Genetic variation at the PAT loci: associations, gene-diet interactions and obesity risk.

A dissertation

submitted by

Kris Richardson

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Genetics

Tufts University

Sackler School of Graduate Biomedical Sciences

(August 2011)

Adviser: Jose M. Ordovas

ABSTRACT:

Currently in the United States 66% of Americans aged twenty or greater are either overweight or obese. However, both medical and nutritional recommendations derived from population-based studies have had a limited impact on effective and sustained weight loss. This intra-individual difference under similar environmental conditions has been attributed to genetic factors, or the interaction of both genetic and environmental factors.

We investigated whether genetic variation in the human PAT candidate gene family can modulate anthropometric related traits alone, or in response to dietary fat. We identified several common SNPs modulating anthropomorphic traits. Furthermore, several of these variants show interaction with dietary fat to modulate anthropometric traits, the most convincing being the *PLIN4* SNP rs8887. We demonstrate the minor A allele of rs8887 created a seed site for microRNA-522. Using ex-vivo techniques we demonstrated that rs8887 is able to modulate PLIN4 levels through this microRNA site. Phylogenetic analysis suggests the microRNA-522 target site created by the rs8887 minor allele is undergoing genetic drift among human populations.

To determine if the effect of rs8887 is a rare occurrence or an example of a more common form of functional SNP that can affect miR binding, we utilized newly released SNP data from the 1000 Genomes Project to perform a genome-wide scan of SNPs that abrogate or create microRNA recognition element seed sites (MRESS). We estimate the percent of SNPs falling within both validated (5%) and predicted miR seed sites (3%) and determine the number appreciable in the human genome. We identified 87 predicted miR seed site SNPs are listed in GWAS association studies, or in strong LD with a GWAS

SNP and that may represent the functional variants of identified GWAS SNPs. We also gather previously published co-expression and eQTL data supporting a functional role for four of these SNPs shown to associate with disease phenotypes.

Our data provide further insight into the link between variation in the PAT gene family and anthropometric related traits in humans. Furthermore, with relevance to human disease we show that publicly available resources can be used to identify a class of high priority candidate SNPs for functional studies.

Table of Contents:

List of Tables:

List of Figures:

Chapter 1. INTRODUCTION

The World Health Organization (WHO) estimates 1.6 billion people are overweight and 400 million obese (www.who.int). Over the last two decades in the United States the prevalence of obesity has doubled, such that 66% of Americans aged twenty or greater are either overweight or obese. Similar increases in excess weight have been observed in children suggesting that the obesity pandemic will become more widespread in the coming years (Bartel 2009). Those affected are at increased risk for occurrence of cardiovascular disease (CVD), type 2 diabetes (T2D), metabolic syndrome (MetS), dyslipidemia and other chronic conditions that reduce both quality of life and life expectancy (Fontaine et al. 2003; Haslam and James 2005; Ogden 2006) (**Figure 1**).

A boon, or consequence, of the industrialized world has been the over abundance of calorie rich food sources as well as a high prevalence of sedentary working and leisure activities (French et al. 2001). Therefore, a common explanation for the cause of this pandemic seems easily diagnosed; too many calories and too little exercise induce biochemical and physiological changes leading to weight gain. However, both medical and nutritional recommendations derived from population-based studies have had a limited impact on effective and sustained weight loss (Kaput et al. 2005; Weiss et al. 2007). The observation that the individual response of some is hyper-sensitive, while the response of others is insensitive to similar weight-loss treatments suggests more than simply one's environment determines susceptibility to obesity.

There is no question that environmental factors such as diet and exercise influence obesity susceptibility; however a significant contribution has been attributed to genetic

