFs, we have narrowed the field of potential obesity candidate genes.

Francis Collins compares the candidate gene approach to finding a set of keys lost in the dark and turning on one light to find them [184]. In his analogy a genome-wide approach is like turning on all the lights at once to find the keys. This, of course, requires turning on many lights which provide no additional information about the location of the keys. In this approach, the probability for false positives is much higher. Our approach of identifying obesity candidate genes combines candidate selection with genomic scale publicly available data to direct the identification of TFs related to obesity.

Several computational approaches have been developed to identify TFs involved in disease, including: clustering of co-regulated genes, multipoint linkage analysis prediction of regulatory modules, and pathway enrichment analysis [190]. These methods have been used successfully to inform our knowledge of disease pathobiology. However, there is greater power in a computational process which combines multiple aspects of these approaches to identify TFs associated with disease. Clustering of co-regulated genes involves correlating gene expression changes between pairs of genes and creating clusters of genes with similar expression patterns [193]. The outcome is a cluster of genes, which by itself provides no biological context. Multipoint linkage analysis examines the association of a phenotype with multiple markers [194, 195]. This type of analysis specifically requires marker and disease data to perform. Prediction of regulatory modules involves clustering of genes differentially expressed in concert by shared regulation [196]. This method has shown limited success, in part due to the necessary assumption that expression levels of regulators are linked to activity levels of the regulators. Biologically this is often not the case [196]. Pathway enrichment analysis tests for differential expression of predefined sets of genes, rather than individual genes [197, 198]. However, to identify the TF responsible for the coordinate changes of genes in the differentially expressed set, a further step is necessary. This step involves the identification of transcription regulatory elements common to differentially expressed genes in the set. This project seeks to combine aspects of pathway enrichment analysis and prediction of regulatory modules to identify the TFs that regulate genes involved in fat storage.

To computationally identify TFs involved in obesity we need a comprehensive dataset where we can extract TF networks. From this dataset we first want to identify groups of genes potentially under common transcriptional control. We form these groups by clustering genes according to Gene Ontology (GO), a controlled vocabulary used to describe the biological process, molecular function, and cellular compartment associated with a gene. Next we search grouped genes for the presence of a shared transcription factor binding site (TFBS). To expedite the identification of TFs involved in obesity, we only search for the presence of known motifs. This ensures that we can trace overrepresented TFBS directly to the TF, and in this way identify TFs involved in obesity.

Alzheimer's disease (AD). In 1993, the first presenilin was discovered in C. elegans. Two years later, mutations in the human presenilin-1 gene were associated with early-onset familial AD. A remarkable functional conservation between C. elegans and human could be demonstrated: expression of the human presenilin-1 in C. elegans could rescue neuronal deficiencies of C. elegans sel-12 presenilin mutants. C. elegans research has further advanced the understanding of AD by identifying the presenilins as components of the gamma-secretase complex, an important target in AD.
Diabetes type 2. In 1997, genetic studies in C. elegans identified negative regulators of the insulin signaling pathway. One of these genes, daf-16, encodes the C. elegans orthologue of the forkhead transcription factor FOXO. Five years later, FOXO loss-of-function was found to rescue the diabetic phenotype of insulin-resistant mice.
Depression. C. elegans is not only an established genetic model but can also be used to investigate the underlying mechanisms of whole-animal pharmacology. For example, the antidepressant fluoxetine has been shown to increase serotonergic signalling in C. elegans by inhibiting the C. elegans orthologue of the serotonin re-uptake transporter SERT. This has stimulated a number of investigations to identify additional mode of actions of fluoxetine and to further elucidate the molecular mechanism of depression.

Table 4 Key biomedical discoveries enabled by C. elegans research

The first step is to identify an obesity related dataset to use in the computational analysis. We selected a dataset where gene disruption consequences on fat storage in

Caenorhabditis elegans (*C.elegans)* were assessed. Data from the worm is of interest for several reasons: many important biomedical discoveries were first made in the worm (Table 4), genetics and diet can be controlled in model organisms, the first multicellular animal to have its genome completely sequenced was *Caenorhabditis elegans*, *C.elegans* orthologs have been identified for 60-80% of human genes, and many human orthologs to worm genes are functionally orthologous as well [199-201].

On average orthologous human and worm genes have 69.3% similarity [202]. For this approach to identify relevant human transcription factors, we first need to find human orthologs to worm genes using several databases designed to identify orthologous genes among species, such as the NCBI tools, BLASTp [203] and HomoloGene [204]. BLASTp analyzes a user input protein sequence to search genomes for similar proteins. HomoloGene is a database that utilizes BLASTp, taxonomy, and BLASTn information to create clusters of orthologous genes. For worm genes another database, Wormbase, is also available. Orthologs in Wormbase are calculated based on groups generated by the InParanoid algorithm which generates clusters seeded by a reciprocally best-matching ortholog pair [205].

C.elegans is an excellent, simple eukaryotic system for performing large scale-genetic experiments. Ashrafi et al.[206] and Kamath et al.[206] performed single gene disruption of 16,757 worm genes (86% of all worm genes). They identified 305 gene inactivations that cause reduced body fat in the worm, and 112 gene inactivations that cause increased fat storage, many of which have mammalian orthologs. Additionally 261 were observed to decrease fat content, but also lead to death or sterility. Included in this fat storage gene set are pathway-specific fat regulators. This is the first time that an

entire genome sequence has been turned into functional information by RNAi disruption of every gene; already the first analysis gives a glimpse of information about obesity.

After selecting an appropriate dataset, the next step is to group the genes in obesity related GO classifications [207, 208]. Overrepresentation of genes in a specific gene ontology classification hints at regulation by the same transcription factor. Mootha et al.[189] discovered a cluster of coregulated genes in diabetic tissue samples that correspond to 2/3 of the genes in the OXPHOS gene set. Looking for overrepresented GO categories we not only can identify new processes involved in obesity, but these gene sets are also potentially coregulated in obesity.

After identifying gene sets within our obesity dataset, we will search the genes in each set for common TFBS. In the previous study by Mootha et al. peroxisome proliferator-activated receptor γ coactivator 1α (PGC-1α) was identified as the mechanism of regulation for the subset of OXPHOS genes based on the presence of transcription factor binding sites [189]. We will use this same approach and search our obesity related gene sets for TFBS. In the end we will yield a list of obesity related TFs that bind those sites.

Results

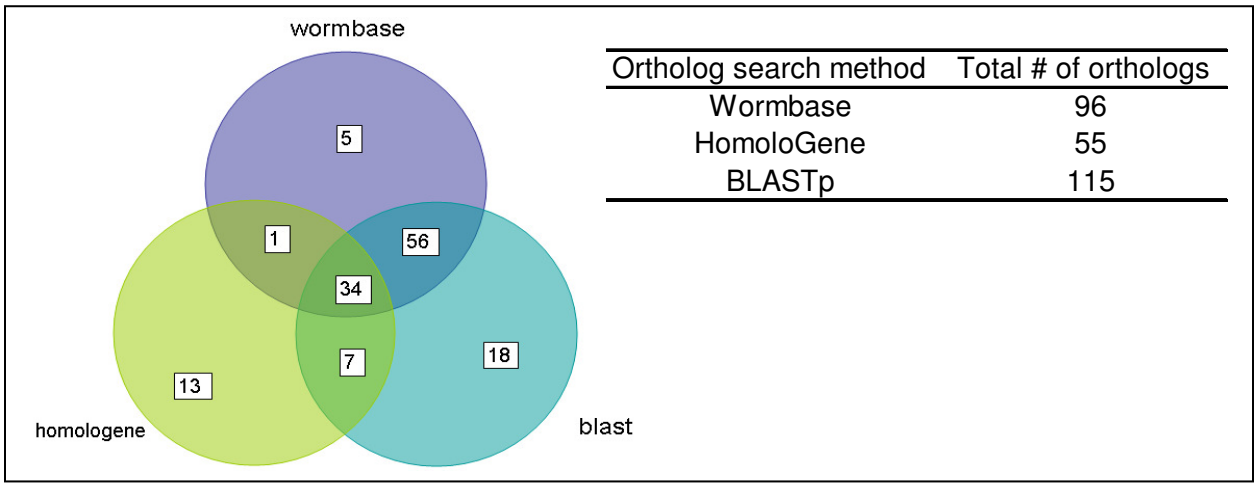
Ortholog Search

In order to identify TFs involved in obesity, we computationally search genomic data for TFBS overrepresented in gene groups. The dataset we choose is a list of genes from *C.Elegans*, whose knock-down affects fat storage content in the worm. This information is publicly available, and our compiled list represents 305 gene inactivations that cause a reduction in body fat, and 112 gene inactivations that cause an increase in body fat. Our list does not include the 261 genes that impact fat storage, but also impact viability or fertility.

We used three approaches to find the human genes orthologous to worm genes that affect fat storage. The first, a basic local alignment search tool (BLASTp), identifies short matches between two protein sequences, propagates the match along the sequence, and creates an alignment. A BLASTp search of the human genome was conducted with protein sequences for the 417 worm genes identified in the fat screen. Of the 417 genes 213 returned a BLAST match, and of those 115 returned a match with an E-value below the threshold $1*10^{-50}$ (Supplementary Table 1). Nine of the human genes returned in this search are identified by the human obesity gene map as obesity related genes [209].

Although informative, using BLASTp to find human orthologs to worm genes that affect fat storage is time consuming. Therefore we also extracted ortholog information from two databases, HomoloGene and Wormbase, with the goal of increasing the efficiency of ortholog retrieval without sacrificing accuracy. The

HomoloGene database was compiled by NCBI using a combination of BLASTp, BLASTn, taxonomy, and nucleotide sequence similarity. Using the HomoloGene database at NCBI 55 human orthologs were identified for 55 worm genes (Supplementary Table 2). This approach is more rapid than performing individual BLASTp searches, but is extremely conservative.



Ortholog search method	Total # of orthologs
Wormbase	96
HomoloGene	55
BLASTp	115

Figure 6 Overlap of C.elegans ortholog identification between methods. Searches for human orthologs to worm genes that impact fat storage were conducted using Wormbase, HomoloGene, and BLASTp. Wormbase results are shown in the purple circle. BLASTp results are shown in the blue circle. HomoloGene results are shown in the green circle. The total number of orthologs identified using each method is shown in the inset table.

On the other hand, Wormbase is a compilation of several databases all containing data on *C.Elegans*. To identify orthologs for the worm fat storage genes, we utilized the Wormbase ortholog data. These data are compiled mainly from Ensembl Compara [210], in contrast to the NCBI database used by HomoloGene. Wormbase returned 244 human orthologs for 97 worm genes (Supplementary Table 1). The number of orthologs retrieved from Wormbase is similar to individual BLASTp searches. A summary of the overlap between the human orthologs identified for the worm fat storage genes in all three databases can be seen in Figure 6.

In the next step towards identification of TFs involved in obesity, we used 293 human genes orthologous to worm genes that alter fat storage content when knocked down. These 293 human genes are compiled from all three ortholog retrieval methods described above. An estimated 50% of the worm genes have one human ortholog. Approximately 40% of the worm genes have 2 human orthologs, and roughly 10% have 3 or 4 orthologs (Figure 7).

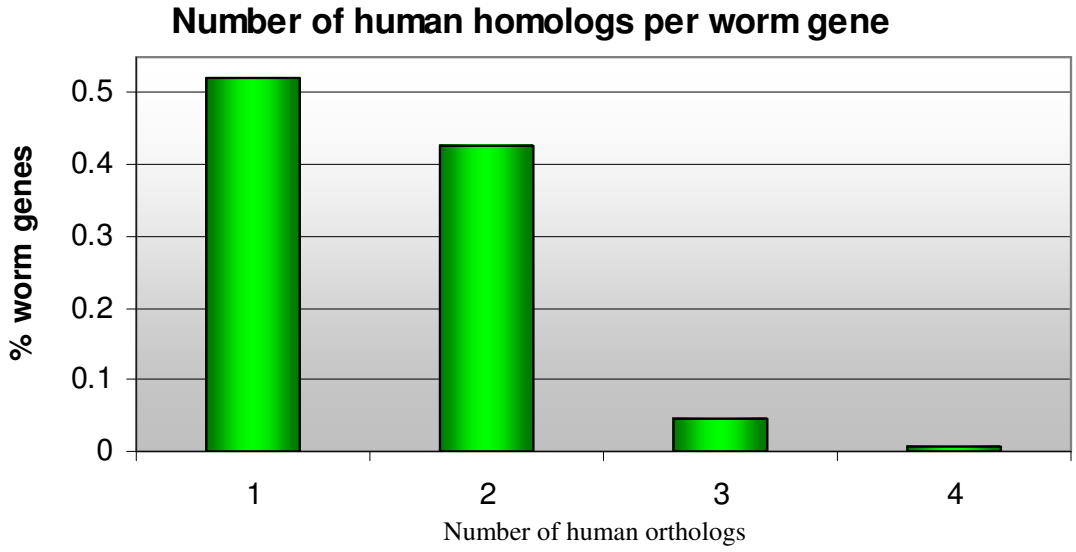


Figure 7 One-to-one and one-to-many relationships between worm genes and human orthologs. Each bar represents the percent of worm fat storage genes with the number of corresponding human orthologs.

Gene Ontology Overrepresentation

In order to identify transcription factors associated with obesity, the human genes from the ortholog screen were examined to find gene ontology classifications (GO) of coordinately expressed genes. The web-based program WebGestalt was used to analyze the human ortholog fat storage set for overrepresented GO groups. To generate a statistic

for overrepresentation Webgestalt requires a list of interest and a background set. Our list of interest is generated from human orthologs of worm fat storage genes. An appropriate background for this list is the collection of all human genes with a worm ortholog.

The list of 293 human genes was entered into Webgestalt along with the background list of 16,365 human genes which have worm orthologs. Genes from the ortholog list occur in 112 GO categories more often than expected by random chance (Figure 8 &Figure 9). For example, 12 genes from our list are involved in the GO biological process Lipid Metabolism. By random chance we would expect 4.15 genes ($p=0.000836$). Supplemental table 2 lists the GO categories overrepresented in the list of human genes orthologs to worm genes that impact fat storage.

TFBS searching

Next we look for shared TFBS in the groups of genes created by the GO search. Since we would like to identify TFs involved in obesity, we will use a program that searches for known motifs. Based on findings about the distribution of TFBS from previous research, we chose to scan the 2kb region upstream of the start site of each gene [177, 178].

To identify TFBS in our obesity related GO groups, there are a number of programs available. Table 5 shows a list of programs that were tested using a positive control set [211-213]. Of the four programs that were tested only F-Match and TFM-

Explorer identified the correct motif as the most overrepresented. WebMOTIFS has a limited number of search motifs available, and our positive control motif is not one of those motifs. Due to the limited number of motifs, WebMOTIFS was considered unsuitable for this analysis. TOUCAN searches genes for TFBS and returns a handy visual output, however it fails to distinguish between two hits in the same gene and one hit each in two genes. F-Match was easy to use and straight forward; however it was slow to run and returns a different number of binding sites from run to run. In addition, background information on how F-Match determines hits is unavailable. In the end the program TFM-Explorer was selected for the analysis. TFM-Explorer correctly identified our positive control motif, has publicly available methodology, and contains TFBS info from Transfac, which contains 960 human motifs, and JASPAR, which contains 138 motifs from multicelluar organisms [214, 215]. Therefore we will use TFM-Explorer.

Program	Ided positive control motif	reproducible results?	matches/gene	Method Available	Based on Transfac/JASPAR
WebMOTIFS	no	yes	yes	yes	no
TOUCAN	no	yes	no	yes	no
F-Match	yes	no	?	no	yes
TFM-Explorer	yes	yes	yes	yes	yes

Table 5 Analysis of potential programs for TFBS overrepresentation.

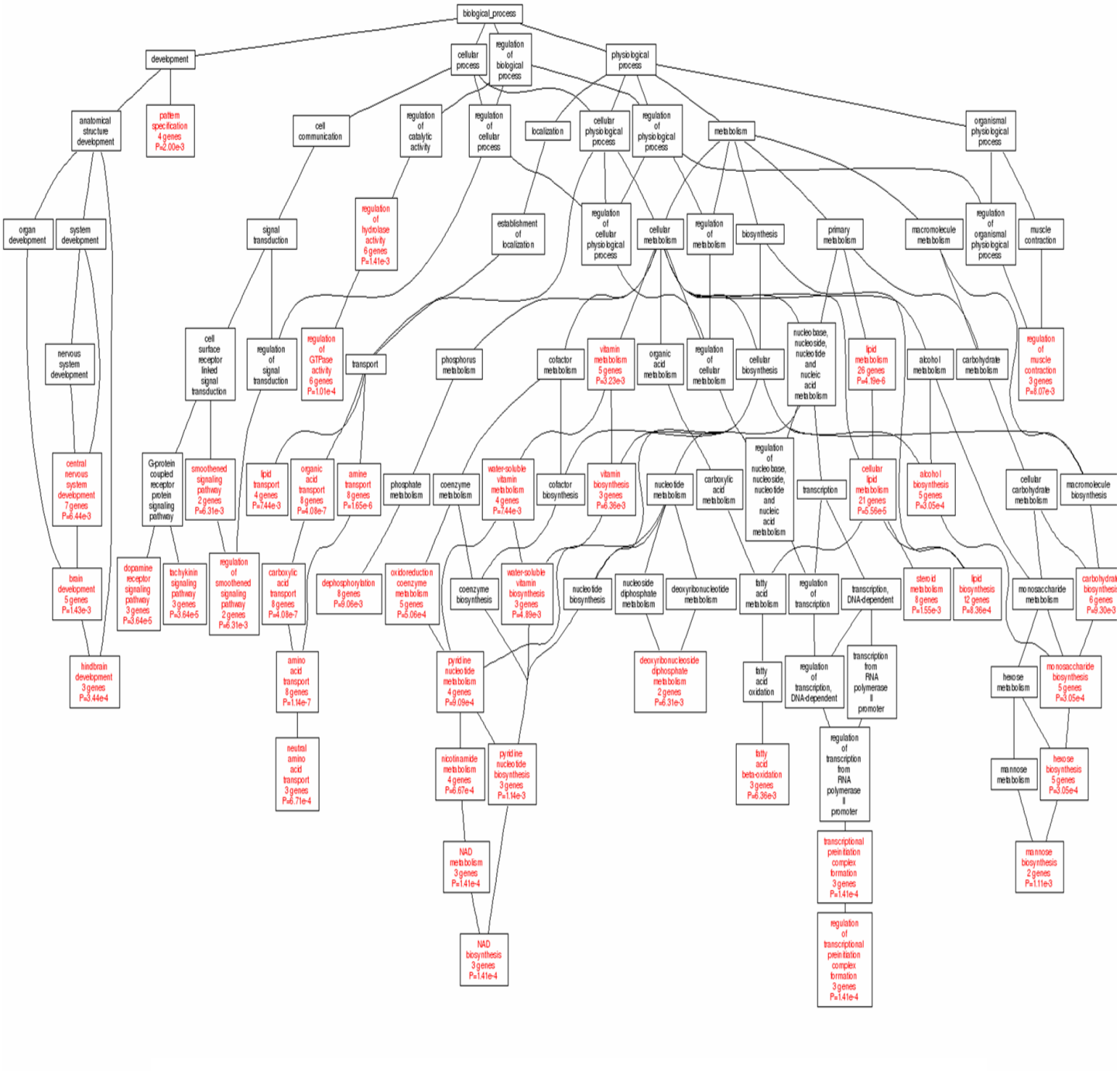


Figure 8 Enriched GO tree diagram. Biological process branch of the GO tree containing significantly overrepresented categories are shown. Enriched categories are shown in red ($p<0.01$).

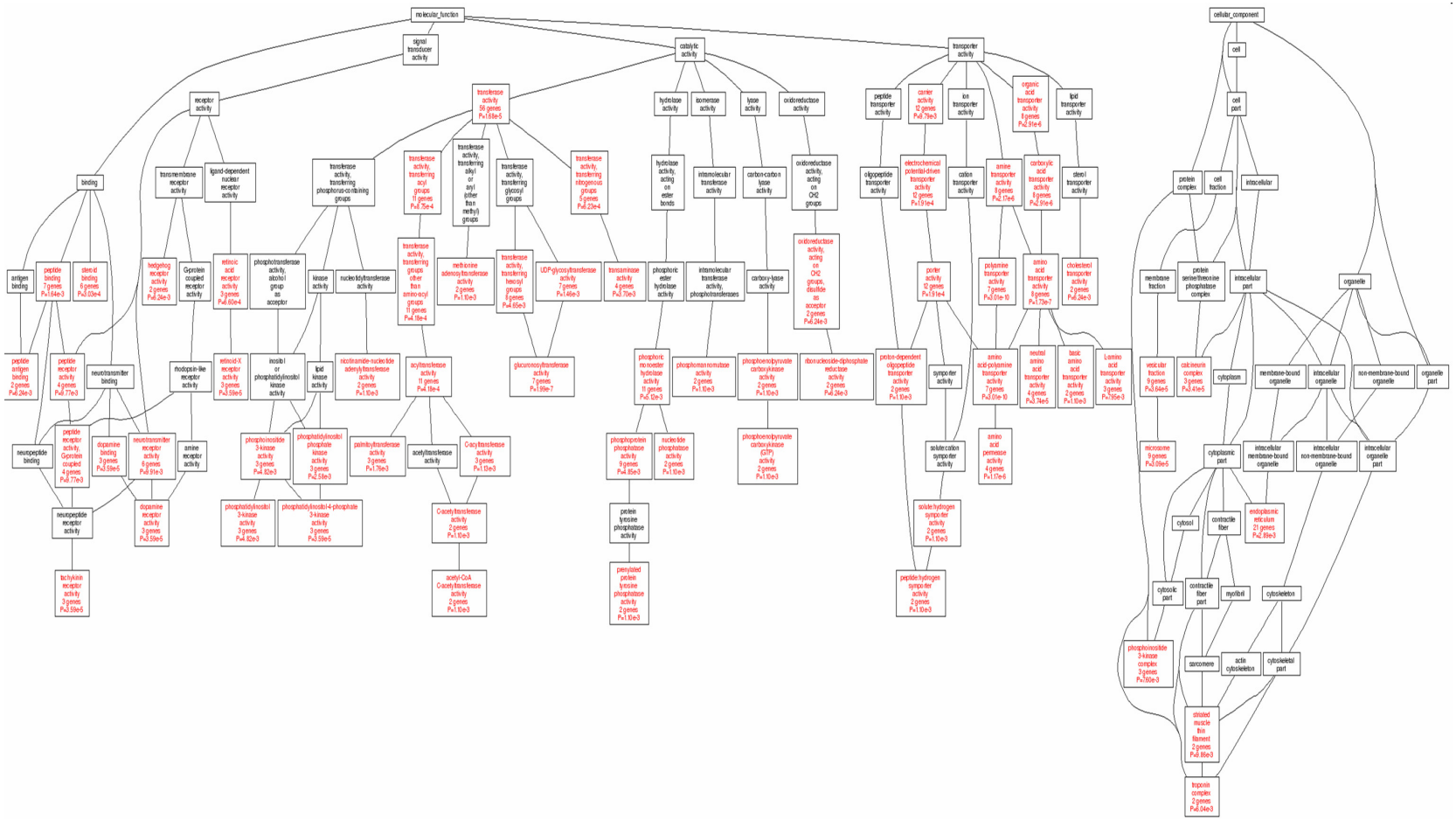


Figure 9 Enriched GO tree diagram. Molecular function and cellular compartment branches of the GO tree containing significantly overrepresented categories are shown. Enriched categories are shown in red ($p<0.01$).

Now that we have selected a program that accurately identifies TFBS, we apply it to search for common sites between genes from our overrepresented GO groups. Using TFM-Explorer we identified the most overrepresented TFBS among the genes in each obesity related GO category. As an example, we searched the upstream regions of the 12 genes from the GO category Lipid Metabolism. The search yields 7 TFBS matrices that are overrepresented in Lipid Metabolism genes at a p-value cut-off of 10^{-6}, a cut-off empirically determined by TFM-explorer as a low probability of false positives (Table 6). A similar search was performed on the gene sets from all 112 enriched GO categories.

Rank	Factor Family	Matrix ID	Location	Sequences	P-Value
1	ETS	SPI-1	[-0321:-0016]	6 (66%)	1.74E-10
2	NF-Y	V$NFY_01	[-0338:-0056]	4 (44%)	3.93E-07
3	Evi-1	V$EVI1_04	[-1799:-1715]	8 (88%)	1.96E-06
4	NF-Y	V$NFY_Q6	[-0372:-0104]	6 (66%)	2.59E-06
5	Freac-3	V$FREAC3_01	[-1685:-0887]	6 (66%)	3.54E-06
6	GATA-1	V$GATA1_02	[-1489:-0123]	8 (88%)	6.11E-06
7	XFD-2	V$XFD2_01	[-1452:-0833]	7 (77%)	8.57E-06
8	Pax-4	V$PAX4_02	[-1800:-0978]	8 (88%)	1.13E-05
9	NF-Y	V$NFY_C	[-0980:-0094]	8 (88%)	1.14E-05
10	HNF-3beta	V$HNF3B_01	[-1488:-0435]	9 (100%)	1.26E-05

Table 6 TFMexplorer hits from the GO category Lipid Metabolism. Rows in gray fall outside of the cut-off p-value of 1x10-6.

In total, this method yields 53 unique TFs that are potentially involved in fat storage (Table 7). To validate our computational approach we tested gene networks in adipogenesis. This assay is described in greater detail in the next section.

Candidate TFs					
CDX1	FOXF1	Gfi	HOXA5	Pax4	Sp1
cEBP	FOXF2	Gfi-1	Jun	Pax6	SPI-1
E4BP4	FOXI1	Gklf	MEF2A	Pbx1	SRF
EGR4	FOXJ2	GR-alpha	Msx-1	POU2F1	SRY
EVI1	FOXL1	GR-beta	MYEF2	POU3F1	TCF11-MafG
Evi-1	FOXQ1	HFL	NFE2L1	POU3F2	TEF
FOXA2	GATA1	hnf1	NFIL3	PRRX2	TFAP4
FOXC1	GATA2	HNF-1A	NFY	RORA	YY1
FOXD1	GATA3	HNF4A	NKX2-5	Sox5	
FOXD3	GC box elements	HOXA3	Pax2	Sox9	

Table 7 Candidate TFs for involvement in obesity.

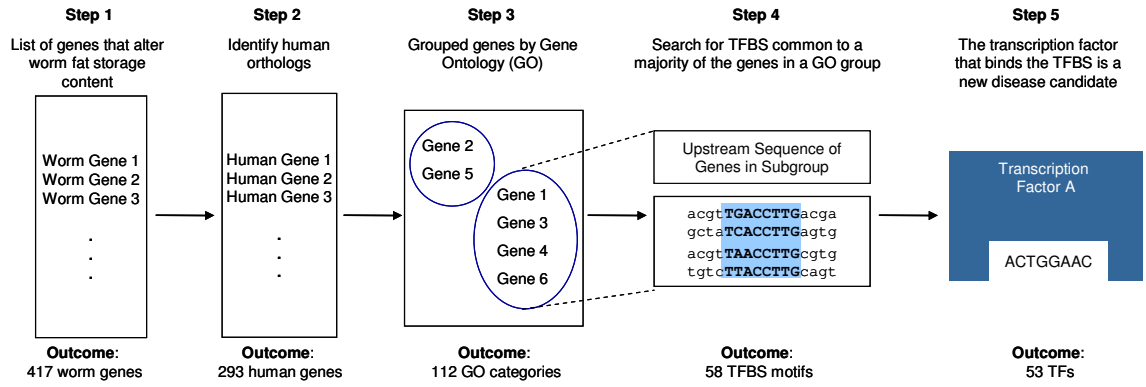


Figure 10 Schematic representation outlining the computational method and findings.

Discussion

This study marks the first time the transcriptional network of a genome-wide RNAi obesity gene screen has been mapped. We teased out the TFs responsible for regulation in human genes orthologous to worm genes which alter fat storage content. Our approach combines grouping genes according to function (GO) and scanning those grouping for common TFBS, and tracing the TFBS to the TF that binds the motif. Results from regulatory motif analysis of our GO groups provided evidence for putative binding site for 53 TFs in genes association with fat storage changes in the worm. There is evidence for 30 of these TFs related to obesity in the literature [216-223]. Among the 30 TFs with a known connection to obesity, we find genes such as GATA2 and GATA 3. GATA 2 and 3 both have a well defined role in suppressing adipocyte formation [166, 224, 225]. Recently it has even been shown that treatment with botanical alkaloids inhibits adipogenesis through upregulation of GATA2 and GATA3, making them excellent obesity treatment targets [226]. Although further experimental work is needed

to determine if our other 23 candidates are associated with obesity, these results suggest our candidate TFs constitute an important part of the pathobiology of obesity.

We found that Wormbase is a comprehensive and easy to use database for the identification of human genes with orthology to worm genes. Individual BLASTp searches using protein sequences is the most common ortholog identification method, when looking for updated ortholog mapping [227]. These searches can be time-consuming; using wormbase expedites the process of ortholog identification. Wormbase, however, only contains worm gene ortholog data. Other databases need to be verified to be suitable for identification of orthologs among other species.