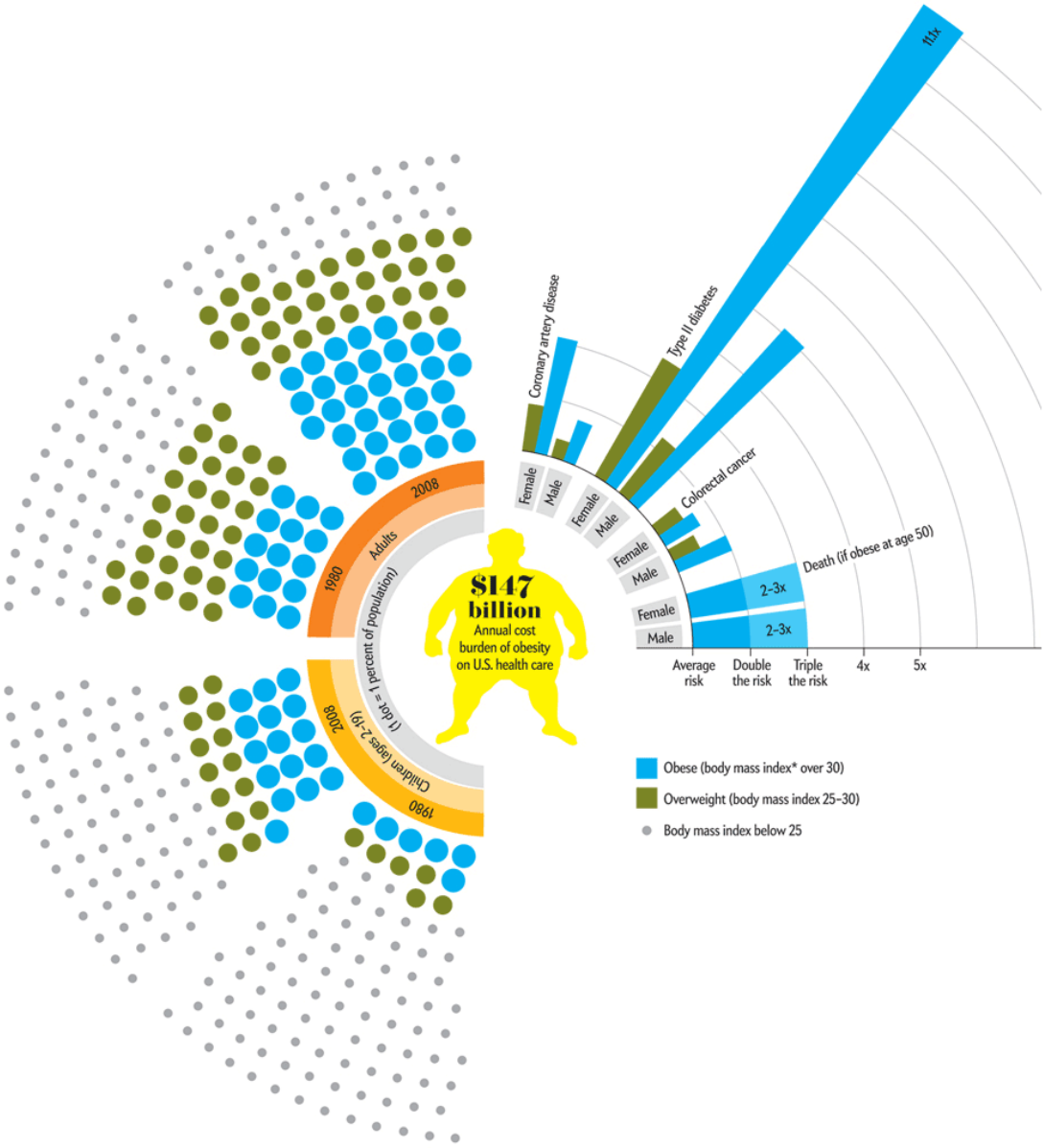


Figure 1. Prevalence of Obesity and Overweight in the US.
This slide shows the prevalence of Overweight and obesity among US adults over the last 30 years. In addition to the added risk to related conditions such as CVD and T2D, it contributing to increases in health care cost.

An image reproduced from an article appearing in Scientific American (http://www.scientificamerican.com/article.cfm?id=dying-to-eat).

factors. A seminal study examined identical twins some of whom were reared together and others apart. The results of this study showed that the correlation coefficient for BMI between identical twins raised apart was significant at 0.70. This value was only slightly less than that of twins raised together, indicating a significant genetic contribution to variability in anthropometrics (Stunkard et al. 1990). An additional study investigated body weight and size among adopted children compared to their biological parents. This work demonstrated that the adopted children were more similar to their biological parents than the adopted parents for various measures of body size and weight indicating that heritability is a more powerful determinant of anthropometrics (Stunkard et al. 1986). This work and other family and twin studies have lead to the estimation that 40-60% of variability in human obesity related phenotypes is heritable (Arner 2000; Bell et al. 2005; Lyon and Hirschhorn 2005).

While rare autosomal dominant mutations explaining severe forms of obesity have been well documented in loci such as the leptin receptor (LEPR) and melanocortin-4 receptor (MC4R), locating single nucleotide polymorphisms (SNPs) contributing significantly to common polygenic obesity has been less successful. It is hypothesized that common complex diseases, such as obesity, occur as a consequence of common genetic variation - the common disease, common variant hypothesis (Reich and Lander 2001). In this scenario, risk for disease is dependent on the collective contribution of genetic variants with small to moderate effect size, which an individual happens to carry. The search for these factors in genome-wide association studies (GWAS) and candidate-gene studies has identified numerous SNPs influencing obesity risk; however many of these associations have not been convincingly replicated in follow-up studies. This may

be due to a number of reasons, such as inadequate sample size for detection of small effect size, poor coverage of genomic regions of interest or the reporting of false positives (Lohmueller et al. 2003). Even the most recent GWAS with a sample size of more than 200,000 individuals and over 2 million SNPs investigated identify only 6-11% of the heritability of BMI (Speliotes et al. 2010).

One explanation, often not considered in epidemiological population-based studies, for both the lack of reproducibility and limited discovery of obesity modulating SNPs is that the attributable risk of common SNPs for obesity is manifested through interactions with environmental factors (GxE interaction), the most obvious being diet (Qi and Cho 2008). A GxE interaction occurs when the response of a phenotype (eg. BMI) to an environmental stimulus (eg. Diet) is dependent on the individual's genetic make-up. For example, it may be that the minor, or major, allele of a particular SNP is found to associate with an increase in disease risk only in the presence of a particular environmental stimulus that is above, or below, a particular threshold. Furthermore, this association may be missed in other studies that do not account for the relevant environmental factors in their tests for association. In support of these hypotheses, a growing number of SNPs have been identified, modulating obesity related phenotypes in response to dietary factors (Ordovas and Corella 2004; Corella and Ordovas 2005).

An example of a GxE interaction was noted for variants in the *ABCA1* gene, which belongs to the family of ABC transporters and it is involved in cholesterol homeostasis. Variants within the *ABCA1* locus have been associated with atherosclerosis, HDL metabolism and postprandial lipid metabolism. A recent intervention study demonstrated that subjects homozygous for the major alleles of rs4149272 and rs2575875 had lower

postprandial lipemia responses compared to minor allele carriers after consumption of a high fat meal. In this case those subjects homozygous for the major allele have reduced risk for CVD compared with carriers of the minor allele (van Herpen and Schrauwen-Hinderling 2008; Delgado-Lista et al. 2010).

Ultimately, the collection and cataloging of the genetic factors modulating phenotypic response to different dietary factors may provide health care professionals with a better understanding of disease risk and may lead to more effective therapeutic strategies in the form of personalized nutrition (Laclaustra et al. 2007).

The lipid storage droplet (LSD) proteins: gate-keepers of lipid storage and turnover.

The basis for obesity is the inability of the individual to maintain the balance between energy uptake, storage and expenditure. The adipocyte plays a critical role in this complex equilibrium as well as protecting against the potential lipo-toxic damage of circulating free fatty acids (FFA) by acting as an intracellular sink for triglycerides (TAG) in LSDs. Traditionally, the adipocyte has been looked at as a simple fat storage depot with few intrinsic metabolic properties, providing mechanical cushioning, heat insulation and storage for fats in the form of TAGs. The cells of the adipose tissue are now recognized to produce metabolically active substances, or adipokines, that initiate feedback mechanisms regulating appetite, food-intake, glucose-utilization and energy-expenditure (Koutsari 2006). It has been hypothesized that some of the adverse metabolic consequences related to obesity are the result of overwhelming the buffering capacities of the adipose tissue, resulting in an overflow of FFAs toward other non-

adipose tissues, a process which has been associated with insulin resistance and decreased clearance of TAG rich particles (Unger 2003; van Herpen and Schrauwen-Hinderling 2008; Vanherpen and Schrauwenhinderling 2008) (**Figure 2**). This idea is supported by work in model organisms in which the ability to store fat has been compromised. For example the phenotypes of insulin resistance, hyperglycemia and liver steatosis affecting lipoatrophic mice are rescued upon surgical implantation of adipose tissue (Gavrilova et al. 2000). As the role of the adipocyte in disease becomes more apparent, studies are focusing on identifying factors regulating the pathways behind FFA storage and turnover.