For the worm genes that alter fat storage content approximately 25% have human orthologs. Previous reports indicate 60-80% orthology between human and worm [199, 228]. However, Tamas et al. uses a much stricter ortholog definition and demonstrates a lower rate of ortholog identification [201]. Therefore, our definition of ortholog may be more stringent that previous reports. A more likely scenario is the absence of essential genes from our dataset. Ashrafi et al. [206] identified 261 genes that affect fat storage and fertility or viability. These genes are not included in our fat storage gene list. Amsterdam et al. reports higher orthology among essential genes [229]. It is likely that our dataset, which excludes essential genes, has decreased orthology.

Roughly 50% of the human orthologs we identify appear to be inparalogues. Inparalogues are gene duplications that occur after a speciation event [199]. Interestingly inparalogues often split the function of the ancestral gene. For example, the worm gene *Gana-1* performs the dual functions of its human orthologs, alpha-galactosidase A and alpha-N-acetylgalactosaminidase [199]. It is beyond the scope of the current project, but

it would be interesting to access the combined functional impact of each ortholog cluster on human obesity.

After grouping the human orthologs to worm fat storage genes according to GO, we find 112 overrepresented categories. Many of these categories reflect our current knowledge of obesity. Lipid metabolism and fatty acid oxidation are key processes in obesity [230-232]. We identify them as overrepresented in our gene list. In addition, the brain plays an important role in central regulation of obesity [233-236]. From our list of human orthologs, we identify brain development as a process involved in fat storage. In fact, using an advanced literature search engine, Chilibot [237], we see at least 75% co-occurrence between our overrepresented GO categories and obesity-related terms in PubMed abstracts. This indicates that our computational method is effective in identifying GO categories that are relevant to obesity.

After successfully clustering our human orthologs into obesity related GO groups, we used the program TFM-Explorer to search for known TFBS in each cluster of genes. We used the program TFM-Explorer, because it correctly identified the common TFBS in our positive control set. However, there are still limitations to TFBS identification using any program. Some of these limitations come from an incomplete knowledge of TF binding at the biological level [238]. We have not yet described the binding sites for all TFs in all tissues and developmental time points. As well, our understanding of the impact of neighboring binding sites on one another is incomplete [173, 238]. Therefore computational identification of binding sites is only as good as our knowledge of TFBS biology. Despite these limitations, our positive control test set demonstrates that programs can accurately identify the common TFBS in a set of genes.

New programs for known TFBS searches are being developed constantly. The current generation of TFBS search engines has begun to incorporate 3-D DNA structure data and neighboring as well as non-neighboring base dependencies [238-240]. In the future we can continually improve the accuracy of our computational approach by implementing emerging binding site search engines.

Similar approaches for computational mapping of regulatory networks have been used before. Boily et al. [241], Wang et al [242], and Romero et al. [243] have all combined gene expression clustering techniques with GO grouping and TFBS searches. Boily et al. [241] confirms our hypothesis that GO and TFBS groupings significantly overlap with co-regulated gene sets. Wang et al. [242] identifies the known immune response associated TF NFKB1 as the overrepresented TFBS in a cluster of genes modulated by bacterial infection. In addition they identify immune response related GO terms as upregulated in their cluster of genes [242]. The positive results of this study demonstrate the validity of this combined computational approach. However, what distinguishes our approach from these studies is the input data and the aim of the study. The studies mentioned above all generate their own experimental gene expression data specifically tailored for their analysis. Our method utilizes data from a whole genome RNAi screen, providing an initial dataset where a phenotype is produced in a whole animal. Therefore there is a direct link between our input data and the phenotype of interest. In addition, the aim of our study is to analyze publicly available data to determine the TFs involved in the regulatory networks of obesity. Using our method to increase our knowledge of gene regulation gives us a novel approach to discovering candidate genes involved in obesity.

This methodology of candidate gene discovery has not only revealed 23 novel therapeutic targets for obesity, but it also provides a candidate gene discovery platform to aid in uncovering the heritability of traits and disease. Table 2 demonstrates the gap in heritability explained by GWAS [107]. Several groups have examined the utility of combining prior knowledge with GWAS studies to test a directed panel of SNPs [120, 244-248]. These studies demonstrate an increased capacity to identify variants associated with a variety of phenotypes, while drastically reducing the number of statistical test performed. This lowers the chance of false positives and reduces the total cost of genotying. Using the SNP selection scheme outlined by Chen et al. [248] or Xu et al. [120] informed by our 53 candidate TFs a directed screen for obesity associated variants can be conducted. Additionally, our method can be applied to other phenotypes and datasets to provide candidates to generate informed SNP screens across a variety of diseases and phenotypes.

Looking forward, to validate the candidate obesity TFs identified by our method, we will test their function in adipogenesis. We will use differentiating OP9 cells as a model for adipogenesis [167]. This model was chosen because of the availability of the system, relevance to the original fat storage phenotype, and involvement of numerous transcription factors in the process. Development of a high-throughput RNAi screen to determine the effect of depletion of our candidate TFs on adipogenesis is expected to be finished in the near future.

Conclusion

Using our new approach of grouping genes according to function and searching them for common TFBS, we have extracted 58 TFs with a potential role in obesity. Through this analysis we also identified Gene Ontology categories involved in obesity. Our method started with a publicly available dataset, which we analyzed in a novel way. Although we applied our method specifically to the problem of obesity, the approach can be broadly applied to identifying TFs in a variety complex diseases and phenotypes.

Methods

BLASTp searches: For BLASTp searches (http://blast.ncbi.nlm.nih.gov/Blast.cgi), human orthologs were identified by searching the human reference proteins (refseq_protein) database using the worm protein reference sequence NCBI identifier in BLASTp. The search was performed using the blastp algorithm with default parameters. A threshold E-value of $\leq 10^{-50}$ was used to select orthologous proteins. The search was repeated for each worm gene**.** The identified orthologs were used for subsequent analysis.

Gene Ontology Overrepresentation: WebGestalt Gene Ontology Search: Gene Ontology (GO) overrepresentation was determined using the web-based tool, Webgestalt (http://bioinfo.vanderbilt.edu/webgestalt/). Webgestalt automatically calculates the overrepresentation in all three branches of the GO tree. Gene identifiers were entered

using the Entrez Gene ID. The following parameters were selected: organism(hsapiens), gene ID type (hsapiens_entrezgene), analysis (Enrichment analysis, GO analysis), statistical method (hypergeometric), multiple test adjustment (BH), significance level (.01), minimum number of genes for a category (3), reference set (user curated set consisting of all human genes with worm orthologs).

TFM-Explorer: Genes were searched for the presence of known transcription factor binding sites using the web-based tool TMF-Explorer (bioinfo.lifl.fr/TFM/TFME/). A list of RefSeq identifiers was entered representing gene sets from GO categories. Genes with >2 RefSeq IDs were eliminated from the analysis. Each gene was searched from -2000:200bp from the transcription start site by the automated TFM-Explorer software. The JASPAR and TRANSFAC matrices were used. Maximum 25 clusters were displayed, with a maximum P-value of 0.0001, and a ratio value of 3.

Chapter 3: Adipogenesis in OP9 cells

Introduction

Obesity is an excessively high amount of body fat or adipose tissue. Adipose tissue provides a critical link in maintaining systemic energy balance, sending hormone signals like Leptin directly to the brain [136, 166]. One aspect of adipose cell biology, adipogenesis, has been particularly well-studied. Immortalized preadipocyte cell lines have provided an in vitro system for the study of fat cell development. Mouse cell lines designated 3T3-L1, 3T3-F442A, and OP9 readily differentiate into adipocytes [166, 167]. Many features of these cells lines recapitulate the process in vivo, including morphological changes, lipid accumulation, and sensitivity to most or all of the key hormones that affect adipocytes. Through research based on these cell lines, mouse studies, and other techniques, three classes of transcription factors have been identified that directly influence adipose cell development: PPARγ [168], C/EBPs [169], and the ADD1/SREBP1 basic helix-loop-helix family [166, 170] (Figure 4).

Cells models of adipogenesis are useful models for testing the involvement of candidate TFs in obesity. These models are advantageous because of the availability of the system, the relevance of adipocytes to the original fat storage phenotype, and the involvement of numerous transcription factors in the process. A role for SREBP1, PPARγ, and C/EBPs has been described, but current data reveals the need for other

branches of the adipogenic transcriptional network to account for all of the available data [166].

The adipogenesis model is a useful method for screening a large number of candidate TFs. The method is cost-effective, easy to scale-up, and already in use by local labs. Transcript levels of candidates TFs will be measured in differentiating preadipocyte OP9 cells, due to the rapid nature of OP9 cell adipogenesis.

OP9 cells are a line of bone marrow derived mouse stromal cells that rapidly differentiate into adipocytes [167]. Cells are induced to differentiate by treatment with low serum media containing insulin and oleic acid. Upon stimuli OP9 cells accumulate triacylglycerol, display adipocyte morphology, and express adipocyte late marker proteins such as Glut4 and adiponectin (Figure 11). OP9 cell differentiation takes three days, as opposed to the 10 day period needed for differentiation in 3T3-L1 cells. This difference makes transient transfection experiments possible in OP9 cells. In addition, OP9 cells are immortalized and can be passaged for many generations without a decrease in adipogenic potential. Lastly, unlike 3T3-L1 cells, OP9 cells are readily transfectable (see results).

To determine if a transcription factor is involved in adipogenesis we study the expression levels of the regulatory network of OP9 cells. Microarrays simultaneously measure the expression level of tens of thousands of genes. We use microarrays on differentiating OP9 cells to globally measure expression.

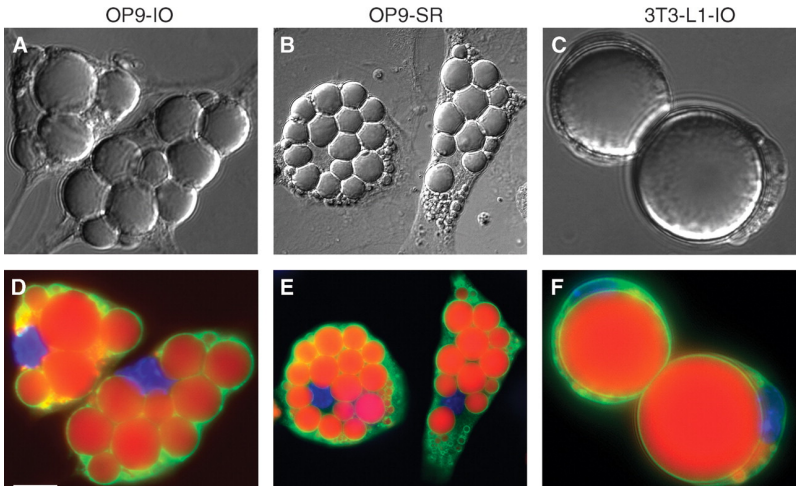


Figure 11 Morphology of OP9 and 3T3-L1 adipocytes. A-C. Differential interference contrast images. D-F. Fluorescently labeled fatty acids (red) and Perilipin (green). OP9 and 3T3-L1 cells form lipid droplets coated with Perilipin. OP9 cells are shown differentiated with either insulin oleate (IO) or serum replacement (SR). 3T3-L1 cells differentiated with insulin oleate media. [167]

Results

Clonal OP9 cell lines differentiate rapidly and with high efficiency

To study the effect of gene knockdown on adipogenesis we created a highly efficient preadipocyte clonal OP9 cell line. Previously OP9 cells plated at 5,000cells/cm2 were shown to differentiate upon treatment with insulin oleate [167]. In our hands OP9 cell differentiate with low efficiency. We do, however, see a slight increase in TG accumulation after 72 hours, which indicates some cells within our population have the ability to differentiate (Figure 12, parental).

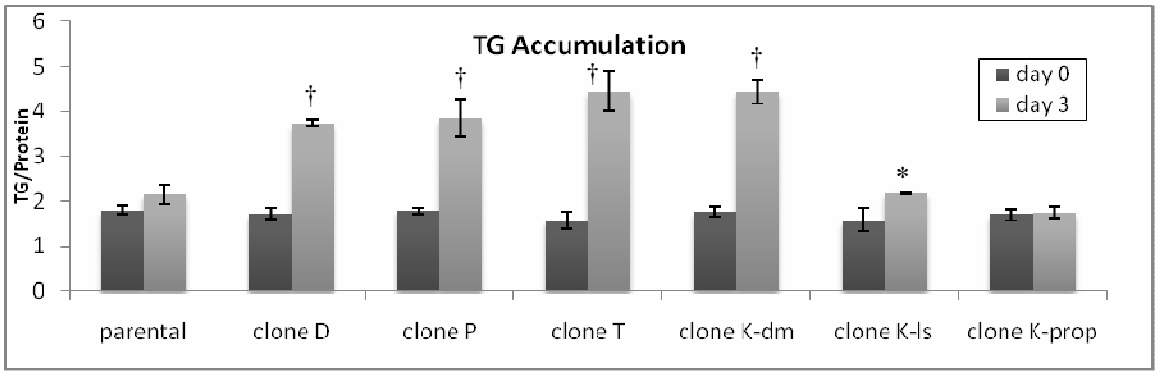


Figure 12 Clonal OP9 cells differentiate with varied efficiency. Triglyceride accumulation of clonal OP9 cells treated with IO media for 0 or 72 hrs. Error bars represent standard deviation. * indicates significance <0.05, † <0.001

The OP9 cell line is a mixed population. To separate the population of cells capable of efficient differentiation, we created clonal OP9 cell lines. Clonal OP9 cell lines which grew rapidly and had a spindly morphology failed to differentiate (data not shown). We identified four clonal populations with the ability to differentiate at various efficiencies. Triglyceride (TG) content in parental OP9 cells increased 10% after 3 days in differentiation media (Figure 12). All four clonal cells lines accumulate more TG than the parental line (Figure 12). Differentiated clone K OP9 cell lines contain 2-fold more triglyceride than undifferentiated clone K OP9 cells.

OP9 cells were also stained with a lipophilic stain, Nile Red, to assess lipid content. Microscopy revealed that a handful of OP9 parental cells in each plate accumulate lipid droplets (data not shown). All four clonal cell lines showed a significant increase in Nile Red staining between day 0 and day 3. After treatment with IO media all four clonal cell lines visibly accumulate lipid droplets (Figure 13). Differentiated clone K and T cells also display hallmark adipocyte morphology: large lipid droplets and a peripherally located nucleus (Figure 13). Additionally several clone

K cells display a mature adipocyte phenotype consisting of one large coalesced lipid droplet (arrow in Figure 13). Greater than 90% of OP9 clone K cells contain lipid droplets after 3 days in IO media and this potential does not wane with time. Therefore OP9 clone K cells were selected for further characterization.

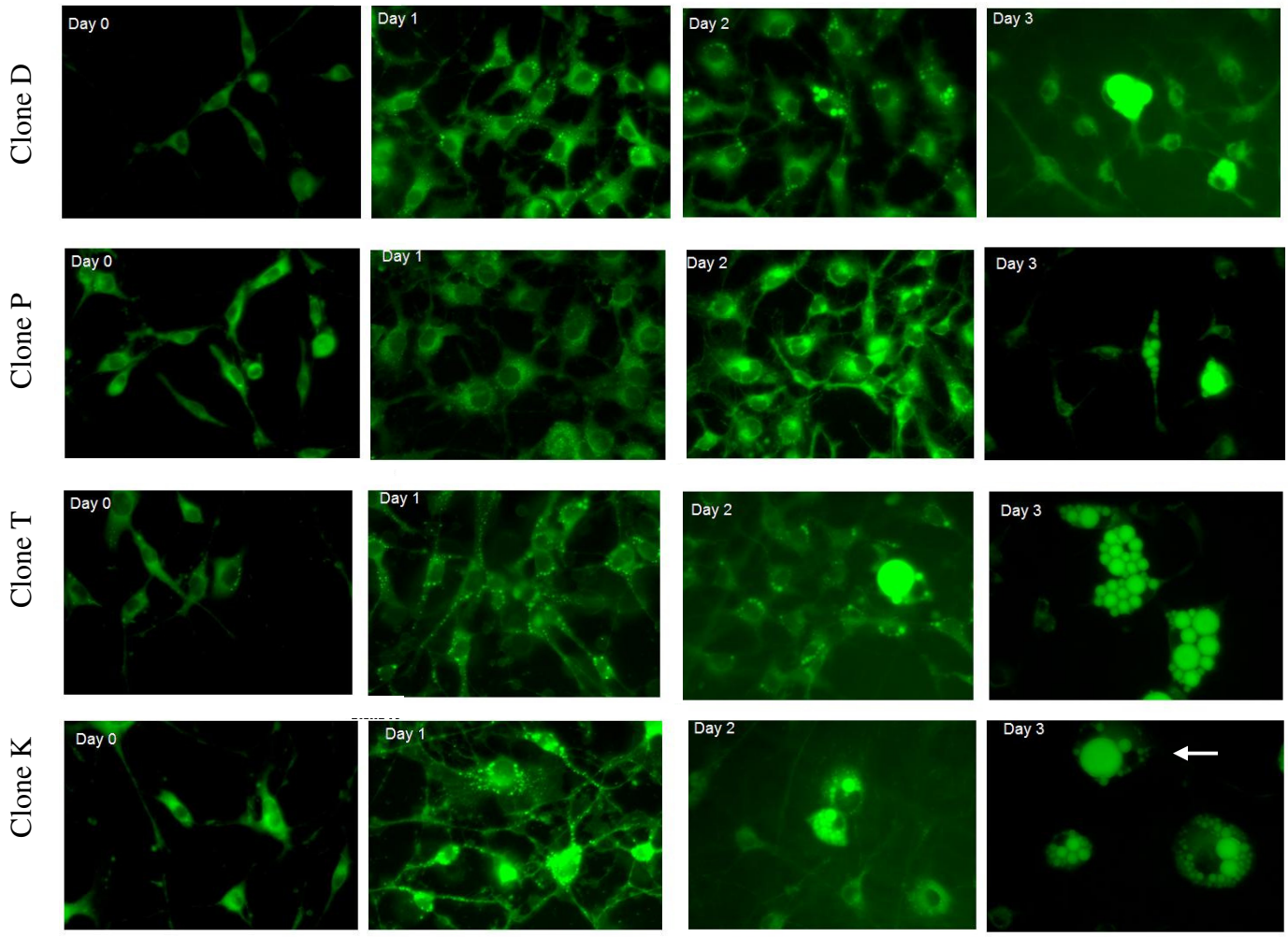


Figure 13 Lipid staining of clonal OP9 cells. Nile red lipid staining of clonal populations.. Images are 20x. Arrow indicates a mature adipocyte.

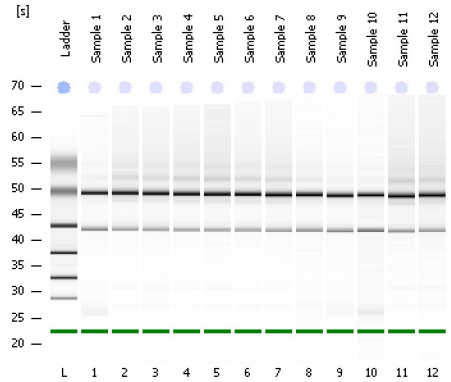
Response of the OP9 clone K transcriptome to adipogenesis

To evaluate if OP9 cells undergo a differentiation process similar to other adipocyte models, a microarray analysis was performed on RNA extracted from OP9 clone K cells during the course of differentiation. To test for reproducibility frozen clone K cells from passage 30 were differentiated as well as clone K cells from passage 34 that were never frozen. RNA from clone K cells treated for 0, 24, 48, or 72 hours with IO media was collected (day 0, 1, 2, and 3). The concentration of RNA extracted ranged from 25ng/ul – 77ng/ul with an average concentration of 55ng/ul (Table 8). RNA samples were sent for quality analysis. Output from the Bioanalyzer shows intact and pure RNA (RIN>7) (Figure 14).

Frozen?	Sample	conc
No	0-1	58.9157
No	0-2	42.6059
No	0-3	63.6073
No	1-1	57.2369
No	1-2	50.9813
No	1-3	26.0841
No	2-1	49.5923
No	2-2	64.7902
No	2-3	53.6485
No	3-1	52.5029
No	3-2	68.2301
No	3-3	32.5029
Yes	0-1	61.6448
Yes	0-2	68.8956
Yes	0-3	73.86
Yes	1-1	69.4289
Yes	1-2	76.9822
Yes	1-3	64.9219
Yes	2-1	66.5059
Yes	2-2	53.521
Yes	2-3	39.5318
Yes	3-1	24.9997
Yes	3-2	60.2008
Yes	3-3	52.945

Table 8 Concentration of RNA extracted from OP9 cells in a total volume of 100uL. Two rounds of differentiation were performed, once on cells frozen and thawed, and once on cells never frozen. Sample name indicated Day-Replicate of each sample. Concentration is in ng/uL.

A



B

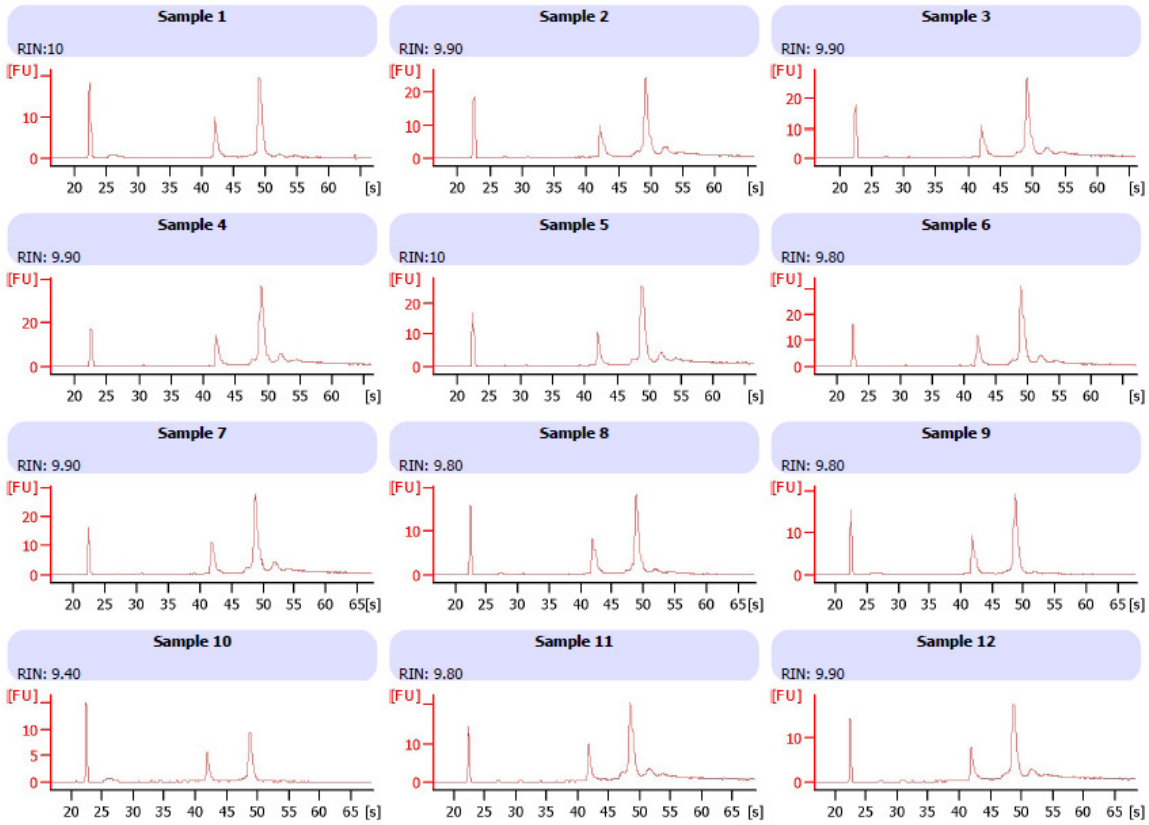


Figure 14 Bioanalyzer RNA integrity results. A) Electrophoresis gel image B) Electropherograms showing the 18s and 28s ribosomal peals. RNA integrity number is shown for each sample. Sample 1-12 represents 3 biological replicates from Day 0, Day 1, Day 2, and Day 3 respectively.

For analysis of the OP9 transcriptome, samples were hybridized to the Affymetrix GeneChip® Mouse Gene 1.0 ST Array and analyzed using Expression Console from Affymetrix and the R Limma package. Microarray data was normalized using RMA

(Figure 15a). The log expression values range from 1-16 (Figure 15b). Data was filtered at a cutoff of 3.5, the background threshold for the mouse ST 1.0 array. To check the variability across replicates, sample cluster analysis was run on the data. Biological replicates cluster together (Figure 15c). We performed Between Group Analysis (BGA) to determine if gene expression changes happen gradually or in intervals. BGA measures and orders samples according to correspondence between the samples. Gene expression data from day 0 and day 1 cells form a distinct group from day 2 and day 3 cells (Figure 15d). There are 4761 probes with significant gene expression changes across all days relative to day 0 (Figure 15e). The greatest overlap is between day 2 and day 3, and the least between day 1 and day 3. Supplemental tables 3 and 4 show the 25 most up/down regulated genes from each time point.

To verify the results of the microarray, quantitative PCR (qPCR) was performed on four known adipose marker genes, Plin, Gata2, C/Ebpa and C/Ebpb (Figure 16a). Microarray and qPCR demonstrate C/Ebpa and Plin are upregulated and Gata2 is downregulated during OP9 adipogenesis (Figure 16b). The early adipogenesis marker gene, C/Ebpb, is expressed in OP9 cells during adipogenesis; however, levels do not significantly change. This corroborates protein expression data from Wolins et al. suggesting that OP9 cells are later stage preadipocytes [167].

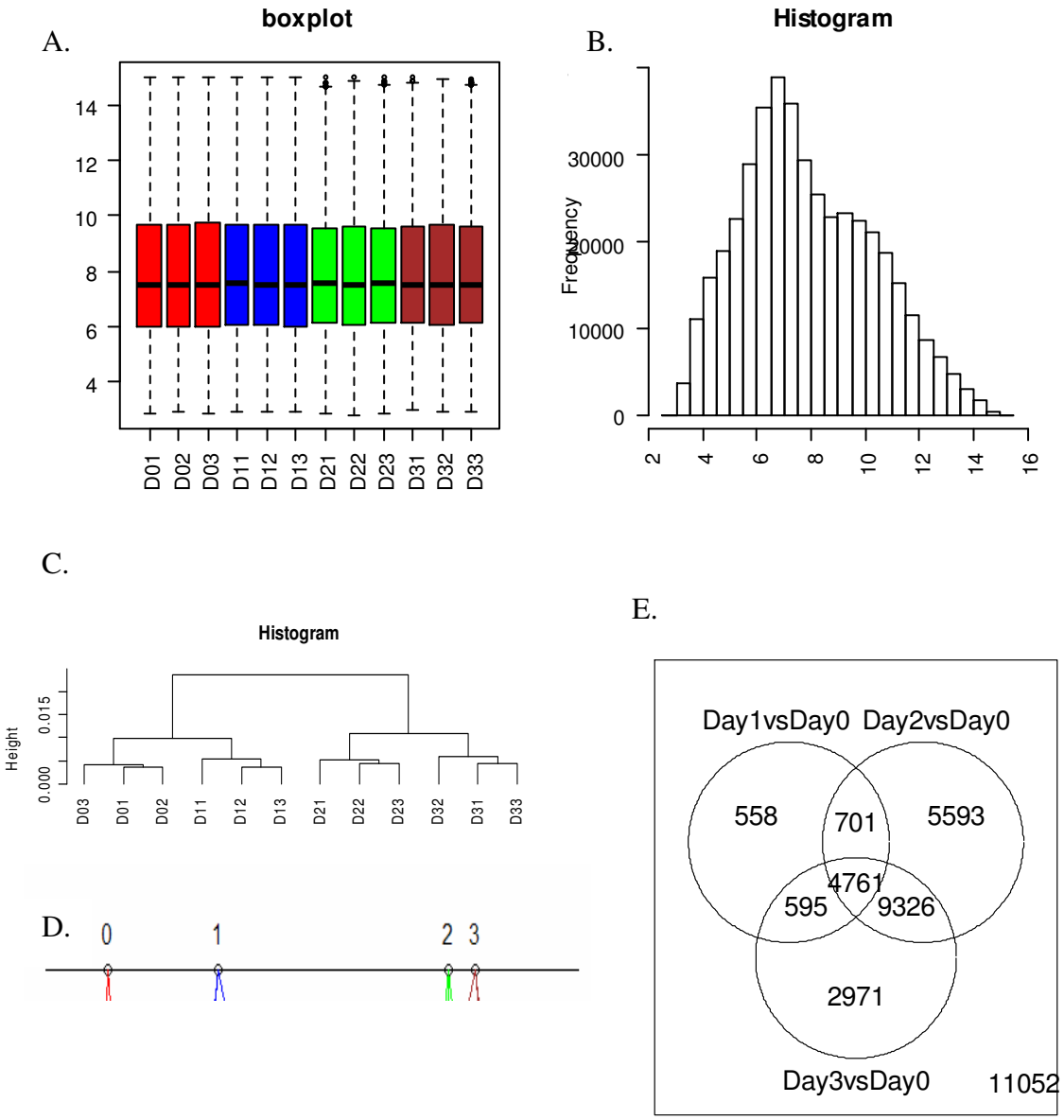


Figure 15 Expression values of differentiating OP9 cells. Samples are labeled with day and biological replicate (D01=day 0, replicate1). A) Log expression values after Robust Multichip Average (RMA) normalization. B) Histogram of expression values across all 12 samples. C) Hierarchical clustering of samples. D) Between group analysis (BGA) supervised classification of groups shows clusters with relative distance on a line. E) Overlap between probes with significant gene expression change.

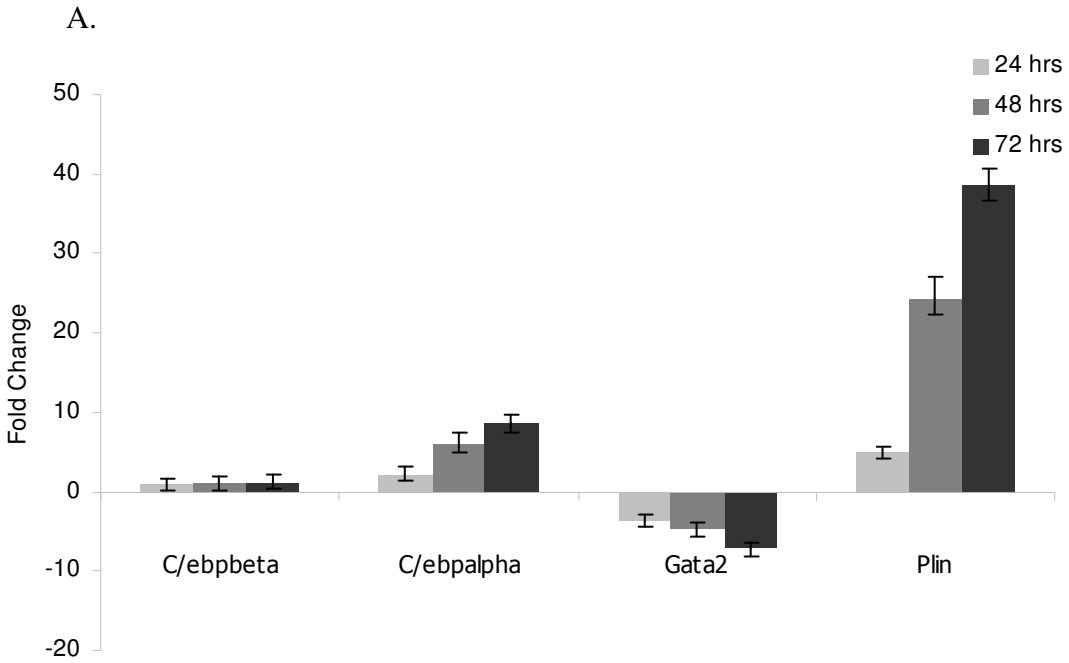


Figure 16 Mid and late stage adipocyte markers are upregulated during OP9 cell adipogenesis. Quantitative PCR fold change of adipocyte marker genes at 24, 48, and 72 hrs after addition of IO media relative to 0 hrs. Standard deviation is derived from 3 replicates.

	Fold Change (p-value)		
Symbol	Day 1	Day 2	Day 3
Plin	6.88 (1.5E-10)	16.63 (2.75E-12)	21.81 (9.85E-13)
Cebpa	1.88 (1.7E-7)	2.89 (1.18E-9)	3.42 (1.36E-10)
Cebpb	1.07 (0.200)	-1.04 (0.314)	-1.18 (0.002)
Gata2	-1.33 (0.008)	-1.5 (3.83E-4)	-1.79 (2.15E-5)

Table 9 Microarray analysis profile of probes for adipocyte marker genes. Expression values are RMA normalized and log transformed. Fold change is relative to Day 0.

To determine which biological processes are involved in OP9 adipogenesis, the 250 most differentially regulated genes at each time point were analyzed for pathway enrichment using GProfiler [249]. Adipose and bone cells share a common progenitor and differentiation is mutually exclusive [250]. We see this confirmed in our samples;

Day 1 OP9 cells down regulate cartilage development, signaling to bone morphogenic proteins, and proteinaceous extracellular matrix (Figure 17). Furthermore, adipocytes are mitotically quiescent cells [251]. We see down regulation of cell cycle components during day 2 and day 3 of OP9 adipogenesis (Figure 17). We also see downregulation of transcripts associated with miR 568, miR 374, and miR 590-3p.

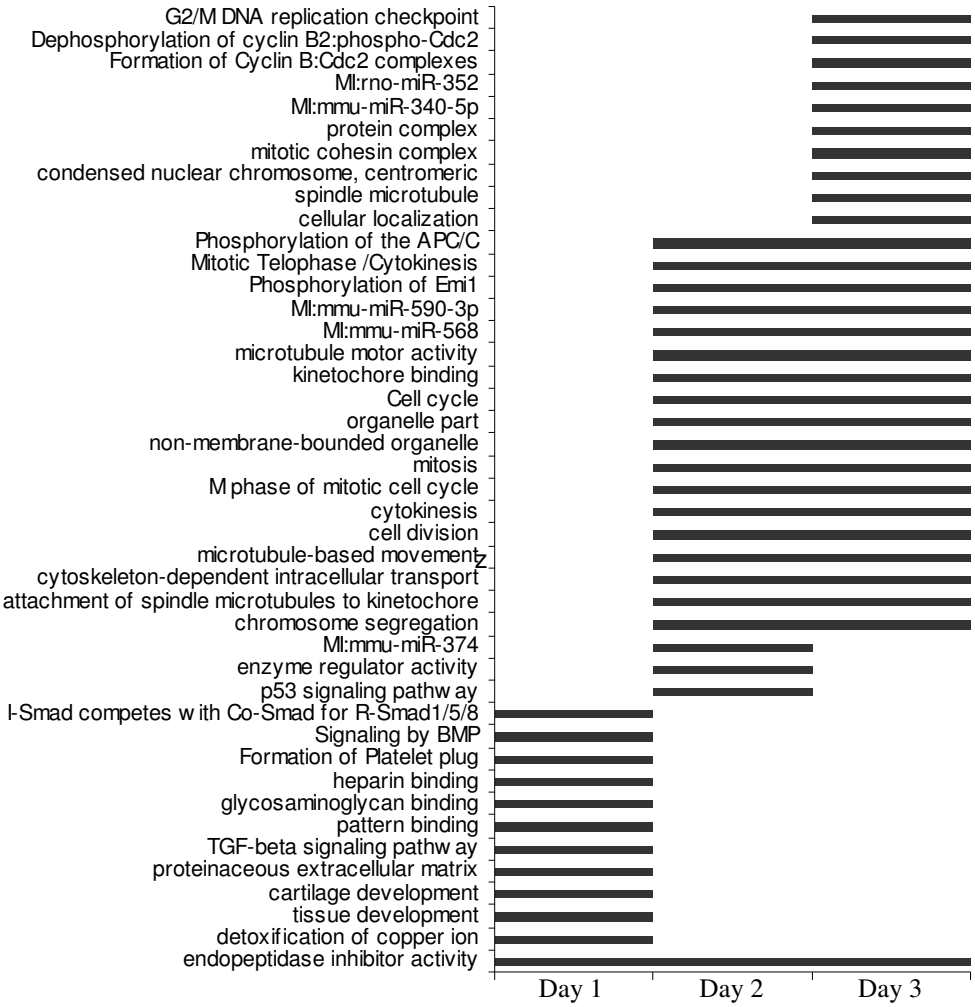


Figure 17 Functional profile of the 250 most down-regulated genes during OP9 adiopgenesis

During the entire course of fat cell differentiation the following processes are upregulated; lipid metabolism, PPAR signaling, and triacylglyceride biosynthesis (Figure 18). The only process upregulated on day 1 alone is transmembrane receptor protein tyrosine kinase adaptor protein activity. Many pathways are upregulated on day 2 and remain upregulated. These pathways include cellular alcohol metabolism, sterol metabolism, and cholesterol metabolism. By day three the following processes are upregulated; vitamin A metabolism, cellular carbohydrate metabolism, steroid biosynthesis, glucose metabolism, cholesterol biosynthesis, and miR 516a-5p. Interestingly the nuclear envelope-endoplasmic reticulum network is also enriched in day 2 and day 3 differentiating cells.

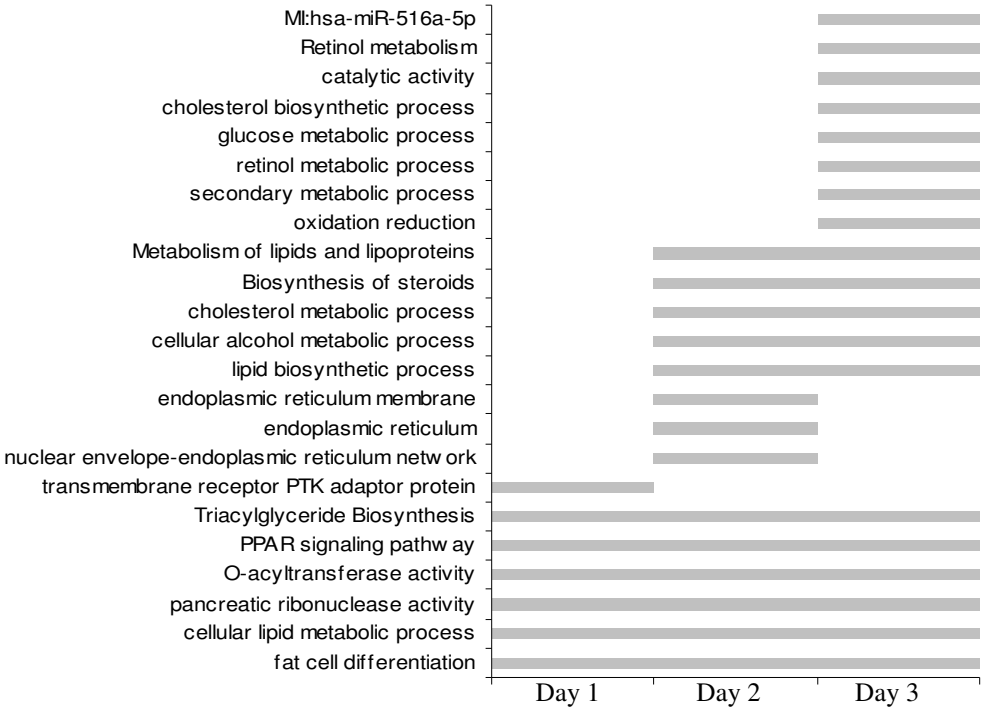
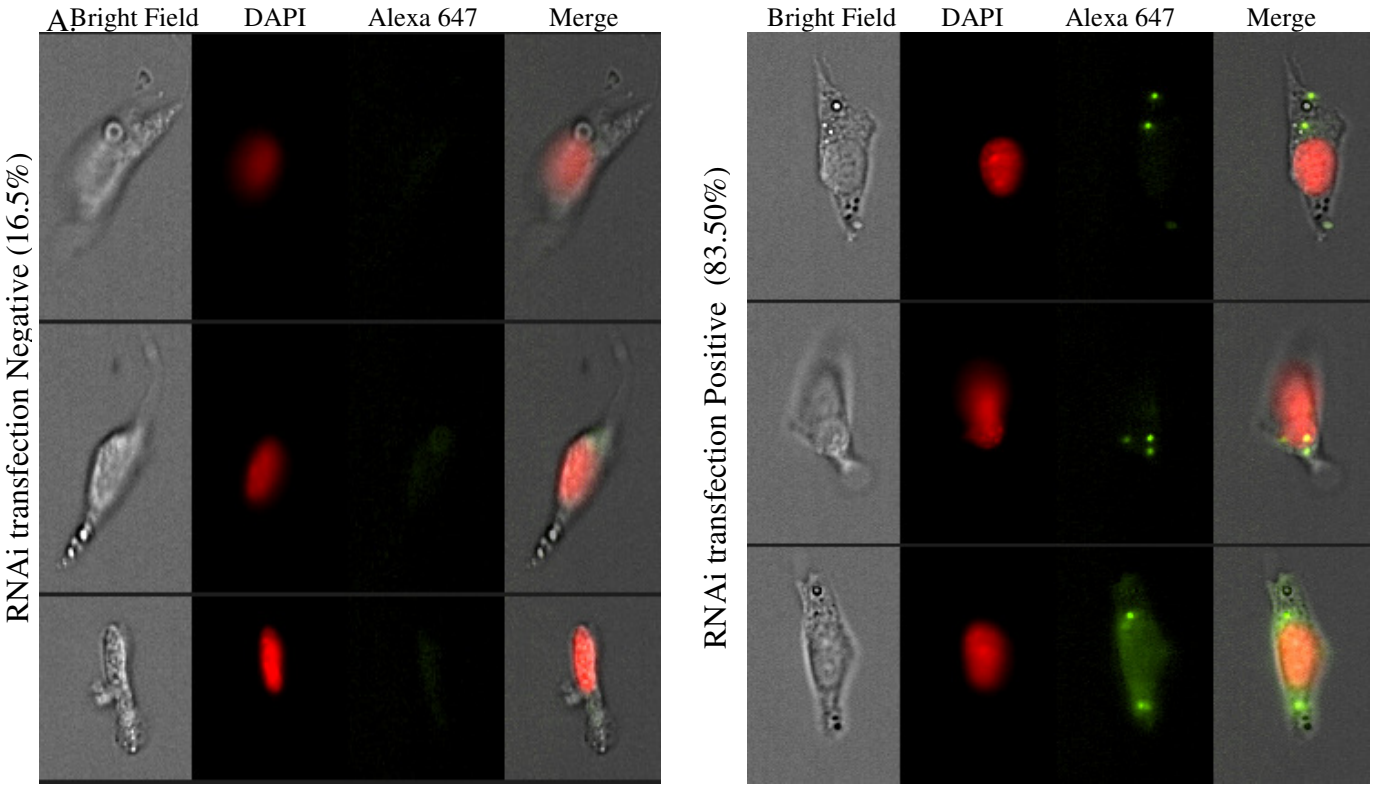


Figure 18 Functional profile of the 250 most up-regulated genes during OP9 adipogenesis

High-throughput knock-down of target genes in OP9 cells

After characterizing the transcriptome of OP9 clone K cells, we conclude OP9 adipogenesis is similar to other cell models. The main difference between 3T3-L1 cells and OP9 cells is the differentiation period. OP9 cells differentiate in 72 hours. To test if high-throughput knock-down of target genes is possible in OP9 cells we tested OP9 clone K transfectability, the efficiency of gene knock-down, and the effect of knocking-down genes essential for differentiation.

To assess if OP9 cells are readily transfectable an Alexa 647 control RNAi was transfected into OP9 clone K cells. Cells imaged 48 hours after transfection have green foci, indicating transfection (Figure 19a). To quantify the percent of transfected cells we used the Amnis ImageStream, multispectral imaging flow cytometer. Roughly 83.5% of the cells contain foci (Figure 19a). To study the ability of RNAi to knock-down transcripts in OP9 cells we transfected RNAi directed against GAPDH. We observed an 8-fold reduction in GAPDH expression level in cells transfected with GAPDH RNAi (Figure 19b). We also transfected OP9 cells with a positive cell death phenotype control RNAi, and saw a marked reduction in the number of cells in positive cell death RNAi transfected wells compared to negative control RNAi (data not shown). These experiments demonstrate that OP9 cells are readily transfectable.



B.

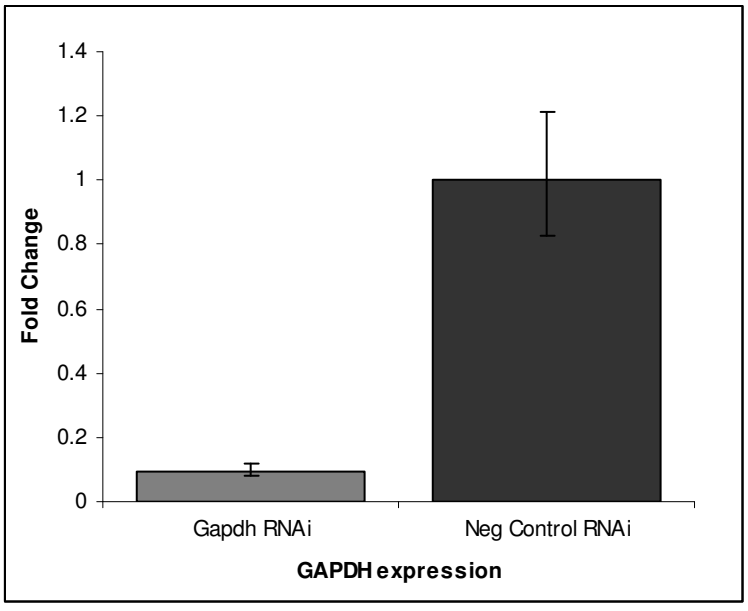


Figure 19 Transfection efficiency and efficacy in OP9 clone K cells A. Amnis imagestream images of OP9 clone K cells transfected with Alexa 647 tagged RNAi. B. Relative expression of GAPDH. qPCR analysis of GAPDH expression level in two GAPDH RNAi samples and two negative control RNAi samples. Data are shown relative to GAPDH knock-down sample 2.

To create a high-throughput method to quantify differentiation, images of differentiating cells were obtained on the ImageExpress machine and an analysis scheme was created using Acquity software. The system allows us to compare nile red staining intensity (threshold NR intensity) of cells with different treatments. Preliminary analysis of differentiating OP9 cells using the ImageExpress and Acquity software shows a correlation between the integrated intensity measurements and lipid accumulation (Figure 20).

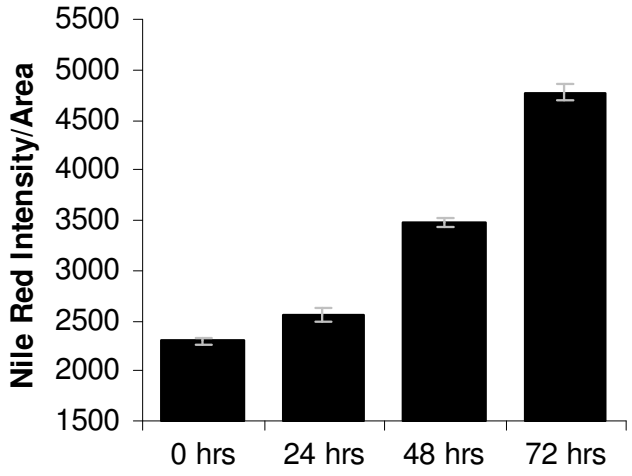


Figure 20 High-throughput measurement of OP9 cell lipid accumulation. OP9 clone K cells were differentiated for 72 hours. Samples were collected every 24 hours, fixed, stained, and scanned using ImageExpress. The intensity per area of nile red staining increases during the course of differentiation.

Next we knocked down PPARgamma and SPI1 in OP9 cells to test the effect of TFs on adipogenesis. It is expected that PPARgamma RNAi will prevent adipogenesis in OP9 cells [167]. As expected, compared to negative control cells, PPARG RNAi cells do not differentiate (Figure 21, $p<0.0001$). Additionally, PPARG RNAi cells do not respond to the PPARG agonist ciglitizone ($p<0.0001$). Cells treated with SPI1 RNAi

accumulate more lipid than negative control cells ($p<0.0001$). However, there is no significant increase in lipid accumulation in SPI1 RNAi cells treated with ciglitizone verses negative control cells (Figure 21).

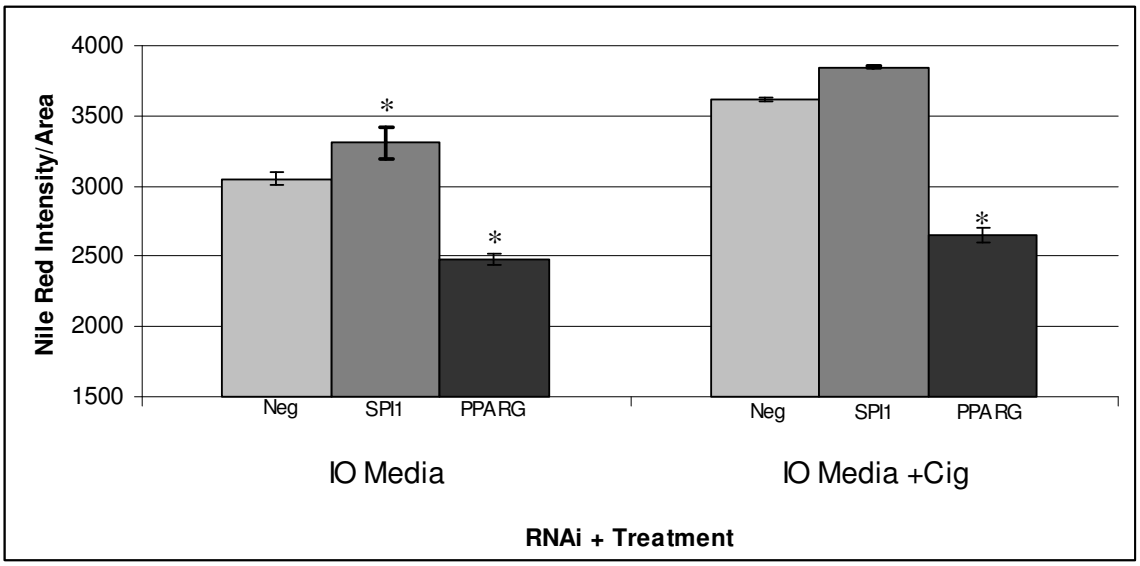


Figure 21 Inhibition of adipogenesis by gene knockdown. OP9 cells transfected with RNAi directed against negative control, SPI1, or PPARG. Cells are treated with insulin oleate differentiation media with and without the PPARG agonist, ciglizitone, for 72 hours. Cells are fixed and stained with DAPI and Nile Red. Each well is scanned by ImageExpress and the intensity of the Nile Red stain is calculated using acquity software. Intensity values are shown are the average of 72 images. Error bars represent the standard error of the mean. * indicated significantly different from negative control at $P<0.05$

Discussion

We set out to establish OP9 cells as a model of adipogenesis suited for gene-depletion studies. The gene expression profiles of differentiating OP9 clone K cells were studied to determine if classical adipogenic marker genes and processes identified in other cell lines are important for OP9 cell differentiation as well. After confirming OP9

clone K cells as a model of adipogenesis, we establish a method for evaluating gene-depletion effects in differentiating OP9 cells, representing mid to late stage adipogenesis.

The transcriptome of OP9 clone K cells shows a number of interesting genes which are differentially regulated. Lipocalin 2 (Lcn2) is highly upregulated during differentiation. Lcn2 inhibits S-phase of the cell cycle, a process which is downregulated in adipogenesis. Lcn2 is also an adipokine and regulates inflammation in adipocytes and macrophages [252, 253]. Protein expression levels of lipocalin-2 are upregulated in visceral adipose tissue of obese subjects [254]. Lcn2 is an excellent candidate for a genotype phenotype association study.

Adipogenic marker gene expression data shows that OP9 clone K cells are more committed than 3T3-L1 cells, displaying patterns of mid to late adipogenesis. 3T3-L1 cells upregulate C/Ebpb expression during the exponential growth and clonal expansion phases, the earliest phase of 3T3-L1 differentiation [255]. C/Ebpb is highly expressed in OP9 clone K cells prior to and during differentiation. We also show an increase of expression of mid and late stage adipogenesis genes C/Ebpa and Plin throughout differentiation. In 3T3-L1 cells, C/Ebpa and Plin are upregulated during the terminal growth arrest stage, roughly 4 days into the induction phase of 3T3-L1 cell differentiation [256]. Expression of Gata2 inhibits adipogenesis, and is downregulated when 3T3-L1 cells reach confluency after the exponential growth phase [257]. We see Gata2 downregulated in OP9 adipogenesis. OP9 cells treated with insulin oleate media differentiate without reaching confluency or undergoing clonal expansion. Therefore we see gene expression patterns in OP9 cells that reflect the behavior of 3T3-L1 cells after clonal expansion. An interesting experiment to determine how committed OP9

preadipocytes are, would be to measure expression of a preadipocyte marker such as monocyte chemoattractant protein 1 (MCP-1) [258]. As well we could expose the OP9 cells to osteoblast differentiation media to see if they can still commit to either lineage, similar to mesenchymal stem cells [259].

Our data also demonstrate that OP9 clone K cells undergo similar functional processes as other adipogenic models. Functional profiling results confirm that differentiation in OP9 cells upregulates triacylglyceride biosynthesis, PPAR signaling, and cholesterol metabolism [260-265]. These processes are hallmark processes of adipogenesis. Another hallmark process of terminal differentiation is cell-cycle arrest [266, 267]. Cell cycle, mitosis, cell division, and microtubule motor activity are all down-regulated as a result of OP9 differentiation, indicating terminal differentiation in these cells. The results highlight the similarities between differentiation of OP9 cells and other adipogenic cell models.

We have found that OP9 clone K cells differentiate after 72 hours of exposure to insulin oleate media, displaying hallmarks of mature adipocytes, with large lipid droplets and peripheral nuclei [268]. This is in contrast to 3T3-L1 cells, which require confluency, three stages of specialized media, and 10 days to achieve differentiation [269]. This difference is important, because the effects of transient RNAi transfection lasts ~96 hrs. The rapid differentiation of OP9 cells opens up countless possibilities for new methods of high-throughput discoveries in adipose biology.

One of the high-throughput screening possibilities with OP9 cells is gene knockdown. For two reasons this has not been possible in 3T3-L1 cells. First, 3T3-L1 cells are notoriously hard to transfect with transient transfection efficiencies between 10-

50% [270]. In contrast, results from our study show that OP9 cells are readily transfectable, achieving over 80% efficiency with Optifect reagent. The potential reason for this difference is the ability of pre-confluent OP9 cells to differentiate, allowing for a target population of cycling pre-adipocyte cells. In addition to differences in transfection efficiency, 3T3-L1 cells require a minimum of 4 days post-confluency to begin to accumulate lipid droplets [271]. Therefore only early-stage adipogenic genes can be assayed using transient RNAi transfection, as the effect lasts roughly 96 hours. In this study we assayed the effect of Gapdh knock-down 72 hours after transfection, and find persistent, strong knockdown. Consequently, the OP9 cell model of adipogenesis has both the transfectability and differentiation time-scale needed to assay the effects of gene knock-down on mid to late adipogenesis. Chung et al. published a method for automated genome-scale siRNA screening [272]. Using their method, Chung et al. can screen eighty 384 well plates (in triplicate) in 8 hours time. Their study focused on mitomycin C sensitivity; however, it is feasible to use this system to screen their library of 22,108 siRNAs for genes that inhibit or accelerate adipogenesis. This would be the first mammalian system-wide RNAi screen of mid to late stage adipogenesis.

A screen for essential adipogenic genes is important to discover novel therapeutic targets for obesity. Hyperplasia and hypertrophy of adipose cells leads to obese adipose tissue [162, 273]. Obese adipose tissue secretes cytokines that greatly increase the risk of developing cardiovascular disease [162, 273-275]. Due to the central role of adipocyte hyperplasia in the pathogenesis of obesity, discovering genes which inhibit or induce adipogenesis provides important drug targets. Suwa et al, discusses the insufficiency of

current drug therapies for obesity, and suggests focusing on targets involved in adipocyte hyperplasia.

Using our system, we demonstrate SPI1 mRNA depletion increases lipid accumulation in adipogenesis. This is consistent with previous reports that overexpression of SPI1 inhibits adipogenesis [276]. As well SPI1 colocalizes with PPARG at binding sites of macrophage specific genes [277], indicating a repressive role for SPI1 in adipogenesis. This result also validates our best candidate TF from our computational screen (see Chapter 2). SPI1 binding sites cluster in many of the gene networks identified in our fat storage dataset. Here we demonstrate the validity of using OP9 cell adipogenesis to test for functional consequences of gene knockdown, as well as the validity of our computational screen.

Conclusion

This is the first time the transcriptome of OP9 preadipocyte cells has been investigated. Much is already known about the process of adipogenesis from 3T3-L1 cell models. However, we have shown here that OP9 cells behave similarly, but differentiate more rapidly. This allows transient knock-down of genes of interest in OP9 cells, and we are in the process of making that a high-throughput process. Additionally we validate our strongest candidate TF from our computational screen, and we will use the high-throughput approach to validate the other TFs of interest from our computational screen.

Methods

RNAi knockdown: Cells were plated 6 well plates at 3,000 cells/cm2. 24 hours later cells were transfected according to the Optifect protocol. In brief, 125uL of optimem media was mixed with 30nM of target RNAi. Additionally, 125uL optimem media was incubated with 10uL of optifect transfection reagent. Both mixes were incubated for 5 minutes and then combined. Mixture was then incubated at room temp for 25 minutes. In the mean time OP9 cells were washed 1x with 1ml HBSS. Wash was aspirated and 1mL of antibiotic free propagation media was added. Finally, the optimem, optifect, RNAi mix was added to the cells and incubated overnight.