LSDs are found in a wide range of organisms from fungi to insects to mammals. LSDs store FFA within the cell for later use as an energy source, for signaling factors and for membrane components. As noted above, the dysregulation of FFA metabolism is associated with obesity, CVD and T2D. Therefore variants in genes involved in the uptake and turnover of lipids in the cell are candidates for investigation as contributors to disease risk. It is for this reason that recent studies have been aimed at discovering and understanding the biochemical properties and functions of LSD proteins (BECKMAN 2006). Indeed, accumulating evidence now points to the proteins around and within the LSDs of the adipocyte, as playing an important role in lipid metabolism and energy balance by regulating the uptake and release of lipids in the cell. These proteins include the PAT family; *perilipin(PLIN)/PLIN1, adipose differentiation related protein(ADRP)/PLIN2, tail interacting protein 47(TIP47)/PLIN3, S3-12/PLIN4 and Lipid Storage Droplet Protein 5 (LSDP5)/PLIN5* (Brasaemle 2007; Ducharme and Bickel 2008).

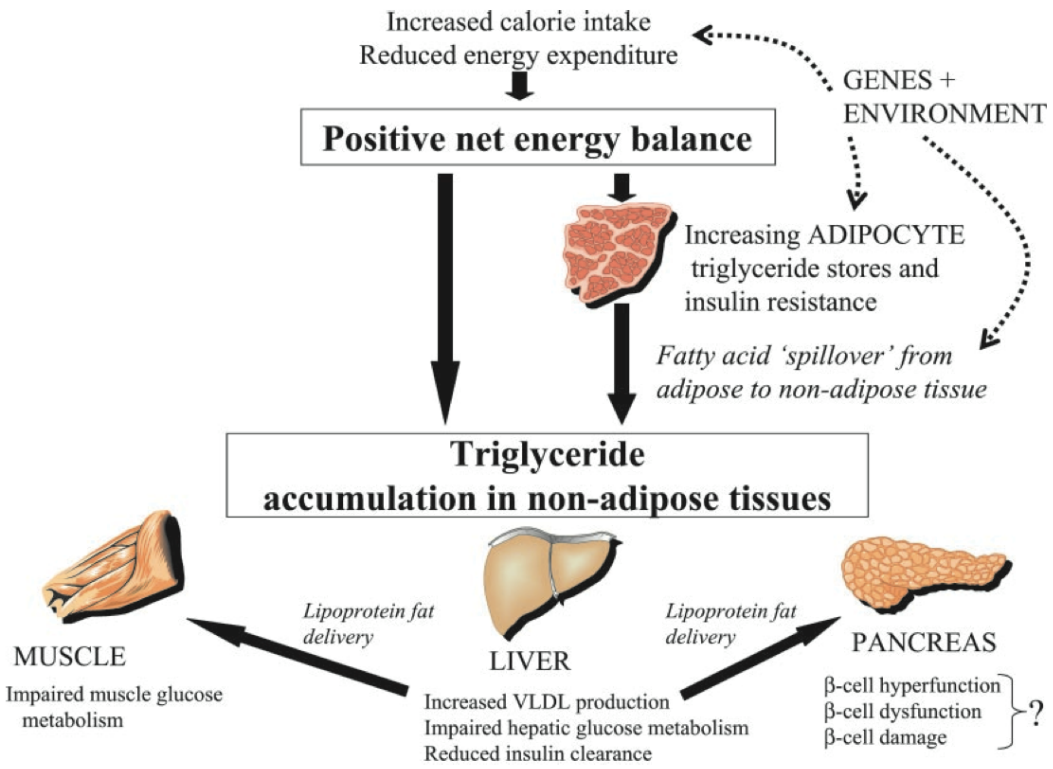


Figure 2. Consequences of a net positive energy balance in the obese.
This illustration demonstrates the consequences of the over abundance of energy in the obese. Through genetic and environmental factors, the adipose tissue may become saturated, resulting in overflow of lipids into non-adipose tissue. This is associated with occurrence of multiple disease states. *A figure from the review paper by Vanherpen et al* (van Herpen and Schrauwen-Hinderling 2008).

The PAT proteins associate with LSDs in adipocytes and other tissues throughout the body. They share some common characteristics, including a conserved N-terminal protein domain (PAT-1) that is found in four of the five members of the group, PLIN1, PLIN2, PLIN3 and PLIN4; and an 11-mer lipid binding helical repeat in their central amino acid sequence. This 11-mer motif is also found in a number of other lipid binding proteins, including the apolipoproteins (Londos et al. 2005).

Using freeze-fracture microscopy, Robenek et al. showed that the PAT proteins are found dispersed throughout the plasma membrane in un-stimulated adipocytes. Upon stimulation with FFAs they show that PLIN1, PLIN2 and PLIN3 form raised clusters in the plasma membrane that are in close contact with underlying LSDs (Robenek et al. 2005). They hypothesized that PAT proteins respond to anabolic stimulation by depositing FFAs into new LSDs. All PAT proteins, except PLIN1, are thought to contain a conserved hydrophobic cleft in their C-terminal region (Londos et al. 2005; Dalen et al. 2007). Interestingly, deletion of this hydrophobic cleft in PLIN3 was shown to abolish its localization to LSDs. These data suggest that binding of FFAs, or other interacting molecules, to the cleft induces a conformational change exposing the 11-mer lipid-binding helical repeat modulating the affinity of PLIN3 to LSDs (Londos et al. 2005; Ohsaki et al. 2006). It is believed that the other PATs may be activated in this way. PAT gene expression profiles vary across most tissue types, indicating they may have evolved specific functions in specific tissues. The genetic structure of the PAT family is also quite interesting. The comparison of gene structures revealed that the mammalian PAT proteins and those in lower organisms were derived from a common ancestral gene, suggesting an essential role for these gene products in lipid storage and

mobilization

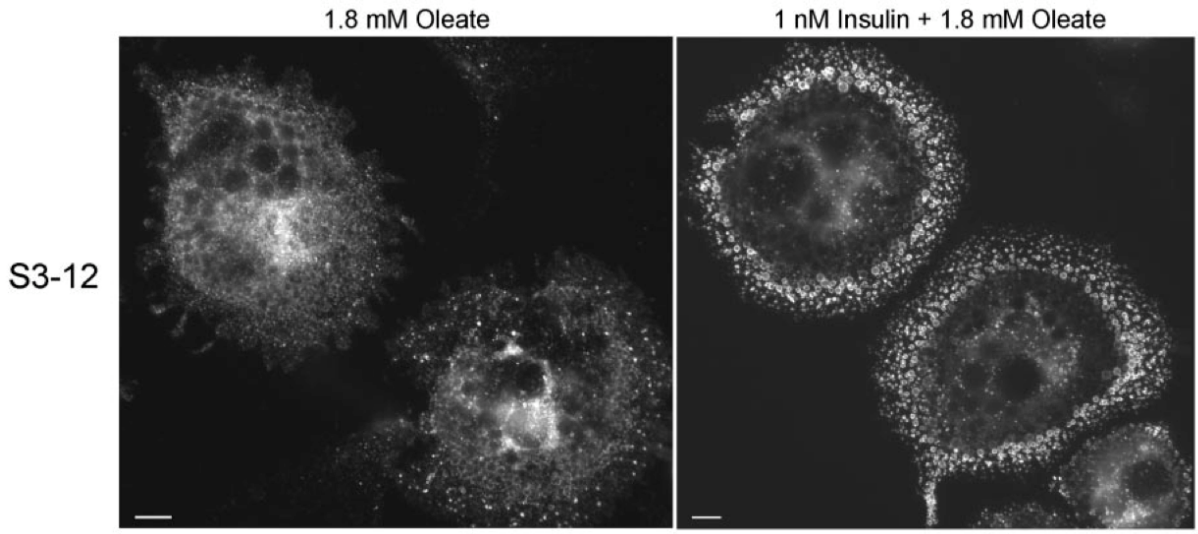


Figure 3. PLIN4/S3-12 relocalization to forming lipid droplets.
This figure was adapted from the paper by Wolins et all (Wolins et al. 2003). Mouse PLIN4/S3-12 in 3T3-L1 cells redistributes from a scattered distribution in the cytoplasm to forming lipid droplets in response to treatment with oleate and insulin.

(Miura et al. 2002).

The best characterized PAT member PLIN1, is found primarily in adipocytes, and interacts with the fat metabolizing enzyme Hormone Sensitive Lipase (HSL). In mature adipocytes unphosphorylated PLIN1 blocks lipolysis of TAG stores within LSDs. Upon catabolic stimulation, PLIN1 is phosphorylated by protein kinase A (PKA) and facilitates interaction of cellular lipases with TAG stores. The mutation to alanine of serine residues within the three most N-terminal sites for PKA in PLIN1 eliminates the enhanced lipolysis observed under PKA stimulation (Brasaemle et al. 2000). Other *in-vitro* studies support the relevant role of PLIN1 in the regulation of LD TAG stores of adipocytes (Brasaemle et al. 2000; Martinez-Botas et al. 2000; Tansey et al. 2001a; Tansey 2003). The *PLIN1* null mouse was shown to consume more food than controls, but maintained a normal body weight with elevated rates of lipolysis. Furthermore, the PLIN1 null mouse rescues the extreme obesity phenotype observed in mice also null for the Leptin Receptor (Tansey et al. 2001b). These mice maintain the phenotype of insatiability; however their ability to store lipids in the adipose tissue is severely compromised.

PLIN2 is expressed ubiquitously, and its expression is related to the amount of lipid that is contained in the cell (Brasaemle et al. 1997). Several studies have demonstrated that PLIN2 stimulates lipid accumulation and LSD formation when over-expressed in murine cell lines (Gao and Serrero 1999; Imamura et al. 2002). Additionally, reports have shown that PLIN2 is up-regulated in the liver of humans and mice fed diets high in fat. This effect is promoted by PPAR-alpha, a known transcriptional regulator of PLIN2

(Targett-Adams et al. 2005a; Dalen et al. 2006; Motomura et al. 2006). PLIN2 null mice fed diets high in fat are protected against the development of fatty liver; however, these mice show no alterations in adipogenesis or lipolysis in other tissues (Chang et al. 2006). Most interesting, a series of studies have demonstrated *in vitro*, that PLIN2 is highly expressed in human macrophages (Wang et al. 1999; Persson et al. 2007). PLIN2 was shown to reduce lipid efflux and thereby is proposed to contribute to the formation of foam cells and atherosclerotic lesions (Larigauderie 2004; Nuotio et al. 2007). Moreover, knockout of PLIN2 in mice null for ApoE showed significant impairment of foam cell formation, compared to PLIN2-wildtype/ApoE-null mice (Paul et al. 2008). These data demonstrate a role for PLIN2 in the accumulation and storage of lipid in LSDs.

Like PLIN2, PLIN3 is a ubiquitously expressed LSD associating protein. Similar to PLIN4, PLIN3 exists in a cytoplasmic pool in un-stimulated cells, but undergoes a re-localization upon stimulation with lipid in the media (Wolins et al. 2001). Furthermore, siRNA knock down of *Plin3* inhibits LSD maturation by limiting incorporation of TAGs in mouse Op9 cells and alters LSD formation in a fibroblastic cell line isolated from *Plin2* null mice (Sztalryd et al. 2006). Over-expression of PLIN3 in human monocytic THP-1 cells correlates with an increase in LSD TAG accumulation (Buers et al. 2009). In addition, it has been demonstrated that abnormalities in LSD maturation and increases in lipolytic rate and insulin resistance occur after knock down of *Plin2* & *Plin3* in murine AML12 cells (Bell et al. 2008). Interestingly, a recent paper has demonstrated that Plin3 is important for facilitating the short-term storage of lipids in mouse enterocytes in response to an acute high fat diet. Authors speculate that PLIN3 acts to prevent the overloading of chylomicrons at the intestine and the subsequent toxic effects of

lipodistrophy (Lee et al. 2009). Taken together, these data suggest a role for *PLIN3* in LSD maturation and utilization.