Propagation of OP9 cells: OP9 cells were grown in propagation media, MEM-alpha with 20% premium fetal bovine serum, .02mM L-glutamine, and 500 Units of penicillin/strep. Cells were split at 80% confluence by rinsing once with Hank's balanced salt solution, incubating for with Trypsin-EDTA for 3 minutes, and then adding propagation media. Cells were split between 1:2 and 1:4. Cells were frozen in 20% FBS, 5% DMSO, and 75% Prop Media

Differentiation of OP9 cells: Cells were differentiated using Insulin Oleate media. Preparation of insulin oleate media is described by Wolins et al. in detail [167]. Briefly, cells were plated at 12,500 cell/cm2 in 96 well plates. 24 hours later propagation media was replaced with insulin oleate media (MEM-α, 0.2% FBS, 175 nM insulin, 900 μM oleate:albumin, and 500 U penn/strep.

Triglyceride Assay: Triglyceride assay was performed according to the manufacturer's protocol. The Biovision TG assay (Cat #K622-100) was used in parallel with the BCA protein assay for standardization. Cells were lysed in 100uL of a solution of 5% Triton-x in water. The solution was pipetted up and down twice to assure total lysis. The plate was then covered and incubated in an 80degree water bath for 5 minutes. Plates were cooled to room temperature and then heated in the water bath again. 10uL of cell lysis suspension was used in the TG assay.

Microscopy: We ordered DAPI from Sigma : Cat# D8417 and black costar plates cat# 3603. Dilute DAPI in water to 700uM , store at -20. Fix cells for 20 minutes at room temperature in 4% Paraformaldehyde (PFA). Pipette gently as differentiated adipocytes are easy to dislodge. Remove PFA and replace with 100uL of 2uL/mL of stock DAPI in 1:1 glycerol:pbs. Incubate for 5 minutes at RT. Remove the DAPI stain and add 100uL of a 1:10000 Nile Red solution in PBS (prepared from 1mg/ml stock). Incubate for a minimum of 25 minutes at RT. Cells can be imaged in the nile red solution. Plates are imaged in the ImageExpress robotic microscope in the SCIID facility. The images were analyzed using the journal program located in appendix 1.

Quantitative PCR: cDNA was generated using the Applied Biosystems TaqMan Reverse Transcription Reagents (P/N N808-0234) according to the manufacturers protocol. Quantitative PCR reactions were performed in 10uL in 384 well plates using 2x sybr green, 50nM forward primer, 50nM reverse primer, 5ng cDNA, and 3uL of water.

Reactions were cycled at 95 degrees for 10min, followed by 40 cycles of 95 degrees for 15 seconds and 60 degrees for 1min. Data were analyzed using relative quantification.

RNA extraction: RNA was extracted using the Ambion Ribopure kit and 15mL heavy phase lock gel from 5 Prime. The Ribopure protocol was followed with the exception of addition of the phenol extraction step to phase lock gel. The organic and inorganic phases were separated on phase lock gel, and a barrier was formed between the two. The organic phase was collected and the Ambion protocol was followed.

Generation of clonal cell lines: OP9 parental cells were trypsinized and resuspended at .5 cells/100uL of propagation media. 100uL of cell suspension was plated in each well of a 96 well plate. Plates were checked every 48 hours for the appearance of visible colonies.

Microarray analysis: Total RNA was extracted as described above and hybridized to Affymetrix Mouse Gene 1.0 ST Arrays. Data were analyzed using Expression Console from Affymetrix and the R Limma package (R code shown in appendix 2). Microarray data was normalized using the Robust Multichip Average (RMA). Data was filtered at a cutoff of 3.5, the background threshold for the mouse ST 1.0 array. Fold change data, p-values, and FDR adjusted p-values were generated by the R limma function "TopTable". Gene ontology overrepresentation was calculated using the g:Profiler (http://biit.cs.ut.ee/gprofiler/). Microarray probes were sorted by fold change for each time point and the 250 genes with the largest positive fold change, and the 250 genes

with the largest negative fold change were selected for analysis. G:Profiler was run using default parameters, organism (*mus musculus)*, and probe IDs as input.

Chapter 4: Association of variation in SPI1 with obesity phenotypes

Introduction

Obesity is largely a preventable disease; however, current preventive public health recommendations are population-based and may be ineffective for a given individual. Phenotypic responses to dietary modifications present dramatic interindividual variability [61, 278-281]. There is strong evidence suggesting that this variability has a significant genetic component. There are numerous examples of TF SNPs that are associated with obesity [282, 283]. Additionally, there are many examples of gene-diet interactions in obesity [284-288]. We seek to investigate variation in the transcription factor SPI1 for association with obesity, and gene-diet interactions.

The transcription factor SPI1, spleen focus forming virus (SFFV) proviral integration oncogene, was chosen for genotyping for a variety of reasons. SPI1 has a well studied role in myeloid and B cell development [289-291]. Through studies in hematopoietic cells, SPI1 has been shown to interact with gene products that play a key role in adipogenesis. C/Ebpa is regulated by SPI1 in mouse erythroleukemia (MEL) cells [290]. In hepatocytic cells, SPI1 cooperates with C/Ebpb to induce transcription of MD-2 [292]. Additionally the SPI1 protein interacts with GATA-2 both in-vitro and in-vivo [252]. Lefterova et al. demonstrates co-localization of SPI1, C/Ebpb, and PPARG binding sites in macrophage cells [277].

Wang et al. hypothesized that SPI1 also plays a role in adipogenesis. Their results show that SPI1 is expressed in white adipose tissue, and increased in adipose tissue of

obese mice. Adipogenesis is inhibited by forced over expression of SPI1 [276]. From the Framingham Heart Study (FHS) GWAS, there are 5 high quality signals within 100kb of SPI1, all for mean HDL cholesterol (Ordovas Lab data). The quantitative trait locus HYPLIP2 (for combined hyperlipidemia and increased atherosclerosis) spans the area containing SPI1. As described in chapter two of this thesis, our bioinformatic screen for TFs involved in obesity identified SPI1 as a top candidate. Evidence points to a potential role of the transcription factor SPI1 in obesity. Here we investigate the impact of variation in SPI1 on obesity phenotypes.

SPI1 interacts with several genes that have been shown to respond to N-3 polyunsaturated fatty acids (PUFAs) and saturated fatty acids [293-295]. Therefore we also specifically looked for interactions between SPI1 SNPs and fatty acids (FA).

RESULTS

SPI1 is located at 11p11.2 in the genome and is 23,719 bp long with 5 exons. There are two reported splice variants of SPI1 in the GenBank database. Based on the size and haplotype structure of SPI1 as shown in HAPMAP (Figure 22) we need four SNPs for adequate coverage. SNPs were selected based on the following criteria: minor allele frequency≥ 10%, function as a tagging SNP in HAPMAP, and potential molecular function. SPI1 has three known coding SNPs. However, all three have an unknown allele frequency, making them unsuitable for this study. Using the ABI SNP browser we selected four candidate SNPs (Table 10). To assess potential molecular function of each SNP, we employed the FASTSNP function analysis tool. FASTSNP is a webserver that compiles information from a variety of tools to predict the functional impact of a SNP.

Using FASTSNP, we predict the T allele of rs10769258 abolishes a binding site for the TF AP-1.

Gene	rsID	Lab SNP ID	Tagging?	Position in gene	Position from TSS of NM_0031020.2	Major	Minor	CEU Minor	Function
SPI1	rs10769258	I9089	Tagging SNP	Intron 2	9089	T	C	0.38	T allele abolishes potential AP-1 TFBS
SPI1	rs3740689	I19535	Tagging SNP	Intron 3	19535	A	G	0.3	
SPI1	rs4752829	I3474	Tagging SNP	Intron 2	3474	G	A	0.2	
SPI1	rs896817	I5823	Tagging SNP	Intron 2	5823	C	T	0.42	

Table 10 Characteristics of SPI1 variants selected for genotyping.

The selected SNPs were genotyped in 854 subjects from the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) Study. The GOLDN cohort participants were recruited from an ongoing study in Utah and Minnesota, and represent individuals of European ancestry. The GOLDN study is described in greater detail by Lai CQ [296]. Genotypic and phenotypic characteristics of the subjects at baseline are shown in Table 11.

	Men (n=421)	Women (n=433)	
	mean±sd	mean±sd	P-value
Diastolic Blood Pressure, *mm Hg*	70.84 ± 9.21	65.74 ± 9.11	*<0.001*
Systolic Blood Pressure, *mm Hg*	118.50 ± 15.10	112.89 ± 17.76	*<0.001*
Total Cholesterol, *mg/dL*	180.25 ± 34.86	188.89 ± 36.49	*<0.001*
HDL Cholesterol, *mg/dL*	42.98± 10.56	53.66 ± 14.19	0.4391
LDL Cholesterol, *mg/dL*	116.29 ± 29.47	118.25 ± 30.43	0.3375
Physical Activity Score	35.12 ± 7.19	33.25 ± 5.16	*0.01214*
Triglycerides, *mg/dL*	142.94 ± 97.21	121.57 ± 76.07	0.928
BMI, *kg/m^2*	28.44 ± 4.72	28.03 ± 6.10	0.0712
Wasit/Hip Ratio	0.948 ± 0.08	0.85 ± 0.08	*<0.001*
Waist, *cm*	100.51 ± 14.26	92.82 ± 16.69	*<0.001*
waist>102cm in men >88 in women, n (%)	170 (40.4%)	243 (56.1%)	*<0.001*
Current Drinker, n (%)	194 (46.1%)	206 (47.6%)	0.7122
Current Smoker, n(%)	32 (7.6%)	26 (6%)	*0.01253*
rs10769258, n (%)			0.6473
T/T	177 (42.1%)	190 (44.2%)	
T/C	189 (45%)	193 (44.9%)	
C/C	54 (12.9%)	47 (10.9%)	
rs896817, n (%)			0.2304
C/C	185 (44.2%)	165 (38.4%)	
C/T	181 (43.2%)	204 (47.4%)	
T/T	53 (12.6%)	61 (14.2%)	
rs3740689, n (%)			0.2012
A/A	139 (33.8%)	164 (38.6%)	
A/G	207 (50.4%)	188 (44.2%)	
G/G	65 (15.8%)	73 (17.2%)	
rs4752829, n (%)			0.421
G/G	204 (48.8%)	230 (53.1%)	
G/A	176 (42.1%)	164 (37.9%)	
A/A	38 (9.1%)	39 (9%)	

Table 11 Baseline characteristics of the GOLDN population.

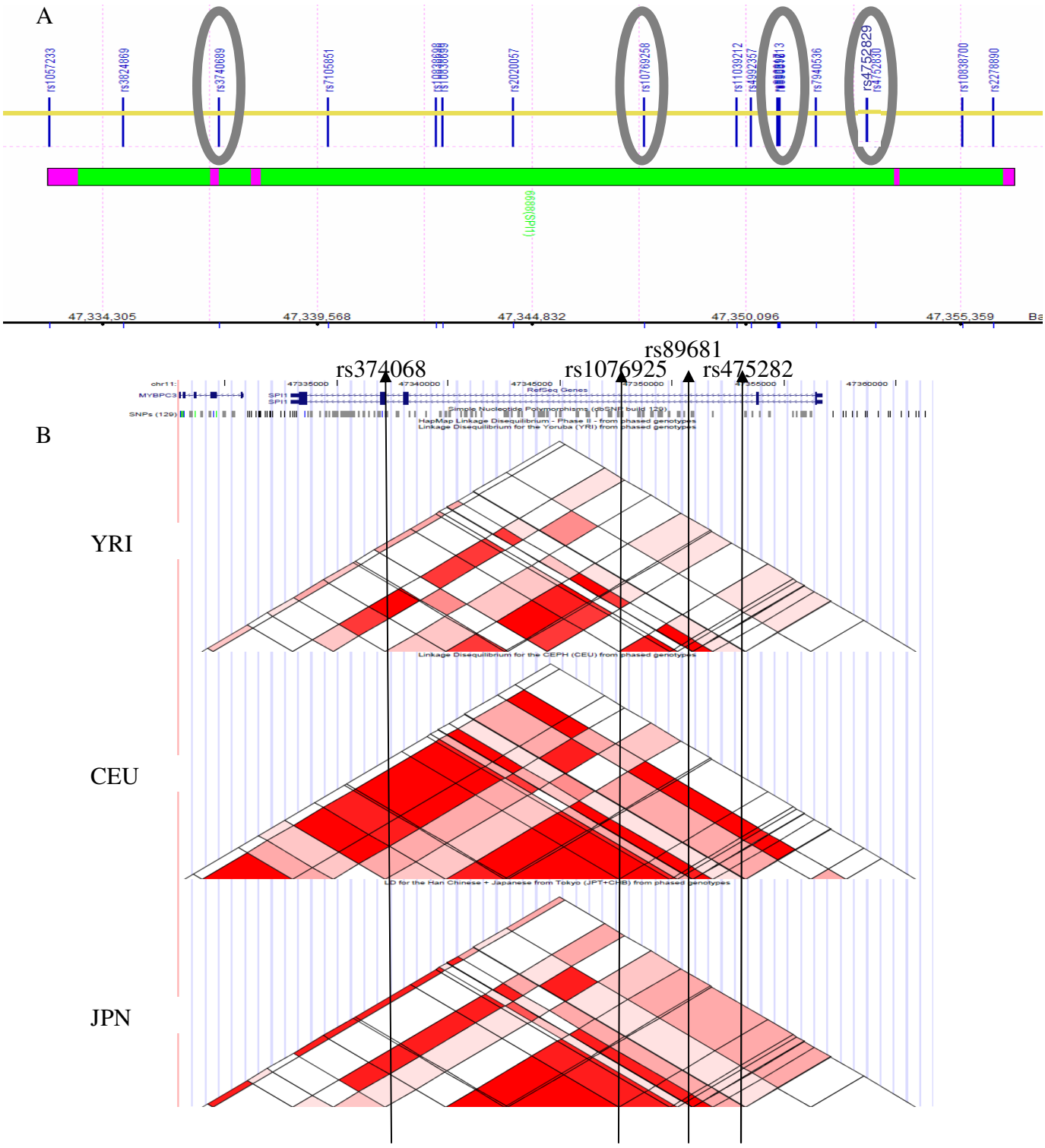


Figure 22 SPI1 gene and variation structure. SPI1 is located on the reverse strand of chromosome 11. **A.** SPI1 has five exons (in pink) and 4 introns (in green). The locations of rs3740689 (intron 2), rs10769258 (intron 3), rs3740689 (intron 3), and rs4752829 (intron 3) are shown. **B.** Linkage structure of SPI1 is shown in all three HapMap populations. Color scale represents R^2 values, the deeper the red the higher the R^2.

To test for an association between variation in SPI1 and obesity related phenotypes, we genotyped our SNPs in the GOLDN population. SPI1 genotype frequencies do not deviate from Hardy-Weinberg equilibrium (Table 12). Nor do genotype frequencies different by gender, therefore we will analyze men and women together (Table 11). The minor allele frequencies for rs10769258, rs896817, rs3740689, and rs4752829 are 0.397, 0.344, 0.282, and 0.369 respectively. These four variants are not in linkage disequilibrium with one another (based on r^2) (Figure 23). All statistical analyses were performed using R 2.7.2.

	Expected	Obs-Exp	HWE Chi^2	P-value
rs10769258	416.047	3.953	0.238	0.621
	548.905	-7.905		
	181.047	3.953		
rs3740689	91.32	6.68	0.949	0.304
	464.361	-13.361		
	590.32	6.68		
rs4752829	456.134	4.866	0.381	0.568
	533.733	-9.733		
	156.134	4.866		
rs896817	135.459	-9.459	1.534	0.239
	517.082	18.918		
	493.459	-9.459		

Table 12 Chi Squared test of Hardy-Weinberg equilibrium for all four genotypes.

We created a regression model to test if SPI1 variance conveys an increased obesity risk. We tested for association with BMI, waist circumference, and waist/hip ratio (W/H). Potential confounding factors included age, gender, smoking status, alcohol, physical activity, triglycerides, systolic and diastolic blood pressure, diabetes, HOMA, LDL and HDL cholesterol. Minor allele carriers of rs4752829 have significantly higher BMI than major allele homozygotes.

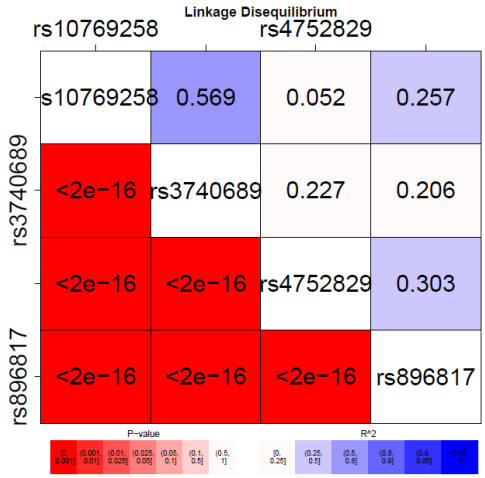


Figure 23 Linkage Disequilibrium between the four genotyped SNPs in SPI1. Shades of blue represent R^2 and shades of red represent p-value.

SNP rs4752829 showed significant association with BMI in a multivariate adjusted model (Table 13 and Table 14). Minor allele homozygotes exhibit increased BMI compared to carriers of the G allele. Using the leaps package to perform all subsets analysis the model was adjusted for age, gender, alcohol, triglycerides, systolic blood pressure, HOMA, and LDL and HDL cholesterol. Examination of the model showed residuals evenly distributed around 0 (Figure 24). There appear to be a handful of points with considerable leverage on the model (Figure 24). Figure 24 shows leverage of each data point on the model. Data points 288, 550, 912, and 977 have a large Cook's distance (Figure 24). The same three points are shown to have a large amount of leverage on the

model according to the leverage plot (Figure 24). Based on the large residuals and high leverage of data points 288, 550, 912, and 977 we decided to remove them from the model, in order to create a model that fits the data as opposed to a few outliers. Figure 24 shows the fit of the model after removal of points 288, 550, 912, and 977. Using ANOVA we see that the regression model including SNP rs4752829 explains the data significantly better than a model without (p<-.001). No association was found between the other SNPs tested and obesity related phenotypes.

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.500658	0.213908	11.69	<2.00E-16	***
AGE	0.001201	0.000349	3.442	0.000606	***
SEXfemale	0.030016	0.011219	2.675	0.00761	**
Current drinker	0.024374	0.010485	2.325	0.020331	*
logTRR	0.018749	0.011192	1.675	0.094259	.
logSYSBP	0.167346	0.040754	4.106	4.42E-05	***
logHOMA	0.145901	0.010512	13.879	<2.00E-16	***
LDLCV0	0.000682	0.000177	3.854	0.000125	***
logHDL	-0.09279	0.023961	-3.873	0.000116	***
rs4752829G/A	-0.04147	0.018855	-2.199	0.028124	*
rs4752829G/G	-0.03074	0.018445	-1.667	0.09594	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Adjusted R-squared: 0.3918

Table 13 Regression model for rs4752829 with logBMI as dependent.

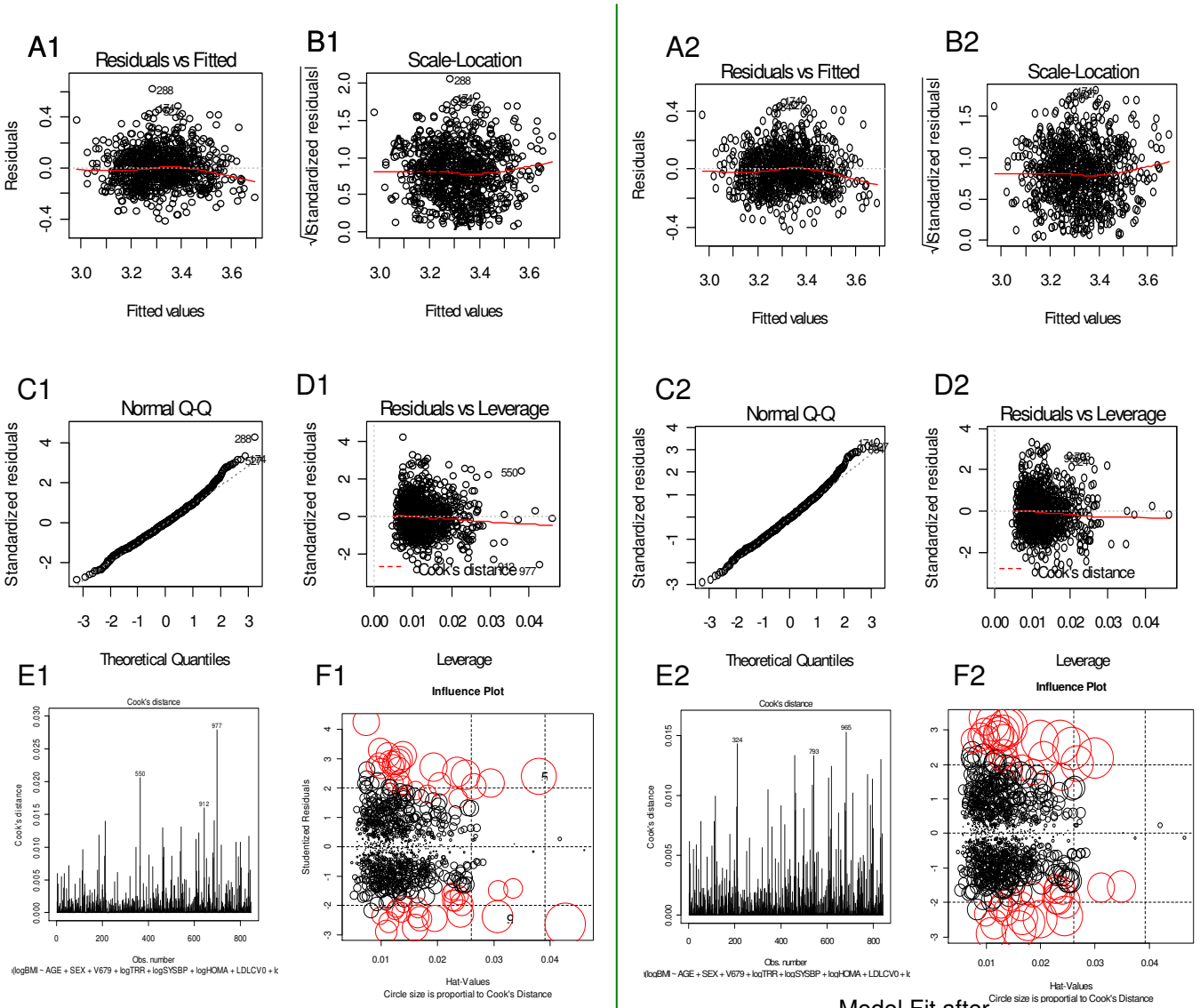


Figure 24 The fit of the regression model for SNP rs4752829 is improved after outlier removal. Diagnostic models shown before (left) and after (right) removal of outliers (observations 288, 550, 912, and 977)**. A.** Residuals vs. Fit plot. Shows the residual of each point plotted against the predicted value. **B.** Scale location plot. Shows the square root of each residual plotted vs. predicted value. The square root of the residual diminishes skewness. **C.** Normal Q-Q plot shows the fit of the data vs. the normal distribution. **D.** Residuals vs. Leverage plot. Plot of the residuals of the fitted model vs. the Cook's distance, which checks for extreme x-values. **E.** Plot of the Cook's distance for each observation. Cook's distance is a measure of influence. **F.** Influence Plot. Standardized residuals of the model are plotted vs. the hat values, a common measure of extreme x-values. The size of the circle is proportional to the Cook's value.

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.415862	0.214035	11.287	<2E-16	***
AGE	0.001288	0.000343	3.755	0.000185	***
SEXfemale	0.028345	0.011018	2.573	0.010268	*
Current drinker	0.019773	0.010315	1.917	0.0556	.
logTRR	0.017481	0.011003	1.589	0.112489	
logSYSBP	0.17972	0.040614	4.425	1.09E-05	***
logHOMA	0.14777	0.010433	14.164	<2E-16	***
LDLCV0	0.000638	0.000174	3.665	0.000263	***
logHDL	-0.08279	0.023672	-3.497	0.000495	***
rs4752829G/A	-0.04622	0.018601	-2.485	0.013146	*
rs4752829G/G	-0.03634	0.018185	-1.999	0.04598	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Adjusted R-squared: 0.4033

Table 14 Regression model for rs4752829 with logBMI as dependent after regression diagnostics removal of discrepant points.

To determine if SPI1 SNPs modulate the effect of diet on obesity phenotypes, we tested for interactions. The GOLDN dataset contains information on dietary FA intake, as well as membrane FA composition. Using saturated fat as a continuous variable, we find individuals with the rs3740689 GG genotype show a greater association between dietary saturated fat and BMI (B=-0.71, p for interaction=0.013). With an increase in dietary saturated fat we see a decrease in BMI for individuals with the GG genotype. We also assessed this interaction stratified by mean saturated fatty acid intake (< 24.33 g/d low and ≥24.33 g/d high) (Figure 26). We find a significant interaction between rs3740689 and saturated fatty acid as a categorical variable (G/A p=0.615 G/G p=0.004).

Individuals homozygous for the G allele in the low saturated FA strata have increased BMI compared to the high saturated FA strata.

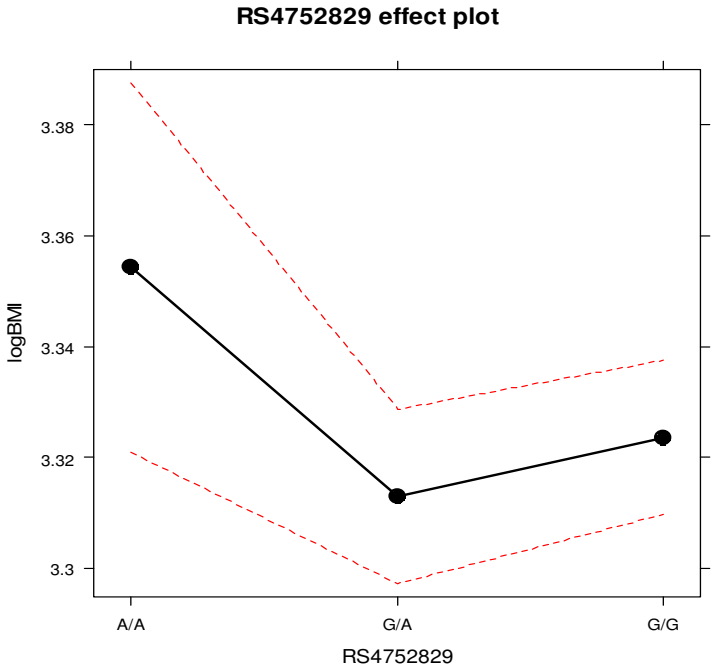


Figure 25 BMI varies by genotype of RS4752829**.** Adjusted means are shown for the estimated logBMI from the regression model. Black dots represent adjusted means and the red dotted line shows the confidence integral.

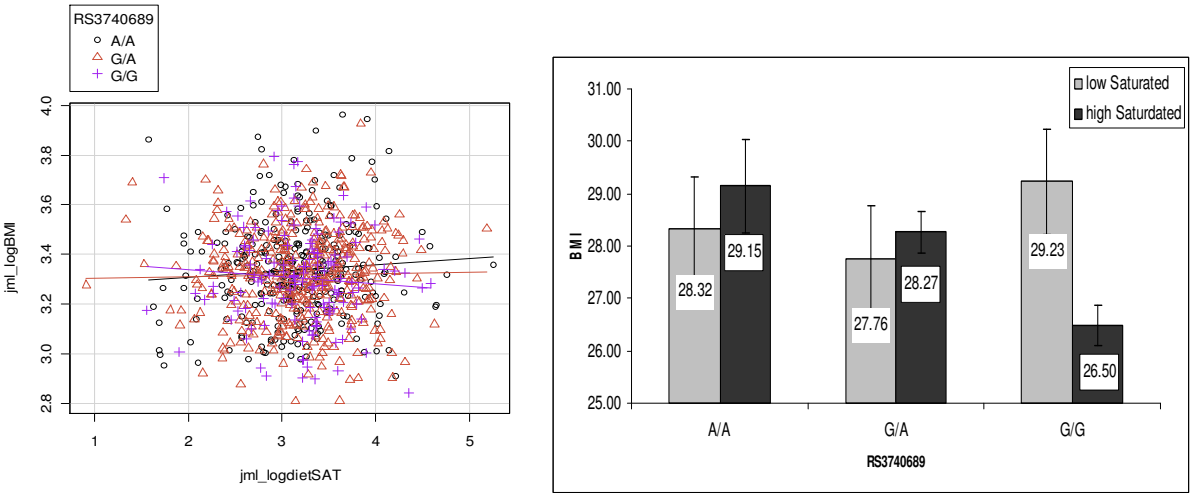


Figure 26 Interaction between SPI1 SNP rs3740689 and dietary saturated FA in relation to body mass index**. A.** Log values of BMI plotted vs. log values saturated FA intake by rs3740689 genotype. Line represents the linear regression line. Symbols are as follows: ○ = A/A △= G/A + = G/G. **B.** Means of BMI according to genotype and strata of dietary saturated FA intake (< 24.33 g/d low and ≥24.33 g/d high).

When we used erythrocyte membrane N3-PUFA as a continuous variable we see an interaction with SPI1 SNP rs3740689. Individuals homozygous for the G allele show a greater association between erythrocyte membrane N3PUFA and BMI (B=0.186, p for interaction=0.0026). With an increase in erythrocyte membrane N3-PUFA we see an increase in BMI for individuals with the GG genotype. When we test the interaction using erythrocyte membrane PUFA as a stratified variable, we find a significant interaction (<3.55% low and ≥ 3.55% high) (G/A p=0.041 G/G p=0.066). Those with a G/A genotype exhibit a lower BMI with high erythrocyte membrane N3 content, and those with a G/G genotype exhibit a lower BMI with low erythrocyte membrane N3 content (Figure 27).

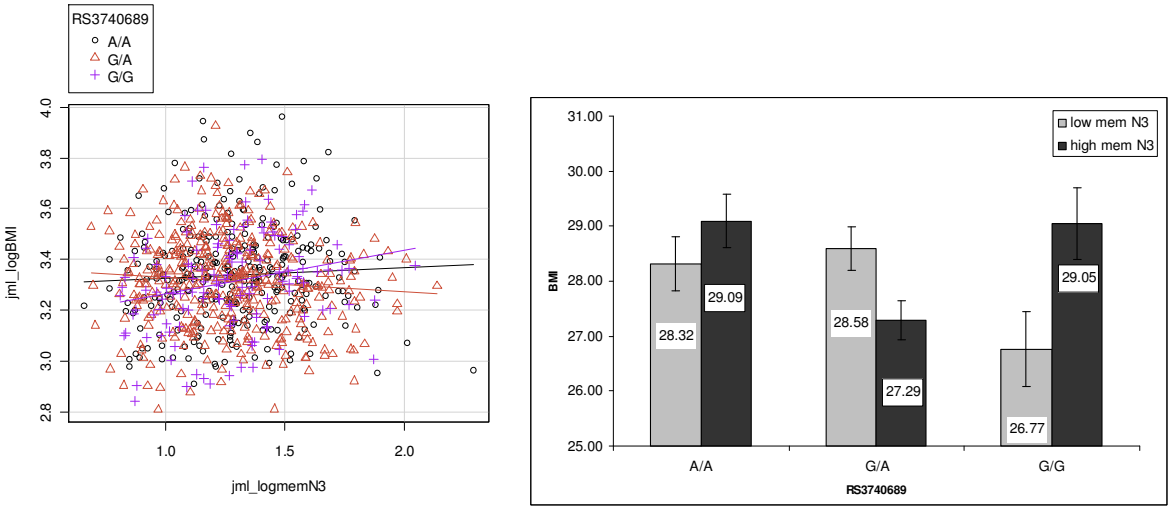


Figure 27 Interaction between SPI1 SNP rs3740689 and erythrocyte fatty acid composition in relation to body mass index. **A.** Log values of BMI plotted vs. log values of erythrocyte membrane N3-PUFA by rs3740689 genotype. Line represents the linear regression line. Symbols are as follows: ○ = A/A △= G/A + = G/G. **B.** Means of BMI according to genotype and strata of membrane N3-PUFA content (<3.55% low and ≥ 3.55% high).

BMI				
		AA	GA	GG
MAIN EFFECTS				
rs4752829	adj mean	28.62516	27.46235	27.75852
	lower CI	27.6842	27.03269	27.37606
	upper CI	29.59807	27.89884	28.1463
	p-value	n/a	0.028124*	0.09594
INTERACTION				
rs3740689:dietary SAT FA	p-value	n/a	0.990568	0.013488*
rs3740689:erythrocyte N3 FA	p-value	n/a	0.072602	0.002589*

Table 15 Summary of association data.

Discussion

We hypothesize that variation in SPI1 is associated with obesity. SPI1 expression is increased in the white adipose tissue of obese mice [276, 297]. In addition, overexpression of SPI1 inhibits adipogenesis [276]. For the first time, we demonstrate the influence of a common genetic variant in SPI1 on body mass index in individuals of European descent. Carriers of SNP rs4752829 have increased BMI compared to major allele homozygotes. This effect is driven by SNP rs4752829 or a causal variant in LD.

We believe the association seen between BMI and SNP rs4752829 is due to a functional SNP in the 3'UTR. Hikami et al. recently demonstrated linkage between rs4752829 in intron 2 and rs1057233 in the 3'UTR of SPI1 [298]. SNP rs1057233 abolishes the binding site for miR569, leading to a release of repression of SPI1 mRNA [298]. This is consistent with abdominal obesity QTL data centered on 3q27, the chromosomal location of miR569 [299]. Taken together these studies indicate the association between rs4752829 and BMI is most likely due to the causal variant rs1057233 in LD.

To strengthen these findings, we need to validate this association in another population of European descent. Based on availability of the samples, we suggest the Framingham Offspring Study (FOS) [300]. In addition, we can genotype SPI1 in the Boston-Puerto Rican Centers on Population Health and Health Disparities Study, to access the impact of variation in SPI1 in a population with admixture [301].

This study is also the first to show modulation of the effect of saturated FA and N3-PUFA on BMI by variation in SPI1. Our results indicate that minor allele

homozygotes for SNP rs3740689 have lower BMI than major allele carriers in individuals with a diet low in saturated fats. We also found modulation of the effect of N3 PUFAs in BMI by rs3740689. Minor allele homozygotes have a lower BMI than major allele carrier in individuals with a diet low in N3 PUFA. Additionally individuals heterozygous at rs3740689 with a diet high in N3 PUFA have lower BMI than individuals with a diet low in N3 PUFA. This interaction demonstrates a pattern known as heterozygote advantage, where heterozygotes survive better than homozygotes under specific selective pressure [302]. In this case rs3740689 heterozygotes maintain a lower BMI under conditions of high N3 PUFA.

It has previously been shown that SPI1 is responsive to dietary components. Treatment of cells with metabolites of vitamin A (retinoic acid) increases expression of SPI1 [303]. Our data suggest that SPI1 is influenced by dietary FAs. In addition the network of genes shown to interact with SPI1 is also influenced by FAs. This network includes inducible cAMP early repressor (ICER) [304]. Treatment of cells with the saturated fatty acid, palmitate, leads to overexpression of ICER [293]. Saturated fat induced overexpression of ICER leads to beta-cell secretory dysfunction [293]. We therefore predict the genetic interaction between rs3740689 and saturated FA to function through the relationship of SPI1 with ICER. SPI1 interacts with GATA2 [252], and in leukocytes GATA2 is upregulated by fish oil (N3 PUFA) supplementation [295]. It seems likely that N3-PUFA could impact SPI1 through interaction with GATA2.

These interactions between genetics and dietary fat can be used to change the behavior in high risk populations. Arkadianos et al. show greater long-term BMI reduction in a group of individuals given a nutrigenetically tailored diet compared to a

control group [305]. Reinehr et al. demonstrate successful weight loss after intervention only in children without variations in the MC4R gene [306]. These studies indicate that variants change the effect of behaviors. Therefore, understanding how genetics impacts behavioral factors is important for risk reduction.

In addition to providing novel information for risk reduction, our study adds to the body of knowledge linking SPI1 to obesity. We hypothesize that SPI1 is involved in the balance between hematopoiesis and adipogenesis. In bone marrow cells there is a well-documented reciprocal role between adipocytes and hematopoiesis [307-309]. Under conditions of stress or aging adipocytes expand in bone marrow and hematopoiesis is drastically reduced [308, 309]. SPI1 also has a well studied role as a master regulator for the development of myeloid, lymphoid, and blood stem cells [310-312]. Proper spatial and temporal expression of SPI1 is critical for differentiation of various hematopoietic progenitors [312].

It would appear that SPI1 also plays a role in inhibition of adipogenesis. In macrophage cells, Lefterova et al. demonstrated colocalization of SPI1 with PPARG at sites of active gene transcription [277]. In the same experiment SPI1 no longer colocalizes with PPARG in adipocyte cells [277], suggesting that the presence of SPI1 is repressive for adipogenesis. We have shown variation in SPI1 associates with obesity. Potentially this functions through modulation of expression levels of SPI1, and a disruption in the balance between hematopoiesis and adipogenesis, leading to adipocyte hypertrophy.

Conclusion

This study is the first to demonstrate an impact of variation in SPI1 on the individual risk of developing obesity. If these finding are replicated nutritional FA intake recommendations based on SPI1 genotype can be possible. In addition, these studies indicate necessary research into the functional consequences of variation on anti-sense RNA regulation of SPI1 mRNA translation.

Methods

Study design: SNPs were genotyped in subjects from the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) Study. The mainly Caucasian participants were recruited from an ongoing study in Utah and Minnesota. The study has 1,083 subjects (568 women and 515 men) age 17-92.

Fatty Acid Dietary Information: In the GOLDN study the Diet History Questionnaire was used to measure diet.

Genotyping: DNA was extracted using a QIAamp DNA Blood Maxi Kit. SPI1 SNPs were genotyped using Taqman pre-designed probes and the ABI 7900HT. Blank wells were includes for quality control.

Data Analysis: All statistical analysis were performed using R. Continuous variables were checked for normalcy and appropriate transformations were applied. Associations

between genotype and phenotypes were tested using multiple linear regression models in R. To test for effects of genotype on diet an interaction model was used. Fatty acid was used continuous variable or split into a dichotomous variable using the median.

Chapter 5: Discussion and Conclusions

Discussion

The overall goal of this work was to investigate the genetic factors involved in body weight. In this project we have considered three different approaches to identifying novel TFs regulating body weight. We used a computational method to tease apart the regulatory network of genes affecting fat storage in the worm. Using our methodology we identified 58 TFs controlling networks of genes involved in fat storage, 30 of which are novel. Our study adds to the body of research successfully analyzing genome scale data to understand regulatory networks [241-243]. Additionally our computational process is useful for informing GWAS. Several groups have reported increased GWAS success by limiting the variation investigated, and consequentially lowering the cost and false positive rates [244-248]. Our method identified a set of candidate TFs that can be used to direct a GWAS study.

In order for our computational approach to remain relevant, new tools for ortholog identification, TFBS searching, and GO descriptions should be incorporated as they become available. Many of the limitations from a computational approach come from an incomplete understanding of the biology being modeled. It is essential to take advantage of each new piece of information to eliminate false positives and negatives.

We also established a cell culture model of adipogenesis, OP9 cells, for the purpose of screening for critical adipogenic genes. We show highly efficient and reproducible differentiation over 72 hours in our clonal population of OP9 cells. By

analyzing the transcriptome of our clonal population we show OP9 differentiation is comparable to other well-studied models of adipogenesis [271]. According to our results OP9 cells are readily transfectable. We also established an automated microscopy process to quantify the level of lipid in differentiating cells. This allowed us to screen for the effects of gene knock-down on OP9 adipogenesis. We used this system to validate the top candidate TF from our computational analysis, SPI1. Gene knock-down of SPI1 increased lipid content in differentiating cells. This is consistent with reports from Wang et al. where forced overexpression of SPI1 inhibits adipogenesis [276]. We can use the OP9 assay to assess the role of all our candidate TFs in adipogenesis. Of note, this study also lays the groundwork for a knockdown screen of genes involved in adipogenesis that can increase our knowledge of adipose cell biology and provide novel drug targets.

Performing an association study of variation in SPI1 with obesity we discovered variation in SPI1 associated with BMI. Additionally we detected variation which modulated the effect of dietary FAs on BMI. These associations must be replicated in another population. As well, our approach can be improved by including more samples, thus increasing our power to detect small effects. Of interest would be genotyping SPI1 in a population of African descent. The haplotype block structure between European and Caucasian populations is drastically different. If these results validate, it will be interesting to include variation in SPI1 in panels of SNPs used to provide dietary recommendations.

Personalized dietary recommendations can be used to change the behavior of high risk populations. To date very few studies have demonstrated success using genetic information to inform diet [305]. A review conducted by Janssens et al. concluded there

is insufficient evidence for assessing disease risk or creating personalized diets using current commercially available genetic tests [313]. The results of this work can help lead to personalized nutrition by increasing our knowledge of genes with dietary interactions and identifying future candidates for genotyping and directed GWAS studies.

Conclusion

We have created a novel method that identified 58 TFs involved in obesity, and created an approach that can be generalized to identify TFs involved in a diverse array of diseases and disorders. As well we described the transcriptome of OP9 adipogenesis, and set the stage for a high-throughput gene knockdown screen. Lastly, we identified specific variation in SPI1 associated with BMI that can assist in assessing obesity risk for subgroups of the population. In total this work has increased our knowledge of the transcriptional cascades involved in obesity and created computational and cell based models for future discovery.

Supplementary tables

Supplementary Table 1 Full list of human homologs identified for worm fat storage genes. Bl=blast Hg=HomoloGene Wb=wormbase

database found	Sequence Name (Gene)	hs Symbol	**BLAST tophit Evalue**
bl	T28D6.2	K-ALPHA-1	0
bl	T28D6.2	TUBA3	0
bl	Y65B4B_10.a	WWP1	0
bl	Y65B4B_10.e	WWP1	0
bl	Y65B4B_10.a	WWP2	0
bl	Y65B4B_10.e	WWP2	0
bl	Y4C6A.h	GRM3	9.00E-166
bl	Y4C6A.h	GRM2	2.00E-161
bl	F28F8.2	FLJ20920	1.00E-124
bl	T12A2.1	MGC35366	2.00E-116
bl	Y51H7C_255.	COG4	8.00E-108
bl	F28H6.3	KAT3	4.00E-99
bl	C14A4.1	HLRC1	4.00E-90
bl	Y49F6B.n	IK	2.00E-84
bl	Y37A1A.1	SETMAR	3.00E-81
bl	F38A5.1	FLJ11200	3.00E-79
bl	B0513.7	SETMAR	5.00E-79
bl	T02G5.4	ACAT1	1.00E-76
bl	C24A11.9	TPRT	6.00E-73
bl	M01B12.5	RIOK3	4.00E-67
bl	C07A9.2	G10	9.00E-64
bl	R01H2.3	LRP1B	2.00E-59
bl	T19B4.7	NOPE	6.00E-58
bl	W04A4.5	INT4	3.00E-56
bl	C15B12.7	SLC30A1	6.00E-56
bl	ZK131.8	HIST1H4C	2.00E-53
bl	H25K10.1	FLJ16165	3.00E-51
bl	C47C12.3	ZIC1	6.00E-51
bl	K08F8.1	MAPKAPK5	7.00E-51
bl	T05C12.1	TTBK1	7.00E-51
bl	C47C12.3	ZIC2	1.00E-50
bl	Y105E8B.a	HPRT1	7.00E-50
bl	Y47G6A_245.	KIAA1109	7.00E-50
bl	C38C10.1	TACR1	7.00E-50
bl	C47C12.3	ZIC3	7.00E-50
hg	F52C12.2	C16orf42	
hg	F23H11.9	CRLS1	
hg	Y41E3.10	EEF1D	

hg	ZK909.3	HDDC3
hg	worm gene	hs_gene
hg	W04A4.5	INTS4
hg	C09G5.8	KIAA1005
hg	C05E11.1	KIAA1715
hg	F26H9.4	NMNAT2
hg	T21D12.3	PQBP1
hg	K02D7.5	RAG1AP1
hg	F25H8.1	RG9MTD2
hg	B0041.5	SLC35F5
hg	W02C12.3	TFEB
hg	D1007.5	TMEM39A
hg	C30F12.1	ZC3H12C
wb	C34G6.4	ABCB10
wb	C34G6.4	ABCB11
wb	C34G6.4	ABCB5
wb	C34G6.4	ABCB8
wb	T04B2.2	ABL1
wb	T04B2.2	ABL2
wb	K05F1.3	ACAD8
wb	K05F1.3	ACAD9
wb	K05F1.3	ACADSB
wb	K05F1.3	ACADVL
wb	T02G5.4	ACAT2
wb	C34G6.4	ACCN3
wb	T09B4.8	AGXT2L2
wb	R11A5.1	AP1B1
wb	F53H8.1	AP1M2
wb	R11A5.1	AP2B1
wb	F15A8.6	BCHE
wb	F38A5.1	C4orf20
wb	F28H6.3	CCBL1
wb	Y17G7B.15	CENTB1
wb	F15A8.6	CES1
wb	R11A5.1	CPEB1
wb	K07C6.4	CYP2A6
wb	K07C6.4	CYP2A7
wb	K09D9.2	CYP2B6
wb	K07C6.5	CYP2C18
wb	K09D9.2	CYP2C18
wb	K07C6.5	CYP2C19
wb	K09D9.2	CYP2C19
wb	K09D9.2	CYP2F1
wb	K07C6.4	CYP2U1
wb	C14A4.1	DOHH
wb	T14E8.3	DRD2
wb	T14E8.3	DRD3
wb	ZK757.3	EIF2C3

wb	ZK757.3	EIF2C4
wb	T04C10.2	EPN3
wb	C43H6.9	GRIA1
wb	C43H6.9	GRIK1
wb	E04F6.3	HSD17B4
wb	K05F1.3	IVD
wb	C32C4.1	KCNA4
wb	C32C4.1	KCNC1
wb	C32C4.1	KCNC4
wb	C32C4.1	KCNS1
wb	C32C4.1	KCNS2
wb	C32C4.1	KCNS3
wb	C32C4.1	KCNV1
wb	Y17G7B.15	KCTD11
wb	C06G3.2	KIF4A
wb	C06G3.2	KIF5A
wb	C06G3.2	KIF5C
wb	C47D12.7	KLHL17
wb	C47D12.7	KLHL18
wb	C47D12.7	KLHL20
wb	C47D12.7	KLHL5
wb	ZK6.7	LIPL1
wb	ZK6.7	LIPL3
wb	ZK6.7	LIPL4
wb	R01H2.3	LRP1
wb	R01H2.3	LRP2
wb	ZC504.4	MINK1
wb	F15A8.6	NLGN2
wb	F15A8.6	NLGN4Y
wb	F11C1.6	NR5A1
wb	F11C1.6	NR5A2
wb	T04C9.1	OPHN1
wb	F14H8.1	OSBP
wb	F14H8.1	OSBPL3
wb	F14H8.1	OSBPL6
wb	F14H8.1	OSBPL7
wb	C24A11.9	PDSS1
wb	T28D6.2	PEX26
wb	F39B1.1	PIK3C2G
wb	F39B1.1	PIK3CA
wb	F39B1.1	PIK3CB
wb	F39B1.1	PIK3CD
wb	F39B1.1	PIK3CG
wb	F45H7.4	PIM2
wb	C06A1.3	PPP1CB
wb	C06A1.3	PPP2CA
wb	C02F4.2	PPP2CB
wb	C06A1.3	PPP2CB

wb	C02F4.2	PPP3CB
wb	C02F4.2	PPP3CC
wb	C06A1.3	PPP4C
wb	C06A1.3	PPP6C
wb	T19B4.7	PRTG
wb	W03C9.3	RAB9A
wb	F28H6.3	RBMXL1
wb	T19B4.7	SDK1
wb	T19B4.7	SDK2
wb	M05B5.4	SLC12A4
wb	C13D9.7	SLC24A6
wb	F52H2.2	SLC7A10
wb	F52H2.2	SLC7A11
wb	F52H2.2	SLC7A13
wb	F23F1.6	SLC7A14
wb	F23F1.6	SLC7A3
wb	F52H2.2	SLC7A7
wb	F52H2.2	SLC7A8
wb	F52H2.2	SLC7A9
wb	C06A1.3	SPDYA
wb	F43H9.2	SPTLC2L
wb	T28D6.2	TUBA1
wb	T28D6.2	TUBA2
wb	T28D6.2	TUBA8
wb	W08D2.1	WNT2
wb	W08D2.1	WNT2B
wb	W08D2.1	WNT7A
wb	C30F12.1	ZC3H12A
wb, bl	C34G6.4	ABCB1
wb, bl	C34G6.4	ABCB4
wb, bl	C37H5.3	ABHD5
wb, bl	K05F1.3	ACADM
wb, bl	K05F1.3	ACADS
wb, bl	C36A4.9	ACSS1
wb, bl	T09B4.8	AGXT2L1
wb, bl	R11A5.1	AP3B1
wb, bl	R11A5.1	AP3B2
wb, bl	F53H8.1	AP3M1
wb, bl	F53H8.1	AP3M2
wb, bl	T04C9.1	ARHGAP10
wb, bl	T04A8.16	CAPN7
wb, bl	Y17G7B.15	CENTB2
wb, bl	Y17G7B.15	CENTB5
wb, bl	Y49A3A.1	CHPT1
wb, bl	B0286.4	CNOT2
wb, bl	K07C6.4	CYP2C8
wb, bl	K07C6.5	CYP2C8
wb, bl	K09D9.2	CYP2C8

wb, bl	K07C6.5	CYP2C9
wb, bl	K09D9.2	CYP2C9
wb, bl	T19B4.7	DCC
wb, bl	W01C9.4	DECR1
wb, bl	T08B2.7	EHHADH
wb, bl	ZK757.3	EIF2C1
wb, bl	ZK757.3	EIF2C2
wb, bl	T04C10.2	EPN1
wb, bl	T04C10.2	EPN2
wb, bl	T04B2.2	FER
wb, bl	T04B2.2	FES
wb, bl	F46G11.3	GAK
wb, bl	K10B3.7	GAPDH
wb, bl	K10B3.7	GAPDHS
wb, bl	C43H6.9	GRIK2
wb, bl	C43H6.9	GRIK3
wb, bl	H08M01.2	GRLF1
wb, bl	K10C3.6	HNF4A
wb, bl	K10C3.6	HNF4G
wb, bl	Y41E3.11	HNRPU
wb, bl	F40H3.5	HS3ST1
wb, bl	F40H3.5	HS3ST4
wb, bl	C32C4.1	KCNB1
wb, bl	C32C4.1	KCNB2
wb, bl	F20H11.2	KIAA0963
wb, bl	C06G3.2	KIF15
wb, bl	C06G3.2	KIF5B
wb, bl	C47D12.7	KLHL2
wb, bl	C47D12.7	KLHL3
wb, bl	M05B5.4	LCAT
wb, bl	F14D12.2	LIMS1
wb, bl	F14D12.2	LIMS2
wb, bl	ZK6.7	LIPA
wb, bl	ZK6.7	LIPF
wb, bl	ZC504.4	MAP4K4
wb, bl	K08F8.1	MAPKAPK2
wb, bl	K08F8.1	MAPKAPK3
wb, bl	C06E7.3	MAT1A
wb, bl	F20D12.2	MCM3AP
wb, bl	ZC302.1	MRE11A
wb, bl	T19B4.7	NEO1
wb, bl	F15A8.6	NLGN3
wb, bl	F15A8.6	NLGN4X
wb, bl	F14H8.1	OSBPL1A
wb, bl	F14H8.1	OSBPL2
wb, bl	C34C6.6	PEX5
wb, bl	C34C6.6	PEX5L
wb, bl	F39B1.1	PIK3C2B

wb, bl	F45H7.4	PIM1
wb, bl	F45H7.4	PIM3
wb, bl	Y53C12A.1	PKMYT1
wb, bl	F52B11.2	PMM1
wb, bl	F13D12.6	PPGB
wb, bl	C02F4.2	PPP1CA
wb, bl	C06A1.3	PPP1CA
wb, bl	C02F4.2	PPP1CC
wb, bl	C06A1.3	PPP1CC
wb, bl	C02F4.2	PPP3CA
wb, bl	ZK675.1	PTCH
wb, bl	ZK675.1	PTCH2
wb, bl	H27A22.1	QPCTL
wb, bl	F11A5.3	RAB2
wb, bl	F11A5.3	RAB2B
wb, bl	W03C9.3	RAB7
wb, bl	W03C9.3	RAB9B
wb, bl	C18H9.7	RAPSN
wb, bl	C03D6.3	RNGTT
wb, bl	F20H11.2	SBNO1
wb, bl	K04E7.2	SLC15A1
wb, bl	K04E7.2	SLC15A2
wb, bl	F20D1.9	SLC25A18
wb, bl	F23F1.6	SLC7A2
wb, bl	F23F1.6	SLC7A4
wb, bl	F52H2.2	SLC7A5
wb, bl	F52H2.2	SLC7A6
wb, bl	F43H9.2	SPTLC2
wb, bl	C54H2.5	SURF4
wb, bl	ZC504.4	TNIK
wb, bl	C24F3.1	TRAM2
wb, bl	C47D12.1	TRRAP
wb, bl	T28D6.2	TUBA6
wb, bl	C46F11.1	UNC93A
wb, bl	C37A5.1	VMD2
wb, bl	C37A5.1	VMD2L1
wb, bl	C37A5.1	VMD2L2
wb, bl	T21C9.2	VPS54
wb, bl	W08D2.1	WNT16
wb, bl	W08D2.1	WNT7B
wb, bl, hg	C37H5.3	ABHD4
wb, bl, hg	C36A4.9	ACSS2
wb, bl, hg	T09B4.8	AGXT2
wb, bl, hg	C56C10.10	AIP
wb, bl, hg	T04C9.1	ARHGAP26
wb, bl, hg	H08M01.2	ARHGAP5
wb, bl, hg	Y49A3A.1	CEPT1
wb, bl, hg	R07E5.1	GPATC1
wb, bl, hg	T08B2.7	HADHA

wb, bl, hg	F08F8.2	HMGCR
wb, bl, hg	Y41E3.11	HNRPUL1
wb, bl, hg	R11H6.5	ILF2
wb, bl, hg	F46C5.6	KIAA1622
wb, bl, hg	C15H9.7	KYNU
wb, bl, hg	M05B5.4	LYPLA3
wb, bl, hg	C06E7.3	MAT2A
wb, bl, hg	C07E3.2	NOC2L
wb, bl, hg	F39B1.1	PIK3C2A
wb, bl, hg	ZK930.1	PIK3R4
wb, bl, hg	F52B11.2	PMM2
wb, bl, hg	F21D5.5	PNKP
wb, bl, hg	H27A22.1	QPCT
wb, bl, hg	M01B12.5	RIOK1
wb, bl, hg	B0280.3	RPIA
wb, bl, hg	T22D1.4	RPN1
wb, bl, hg	F41D9.1	RUTBC3
wb, bl, hg	F23F1.6	SLC7A1
wb, bl, hg	C24F3.1	TRAM1
wb, bl, hg	T27F7.1	VPS24
wb, hg	T12A2.1	AMDHD1
wb, hg	C07A9.2	BUD31
wb, hg	C32C4.1	KCNF1
wb, hg	C37A5.1	VMD2L3

Supplementary table 2. Categories overrepresented in human homologs of worm fat storage genes

GO Cat	Obs	Exp	R-val	P-val	Branch
fatty acid beta-oxidation	3	0.4	7.5	6.36E-03	BP
dopamine receptor signaling pathway	3	0.1	30	3.64E-05	BP
pyridine nucleotide biosynthesis	3	0.23	13.04	1.14E-03	BP
NAD biosynthesis	3	0.13	23.08	1.41E-04	BP
vitamin biosynthesis	3	0.4	7.5	6.36E-03	BP
water-soluble vitamin biosynthesis	3	0.37	8.11	4.89E-03	BP
NAD metabolism	3	0.13	23.08	1.41E-04	BP
neutral amino acid transport	3	0.2	15	6.71E-04	BP
transcriptional preinitiation complex formation	3	0.13	23.08	1.41E-04	BP
regulation of transcriptional preinitiation complex formation	3	0.13	23.08	1.41E-04	BP
hindbrain development	3	0.17	17.65	3.44E-04	BP
tachykinin signaling pathway	3	0.1	30	3.64E-05	BP
lipid transport	4	0.8	5	7.44E-03	BP
pyridine nucleotide metabolism	4	0.47	8.51	9.09E-04	BP
nicotinamide metabolism	4	0.43	9.3	6.67E-04	BP
water-soluble vitamin metabolism	4	0.8	5	7.44E-03	BP
pattern specification	4	0.56	7.14	2.00E-03	BP
oxidoreduction coenzyme metabolism	5	0.7	7.14	5.06E-04	BP
vitamin metabolism	5	1.03	4.85	3.23E-03	BP
brain development	5	0.86	5.81	1.43E-03	BP
alcohol biosynthesis	5	0.63	7.94	3.05E-04	BP
monosaccharide biosynthesis	5	0.63	7.94	3.05E-04	BP
hexose biosynthesis	5	0.63	7.94	3.05E-04	BP
carbohydrate biosynthesis	6	1.83	3.28	9.30E-03	BP
central nervous system development	7	2.23	3.14	6.44E-03	BP