Like PLIN1, PLIN4 is highly expressed in adipocytes. Wolins et al. showed PLIN4 re-locates to forming lipid droplets from a scattered distribution in the cytoplasm of 3T3-L1 cells when stimulated with media containing insulin and either linoleate, palmitate or oleate. Upon removal of this media, PLIN4 returns to its basal state location in the cell periphery suggesting PLIN4 facilitates uptake of FFAs from the serum to the LSD, in response to the nutritional state of the cell (Wolins et al. 2003; Wolins et al. 2005) (**Figure 3**). Dalen et al., demonstrated levels of PLIN1 and PLIN4 are altered in the Zucker rat, a model of obesity (Dalen et al. 2004).

PLIN5**,** the most recent PAT family member identified, is expressed in tissues that actively oxidize fatty acids, such as heart, red muscle and liver, in addition to adipose. When PLIN5 is ectopically expressed in CHO cells it significantly reduces lipolysis rates and when over-expressed in Cos-7 cells it increases long chain fatty acid oxidation (Wolins et al. 2006; Dalen et al. 2007). Fasting induces elevated expression of PLIN5 in C57/BL mice, an effect that was reproduced by feeding these same mice PPARa agonists (Yamaguchi et al. 2006). Analysis of 85 samples of subcutaneous adipose tissue from non-diabetic human subjects with varying levels of obesity showed a significant correlation of reduced expression of PLIN5 and increased BMI. The authors of this work suggest that PLIN5 modulates the "flexibility" of lipid stores, and mis-regulation of PLIN5 may contribute to increased accumulation of lipid in adipose and non-adipose tissue, perhaps contributing to development of insulin resistance (Wolins et al. 2006).

In summary, the notion of the adipocyte and its role in lipid metabolism has evolved from a mere passive storage of fat to a highly organized and metabolically active organ. As a result, it is essential that we gain a more complete understanding of its inner workings, including the generation and maintenance of LSDs. Identifying the genes responsible for lipid uptake and turnover within adipose and other tissues will be key to understanding the factors defining an individual's risk for both obesity and related conditions.

PAT gene expression

The PAT family appears to respond to dietary FFA in both transcriptional response and the protein levels. The promoter regions of *PLIN1* and *PLIN4* contain conserved and functional peroxisome proliferator-activated receptor (PPAR) response-elements (PPREs) (14). Additionally, PLIN2 and PLIN5 are also upregulated by PPAR stimulation (Londos et al. 2005). PPARs are a family of nuclear-receptor transcription factors that modulate energy and lipid metabolism (15). Polyunsaturated fatty acids (PUFA) are known ligands for PPAR receptors suggesting PAT genes respond to dietary lipids at the transcriptional level. Furthermore, Monounsaturated fatty acids (MUFA) and saturated fatty acids (SFA) are also agonists for PPARs, although weaker. These data support the notion that the PAT genes may modulate uptake and release of TAGs from LSDs, and furthermore that some dietary FFAs may modulate PAT gene expression.

PLIN1 and Human Obesity

The chromosomal location of *PLIN1* in the human genome is 15q26.1, near a

previously reported susceptibility locus for obesity and diabetes. Also, one study demonstrated that obesity and high lipolysis rates are independently associated with lower PLIN1 protein levels in women, whereas another demonstrated reduced levels of both *PLIN1* mRNA and protein in the adipose tissue of obese compared to non-obese subjects (Mottagui-Tabar et al. 2003; Wang et al. 2003).

Given that the PAT proteins contribute to the processes of lipogenesis and lypolysis, and therefore lipid storage and release, investigators from our laboratory hypothesized that genetic variation at the *PLIN1* locus could be associated with variability in anthropomorphic measures in human populations. A series of association studies, in three distinct human populations demonstrated variation at the *PLIN1* locus associates with anthropometric phenotypes primarily in female subjects (Qi et al. 2004a; Qi et al. 2004b; Qi et al. 2005). Interestingly, other studies have demonstrated GxE interactions in humans for *PLIN1* SNPs; influencing weight in response to the PPAR agonist Rosiglitazone, and insulin resistance levels for women consuming diets high in saturated fat (Corella 2006; Kang 2006). Taken together, these results suggest a role for the *PLIN1* locus as gender dependent modulator of obesity risk in humans.

SNPs within microRNA (miR) binding sites of target mRNAs as Functional candidates

It is thought that variants modulating gene transcription are likely candidate loci for phenotypic variability and disease susceptibility (Dixon et al. 2007). Although there have been numerous functional investigations into variants located in the intronic, exonic and promoter regions of candidate genes little attention has been given to variants falling in the 3'UTR where microRNAs (miRs) may bind. miRs regulate protein output, and

individual miR-to-target mRNA interactions may act to dampen mRNA translation often by 33% or less (Stark et al. 2005).

The most critical region for binding and repression of mRNA by a miR are positions 2-7 of the MRE, referred to as the seed site. Although there are examples of miRs targeting mRNAs without perfect Watson-Crick complementarity to the MRESS, a collection of evidence supports the MRESS as the most important feature for prediction and function. In some cases single 7mer seed sites are sufficient for a miR to repress translation, and *ex-vivo* experiments have shown single point mutations in the MRESS may reduce effectiveness or abolish miR mediated repression (Brennecke et al. 2005). Given the importance of the MRESS, it has been proposed that SNPs mapping within the MRESS, or which create novel MRESS (CNM), may have functional consequences resulting in phenotypic variation (Sethupathy and Collins 2008a). Furthermore, there are several examples of miR SNPs associating with disease risk by modulating mRNA levels through abrogation of MRESS sites in the literature (Sethupathy and Collins 2008a). These SNPs may modulate gene transcript and protein levels relevant to a phenotype of interest, generally, or under the influence of particular environmental conditions. Currently, no miRs have been reported to target and regulate a PAT gene.

Main objectives of this thesis

Considering the *in-vivo* and *in-vitro* data described above we hypothesize that variation at the remaining human PAT gene loci may influence obesity related phenotypes in humans. Furthermore, given that the PAT genes are regulated by PPARs and dietary FFAs are known ligands for PPARs, we hypothesized that variation at the

PAT gene loci may modulate individual response to dietary fat.

The specific aims of this thesis were to 1) identify common SNPs at the four candidate PAT gene loci that may modulate obesity related phenotypes in humans, 2) to determine if these SNPs may modulate obesity related phenotypes as main effects or through interaction with dietary fatty acids and 3) to perform functional investigations on SNPs showing strong association and a plausible biological mechanism.

Chapter 2: Materials and Methods.

SNP selection and genotyping

To identify common SNPs in the remaining human PAT loci, we performed a search of the CEPH population in the HapMap PHASE II database for polymorphic alleles with a minor allele frequency >= 5%. The CEPH population is the closest match to the FOS and GOLDN populations, all being of European ancestry. Each locus was defined as 5000bp upstream of the predicted start codon, and 2000bp downstream of the termination codon. This region includes the 5' UTR, the core promoter elements and potential upstream enhancers. This region also includes coding regions and intronic sequences, especially those that may lie in regions of splicing control and the 3'UTR.

For *PLIN2* nine SNPs were identified using the above criteria. We chose four of these SNPs for genotyping based on potential function; two intronic (rs3802335, rs3824369) and two Promoter (rs3824368, rs2046841) SNPs were chosen. Because analysis in the CEU population of the Hapmap database with the Haploview program detected no tag SNPs, we selected these four SNPs based on location and potential function.

Thirty one SNPs were identified for *PLIN3* using the above criteria. We chose 5 of these SNPs for genotyping. The name and location of these SNPS are; two promoter (rs3760948, rs3760950), one intronic (rs12977684), one exonic mis-sense (rs8289), and one 3' downstream (rs11880629). Analysis in the CEU population of the Hapmap database with the Haploview program determined rs12977684 (rs7250638, rs7246083), rs3760950 (rs4807660), rs3760948 (rs262547) and rs11880629 (16992609) to be tag

SNPs. Given the wide breadth of the *PLIN3* region, to select SNPs of interest for genotyping, we also considered the potential functional consequences. Overall this gave us coverage of 10 SNPs in the *PLIN3* region.

For *PLIN4* thirteen SNPs were identified using the above criteria. We chose seven of these SNPs for genotyping based on potential function and tag properties; two promoter (rs884164 and rs1609717), one exonic missense (rs7250947), one 3'UTR (rs8887) and three intronic (rs8102428, rs892158, and rs11673616). Analysis in the CEU population of the Hapmap database with the Haploview program determined rs892158 (*rs7260518 and rs10406797*), rs7250947 (rs*8102428 and rs884164*), rs11673616 (*rs4991027*) and rs1609717 (*rs4807598*) to be tag SNPs capturing variants in LD with an r2 >0.8, predicting coverage over eleven of thirteen SNPs in the region (Barrett et al. 2005).

Thirteen SNPs were identified in *PLIN5* using the above criteria. We chose four of these SNPs for genotyping based on potential function and tag properties; two promoter (rs4806985, rs7254885), one intronic (rs1666660), and one exonic mis-sense (rs1610090). Analysis in the CEU population of the Hapmap database with the Haploview program determined rs11666660 (rs10416790, rs11665704, rs11085080, rs10406628, rs892157, rs11665666 and rs11670485) as a tag SNP in addition to rs161090 (rs4806985). These four SNPs, according to HAPMAP, provide coverage over twelve SNPs identified in PLIN5.

Ready-made 5' nucleic allelic discrimination assays were available from Applied Biosystems for all SNPs except two. We used the Applied Biosystems Custom Assay design web tool to generate functional assays for SNPs rs1609717 and rs7250947 for

which no ready-made assay was available (appliedbiosystems.com).

DNA was isolated from blood samples using DNA blood Midi kits (Qiagen, Hilden, Germany) according to the vendor's recommended protocol. We genotyped all SNPs using the Taqman assays listed above on the ABIPrism 7900HT Sequence Detection System (Applied Biosystems). A list of SNPs selected and genotypes can be seen in **Table 1.**

Identify all known and putative transcription factor binding sites for PAT genes via literature searching and bioinformatic sequence analysis

We hypothesized that SNPs that fall in regulatory regions of our candidate genes are more likely to have functional consequence. Therefore, in addition to looking at tagging potential, we guided our SNP selection for genotyping based on the potential for being functional. We first determined if any SNP lay in regions of our candidate genes known to have a regulatory function. To accomplish this a literature search looking for responsive transcription factor (TF) binding sites was performed. As noted in the introduction, we found several studies relating PAT expression to PPARs. In fact, a functional PPRE has been found for *PLIN1*, *PLIN2*, *PLIN4*, and *PLIN5* (Dalen et al. 2004; Shimizu et al. 2004; Targett-Adams et al. 2005b; Dalen et al. 2006; Wolins et al. 2006; Dalen et al. 2007). Also, an ETS/AP-1 site was identified in *ADRP* and shown to regulate its expression in macrophages (Wei et al. 2005). However, no SNPs listed in HAPMAP where contained in these published sites.