Dephosphorylation	8	2.96	2.7	9.06E-03	BP
steroid metabolism	8	2.23	3.59	1.55E-03	BP
amine transport	8	0.9	8.89	1.65E-06	BP
amino acid transport	8	0.66	12.12	1.14E-07	BP
organic acid transport	8	0.76	10.53	4.08E-07	BP
carboxylic acid transport	8	0.76	10.53	4.08E-07	BP
lipid biosynthesis	12	4.15	2.89	8.36E-04	BP
cellular lipid metabolism	21	8.17	2.57	5.56E-05	BP
lipid metabolism	26	9.87	2.63	4.19E-06	BP
palmitoyltransferase activity	3	0.26	11.54	1.76E-03	CC
dopamine binding	3	0.1	30	3.59E-05	CC
dopamine receptor activity	3	0.1	30	3.59E-05	CC
L-amino acid transporter activity	3	0.43	6.98	7.95E-03	CC
C-acyltransferase activity	3	0.23	13.04	1.13E-03	CC
phosphatidylinositol phosphate kinase activity	3	0.3	10	2.58E-03	CC
phosphatidylinositol-4-phosphate 3-kinase activity	3	0.1	30	3.59E-05	CC
phosphoinositide 3-kinase activity	3	0.36	8.33	4.82E-03	CC
phosphatidylinositol 3-kinase activity	3	0.36	8.33	4.82E-03	CC
retinoic acid receptor activity	3	0.2	15	6.60E-04	CC
retinoid-X receptor activity	3	0.1	30	3.59E-05	CC
tachykinin receptor activity	3	0.1	30	3.59E-05	CC
neutral amino acid transporter activity	4	0.23	17.39	3.74E-05	CC
amino acid permease activity	4	0.13	30.77	1.17E-06	CC
peptide receptor activity	4	0.86	4.65	9.77E-03	CC
peptide receptor activity\, G-protein coupled	4	0.86	4.65	9.77E-03	CC
transaminase activity	4	0.66	6.06	3.70E-03	CC
transferase activity\, transferring nitrogenous groups	5	0.73	6.85	6.23E-04	CC
steroid binding	6	0.96	6.25	3.03E-04	CC

neurotransmitter receptor activity	6	1.85	3.24	9.91E-03	CC
glucuronosyltransferase activity	7	0.5	14	1.99E-07	CC
UDP-glycosyltransferase activity	7	1.72	4.07	1.46E-03	CC
amino acid-polyamine transporter activity	7	0.26	26.92	3.01E-10	CC
polyamine transporter activity	7	0.26	26.92	3.01E-10	CC
peptide binding	7	1.75	4	1.64E-03	CC
transferase activity\, transferring hexosyl groups	8	2.64	3.03	4.65E-03	CC
amine transporter activity	8	0.93	8.6	2.17E-06	CC
amino acid transporter activity	8	0.69	11.59	1.73E-07	CC
organic acid transporter activity	8	0.96	8.33	2.91E-06	CC
carboxylic acid transporter activity	8	0.96	8.33	2.91E-06	CC
phosphoprotein phosphatase activity	9	3.24	2.78	4.85E-03	CC
phosphoric monoester hydrolase activity	11	4.49	2.45	5.12E-03	CC
transferase activity\, transferring acyl groups	11	3.6	3.06	8.75E-04	CC
transferase activity\, transferring groups other than amino-acyl groups	11	3.3	3.33	4.18E-04	CC
acyltransferase activity	11	3.3	3.33	4.18E-04	CC
carrier activity	12	5.58	2.15	9.79E-03	CC
electrochemical potential-driven transporter activity	12	3.54	3.39	1.91E-04	CC
porter activity	12	3.54	3.39	1.91E-04	CC
transferase activity	56	32.75	1.71	1.68E-05	CC
palmitoyltransferase activity	3	0.26	11.54	1.76E-03	MF
dopamine binding	3	0.1	30	3.59E-05	MF
dopamine receptor activity	3	0.1	30	3.59E-05	MF
L-amino acid transporter activity	3	0.43	6.98	7.95E-03	MF
C-acyltransferase activity	3	0.23	13.04	1.13E-03	MF
phosphatidylinositol phosphate kinase activity	3	0.3	10	2.58E-03	MF
phosphatidylinositol-4-phosphate 3-kinase activity	3	0.1	30	3.59E-05	MF
phosphoinositide 3-kinase activity	3	0.36	8.33	4.82E-03	MF

phosphatidylinositol 3-kinase activity	3	0.36	8.33	4.82E-03	MF
retinoic acid receptor activity	3	0.2	15	6.60E-04	MF
retinoid-X receptor activity	3	0.1	30	3.59E-05	MF
tachykinin receptor activity	3	0.1	30	3.59E-05	MF
neutral amino acid transporter activity	4	0.23	17.39	3.74E-05	MF
amino acid permease activity	4	0.13	30.77	1.17E-06	MF
peptide receptor activity	4	0.86	4.65	9.77E-03	MF
peptide receptor activity\, G-protein coupled	4	0.86	4.65	9.77E-03	MF
transaminase activity	4	0.66	6.06	3.70E-03	MF
transferase activity\, transferring nitrogenous groups	5	0.73	6.85	6.23E-04	MF
steroid binding	6	0.96	6.25	3.03E-04	MF
neurotransmitter receptor activity	6	1.85	3.24	9.91E-03	MF
glucuronosyltransferase activity	7	0.5	14	1.99E-07	MF
UDP-glycosyltransferase activity	7	1.72	4.07	1.46E-03	MF
amino acid-polyamine transporter activity	7	0.26	26.92	3.01E-10	MF
polyamine transporter activity	7	0.26	26.92	3.01E-10	MF
peptide binding	7	1.75	4	1.64E-03	MF
transferase activity\, transferring hexosyl groups	8	2.64	3.03	4.65E-03	MF
amine transporter activity	8	0.93	8.6	2.17E-06	MF
amino acid transporter activity	8	0.69	11.59	1.73E-07	MF
organic acid transporter activity	8	0.96	8.33	2.91E-06	MF
carboxylic acid transporter activity	8	0.96	8.33	2.91E-06	MF
phosphoprotein phosphatase activity	9	3.24	2.78	4.85E-03	MF
phosphoric monoester hydrolase activity	11	4.49	2.45	5.12E-03	MF
transferase activity\, transferring acyl groups	11	3.6	3.06	8.75E-04	MF
transferase activity\, transferring groups other than amino-acyl groups	11	3.3	3.33	4.18E-04	MF
acyltransferase activity	11	3.3	3.33	4.18E-04	MF
carrier activity	12	5.58	2.15	9.79E-03	MF

electrochemical potential-driven transporter activity	12	3.54	3.39	1.91E-04	MF
porter activity	12	3.54	3.39	1.91E-04	MF
transferase activity	56	32.75	1.71	1.68E-05	MF

Supplementary table 3: List of the 25 most upregulated genes in OP9 cell adipogenesis

Time point	Probe ID	Fold Change	P value	Symbol	Gene Name
Overexpressed					
24 hours	10481627	23.13948	1.50E-10	Lcn2	lipocalin 2
	10452316	15.39519	4.40E-12	C3	complement component 3
	10581605	11.30881	1.71E-11	Hp	haptoglobin
	10490913	9.400499	5.17E-07	Car3	carbonic anhydrase 3
	10499899	9.3602	2.73E-09	Sprr1a	small proline-rich protein 1A
	10434747	8.090149	1.89E-10	Adipoq	adiponectin, C1Q and collagen domain containing
	10564795	6.878535	1.50E-10	Plin	perilipin
	10495285	6.776545	1.50E-10	Sort1	sortilin 1
	10458555	6.353926	1.30E-07	Spry4	sprouty homolog 4 (Drosophila)
	10555389	6.316609	1.95E-08	Ucp2	uncoupling protein 2 (mitochondrial, proton carrier)
	10533725	5.900049	1.89E-10	Gpr81	G protein-coupled receptor 81
	10607868	5.896605	1.85E-06	Tlr8	toll-like receptor 8
	10419151	5.83692	2.27E-06	Ear1	eosinophil-associated, ribonuclease A family, member 1
	10465701	5.768146	2.21E-07	Lgals12	lectin, galactose binding, soluble 12
	10504775	5.606649	3.52E-10	Col15a1	collagen, type XV, alpha 1
	10544573	5.168261	7.73E-10	Rarres2	retinoic acid receptor responder (tazarotene induced) 2
	10561031	4.948785	5.17E-10	Lipe	lipase, hormone sensitive
	10350473	4.91318	8.05E-07	B3galt2	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 2
	10358389	4.893413	8.02E-09	Rgs2	regulator of G-protein signaling 2
	10571657	4.820624	1.89E-10	Acsl1	acyl-CoA synthetase long-chain family member 1
	10579812	4.393804	4.47E-08	Ednra	endothelin receptor type A
	10528207	4.333962	1.89E-10	Cd36	CD36 antigen
	10495675	3.997785	1.26E-09	F3	coagulation factor III

	10414262	3.955917	2.95E-06	Ear2	eosinophil-associated, ribonuclease A family, member 2
	10565775	3.840118	1.08E-08	Dgat2	diacylglycerol O-acyltransferase 2
48 hours	10481627	62.79472	6.01E-12	Lcn2	lipocalin 2
	10434747	24.90465	3.17E-12	Adipoq	adiponectin, C1Q and collagen domain containing
	10495285	23.68278	2.18E-12	Sort1	sortilin 1
	10419568	20.11583	1.64E-09	Ear11	eosinophil-associated, ribonuclease A family, member 11
	10452316	19.55941	1.86E-12	C3	complement component 3
	10569870	18.50284	1.83E-10	Retn	resistin
	10465701	17.09226	1.83E-09	Lgals12	lectin, galactose binding, soluble 12
	10581605	16.64087	2.51E-12	Hp	haptoglobin
	10564795	16.62791	2.75E-12	Plin	perilipin
	10458555	15.70282	1.85E-09	Spry4	sprouty homolog 4 (Drosophila)
	10490913	14.46717	2.80E-08	Car3	carbonic anhydrase 3
	10386020	11.71269	6.93E-08	Slc36a2	solute carrier family 36 (proton/amino acid symporter), member 2
	10478109	11.63172	3.43E-09	BC054059	cDNA sequence BC054059
	10607868	11.44284	4.25E-08	Tlr8	toll-like receptor 8
	10561031	11.12231	1.08E-11	Lipe	lipase, hormone sensitive
	10451953	10.68287	1.51E-09	Lrg1	leucine-rich alpha-2-glycoprotein 1
	10350473	9.981247	1.21E-08	B3galt2	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 2
	10381371	9.497424	9.46E-10	Aoc3	amine oxidase, copper containing 3
	10405693	9.402365	1.11E-09	Dapk1	death associated protein kinase 1
	10358389	9.389452	2.06E-10	Rgs2	regulator of G-protein signaling 2
	10499160	9.255629	1.80E-09	Cd1d1	CD1d1 antigen
	10565775	9.131746	9.99E-11	Dgat2	diacylglycerol O-acyltransferase 2
	10504775	8.881676	3.49E-11	Col15a1	collagen, type XV, alpha 1

	10571657	8.094134	1.08E-11	Acsl1	acyl-CoA synthetase long-chain family member 1
	10472820	7.855126	6.52E-08	Itga6	integrin alpha 6
72 hours	10481627	66.08166	3.82E-12	Lcn2	lipocalin 2
	10419568	51.29153	9.18E-11	Ear11	eosinophil-associated, ribonuclease A family, member 11
	10569870	44.40227	1.10E-11	Retn	resistin
	10495285	37.24419	8.70E-13	Sort1	sortilin 1
	10434747	35.48448	9.85E-13	Adipoq	adiponectin, C1Q and collagen domain containing
	10607868	27.99932	2.00E-09	Tlr8	toll-like receptor 8
	10478109	24.25319	2.04E-10	BC054059	cDNA sequence BC054059
	10350473	22.22909	6.11E-10	B3galt2	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 2
	10358389	22.19251	9.55E-12	Rgs2	regulator of G-protein signaling 2
	10564795	21.81242	9.85E-13	Plin	perilipin
	10386020	21.5867	6.81E-09	Slc36a2	solute carrier family 36 (proton/amino acid symporter), member 2
	10465701	21.39054	3.54E-10	Lgals12	lectin, galactose binding, soluble 12
	10381371	20.17728	4.74E-11	Aoc3	amine oxidase, copper containing 3
	10452316	19.43723	8.70E-13	C3	complement component 3
	10472820	18.83687	2.26E-09	Itga6	integrin alpha 6
	10360270	17.44923	4.99E-10	Atp1a2	ATPase, Na+/K+ transporting, alpha 2 polypeptide
	10451953	17.29833	1.36E-10	Lrg1	leucine-rich alpha-2-glycoprotein 1
	10581605	17.04491	9.85E-13	Hp	haptoglobin
	10476493	16.48139	1.04E-10	BC034902	cDNA sequence BC034902
	10405693	15.82566	9.30E-11	Dapk1	death associated protein kinase 1
	10458555	15.68302	6.11E-10	Spry4	sprouty homolog 4 (Drosophila)
	10551836	14.34476	5.80E-11	Cox7a1	cytochrome c oxidase, subunit VIIa 1
	10505438	14.16722	2.15E-10	Orm1	orosomucoid 1

10561031	13.95617	3.82E-12	Lipe	lipase, hormone sensitive
10490838	13.84232	2.15E-10	Fabp5	fatty acid binding protein 5, epidermal

Supplementary table 4: List of the 25 most downregulated genes in OP9 cell adipogenesis

Time point	Probe ID	Fold Change	P value	Symbol	Gene Name
Underexpressed					
24 hours	10404452	-12.0161	5.29E-07	Serpinb9d	serine (or cysteine) peptidase inhibitor, clade B, member 9d
	10408576	-8.61323	2.74E-07	Serpinb9c	serine (or cysteine) peptidase inhibitor, clade B, member 9c
	10404439	-8.25422	2.46E-08	Serpinb9b	serine (or cysteine) peptidase inhibitor, clade B, member 9b
	10404464	-7.68874	3.55E-08	Serpinb9f	serine (or cysteine) peptidase inhibitor, clade B, member 9f
	10574023	-7.58481	7.68E-08	Mt2	metallothionein 2
	10404472	-7.47208	2.39E-08	Serpinb9g	serine (or cysteine) peptidase inhibitor, clade B, member 9g
	10404479	-7.43409	2.24E-08	Serpinb9g	serine (or cysteine) peptidase inhibitor, clade B, member 9g
	10399691	-6.61373	2.46E-08	Id2	inhibitor of DNA binding 2
	10428604	-5.83058	2.33E-08	Tnfrsf11b	tumor necrosis factor receptor superfamily, member 11b (ost
	10404458	-5.80986	2.20E-07	Serpinb9e	serine (or cysteine) peptidase inhibitor, clade B, member 9e
	10538187	-5.56133	2.32E-08	Gpnmb	glycoprotein (transmembrane) nmb
	10372796	-5.12256	3.25E-08	Hmga2	high mobility group AT-hook 2
	10534667	-4.91403	1.90E-08	Serpine1	serine (or cysteine) peptidase inhibitor, clade E, member 1
	10359908	-4.80303	3.55E-08	Rgs4	regulator of G-protein signaling 4
	10420935	-4.65228	1.57E-07	Ephx2	epoxide hydrolase 2, cytoplasmic
	10517677	-4.30678	7.57E-08	Nbl1	neuroblastoma, suppression of tumorigenicity 1
	10574027	-4.0603	2.54E-08	Mt1	metallothionein 1
	10356764	-4.03707	2.45E-07	2310007B03Rik	RIKEN cDNA 2310007B03 gene
	10492091	-3.99191	2.64E-08	Smad9	MAD homolog 9 (Drosophila)
	10387536	-3.89117	4.38E-08	Cd68	CD68 antigen
	10502655	-3.70412	2.52E-08	Cyr61	cysteine rich protein 61

	10586079	-3.69564	1.38E-07	Itga11	integrin alpha 11
	10449284	-3.67579	4.17E-08	Dusp1	dual specificity phosphatase 1
	10344149	-3.57097	4.73E-05	---	---
	10467124	-3.54928	4.41E-07	Acta2	actin, alpha 2, smooth muscle, aorta
48 hours	10407797	-43.8761	2.63E-10	Prl2c3	prolactin family 2, subfamily c, member 3
	10404452	-42.0795	6.09E-09	Serpinb9d	serine (or cysteine) peptidase inhibitor, clade B, member 9d
	10408576	-38.2786	1.69E-09	Serpinb9c	serine (or cysteine) peptidase inhibitor, clade B, member 9c
	10403579	-33.3535	1.04E-08	Prl2c5	prolactin family 2, subfamily c, member 5
	10404464	-25.1368	4.65E-10	Serpinb9f	serine (or cysteine) peptidase inhibitor, clade B, member 9f
	10404472	-24.3282	2.79E-10	Serpinb9g	serine (or cysteine) peptidase inhibitor, clade B, member 9g
	10404479	-23.5975	2.70E-10	Serpinb9g	serine (or cysteine) peptidase inhibitor, clade B, member 9g
	10399691	-16.986	4.65E-10	Id2	inhibitor of DNA binding 2
	10404458	-15.4255	2.30E-09	Serpinb9e	serine (or cysteine) peptidase inhibitor, clade B, member 9e
	10538187	-15.2297	2.70E-10	Gpnmb	glycoprotein (transmembrane) nmb
	10404439	-13.6168	1.57E-09	Serpinb9b	serine (or cysteine) peptidase inhibitor, clade B, member 9b
	10497122	-11.9944	5.63E-09	Depdc1a	DEP domain containing 1a
	10404063	-11.4968	6.17E-09	Hist1h2ab	histone cluster 1, H2ab
	10428604	-10.8978	7.54E-10	Tnfrsf11b	tumor necrosis factor receptor superfamily, member 11b (ost
	10420935	-10.8175	1.69E-09	Ephx2	epoxide hydrolase 2, cytoplasmic
	10563883	-10.5055	5.71E-09	Depdc1a	DEP domain containing 1a
	10411728	-10.0623	3.73E-08	Cenph	centromere protein H
	10586448	-9.45676	2.63E-08	2810417H13Rik	RIKEN cDNA 2810417H13 gene
	10487480	-9.23654	6.81E-09	Bub1	budding uninhibited by benzimidazoles 1 homolog (S. cerevis
	10372796	-9.14193	1.11E-09	Hmga2	high mobility group AT-hook 2
	10462796	-9.08965	1.85E-09	Kif11	kinesin family member 11

	10576883	-8.80593	5.76E-09	Shcbp1	Shc SH2-domain binding protein 1
	10400589	-8.68963	1.51E-09	C79407	expressed sequence C79407
	10342190	-8.64973	0.000174	---	---
	10594251	-8.45425	1.68E-09	Kif23	kinesin family member 23
72 hours	10407797	-57.3822	4.74E-11	Prl2c3	prolactin family 2, subfamily c, member 3
	10408576	-44.2648	4.21E-10	Serpinb9c	serine (or cysteine) peptidase inhibitor, clade B, member 9c
	10403579	-42.2576	2.99E-09	Prl2c5	prolactin family 2, subfamily c, member 5
	10462796	-40.109	2.74E-11	Kif11	kinesin family member 11
	10404452	-39.3858	2.95E-09	Serpinb9d	serine (or cysteine) peptidase inhibitor, clade B, member 9d
	10586448	-34.9801	4.39E-10	2810417H13Rik	RIKEN cDNA 2810417H13 gene
	10497831	-33.8305	1.61E-11	Ccna2	cyclin A2
	10562637	-30.7234	1.28E-10	Ccnb1	cyclin B1
	10515836	-30.5124	2.74E-11	Ccnb1	cyclin B1
	10411739	-30.059	1.61E-11	Ccnb1	cyclin B1
	10557156	-27.8728	8.70E-13	Plk1	polo-like kinase 1 (Drosophila)
	10568714	-27.1664	5.04E-11	Mki67	antigen identified by monoclonal antibody Ki 67
	10359890	-24.9383	1.28E-11	Nuf2	NUF2, NDC80 kinetochore complex component, homolog (S.
	10594774	-24.5463	7.05E-12	Ccnb2	cyclin B2
	10487480	-24.0814	2.29E-10	Bub1	budding uninhibited by benzimidazoles 1 homolog (S. cerevis
	10390707	-23.0709	3.80E-11	Top2a	topoisomerase (DNA) II alpha
	10554445	-20.479	2.86E-10	Prc1	protein regulator of cytokinesis 1
	10474875	-19.9405	8.09E-08	Casc5	cancer susceptibility candidate 5
	10521731	-19.0246	1.71E-10	Ncapg	on-SMC condensin I complex, subunit G
	10538187	-18.954	4.97E-11	Gpnmb	glycoprotein (transmembrane) nmb
	10374083	-18.239	5.65E-12	Aebp1	AE binding protein 1

10404063	-17.5978	7.53E-10	Hist1h2ab	histone cluster 1, H2ab
10420877	-17.1497	2.69E-10	Esco2	establishment of cohesion 1 homolog 2 (S. cerevisiae)
10419323	-16.6682	2.56E-10	Dlgap5	discs, large (Drosophila) homolog-associated protein 5
10394978	-16.4448	1.64E-10	Rrm2	ribonucleotide reductase M2

Appendix 1: Journal Code for ImageExpress

```
Journal for DAPI/Nile Red Analysis (Version 10)
Open Excel workbook to log results
1: Open Summary Log(OPENDDE and OVERWRITEMODE, "")

Using Cell Scoring App to segment cells and records total cell number
using nuclear stain
2: Overwrite "Segmentation" = Cell Scoring(All nuclei = "DAPI - Pos 1",
Positive marker = "TRITC - Pos 4")
TotalCellNumber = CellScoring.TotalCells

Threshold bright Nile Red staining/droplets and measured their area and
integrated intensity of the dye
3: Overwrite "Multiply" = ("TRITC - Pos 4" * "Segmentation") / 1
4: Threshold Image("Multiply", 1, 65535)
5: Show Region Statistics("Multiply", ENTIREIMAGE)
TotalCellArea = ShowRegionStatistics.ThresholdedArea
WholeCellInetgratedIntensity = ShowRegionStatistics.Integrated
6: Threshold Image("Multiply", 750, 65535)
7: Show Region Statistics("Multiply", ENTIREIMAGE)
ThresholdedNileRedArea = ShowRegionStatistics.ThresholdedArea
ThresholdedNileRedIntegratedIntensity = ShowRegionStatistics.Integrated

Reporting measurement to Excel spreadsheet
8: Log Variable(TotalCellArea, NONEWLINE, NO HEADER)
9: Log Variable(WholeCellInetgratedIntensity, NONEWLINE, NO HEADER)
10: Log Variable(ThresholdedNileRedArea, NONEWLINE, NO HEADER)
11: Log Variable(ThresholdedNileRedIntegratedIntensity, NONEWLINE, NO
HEADER)
```

Appendix 2: R Code for Microarray Analysis

```
data <- read.delim("expressionValuesRMA-GENE-DEFAULT - Group
1.TXT",sep="\t",stringsAsFactors=F)
names(data)
names(data)[2:13] <-
c("D01","D02","D03","D11","D12","D13","D21","D22","D23","D31","D32","D3
3")
names(data)
eset <- data[,1:13]
head(eset)
rownames(eset) <- eset$Probe.Set.ID
head(eset)
eset <- eset[,-1]
head(eset)
names(data)
annot <- data[,c(1,14:32)]
names(annot)
rownames(annot) <- annot[,1]
rm(data)
ls()
library(limma)
limmaUsersGuide()
ls()
adjustFC <- function(tabl){#
    index <- which(tabl$FC < 1.0);#
    fc <- (-1 / tabl$FC)[index];#
    tabl[index,"FC"] <- fc;#
    return(tabl);#
}
head(eset)
library(genefilter)
?pOverA
p1 <- pOverA(0.5, 3.5)
ff <- filterfun(p1)
selected <- genefilter(eset, ff)
eSub <- eset[selected,]
dim(eset)
dim(eSub)
p1 <- pOverA(0.5, 5)
ff <- filterfun(p1)
selected <- genefilter(eset, ff)
eSub <- eset[selected,]
dim(eSub)
dir()
targets <- data.frame(Samples = names(eset), Day=
c(rep("0",3),rep("1",3),rep("2",3),rep("3",3)))
targets
TS <- factor(targets$Day, levels = unique(targets$Day))
TS
design <- model.matrix(~0 + TS)
design
colnames(design) <- levels(TS)
design
colnames(design) <- c("Day0","Day1","Day2","Day3")
```

```
design
cont.matrix <- makeContrasts(Day1vsDay0 = Day1 - Day0, Day2vsDay0 =
Day2 - Day0, Day3vsDay0 = Day3 - Day0, levels=design)
cont.matrix
fit <- lmFit(eset, design)
names(fit)
class(fit)
head(fit$genes)
head(eset)
names(fit)
head(fit$Amean)
head(fit$coefficients)
mean(6.9721214, 7.012953, 7.036483)
fit2 <- contrasts.fit(fit, cont.matrix)
names(fit2)
head(fit2$coefficients)
2 ^ 0.17305833
7.180245/7.007187
7.180245 - 7.007187
- (2 ^ .07139733)
2 ^ -0.07139733
-1/0.9517158
?eBayes
fit2 <- eBayes(fit2)
?topTable
topTable(fit2,coef="Day1vsDay0")
head(topTable(fit2, coef="Day1vsDay0",number=20000))
head(topTable(fit2, coef="Day1vsDay0",number=30000))
resultD1vsD0 <- topTable(fit2, coef="Day1vsDay0",number=dim(eset)[1])
dim(eset)
dim(eset)[1]
head(resultD1vsD0)
FC <- 2 ^ resultD1vsD0$logFC;
head(FC)
tail(FC)
adjustFC
resultD1vsD0 <- cbind(resultD1vsD0[,1:2], FC=FC, resultD1vsD0[,3:7])
head(resultD1vsD0)
resultD1vsD0 <- adjustFC(resultD1vsD0)
tail(resultD1vsD0)
resultD1vsD0 <- cbind(resultD1vsD0, annot[rownames(resultD1vsD0),])
head(resultD1vsD0)
sum(which(is.na(resultD1vsD0$Gene.Symbol)))
length(which(is.na(resultD1vsD0$Gene.Symbol)))
head(annot)
names(resultD1vsD0)
resultD1vsD0 <- resultD1vsD0[,1:8]
head(rownames(resultD1vsD0))
11208 %in% rownames(annot)
head(11208)
head(resultD1vsD0)
resultD1vsD0 <- cbind(resultD1vsD0, annot[resultD1vsD0$ID,])
head(resultD1vsD0)
write.table(resultD1vsD0,
file='resultD1vsD0.tsv',sep='\t',row.names=F,quote=F)
pwd
getwd()
```

```
names(resultD1vsD0)[1] <- "Probe.Set.ID"
write.table(resultD1vsD0,
file='resultD1vsD0.tsv',sep='\t',row.names=F,quote=F)
?topTable
resultD1vsD0 <- cbind(resultD1vsD0, eset[resultD1vsD0$Probe.Set.ID,])
names(resultD1vsD0)
resultD1vsD0 <- resultD1vsD0[,-c(35:40)]
names(resultD1vsD0)
write.table(resultD1vsD0,
file='resultD1vsD0.tsv',sep='\t',row.names=F,quote=F)
resultD2vsD0 <- topTable(fit2, coef="Day2vsDay0", number=dim(eset)[1])
resultD2vsD0 <- cbind(resultD2vsD0, FC = 2 ^ resultD2vsD0$logFC)
names(resultD2vsD0)
resultD2vsD0 <- adjustFC(resultD2vsD0)
names(resultD2vsD0)[1] <- "Probe.Set.ID"
resultD2vsD0 <- resultD2vsD0[,c(1,2,8,3:7)]
resultD2vsD0 <- cbind(resultD2vsD0,
annot[resultD2vsD0$Probe.Set.ID,],eset[resultD2vsD0$Probe.Set.ID,c(1:3,
7:9)])
names(resultD2vsD0)
write.table(resultD2vsD0, file='resultD2vsD0.tsv',row.names=F,quote=F)
resultD3vsD0 <- topTable(fit2, coef="Day3vsDay0",number=dim(eset)[1])
?topTable
resultD3vsD0 <- cbind(resultD3vsD0, FC = 2^ resultD3vsD0$logFC)
resultD3vsD0 <- adjustFC(resultD3vsD0)
names(resultD3vsD0)[1] <- "Probe.Set.ID"
names(resultD3vsD0)
resultD3vsD0 <- resultD3vsD0[,c(1,2,8,3:7)]
names(resultD3vsD0)
resultD3vsD0 <- cbind(resultD3vsD0, annot[resultD3vsD0$Probe.Set.ID,],
eset[resultD3vsD0$Probe.Set.ID,c(1:3,10:12)])
names(resultD3vsD0)
write.table(resultD3vsD0,
file='resultD3vsD0.tsv',sep='\t',row.names=F,quote=F)
ls()
write.table(resultD2vsD0,
file='resultD2vsD0.tsv',sep='\t',row.names=F,quote=F)
test <- resultD2vsD0[which(resultD2vsD0$FC > 2.0),]
test <- test[which(test$adj.P.Val < 0.001),]
dim(test)
test <- resultD2vsD0[which(abs(resultD2vsD0$FC) > 2.0),]
test <- test[which(test$adj.P.Val < 0.001),]
dim(test)
head(test)
ls()
resultD1vsD0 <- resultD1vsD0[order(resultD1vsD0$Probe.Set.ID,
decreasing=TRUE),]
head(resultD1vsD0)
resultD2vsD0 <- resultD2vsD0[order(resultD2vsD0$Probe.Set.ID,
decreasing=TRUE),]
resultD3vsD0 <- resultD3vsD0[order(resultD3vsD0$Probe.Set.ID,
decreasing=TRUE),]
fc <- cbind(resultD1vsD0$probe.Set.ID, resultD1vsD0$FC,
resultD2vsD0$FC, resultD3vsD0$FC, resultD1vsD0$Symbol,
resultD1vsD0$Description);
head(fc)
```

```
fc <- cbind(ProbeID=resultD1vsD0$probe.Set.ID, FC1=resultD1vsD0$FC,
FC2=resultD2vsD0$FC, FC3=resultD3vsD0$FC, Symbol=resultD1vsD0$Symbol,
Desc=resultD1vsD0$Description);
head(fc)
names(resultD1vsD0)
fc <- cbind(ProbeID=resultD1vsD0Probe.Set.ID, FC1=resultD1vsD0$FC,
FC2=resultD2vsD0$FC, FC3=resultD3vsD0$FC,
Symbol=resultD1vsD0$Gene.Symbol, Desc=resultD1vsD0$Gene.Description);
fc <- cbind(ProbeID=resultD1vsD0$Probe.Set.ID, FC1=resultD1vsD0$FC,
FC2=resultD2vsD0$FC, FC3=resultD3vsD0$FC,
Symbol=resultD1vsD0$Gene.Symbol, Desc=resultD1vsD0$Gene.Description);
head(fc)
fc <- fc[order(fc$ProbeID, decreasing=F),]
write.table(fc,
file='All_Comparisons.tsv',sep='\t',row.names=F,quote=F)
class(fc)
fc <- data.frame(fc, stringsAsFactors=F)
head(fc)
fc <- fc[order(fc$ProbeID, decreasing=TRUE),]
head(fc)
head(resultD1vsD0$Probe.Set.ID)
fc2 <- cbind(fc[,1:2],resultD1vsD0$adj.P.Val,
fc[,3],resultD2vsD0$adj.P.Val, fc[,4], resultD3vsD0$adj.P.Val,
fc[,4:5])
names(fc2)
names(fc2)[2:8] <-
c("D1vD0_FC","Adj.P.Val","D2vD0_FC","Adj.P.Val","D3vD0_FC","Adj.P.Val",
"Symbol","Description")
names(fc2)
head(fc2)
fc2 <- fc2[,-9]
head(fc2)
names(fc)
fc2 <- cbind(fc2, fc$Desc)
head(fc2)
names(fc2)[9] <- 'Description'
head(fc2)
names(fc)
fc2 <- cbind(fc[,1:2],resultD1vsD0$adj.P.Val,
fc[,3],resultD2vsD0$adj.P.Val, fc[,4], resultD3vsD0$adj.P.Val,
fc[,5:6])
names(fc2)
head(fc2)
fc2 <- cbind(fc[,1:2],Adj.P.Val1=resultD1vsD0$adj.P.Val,
D2vD0_FC=fc[,3],Adj.P.Val2=resultD2vsD0$adj.P.Val, D3vD0_FC=fc[,4],
Adj.P.Val3=resultD3vsD0$adj.P.Val, fc[,5:6])
head(fc2)
names(fc2)[2] <- "D1vD0_FC"
head(fc2)
write.table(fc2,
file='All_Comparisons.tsv',sep='\t',row.names=F,quote=F)
?grep
names(resultDay1vsDay0)
ls()
resultD1vsD0
names(resultD1vsD0)
```

```
test <- grep("transcription",
resultD1vsD0$GO.Biological.Process.Term,ignore.case=TRUE)
head(test)
resultD1vsD0[test,1:5]
length(test)
head(test)
resultD1vsD0[202,]
ls()
head(eset)
Description <- annot[rownames(eset),"Gene.Description"]
head(Description)
gct <- cbind(NAME=rownames(eset),eset)
head(gct)
gct <- cbind(gct[,1],Description,gct[2:])
dim(gct)
gct <- cbind(gct[,1],Description,gct[2:13])
head(gct)
names(gct)[1] <- "NAME"
head(gct)
write.table(gct, file='all_data.gct',sep='\t',row.names=F,quote=F)
dim(eset)
names(annot)
chip <- annot[,c(1:3)]
head(chip)
names(chip) <- c("Probe Set ID","Gene Symbol","Gene Title")
head(chip)
chip <- chip[order(chip[,2], decreasing=F),]
head(chip)
write.table(chip,
file="~/Documents/GSEA/chip_files/MoGeneST_July09.chip",sep="\t"row.nam
es=F,quote=F)
write.table(chip,
file="~/Documents/GSEA/chip_files/MoGeneST_July09.chip",sep="\t",row.na
mes=F,quote=F)
```

References

1. de Onis, M. and J. Habicht, *Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee.* American journal of clinical nutrition, 1996. **64**(4): p. 650.
2. Hubert, H.B., et al., *Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study.* Circulation, 1983. **67**(5): p. 968.
3. Must, A., et al., *The disease burden associated with overweight and obesity.* JAMA: the journal of the American Medical Association, 1999. **282**(16): p. 1523.
4. Pais, R., et al., *Metabolic syndrome and risk of subsequent colorectal cancer.* World Journal of Gastroenterology: WJG, 2009. **15**(41): p. 5141.
5. Yaskin, J., R.W. Toner, and N. Goldfarb, *Obesity management interventions: a review of the evidence.* Popul Health Manag, 2009. **12**(6): p. 305-16.
6. Gade, W., et al., *Failures of feedback: rush hour along the highway to obesity.* Clinical laboratory science: journal of the American Society for Medical Technology, 2010. **23**(1): p. 39.
7. Anand, S., et al., *Risk factors for myocardial infarction in women and men: insights from the INTERHEART study.* European heart journal, 2008.
8. Lanas, F., et al., *Risk factors for acute myocardial infarction in Latin America: the INTERHEART Latin American study.* Circulation, 2007. **115**(9): p. 1067.
9. Wing, R., et al., *Change in waist-hip ratio with weight loss and its association with change in cardiovascular risk factors.* American Journal of Clinical Nutrition, 1992. **55**(6): p. 1086.
10. Grundy, S., *Obesity, metabolic syndrome, and cardiovascular disease.* Journal of Clinical Endocrinology & Metabolism, 2004. **89**(6): p. 2595.
11. Xu, A., et al., *Circulating adipocyte-fatty acid binding protein levels predict the development of the metabolic syndrome: a 5-year prospective study.* Circulation, 2007. **115**(12): p. 1537.
12. Everson, S.A., et al., *Weight gain and the risk of developing insulin resistance syndrome.* Diabetes Care, 1998. **21**(10): p. 1637.
13. Despres, J., *The insulin resistance-dyslipidemic syndrome of visceral obesity: effect on patients' risk.* Obesity research, 1998. **6**: p. 8S.
14. Zamboni, M., et al., *Obesity and regional body-fat distribution in men: separate and joint relationships to glucose tolerance and plasma lipoproteins.* The American journal of clinical nutrition, 1994. **60**(5): p. 682.
15. Vanhala, M., et al., *Obesity type and clustering of insulin resistance-associated cardiovascular risk factors in middle-aged men and women.* International journal of obesity, 1998. **22**(4): p. 369-374.
16. Grundy, S., et al., *Clinical management of metabolic syndrome: report of the American Heart Association/National Heart, Lung, and Blood Institute/American Diabetes Association conference on scientific issues related to management.* Circulation, 2004. **109**(4): p. 551.
17. Antonopoulos, S., *Third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood*

cholesterol in adults (Adult Treatment Panel III) final report. Circulation, 2002. **106**: p. 3143-421.
18. Allison, D., et al., *Annual deaths attributable to obesity in the United States.* Jama, 1999. **282**(16): p. 1530.
19. Roberts, R.E., et al., *Prospective association between obesity and depression: evidence from the Alameda County Study.* International Journal of Obesity, 2003. **27**(4): p. 514-521.
20. Dong, C., L. Sanchez, and R. Price, *Relationship of obesity to depression: a family-based study.* International journal of obesity, 2004. **28**(6): p. 790-795.
21. Petry, N.M., et al., *Overweight and obesity are associated with psychiatric disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions.* Psychosomatic Medicine, 2008: p. PSY. 0b013e3181651651v1.
22. Fontaine, K., S. Bartlett, and I. Barofsky, *Health-related quality of life among obese persons seeking and not currently seeking treatment.* International Journal of Eating Disorders, 2000. **27**(1): p. 101-105.
23. Han, T., et al., *Quality of life in relation to overweight and body fat distribution.* American journal of public health, 1998. **88**(12): p. 1814.
24. Sullivan, M., et al., *Swedish obese subjects (SOS)--an intervention study of obesity. Baseline evaluation of health and psychosocial functioning in the first 1743 subjects examined.* International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity, 1993. **17**(9): p. 503.
25. Stewart, A. and R. Brook, *Effects of being overweight.* American journal of public health, 1983. **73**(2): p. 171.
26. Thompson, D., et al., *Lifetime health and economic consequences of obesity.* Archives of Internal Medicine, 1999. **159**(18): p. 2177.
27. Flegal, K.M., et al., *Prevalence and trends in obesity among US adults, 1999-2008.* Jama, 2010. **303**(3): p. 235-41.
28. Flier, J.S., M. Harris, and A.N. Hollenberg, *Leptin, nutrition, and the thyroid: the why, the wherefore, and the wiring.* J Clin Invest, 2000. **105**(7): p. 859-861.
29. Kopelman, P.G., *Obesity as a medical problem.* Nature, 2000. **404**(6778): p. 635-643.
30. Rissanen, A.M., et al., *Determinants of weight gain and overweight in adult Finns.* Eur J Clin Nutr, 1991. **45**(9): p. 419-430.
31. Prentice, A.M., et al., *Energy expenditure in overweight and obese adults in affluent societies: an analysis of 319 doubly-labelled water measurements.* Eur J Clin Nutr, 1996. **50**(2): p. 93-97.
32. Kim, D., et al., *US state- and county-level social capital in relation to obesity and physical inactivity: a multilevel, multivariable analysis*

Innovations to reduce television and computertime and obesity in childhood. Soc Sci Med, 2006. **63**(4): p. 1045-59.
33. Rolls, B., P. Van Duijvenvoorde, and E. Rowe, *Variety in the diet enhances intake in a meal and contributes to the development of obesity in the rat.* Physiology & behavior, 1983. **31**(1): p. 21-27.

34. Nielsen, S. and B. Popkin, *Patterns and trends in food portion sizes, 1977-1998.* Jama, 2003. **289**(4): p. 450.
35. Rolls, B., L. Roe, and J. Meengs, *Larger portion sizes lead to a sustained increase in energy intake over 2 days.* Journal of the American Dietetic Association, 2006. **106**(4): p. 543-549.
36. Fisher, J.O., et al., *Portion size effects on daily energy intake in low-income Hispanic and African American children and their mothers*

Larger portion sizes lead to a sustained increase in energy intake over 2 days

Portion sizes and the obesity epidemic. Am J Clin Nutr, 2007. **86**(6): p. 1709-16.
37. Kark, M. and F. Rasmussen, *Growing social inequalities in the occurrence of overweight and obesity among young men in Sweden.* Scandinavian journal of public health, 2005. **33**(6): p. 472.
38. Bennett, G., K. Wolin, and D. Duncan, *Social determinants of obesity.* Obesity epidemiology, 2008: p. 342.
39. Caprio, S., et al., *Influence of race, ethnicity, and culture on childhood obesity: Implications for prevention and treatment.* Diabetes care, 2008. **31**(11): p. 2211.
40. Allen, M.L., et al., *Adolescent participation in preventive health behaviors, physical activity, and nutrition: differences across immigrant generations for Asians and Latinos compared with Whites.* Am J Public Health, 2007. **97**(2): p. 337-43.
41. Dessein, P., et al., *Beneficial effects of weight loss associated with moderate calorie/carbohydrate restriction, and increased proportional intake of protein and unsaturated fat on serum urate and lipoprotein levels in gout: a pilot study.* Annals of the rheumatic diseases, 2000. **59**(7): p. 539.
42. Akers, J.D., P.A. Estabrooks, and B.M. Davy, *Translational research: bridging the gap between long-term weight loss maintenance research and practice.* J Am Diet Assoc. **110**(10): p. 1511-22, 1522 e1-3.
43. Poobalan, A., et al., *Effects of weight loss in overweight/obese individuals and long-term lipid outcomes--a systematic review.* Obes Rev, 2004. **5**(1): p. 43-50.
44. Avenell, A., et al., *Systematic review of the long-term effects and economic consequences of treatments for obesity and implications for health improvement.* Health Technol Assess, 2004. **8**(21): p. iii-iv, 1-182.
45. Avenell, A., et al., *What are the long-term benefits of weight reducing diets in adults? A systematic review of randomized controlled trials.* J Hum Nutr Diet, 2004. **17**(4): p. 317-35.
46. Tate, D.F., E.H. Jackvony, and R.R. Wing, *A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program.* Archives of internal medicine, 2006. **166**(15): p. 1620.
47. Weyer, C., et al., *Implications of the traditional and the new ACSM physical activity recommendations on weight reduction in dietary treated obese subjects.* International journal of obesity, 1998. **22**(11): p. 1071-1078.
48. Klem, M.L., et al., *A descriptive study of individuals successful at long-term maintenance of substantial weight loss.* The American journal of clinical nutrition, 1997. **66**(2): p. 239.

49. Epstein, L.H., et al., *Ten-year outcomes of behavioral family-based treatment for childhood obesity.* Health Psychology, 1994. **13**(5): p. 373.
50. Collins, F.S. and M.K. Mansoura, *The Human Genome Project. Revealing the shared inheritance of all humankind.* Cancer, 2001. **91**(1 Suppl): p. 221-5.
51. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
52. Feuk, L., A.R. Carson, and S.W. Scherer, *Structural variation in the human genome.* Nat Rev Genet, 2006. **7**(2): p. 85-97.
53. Weiner MP, H.T., *Introduction to SNPs: discovery of markers for disease.* Biotechniques, 2002.
54. Herbert, A., et al., *A common genetic variant is associated with adult and childhood obesity.* Science, 2006. **312**(5771): p. 279.
55. Meyre, D., et al., *Variants of ENPP1 are associated with childhood and adult obesity and increase the risk of glucose intolerance and type 2 diabetes.* Nature genetics, 2005. **37**(8): p. 863-867.
56. Frayling, T.M., *A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity.* Science, 2007. **316**: p. 889-894.
57. Menzaghi, C., et al., *A haplotype at the adiponectin locus is associated with obesity and other features of the insulin resistance syndrome.* Diabetes, 2002. **51**(7): p. 2306.
58. Scott, L.J., et al., *A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants.* Science, 2007. **316**(5829): p. 1341.
59. Saxena, R., et al., *Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.* Science, 2007. **316**(5829): p. 1331.
60. Horikawa, Y., et al., *Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus.* Nature genetics, 2000. **26**(2): p. 163-175.
61. Ordovas, J.M. and D. Corella, *Nutritional genomics.* Annu Rev Genomics Hum Genet, 2004. **5**: p. 71-7118.
62. Kokubo, Y., et al., *Identification of 108 SNPs in TSC, WNK1, and WNK4 and their association with hypertension in a Japanese general population.* Journal of human genetics, 2004. **49**(9): p. 507-515.
63. Newhouse, S.J., et al., *Haplotypes of the WNK1 gene associate with blood pressure variation in a severely hypertensive population from the British Genetics of Hypertension study.* Human molecular genetics, 2005. **14**(13): p. 1805.
64. Maes, H.H., M.C. Neale, and L.J. Eaves, *Genetic and environmental factors in relative body weight and human adiposity.* Behav Genet, 1997. **27**(4): p. 325-51.
65. Tholin S, et al., *Genetic and environmental influences on eating behavior: the Swedish Young Male Twins Study.* 2005. **81**: p. 564-569.
66. Fabsitz RR, Sholinsky P, and C. D, *Genetic influences on adult weight gain and maximum body mass index in male twins.* 1994. **140**: p. 711-720.
67. Faith MS, et al., *Evidence for independent genetic influences on fat mass and body mass index in a pediatric twin sample.* 1999. **104**: p. 61-67.
68. Allison DB, Faith MS, and N. JS, *Risch/'s lambda values for human obesity.* 1996. **20**: p. 990-999.

69. Allison DB, et al., *The heritability of body mass index among an international sample of monozygotic twins reared apart.* 1996. **20**: p. 501-506.
70. Malis C, Rasmussen EL, and P. P, *Total and regional fat distribution is strongly influenced by genetic factors in young and elderly twins.* 2005. **13**: p. 2139-2145.
71. Rose KM, et al., *Genetic and behavioral determinants of waist-hip ratio and waist circumference in women twins.* 1998. **6**: p. 383-392.
72. Rankinen, T., et al., *The human obesity gene map: the 2005 update.* Obesity (Silver Spring), 2006. **14**(4): p. 529-644.
73. Mencarelli, M., et al., *Rare melanocortin-3 receptor mutations with in vitro functional consequences are associated with human obesity.* Human Molecular Genetics, 2011. **20**(2): p. 392.
74. O'Rahilly, S. and I.S. Farooqi, *Genetics of obesity.* Philos Trans R Soc Lond B Biol Sci, 2006. **361**(1471): p. 1095-105.
75. Roth, C.L., et al., *A novel melanocortin-4 receptor gene mutation in a female patient with severe childhood obesity.* Endocrine, 2009. **36**(1): p. 52-59.
76. Fischer-Posovszky, P., et al., *A new missense mutation in the leptin gene causes mild obesity and hypogonadism without affecting T cell responsiveness.* Journal of Clinical Endocrinology & Metabolism, 2010. **95**(6): p. 2836.
77. Coleman, D. and E. Eicher, *Fat (fat) and tubby (tub): two autosomal recessive mutations causing obesity syndromes in the mouse.* Journal of Heredity, 1990. **81**(6): p. 424.
78. Clément, K., et al., *A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction.* Nature, 1998. **392**(6674): p. 398-401.
79. Montague, C.T., et al., *Congenital leptin deficiency is associated with severe early-onset obesity in humans.* Nature, 1997. **387**(6636): p. 903-908.
80. Strobel, A., et al., *A leptin missense mutation associated with hypogonadism and morbid obesity.* Nature Genetics, 1998. **18**(3): p. 213-215.
81. Farooqi, I.S., et al., *Effects of recombinant leptin therapy in a child with congenital leptin deficiency.* New England Journal of Medicine, 1999. **341**(12): p. 879-884.
82. Ichihara, S. and Y. Yamada, *Genetic factors for human obesity.* Cell Mol Life Sci, 2008. **65**(7-8): p. 1086-98.
83. Geller, F., et al., *Melanocortin-4 receptor gene variant I103 is negatively associated with obesity.* Am J Hum Genet, 2004. **74**(3): p. 572-81.
84. Li, S. and R.J. Loos, *Progress in the genetics of common obesity: size matters.* Curr Opin Lipidol, 2008. **19**(2): p. 113-21.
85. Rosmond, R., et al., *G-308A polymorphism of the tumor necrosis factor alpha gene promoter and salivary cortisol secretion.* J Clin Endocrinol Metab, 2001. **86**(5): p. 2178-80.
86. Li, W.D., et al., *Sequence variants in the 5' flanking region of the leptin gene are associated with obesity in women.* Ann Hum Genet, 1999. **63**(Pt 3): p. 227-34.
87. Hager, J., et al., *A polymorphism in the 5' untranslated region of the human ob gene is associated with low leptin levels.* Int J Obes Relat Metab Disord, 1998. **22**(3): p. 200-5.
88. Mitchell, B.D., et al., *A paired sibling analysis of the beta-3 adrenergic receptor and obesity in Mexican Americans.* J Clin Invest, 1998. **101**(3): p. 584-7.

89. Matsushita, Y., et al., *The Trp(64)Arg polymorphism of the beta(3)-adrenergic receptor gene is not associated with body weight or body mass index in Japanese: a longitudinal analysis.* J Clin Endocrinol Metab, 2003. **88**(12): p. 5914-20.
90. Gagnon, J., et al., *The Trp64Arg mutation of the beta3 adrenergic receptor gene has no effect on obesity phenotypes in the Quebec Family Study and Swedish Obese Subjects cohorts.* J Clin Invest, 1996. **98**(9): p. 2086-93.
91. Altshuler, D., et al., *The common PPAR Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes.* Nature genetics, 2000. **26**(1): p. 76-80.
92. Ioannidis, J.P.A., et al., *Replication validity of genetic association studies.* Nature genetics, 2001. **29**(3): p. 306-309.
93. Kussmann, M., L. Krause, and W. Siffert, *Nutrigenomics: where are we with genetic and epigenetic markers for disposition and susceptibility?* Nutrition Reviews, 2010. **68**: p. S38-S47.
94. Beamer, B.A., et al., *Association of the Pro12Ala variant in the peroxisome proliferator-activated receptor-gamma2 gene with obesity in two Caucasian populations.* Diabetes, 1998. **47**(11): p. 1806-8.
95. Deeb, S.S., et al., *A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity.* Nat Genet, 1998. **20**(3): p. 284-7.
96. Meirhaeghe, A., et al., *Impact of the Peroxisome Proliferator Activated Receptor gamma2 Pro12Ala polymorphism on adiposity, lipids and non-insulin-dependent diabetes mellitus.* Int J Obes Relat Metab Disord, 2000. **24**(2): p. 195-9.
97. Franks, P.W., et al., *Does peroxisome proliferator-activated receptor gamma genotype (Pro12ala) modify the association of physical activity and dietary fat with fasting insulin level?* Metabolism, 2004. **53**(1): p. 11-6.
98. Luan, J., et al., *Evidence for gene-nutrient interaction at the PPARgamma locus.* Diabetes, 2001. **50**(3): p. 686-9.
99. M. Montagnana, C.F., P. M. Nilsson, G. Engström, B. Hedblad, G. Lippi, P. Minuz, G. Berglund, O. Melander, *The Pro12Ala polymorphism of the PPARG gene is not associated with the metabolic syndrome in an urban population of middle-aged Swedish individuals.* Diabetic Medicine, 2008. **Volume 25**(8): p. 902-908.
100. Lindi, V.I., et al., *Association of the Pro12Ala polymorphism in the PPAR-gamma2 gene with 3-year incidence of type 2 diabetes and body weight change in the Finnish Diabetes Prevention Study.* Diabetes, 2002. **51**(8): p. 2581-6.
101. Vincent, S., et al., *Genetic polymorphisms and lipoprotein responses to diets.* Proc Nutr Soc, 2002. **61**(4): p. 427-34.
102. Bogardus, C., *Missing heritability and GWAS utility.* Obesity (Silver Spring, Md.), 2009. **17**(2): p. 209.
103. Altmuller, J., et al., *Genomewide scans of complex human diseases: true linkage is hard to find.* Am J Hum Genet, 2001. **69**(5): p. 936-50.
104. Todd, J.A., *Statistical false positive or true disease pathway?* Nat Genet, 2006. **38**(7): p. 731-3.
105. Hirschhorn, J.N., et al., *A comprehensive review of genetic association studies.* Genet Med, 2002. **4**(2): p. 45-61.

106. *Genome-Wide Association Studies Fact Sheet*. 2010, National Human Genome Research Institute.
107. Manolio, T., et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-753.
108. Seshadri, S., et al., *Genome-wide analysis of genetic loci associated with Alzheimer disease.* Jama. **303**(18): p. 1832-40.
109. Stefansson, H., et al., *Common variants conferring risk of schizophrenia.* Nature, 2009. **460**(7256): p. 744-7.
110. Purcell, S.M., et al., *Common polygenic variation contributes to risk of schizophrenia and bipolar disorder.* Nature, 2009. **460**(7256): p. 748-52.
111. Willer, C.J., *Six new loci associated with body mass index highlight a neuronal influence on body weight regulation.* Nat Genet, 2009. **41**: p. 25-34.
112. Thorleifsson, G., *Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity.* Nat Genet, 2009. **41**: p. 18-24.
113. Loos, R.J., *Common variants near MC4R are associated with fat mass, weight and risk of obesity.* Nat Genet, 2008. **40**: p. 768-775.
114. Meyre, D., *Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations.* Nat Genet, 2009. **41**: p. 157-159.
115. Benzinou, M., *Common nonsynonymous variants in PCSK1 confer risk of obesity.* Nat Genet, 2008. **40**: p. 943-945.
116. Hofker, M. and C. Wijmenga, *A supersized list of obesity genes.* Nature genetics, 2009. **41**(2): p. 139-140.
117. Latham, J. and A. Wilson, *The Great DNA Data Deficit: Are Genes for Disease a Mirage?*
118. McCarroll, S., et al., *Integrated detection and population-genetic analysis of SNPs and copy number variation.* Nature genetics, 2008. **40**(10): p. 1166-1174.
119. Thomas, D., *Gene–environment-wide association studies: emerging approaches.* Nature Reviews Genetics, 2010. **11**(4): p. 259-272.
120. Xu, Z. and J.A. Taylor, *SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies.* Nucleic Acids Research, 2009. **37**(suppl 2): p. W600.
121. Neale, B.M. and S. Purcell, *The positives, protocols, and perils of genome wide association.* American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 2008. **147**(7): p. 1288-1294.
122. Simes, R.J., *An improved Bonferroni procedure for multiple tests of significance.* Biometrika, 1986. **73**(3): p. 751.
123. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.
124. Risch, N. and K. Merikangas, *The future of genetic studies of complex human diseases.* Science, 1996. **273**(5281): p. 1516-1517.
125. Tan, E., *Genome-wide Association Studies: Promises and Pitfalls.* Annals of the Academy of Medicine, Singapore, 2010. **39**(2): p. 77-78.

126. Zhang, L., et al., *Pathway-based genome-wide association analysis identified the importance of regulation-of-autophagy pathway for ultradistal radius BMD.* Journal of Bone and Mineral Research, 2010. **25**(7): p. 1572-1580.
127. Baranzini, S., et al., *Pathway and network-based analysis of genome-wide association studies in multiple sclerosis.* Human molecular genetics, 2009. **18**(11): p. 2078.
128. Lowell BB and S. BM, *Towards a molecular understanding of adaptive thermogenesis.* 2000. **404**: p. 652-660.
129. Dourish, C., W. Rycroft, and S. Iversen, *Postponement of satiety by blockade of brain cholecystokinin (CCK-B) receptors.* Science, 1989. **245**(4925): p. 1509.
130. Minokoshi, Y., et al., *AMP-kinase regulates food intake by responding to hormonal and nutrient signals in the hypothalamus.* Nature, 2004. **428**(6982): p. 569-574.
131. Matson, C.A., et al., *Cholecystokinin and leptin act synergistically to reduce body weight.* American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 2000. **278**(4): p. R882.
132. Donahey, J.C.K., et al., *Intraventricular GLP-1 reduces short-but not long-term food intake or body weight in lean and obese rats.* Brain research, 1998. **779**(1-2): p. 75-83.
133. Balthasar, N., *Genetic dissection of neuronal pathways controlling energy homeostasis.* Obesity, 2006. **14**: p. 222S-227S.
134. Brüning, J.C., et al., *Role of brain insulin receptor in control of body weight and reproduction.* Science, 2000. **289**(5487): p. 2122.
135. Tannenbaum, G., H.J. Guyda, and B.I. Posner, *Insulin-like growth factors: a role in growth hormone negative feedback and body weight regulation via brain.* Science, 1983. **220**(4592): p. 77.
136. Zhang, Y., et al., *Positional cloning of the mouse obese gene and its human homologue.* Nature, 1994. **372**(6505): p. 425.
137. Stehling, O., et al., *Leptin reduces juvenile fat stores by altering the circadian cycle of energy expenditure.* American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 1996. **271**(6): p. R1770.
138. Schwartz MW, et al., *Is the energy homeostasis system inherently biased toward weight gain.* 2003. **52**: p. 232-238.
139. Phillips, S., T. Ciaraldi, and D. Oh, *Adiponectin secretion and response to pioglitazone is depot dependent in cultured human adipose tissue.* American Journal of Physiology- Endocrinology And Metabolism, 2008. **295**(4): p. E842.
140. Skurk, T., et al., *Relationship between adipocyte size and adipokine expression and secretion.* Journal of Clinical Endocrinology & Metabolism, 2007. **92**(3): p. 1023.
141. Wang, B. and P. Trayhurn, *Acute and prolonged effects of TNF- on the expression and secretion of inflammation-related adipokines by human adipocytes differentiated in culture.* Pflügers Archiv European Journal of Physiology, 2006. **452**(4): p. 418-427.
142. Maury, E., et al., *Adipokines oversecreted by omental adipose tissue in human obesity.* American Journal of Physiology-Endocrinology And Metabolism, 2007. **293**(3): p. E656.