GENE	SNP	Location	Chrom Position	Alleles	MAF	Effect
PLIN4	***rs8887***	3' UTR	chr19:4453201	G>A	0.375	hsa-mir-522 site created w/ minor allele in 3'UTR
	rs11673616	intron 4	chr19:4457915	A>G	0.08	ERalpha motif is abolished with minor allele.
	rs892158	intron 4	chr19:4458716	G>A	0.11	none
	rs7250947	exon 3	chr:4461530	G>A	0.11	Non-Synonymous
	rs7260518	exon 3	chr19:4462955	C>T	0.12	Non -Synonomous
	rs7251858	exon 3	chr19:4461560	C>T	0.3	Non -Synonomous
	rs7259625	exon 3	chr19:4463514	G>C	0.312	Synonomous
	rs8105775	exon 4	chr19:4459945	G>A	0.008	Synonomous
	rs4991027	intron 3	chr19:4460059	C>T	0.08	none
	rs4807598	intron 2	chr19:4465135	T>C	0.05	none
	rs8102428	intron 1	chr19:4467982	A>G	0.11	none
	rs1609717	5' upstream	chr19:4470450	C>T	0.058	PPAR site found with minor allele
	rs884164	5' upstream	chr19:4472625	T>C	0.108	NFKB site weaker with minor allele.
PLIN2	rs10963967	3' downstream	chr9:19104407	A>T	0.017	none
	rs12237614	3' downstream	chr9:19105088	T>C	0.017	none
	rs2229536	coding	chr9:19106543	C>T	0.067	synonomous
	rs10811115	intron 7	chr9:19108112	G>C	0.058	none
	rs10963971	intron 5	chr9:19110752	G>A	0.158	none
	rs10757027	intron 3	chr9:19114596	A>T	0.333	none
	rs3802335	intron 1	chr9:19116476	A>G	0.2	2 base pairs from predicted C/EBPgamma site
	rs3824369	intron 1	chr9:19116565	A>G	0.5	8 base pairs from predicted NFkB site
	rs3824368	5' upstream	chr9:19117902	G>C	0.375	HNFa site abolished with minor allele
	rs10963974	5' upstream	chr9:19120611	G>C	0.226	none
	rs2046841	5' upstream	chr9:19122051	T>C	0.2	none
PLIN3	rs16992609	3' downstream	chr19:4788501	C>T	0.167	none
	rs10424978	3' downstream	chr19:4788557	A>C	0.441	none
	rs11880629	3' downstream	chr19:4788789	C>T	0.167	none
	rs2779180	intron 7	chr19:4792151	T>C	0.474	none
	rs10409097	intron 7	chr19:4795152	G>T	0.175	none
	rs17363814	intron 7	chr19:4795633	C>T	0.25	minor allele increases score for Androgen site
	rs9973235	exon 6	chr19:4798713	G>A	0.125	Non Synonomous
	rs10406652	exon 6	chr19:4798868	G>A	0.136	Non Synonomous
	rs7250294	intron 5	chr19:4799243	G>A	0.233	none
	rs10415336	intron 5	chr19:4799396	A>C	0.142	none
	rs7251627	intron 5	chr19:4799593	C>T	0.242	none
	rs12983495	intron 5	chr19:4802850	C>T	0.263	Minor allele abolishes SP1 site
	rs2271058	exon 5	chr19:4803106	C>T	0.133	synonomous
	rs9304915	exon 6	chr19:4803137	G>A	0.12	synonomous
	rs1055919	exon 5	chr19:4803137	G>A	0.413	synonomous
	rs2271057	intron 4	chr19:4803337	T>G	0.229	none
	rs11881908	intron 4	chr19:4803525	C>T	0.133	none
	rs4807657	intron 4	chr19:4804703	T>C	0.483	Minor allele creates site for CEBPa
	rs11668342	intron 4	chr19:4805650	G>C	0.125	none
	rs1809980	intron 4	chr19:4807287	C>T	0.155	none
	rs2656950	intron 4	chr19:4807694	A>G	0.153	none
	rs8289	exon 3	chr19:4810937	G>A	0.25	Non Synonomous
	rs12977684	intron 2	chr19:4812071	G>A	0.198	Minor allele creates HNFa site
	rs262547	intron 2	chr19:4814545	T>C	0.358	none
	rs7246083	intron 1	chr19:4817451	G>A	0.211	none

	rs262560	5’ utr	chr19:4818650	C>A	0.03	none
	rs262558	5' Upstream	chr19:4818701	C>A	0.259	none
	rs3760950	5' Upstream	chr19:4819689	G>T	0.1	Minor allele abolishes site for SP1 site
	rs3760949	5' Upstream	chr19:4819765	C>T	0.017	Minor Allele creates PPARa site.
	rs3760948	5' Upstream	chr19:4819802	G>A	0.312	Minor allele creates PPARg site
	rs4807660	5' Upstream	chr19:4820059	G>A	0.107	none
PLIN5	***rs1610090***	exon7	Chr19:4476046	C>T	0.112	Non Synonomous
	rs11670485	intronN6	Chr194476217	G>A	0.08	none
	rs892157	intron5	Chr194477182	G>C	0.1	none
	rs10406628	intronN3	Chr194481261	A>G	0.119	none
	rs11665666	intron2	Chr194484629	G>C	0.074	none
	rs11665704	intron2	Chr194484731	G>A	0.09	none
	rs11085080	intron2	Chr194485014	G>A	0.108	Minor allele abolishes SP1 site
	rs11666660	intron1	Chr194485279	A>G	0.112	Minor allele creates Estrogen site
	rs10416790	intron1	Chr194485744	T>G	0.082	minor allele creates SP1 site
	rs966384	5' upstream	Chr194489599	C>T	0.356	Minor allele abolishes NFkB site
	rs7254884	5' upstream	Chr194490081	A>G	0.034	
	rs7254885	5' upstream	Chr194490086	A>G	0.153	1 base pair away from PPARa site
	rs4806985	5' upstream	Chr194491429	G>A	0.1	Minor allle abolishes CEBPa site

Table1. SNPs identified in the PAT gene regions.
All SNPs at PAT gene loci identified using criteria described in the Materials and Methods. Those SNPs bolded and italicized where selected for genotyping. The effect column describes the effect of the minor allele on a predicted regulatory motif.

Additionally we searched the literature for TFs that are expressed or active in adipogenesis or lipolysis. We hypothesized that TFs regulating these processes may also regulate expression of our candidate genes. Our search identified a number of potential regulatory factors. The transcription factor C/EBP alpha is up regulated during differentiation of adipocytes. It is thought that C/EBP may regulate PAT gene expression (Prusty et al. 2002). NFkB, was shown to be up regulated in lipolysis and shown to directly influence PLIN and HSL expression levels (Laurencikiene et al. 2007a). Estrogen stimulation was also shown to influence adipogenesis (Dieudonne et al. 2004). Whereas androgen and its response elements are thought to play a role in the lipolysis process (Fan et al. 2005).

We next utilized the web tool Mapper to predict putative TF binding sites in the specified regions of our candidate genes (Marinescu et al. 2005b). These analyses are not conclusive and can report sites that are present and not functional or may miss prediction of a functional site. Mapper's ability to detect true hits (positives) and the % of false hits (negatives) was estimated in the MAPPER publication. Mapper detected between 98-100% of true positives and had a false positive rate between 0-24%. Mapper was shown to provide better sensitivity and specificity for predicting putative sites than similar programs (Marinescu et al. 2005b).

We also wanted to determine if the Mapper score can reflect the effect of a SNP in a functional binding site. To determine the feasibility of this approach, the literature was searched for validated functional SNPs in which the SNP affected, created or abolished the activity of a TF binding site. Seven examples were found, demonstrating an effect with seven different genes and five different TFs. By inputting the same

sequence twice, once with the major allele and again with the minor allele, MAPPER detected and was able to characterize the change (by its score and sequence parameters) of six of the seven examples [31-37] **Table 2**. MAPPER has the highest rate of true positives and the lowest rate of false positives for any freely available TF prediction program. Together, these data suggest we will be able to identify SNPs that have functional consequence, with a relatively low occurrence of identifying false positives.