143. Frayn, K.N., et al., *Coordinated regulation of hormone-sensitive lipase and lipoprotein lipase in human adipose tissue in vivo: implications for the control of fat storage and fat mobilization.* Advances in Enzyme Regulation, 1995. **35**: p. 163-178.
144. Romanski, S.A., R.M. Nelson, and M.D. Jensen, *Meal fatty acid uptake in adipose tissue: gender effects in nonobese humans.* American Journal of Physiology-Endocrinology And Metabolism, 2000. **279**(2): p. E455.
145. Hamosh, M., et al., *Fatty-acid metabolism by human adipose tissues.* Journal of Clinical Investigation, 1963. **42**(10): p. 1648.
146. Minokoshi, Y., et al., *Leptin stimulates fatty-acid oxidation by activating AMP-activated protein kinase.* Nature, 2002. **415**(6869): p. 339-343.
147. Unger, R.H., Y.T. Zhou, and L. Orci, *Regulation of fatty acid homeostasis in cells: novel role of leptin.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(5): p. 2327.
148. Morton, N.M., et al., *Leptin action in intestinal cells.* Journal of Biological Chemistry, 1998. **273**(40): p. 26194.
149. Sanchez-Margalet, V. and C. Martin-Romero, *Human leptin signaling in human peripheral blood mononuclear cells: activation of the JAK-STAT pathway.* Cellular immunology, 2001. **211**(1): p. 30-36.
150. Vaisse, C., et al., *Leptin activation of Stat3 in the hypothalamus of wild-type and ob/ob mice but not db/db mice.* Nature Genetics, 1996. **14**(1): p. 95-97.
151. El-Haschimi, K., et al., *Two defects contribute to hypothalamic leptin resistance in mice with diet-induced obesity.* Journal of Clinical Investigation, 2000. **105**(12): p. 1827-1827.
152. Andersson, U., et al., *AMP-activated protein kinase plays a role in the control of food intake.* Journal of Biological Chemistry, 2004. **279**(13): p. 12005.
153. Martin, T.L., et al., *Diet-induced obesity alters AMP kinase activity in hypothalamus and skeletal muscle.* Journal of Biological Chemistry, 2006. **281**(28): p. 18933.
154. Wang, J., et al., *Overfeeding rapidly induces leptin and insulin resistance.* Diabetes, 2001. **50**(12): p. 2786.
155. Steinberg, G.R., et al., *Leptin increases FA oxidation in lean but not obese human skeletal muscle: evidence of peripheral leptin resistance.* American Journal of Physiology-Endocrinology And Metabolism, 2002. **283**(1): p. E187.
156. Prokesch, A., et al., *Novel Insights into Adipogenesis from Omics Data.* Current Medicinal Chemistry, 2009. **16**(23): p. 2952-2964.
157. Wang, Z., et al., *Leptin resistance of adipocytes in obesity: role of suppressors of cytokine signaling.* Biochemical and Biophysical Research Communications, 2000. **277**(1): p. 20-26.
158. Lin, S., et al., *Development of high fat diet-induced obesity and leptin resistance in C57BI/6J mice.* International Journal of Obesity, 2000. **24**(5): p. 639-646.
159. Kershaw, E.E. and J.S. Flier, *Adipose tissue as an endocrine organ.* Journal of Clinical Endocrinology & Metabolism, 2004. **89**(6): p. 2548.
160. Romero-Corral, A., et al., *Accuracy of body mass index in diagnosing obesity in the adult general population.* International Journal of Obesity, 2008. **32**(6): p. 959-966.

161. Kubota, N., et al., *PPAR [gamma] mediates high-fat diet-induced adipocyte hypertrophy and insulin resistance.* Molecular Cell, 1999. **4**(4): p. 597-609.
162. Nishimura, S., et al., *Adipogenesis in obesity requires close interplay between differentiating adipocytes, stromal cells, and blood vessels.* Diabetes, 2007. **56**(6): p. 1517.
163. Arner, E., et al., *Adipocyte turnover: relevance to human adipose tissue morphology.* Diabetes, 2010. **59**(1): p. 105.
164. Spalding, K., et al., *Dynamics of fat cell turnover in humans.* Nature, 2008. **453**(7196): p. 783-787.
165. Arner, P. and K. Spalding, *Fat cell turnover in humans.* Biochemical and Biophysical Research Communications, 2010. **396**(1): p. 101-104.
166. Rosen, E.D. and O.A. MacDougald, *Adipocyte differentiation from the inside out.* Nat Rev Mol Cell Biol, 2006. **7**(12): p. 885-896.
167. Wolins, N.E., et al., *OP9 mouse stromal cells rapidly differentiate into adipocytes: characterization of a useful new model of adipogenesis.* J Lipid Res, 2006. **47**(2): p. 450-60.
168. Rosen, E.D., et al., *PPAR [gamma] is required for the differentiation of adipose tissue in vivo and in vitro.* Molecular Cell, 1999. **4**(4): p. 611-617.
169. Cao, Z., R. Umek, and S. McKnight, *Regulated expression of three C/EBP isoforms during adipose conversion of 3T3-L1 cells.* Genes & development, 1991. **5**(9): p. 1538.
170. Kim, J.B. and B.M. Spiegelman, *ADD1/SREBP1 promotes adipocyte differentiation and gene expression linked to fatty acid metabolism.* Genes & development, 1996. **10**(9): p. 1096.
171. Heaton, J.M., *The distribution of brown adipose tissue in the human.* J Anat, 1972. **112**(Pt 1): p. 35-9.
172. Djian, P., D. Roncari, and C. Hollenberg, *Adipocyte precursor clones vary in capacity for differentiation.* Metabolism, 1985. **34**(9): p. 880-883.
173. Siddharthan, R., *Parsing regulatory DNA: general tasks, techniques, and the PhyloGibbs approach.* J Biosci, 2007. **32**(5): p. 863-70.
174. Brown, T., *Genomes 3.* detail, 2006.
175. Lodish, H. and S. Zipursky, *Molecular cell biology.* Biochemistry and Molecular Biology Education, 2001. **29**: p. 126-133.
176. Cooper, G. and R. Hausman, *The cell: a molecular approach*. 2000: ASM Press Washington, DC.
177. Tabach, Y., et al., *Wide-scale analysis of human functional transcription factor binding reveals a strong bias towards the transcription start site.* PLoS One, 2007. **2**(8): p. e807.
178. Koudritsky, M. and E. Domany, *Positional distribution of human transcription factor binding sites.* Nucleic Acids Res, 2008. **36**(21): p. 6795-805.
179. Brown, C.T., *Chapter 18 Computational Approaches to Finding and Analyzing cis[hyphen (true graphic)]Regulatory Elements*
Methods in Cell Biology, in *Avian Embryology, 2nd Edition*, D.M. Bronner-Fraser, Editor. 2008, Academic Press. p. 337-365.
180. Lenhard, B., et al., *Identification of conserved regulatory elements by comparative genome analysis.* Journal of Biology, 2003. **2**(2): p. 13.

181. Zhang, Z. and M. Gerstein, *Of mice and men: phylogenetic footprinting aids the discovery of regulatory elements.* Journal of Biology, 2003. **2**(2): p. 11.
182. *Genetic Analysis Tools Help Define Nature and Nurture in Complex Disorders*, National Human Genome Research Institute, NIH.
183. Glazier, A., J. Nadeau, and T. Aitman, *Finding genes that underlie complex traits.* Science, 2002. **298**(5602): p. 2345.
184. Collins, F., *Science Reporters' Seminar on Genome-Wide Association Studies*, NHGRI, Editor. 2007.
185. Guttmacher, A., *Implementing personalized health care: a research agenda?* 2009, National Human Genome Research Institute.
186. Tiffin, N., et al., *Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes.* Nucleic Acids Res, 2006. **34**(10): p. 3067-81.
187. Xiong, M., et al., *A systems biology approach to genetic studies of complex diseases.* FEBS Letters, 2005. **579**(24): p. 5325-5332.
188. Bischof, J.M. and R. Wevrick, *Genome-wide analysis of gene transcription in the hypothalamus*
10.1152/physiolgenomics.00071.2005. Physiol. Genomics, 2005. **22**(2): p. 191-196.
189. Mootha, V.K., et al., *PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.* Nat Genet, 2003. **34**(3): p. 267-273.
190. Ye, C., et al., *Using network component analysis to dissect regulatory networks mediated by transcription factors in yeast.* PLoS Comput Biol, 2009. **5**(3).
191. Scarpulla, R.C., *Nuclear activators and coactivators in mammalian mitochondrial biogenesis.* Biochim Biophys Acta, 2002. **1576**(1-2): p. 1-14.
192. Hughes, J.D., et al., *Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae.* J Mol Biol, 2000. **296**(5): p. 1205-1214.
193. Eisen, M.B., et al., *Cluster analysis and display of genome-wide expression patterns.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(25): p. 14863.
194. Lathrop, G., et al., *Construction of human linkage maps: likelihood calculations for multilocus linkage analysis.* Genetic epidemiology, 1986. **3**(1): p. 39-52.
195. Lathrop, G., et al., *Strategies for multilocus linkage analysis in humans.* Proceedings of the National Academy of Sciences of the United States of America, 1984. **81**(11): p. 3443.
196. Segal, E., et al., *Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data.* Nature genetics, 2003. **34**(2): p. 166-176.
197. Efron, B. and R. Tibshirani, *On testing the significance of sets of genes.* The Annals of Applied Statistics, 2007. **1**(1): p. 107-129.
198. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(43): p. 15545.

199. Kaletta, T. and M.O. Hengartner, *Finding function in novel targets: C. elegans as a model organism.* Nat Rev Drug Discov, 2006. **5**(5): p. 387-98.
200. Hulsen, T., et al., *Benchmarking ortholog identification methods using functional genomics data.* Genome Biol, 2006. **7**(4): p. R31.
201. Tamas, I., et al., *A combined approach exploring gene function based on worm-human orthology.* BMC Genomics, 2005. **6**(1): p. 65.
202. Wheelan, S.J., et al., *Human and nematode orthologs--lessons from the analysis of 1800 human genes and the proteome of Caenorhabditis elegans.* Gene, 1999. **238**(1): p. 163-70.
203. NCBI, *BLAST.*
204. NCBI, *HomoloGene.*
205. O'Brien, K.P., M. Remm, and E.L.L. Sonnhammer, *Inparanoid: a comprehensive database of eukaryotic orthologs*
10.1093/nar/gki107. Nucl. Acids Res., 2005. **33**(suppl_1): p. D476-480.
206. Ashrafi, K., et al., *Genome-wide RNAi analysis of Caenorhabditis elegans fat regulatory genes.* Nature, 2003. **421**(6920): p. 268-72.
207. Beißbarth, T. and T.P. Speed, *GOstat: find statistically overrepresented Gene Ontologies within a group of genes.* Bioinformatics, 2004. **20**(9): p. 1464.
208. Currie, R.A., et al., *Gene ontology mapping as an unbiased method for identifying molecular pathways and processes affected by toxicant exposure: application to acute effects caused by the rodent non-genotoxic carcinogen diethylhexylphthalate.* Toxicological Sciences, 2005. **86**(2): p. 453.
209. Perusse, L., et al., *The human obesity gene map: the 2004 update.* Obes Res, 2005. **13**(3): p. 381-490.
210. Flicek, P., et al., *Ensembl 2008.* Nucleic acids research, 2008. **36**(suppl 1): p. D707.
211. Defrance, M. and H. Touzet, *Predicting transcription factor binding sites using local over-representation and comparative genomics.* BMC Bioinformatics, 2006. **7**: p. 396.
212. Aerts, S., et al., *TOUCAN 2: the all-inclusive open source workbench for regulatory sequence analysis.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. 393-396.
213. Romer, K.A., G.-R. Kayombya, and E. Fraenkel, *WebMOTIFS: automated discovery, filtering and scoring of DNA sequence motifs using multiple programs and Bayesian approaches.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. 217-220.
214. Sandelin, A., et al., *JASPAR: an open-access database for eukaryotic transcription factor binding profiles.* Nucleic Acids Res, 2004. **32**(Database issue): p. 91-94.
215. Wingender, E., et al., *TRANSFAC: an integrated system for gene expression regulation.* Nucleic Acids Res, 2000. **28**(1): p. 316-319.
216. Klar, J., et al., *RAR-related orphan receptor A isoform 1 (RORa1) is disrupted by a balanced translocation t (4; 15)(q22. 3; q21. 3) associated with severe obesity.* European journal of human genetics, 2005. **13**(8): p. 928-934.

217. Thai, M.V., et al., *Myocyte Enhancer Factor 2 (MEF2)-Binding Site Is Required forGLUT4 Gene Expression in Transgenic Mice.* Journal of Biological Chemistry, 1998. **273**(23): p. 14285.
218. Huang, H.Y., et al., *Transcription factor YY1 promotes adipogenesis via inhibiting CHOP-10 expression.* Biochemical and biophysical research communications, 2008. **375**(4): p. 496-500.
219. Lau, P., et al., *The Orphan Nuclear Receptor, ROR , Regulates Gene Expression That Controls Lipid Metabolism.* Journal of Biological Chemistry, 2008. **283**(26): p. 18411.
220. van Tienen, F.H., et al., *Prolonged Nrf1 overexpression triggers adipocyte inflammation and insulin resistance.* Journal of Cellular Biochemistry, 2010. **111**(6): p. 1575-1585.
221. Fukuda, H. and N. Iritani, *Regulation of ATP citrate-lyase gene expression in hepatocytes and adipocytes in normal and genetically obese rats.* Journal of biochemistry, 1999. **126**(2): p. 437.
222. Muharram, G., et al., *Endocrine pancreatic tissue plasticity in obese humans is associated with cytoplasmic expression of PBX-1 in pancreatic ductal cells.* Biochemical and biophysical research communications, 2005. **333**(4): p. 1153-1159.
223. Hsu, L.A., et al., *Effect of obesity on the association between common variations in the HNF1A gene region and C-reactive protein level in Taiwanese.* Clinica Chimica Acta, 2010.
224. Tong, Q., et al., *Interaction between GATA and the C/EBP family of transcription factors is critical in GATA-mediated suppression of adipocyte differentiation.* Molecular and cellular biology, 2005. **25**(2): p. 706.
225. Tong, Q., et al., *Function of GATA transcription factors in preadipocyte-adipocyte transition.* Science, 2000. **290**(5489): p. 134.
226. Hu, Y., et al., *Inhibitory effect and transcriptional impact of berberine and evodiamine on human white preadipocyte differentiation.* Fitoterapia, 2010. **81**(4): p. 259-268.
227. Moreno-Hagelsieb, G. and K. Latimer, *Choosing BLAST options for better detection of orthologs as reciprocal best hits.* Bioinformatics, 2008. **24**(3): p. 319-24.
228. Sonnhammer, E.L. and R. Durbin, *Analysis of protein domain families in Caenorhabditis elegans.* Genomics, 1997. **46**(2): p. 200-16.
229. Amsterdam, A., et al., *Identification of 315 genes essential for early zebrafish development.* Proc Natl Acad Sci U S A, 2004. **101**(35): p. 12792-7.
230. Das, U.N. and A.A. Rao, *Gene expression profile in obesity and type 2 diabetes mellitus.* Lipids Health Dis, 2007. **6**: p. 35.
231. Parekh, S. and F.A. Anania, *Abnormal lipid and glucose metabolism in obesity: implications for nonalcoholic fatty liver disease.* Gastroenterology, 2007. **132**(6): p. 2191-207.
232. Takahashi, N., et al., *Genome science of lipid metabolism and obesity.* Forum Nutr, 2009. **61**: p. 25-38.
233. Ahima, R.S., et al., *Regulation of neuronal and glial proteins by leptin: implications for brain development.* Endocrinology, 1999. **140**(6): p. 2755-62.

234. Ahima, R.S., *Adipose tissue as an endocrine organ.* Obesity (Silver Spring), 2006. **14 Suppl 5**: p. 242S-249S.
235. Udagawa, J., et al., *The role of leptin in the development of the cortical neuron in mouse embryos.* Brain Res, 2006. **1120**(1): p. 74-82.
236. Udagawa, J., et al., *Roles of leptin in prenatal and perinatal brain development.* Congenit Anom (Kyoto), 2007. **47**(3): p. 77-83.
237. Chen, H. and B.M. Sharp, *Content-rich biological network constructed by mining PubMed abstracts.* BMC Bioinformatics, 2004. **5**: p. 147.
238. Reid, J.E., et al., *Variable structure motifs for transcription factor binding sites.* BMC Genomics. **11**: p. 30.
239. Salama, R.A. and D.J. Stekel, *Inclusion of neighboring base interdependencies substantially improves genome-wide prokaryotic transcription factor binding site prediction.* Nucleic Acids Res. **38**(12): p. e135.
240. Fu, W., P. Ray, and E.P. Xing, *DISCOVER: a feature-based discriminative method for motif search in complex genomes.* Bioinformatics, 2009. **25**(12): p. i321-9.
241. Boily, G., et al., *Connections Between ETV6-Modulated Genes: Identification of Shared Features.* Cancer Informatics, 2008. **6**: p. 183.
242. Wang, Y., et al., *Analysis of porcine transcriptional response to Salmonella enterica serovar Choleraesuis suggests novel targets of NFkappaB are activated in the mesenteric lymph node.* BMC genomics, 2008. **9**(1): p. 437.
243. Romero, D.G., et al., *Gene expression profile in rat adrenal zona glomerulosa cells stimulated with aldosterone secretagogues.* Physiological genomics, 2007. **32**(1): p. 117.
244. Visscher, P.M., *Sizing up human height variation.* Nature Genetics, 2008. **40**(5): p. 489-490.
245. Dupuis, J. and C.J. O'Donnell, *Interpreting results of large-scale genetic association studies.* JAMA: The Journal of the American Medical Association, 2007. **297**(5): p. 529.
246. Gorlov, I.P., et al., *GWAS meets microarray: are the results of genome-wide association studies and gene-expression profiling consistent? Prostate cancer as an example.* PloS one, 2009. **4**(8): p. e6511.
247. Li, C., et al., *Prioritized subset analysis: improving power in genome-wide association studies.* Human heredity, 2007. **65**(3): p. 129-141.
248. Chen, R., et al., *FitSNPs: highly differentially expressed genes are more likely to have variants associated with disease.* Genome biology, 2008. **9**(12): p. R170.
249. Reimand, J., et al., *g: Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments.* Nucleic acids research, 2007. **35**(Web Server issue): p. W193.
250. Di Iorgi, N., et al., *Reciprocal relation between marrow adiposity and the amount of bone in the axial and appendicular skeleton of young adults.* J Clin Endocrinol Metab, 2008. **93**(6): p. 2281-2286.
251. Naaz, A., et al., *Loss of cyclin-dependent kinase inhibitors produces adipocyte hyperplasia and obesity.* FASEB J, 2004. **18**(15): p. 1925-1927.

252. Zhang, P., et al., *Negative cross-talk between hematopoietic regulators: GATA proteins repress PU. 1.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(15): p. 8705.
253. Zhang, J., et al., *The role of lipocalin 2 in the regulation of inflammation in adipocytes and macrophages.* Mol Endocrinol, 2008. **22**(6): p. 1416-1426.
254. Catalán, V., et al., *Increased adipose tissue expression of lipocalin-2 in obesity is related to inflammation and matrix metalloproteinase-2 and metalloproteinase-9 activities in humans.* Journal of molecular medicine, 2009. **87**(8): p. 803-813.
255. Mandrup, S. and M. Lane, *Regulating adipogenesis.* Journal of Biological Chemistry, 1997. **272**(9): p. 5367.
256. Arimura, N., et al., *The peroxisome proliferator-activated receptor regulates expression of the perilipin gene in adipocytes.* Journal of Biological Chemistry, 2004. **279**(11): p. 10070.
257. Tong, Q., et al., *Function of GATA transcription factors in preadipocyte-adipocyte transition.* Science, 2000. **290**(5489): p. 134.
258. Shimada, T., et al., *Unexpected blockade of adipocyte differentiation by K-7174: implication for endoplasmic reticulum stress.* Biochem Biophys Res Commun, 2007. **363**(2): p. 355-360.
259. Minguell, J., A. Erices, and P. Conget, *Mesenchymal stem cells.* Experimental biology and medicine (Maywood, NJ), 2001. **226**(6): p. 507.
260. Fruhbeck, G., et al., *The adipocyte: a model for integration of endocrine and metabolic signaling in energy metabolism regulation.* American Journal of Physiology- Endocrinology And Metabolism, 2001. **280**(6): p. E827.
261. Large, V., et al., *Metabolism of lipids in human white adipocyte.* Diabetes & metabolism, 2004. **30**(4): p. 294-309.
262. Brun, R. and B. Spiegelman, *PPAR gamma and the molecular control of adipogenesis.* Journal of Endocrinology, 1997. **155**(2): p. 217.
263. Cowherd, R., R. Lyle, and R. McGehee. *Molecular regulation of adipocyte differentiation*. in *Seminars in Cell and Developmental Biology*. 1999.
264. Rosen, E., et al., *Transcriptional regulation of adipogenesis.* Genes & development, 2000. **14**(11): p. 1293.
265. Lefterova, M. and M. Lazar, *New developments in adipogenesis.* Trends in Endocrinology & Metabolism, 2009. **20**(3): p. 107-114.
266. Shao, D. and M.A. Lazar, *Peroxisome proliferator activated receptor gamma, CCAAT/enhancer-binding protein alpha, and cell cycle status regulate the commitment to adipocyte differentiation.* J Biol Chem, 1997. **272**(34): p. 21473-8.
267. Reichert, M. and D. Eick, *Analysis of cell cycle arrest in adipocyte differentiation.* Oncogene, 1999. **18**(2): p. 459-66.
268. Cook A and C. Cowan, *Adipose BTI - StemBook*. 2009, Harvard Stem Cell Institute DP - 2008.
269. Mackall, J., et al., *Induction of lipogenesis during differentiation in a" preadipocyte" cell line.* Journal of Biological Chemistry, 1976. **251**(20): p. 6462.
270. Singh, R., et al., *Testosterone Inhibits Adipogenic Differentiation in 3T3-L1 Cells: Nuclear Translocation of Androgen Receptor Complex with {beta}-Catenin and T-Cell Factor 4 May Bypass Canonical Wnt Signaling to Down-Regulate Adipogenic Transcription Factors.* Endocrinology, 2006. **147**(1): p. 141.

271. *3T3-L1 fibroblasts differentiate in culture into cells having adipocyte.* - The Journal of biological chemistry, 1976(- 0021-9258 (Print)): p. - 6462-4.
272. Chung, N., et al., *An efficient and fully automated high-throughput transfection method for genome-scale siRNA screens.* Journal of Biomolecular Screening, 2008. **13**(2): p. 142.
273. Suwa, A., T. Kurama, and T. Shimokawa, *Adipocyte hyperplasia and RMI1 in the treatment of Obesity.* FEBS Journal.
274. Suwa, A. and T. Shimokawa, *Emerging targets for the treatment of obesity.* FEBS Journal.
275. Poulos, S.P., M.V. Dodson, and G.J. Hausman, *Cell line models for differentiation: Preadipocytes and adipocytes.* Experimental Biology and Medicine, 2010. **235**(10): p. 1185.
276. Wang, F. and Q. Tong, *Transcription factor PU.1 is expressed in white adipose and inhibits adipocyte differentiation.* Am J Physiol Cell Physiol, 2008. **295**(1): p. C213-20.
277. Lefterova, M., et al., *Cell-specific determinants of peroxisome proliferator-activated receptor {gamma} function in adipocytes and macrophages.* Molecular and cellular biology, 2010. **30**(9): p. 2078.
278. Corella, D., et al., *Association between the APOA2 promoter polymorphism and body weight in Mediterranean and Asian populations: replication of a gene–saturated fat interaction.* International Journal of Obesity, 2010.
279. Phillips, C.M., et al., *Complement component 3 polymorphisms interact with polyunsaturated fatty acids to modulate risk of metabolic syndrome.* The American journal of clinical nutrition, 2009. **90**(6): p. 1665.
280. Huang, T., et al., *MAT1A variants modulate the effect of dietary fatty acids on plasma homocysteine concentrations.* Nutrition, metabolism, and cardiovascular diseases: NMCD, 2010.
281. Corella, D., et al., *Association of the LCT-13910C> T Polymorphism With Obesity and Its Modulation by Dairy Products in a Mediterranean Population.* Obesity, 2010.
282. Barish, G.D., V.A. Narkar, and R.M. Evans, *PPAR delta: a dagger in the heart of the metabolic syndrome.* J Clin Invest, 2006. **116**(3): p. 590-597.
283. Chan, E., et al., *The V227A polymorphism at the PPARA locus is associated with serum lipid concentrations and modulates the association between dietary polyunsaturated fatty acid intake and serum high density lipoprotein concentrations in Chinese women.* Atherosclerosis, 2006. **187**(2): p. 309-315.
284. Corella, D., et al., *APOA2, dietary fat, and body mass index: replication of a gene-diet interaction in 3 independent populations.* Archives of internal medicine, 2009. **169**(20): p. 1897.
285. Perusse, L. and C. Bouchard, *Gene-diet interactions in obesity.* American Journal of Clinical Nutrition, 2000. **72**(5): p. 1285S.
286. Loos, R. and T. Rankinen, *Gene-diet interactions on body weight changes.* Journal of the American Dietetic Association, 2005. **105**(5): p. 29-34.
287. Uusitupa, M., *Gene-diet interaction in relation to the prevention of obesity and type 2 diabetes: evidence from the Finnish Diabetes Prevention Study.* Nutrition, Metabolism and Cardiovascular Diseases, 2005. **15**(3): p. 225-233.

288. Memisoglu, A., et al., *Interaction between a peroxisome proliferator-activated receptor gene polymorphism and dietary fat intake in relation to body mass.* Human molecular genetics, 2003. **12**(22): p. 2923.
289. Desai, S., et al., *PU. 1 regulates positive regulatory domain I-binding factor 1/Blimp-1 transcription in lymphoma cells.* The Journal of Immunology, 2009. **183**(9): p. 5778.
290. Burda, P., et al., *PU. 1 Activation Relieves GATA-1–Mediated Repression of Cebpa and Cbfb during Leukemia Differentiation.* Molecular Cancer Research, 2009. **7**(10): p. 1693.
291. Koschmieder, S., et al., *Role of transcription factors C/EBP and PU. 1 in normal hematopoiesis and leukemia.* International journal of hematology, 2005. **81**(5): p. 368-377.
292. Tissieres, P., et al., *Cooperation between PU.1 and CAAT/enhancer-binding protein beta is necessary to induce the expression of the MD-2 gene.* J Biol Chem, 2009. **284**(39): p. 26261-72.
293. Allagnat, F., et al., *ICER-1gamma overexpression drives palmitate-mediated connexin36 down-regulation in insulin-secreting cells.* J Biol Chem, 2008. **283**(9): p. 5226-34.
294. Zhou, Y., et al., *Overexpression of repressive cAMP response element modulators in high glucose and fatty acid-treated rat islets.* Journal of Biological Chemistry, 2003. **278**(51): p. 51316.
295. Gorjão, R., et al., *Effect of docosahexaenoic acid-rich fish oil supplementation on human leukocyte function.* Clinical Nutrition, 2006. **25**(6): p. 923-938.
296. Lai, C.-Q., et al., *Fenofibrate effect on triglyceride and postprandial response of apolipoprotein A5 variants: the GOLDN study.* Arterioscler Thromb Vasc Biol, 2007. **27**(6): p. 1417-1425.
297. Furukawa, S., et al., *Increased oxidative stress in obesity and its impact on metabolic syndrome.* J Clin Invest, 2004. **114**(12): p. 1752-61.
298. Hikami, K., et al., *Association of a functional polymorphism in the 3 untranslated region of SPI1 with systemic lupus erythematosus.* Arthritis & Rheumatism, 2011. **63**(3): p. 755-763.
299. Kissebah, A.H., et al., *Quantitative trait loci on chromosomes 3 and 17 influence phenotypes of the metabolic syndrome.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(26): p. 14478.
300. Feinleib, M., et al., *The Framingham offspring study. Design and preliminary data.* Preventive Medicine, 1975. **4**(4): p. 518-525.
301. Lai, C., et al., *PPARGC1A variation associated with DNA damage, diabetes, and cardiovascular diseases: the Boston Puerto Rican Health Study.* Diabetes, 2008. **57**(4): p. 809.
302. Carter, C., *Genetics of common disorders.* British Medical Bulletin, 1969. **25**(1): p. 52.
303. Gu, Z.M., et al., *PU. 1 directly regulates retinoic acid-induced expression of RIG-G in leukemia cells.* FEBS letters, 2010.
304. Pongubala, J.M. and M.L. Atchison, *Activating transcription factor 1 and cyclic AMP response element modulator can modulate the activity of the immunoglobulin kappa 3' enhancer.* J Biol Chem, 1995. **270**(17): p. 10304-13.

305. Arkadianos, I., et al., *Improved weight management using genetic information to personalize a calorie controlled diet.* Nutrition Journal, 2007. **6**(1): p. 29.
306. Reinehr, T., et al., *Lifestyle intervention in obese children with variations in the melanocortin 4 receptor gene.* Obesity, 2008. **17**(2): p. 382-389.
307. Naveiras, O., et al., *Bone-marrow adipocytes as negative regulators of the haematopoietic microenvironment.* Nature, 2009. **460**(7252): p. 259-263.
308. Sugimura, R. and L. Li, *Shifting in balance between osteogenesis and adipogenesis substantially influences hematopoiesis.* Journal of Molecular Cell Biology, 2010. **2**(2): p. 61.
309. Snyder, F., *Fatty acid oxidation in irradiated bone marrow cells.* 1965.
310. Bonadies, N., et al., *PU. 1 is regulated by NF- B through a novel binding site in a 17 kb upstream enhancer element.* Oncogene, 2009. **29**(7): p. 1062-1072.
311. Michaels, J.L., V. Naudot, and L.S. Liebovitch, *Dynamic Stabilization in the PU1-GATA1 Circuit Using a Model with Time-Dependent Kinetic Change.* Bulletin of Mathematical Biology: p. 1-20.
312. Leddin, M., et al., *Two distinct auto-regulatory loops operate at the PU. 1 locus in B cells and myeloid cells.* Blood, 2011.
313. Janssens, A., et al., *A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions.* The American Journal of Human Genetics, 2008. **82**(3): p. 593-599.