Analysis of intronic acceptor and donor sequences at exon/intron junctions, in addition to the 50 bp upstream of the 3' end of the intron, which includes the polypyrimidine tract and branch site essential for appropriate splicing.

We next determined if any of the SNPs within the PAT loci may interfere with exon splicing motifs. We investigated the acceptor and donor sites at exon/intron junctions, in addition to the 50 bp upstream of the 3' end of the intron, which includes the polypyrimidine tract and branch site essential for appropriate splicing. SNPs falling in either the acceptor or donor site may abolish or alter efficiency of the site. In addition, SNPs falling in the polypyrimidine tract of the branch point, that are transversions (pyrimidine to purine: T>A or T>G) may alter the pyrimidine richness of the tract possibly altering the efficiency of the branch point.

Gene	TF	SNP	Region	Reference		Mapper Detection
FLT1	p53	C>T	promoter	Menedez et al. (2005). PNAS. 103:1406-11		Yes
	Methods				**Effect**	
	1. Reporter assay				8 fold increase in luciferase activation between T and C alleles.	
	2. CHIP assay with p53 antibody				T allele bound p53 whereas the C allele bound little if any p53.	
	3. RT-PCR levels of FLT1 in presence of cell stress (p53 activator.).				T carrying (HET) cells, induce FLT exp when induced w/ stress.	
XRCC1	SP1	T>C	5' UTR	Hao et al. (2006). Oncogene. 25:3613-20		Yes
	Methods				**Effect**	
	1.EMSA				Both alleles bind SP1, the C allele with greater affinity	
	2.CHIP A549 cells SP1 antibodies				Sp1 was shown to bind in vivo in these cells with SP1 antibody.	
	3. Reporter assay				The C allele was shown to be ~30% less active than the T allele.	
MDM2	SP1	T>G	promoter	Bond et al. (2004). Cell. 119. 591-602.		No
	Methods				**Effect**	
	1. EMSA				G allele greatly enhances SP1 binding.	
	3. Reporter assay				G allele increased reporter significantly, over T allele.	
GRK5	CREB	T>C	intronic	Arawaka et al. (2006). Jour. Neuro. Sci. 26: 9227-38.		Yes
	Methods				Effect	
	1. EMSA w/ CREB antibody.				C allele increases affinity, over C allele.	
	3. Reporter assay				C allele showed 1.5 increase in signal compared to T allele.	
FasL	C/EBP beta	T>C	promoter	Wu et al. (2003). Jour. Immuno. 3: 132-38		Yes
	Methods				**Effect**	

	1. EMSA				C allele show greater affinity than T allele.	
	2. Reporter assay				C allele twice the activity of T allele.	
AGTRL1	SP1	G>A	promoter	Hata et al. (2007). HMG. Advanced Publication. "Functional SNP in a SP!-binding site of AGTRL1 gebe is associated with susceptibility to brain infarction."		Yes
	Methods				**Effect**	
	2. Reporter assay				G allele showed greater reporter activity than A allele.	
	3. EMSA				Strong affinity to G allele, but weak to A allele.	
KLK3	Androgen	G>A	promoter	Lai et al. (2006). Carcinogenesis. Open Access. "PSA/KLK3 AREI promoter polymorphism alters androgen receptor binding and is associateed with prostate cancer."		Yes
	Methods				**Effect**	
	1. EMSA				A allele bound with greater affinity than G allele.	
	2. Reporter assay				A allele showed greater reporter activity than G allele.	

Table 2. The MAPPER TF prediction tool predicts the effect of 6 out of 7 examples from the literature.

Seven examples from the literature of a SNP modulating TF binding to a gene promoter were analyzed with the Mapper webtool. For each example the methods used to demonstrate the SNPs effect on TF binding are provided. The results of these assays are also provided. A "Yes" in the Mapper Detection column indicates if the Mapper tool was able to successfully reproduce the effect observed in the assay(s).

3' UTR analysis, use of bioiniformatic webtools to search for putative miRNA binding motifs

MicroRNAs (miR) are a class of non coding RNA sequences that play important roles in the regulation of a growing number of genes. miRNAs act by binding to complementary regions of 3'UTR mRNA transcripts to repress translation or to induce degradation.

To identify potential target sequences in the 3'UTRs of the PAT genes we used the bioinformatic webtools Targetscan and MicroRNA.org (Grimson et al. 2007; Betel et al. 2008a). Targetscan uses the presence of 7 and 8mer seed sites at the seed region for matching. In addition, sites containing 1 to 2 mismatches in their seed site are further scored based on the potential compensatory ability from 3' pairing. Results from Targetscan are organized by context score, which takes into account the over all fit of the target sequence to the miR based on established guidelines. MicroRNA.org uses a similar algorithm that also incorporates the free energy estimate for the binding of microRNA to potential target.

All SNPs identified in our analysis and those chosen for Genotyping are shown in **Table 1.**

Study design and subjects: Framingham Offspring Study

The design and methods of the Framingham Offspring Study (FOS), which was initiated in 1971, have been reported (Feinleib et al. 1975). Blood samples for DNA were collected between 1987 and 1991. Anthropometric, lipid and dietary intake variables were recorded for subjects who participated in the fifth and sixth examination visits.

Variables used in this study consist of mean values calculated from these two exams, with the exception of fasting insulin and HOMA measurements that were available only for subjects participating in exam 5. Dietary intake was determined with a semi-quantitative food frequency questionnaire (Rimm et al. 1992). Intakes of PUFA (n3 and n6), MUFA and SFA where calculated for each subject and used in our analyses as continuous variables. The Institutional Review Boards (IRB) for Human Research at Boston University and Tufts University/New England Medical Center approved the protocol. All participants provided written informed consent.

Anthropometric and biochemical determinations for FOS

Briefly, weight was measured with the individual dressed in an examining gown and wearing no shoes. The BMI was calculated as weight in kilograms divided by the square of height in meters. Fasting glucose, plasma lipids, and lipoproteins were measured as previously described (Corella et al. 2007). To analyze SAT and VAT variables, data from the Framingham Heart Study Multidetector Computed Tomography Study, a population-based sub-study of the community-based Framingham Heart Study Offspring and Third-Generation Study cohorts were used (Fox et al. 2007).

Study design and subjects: GOLDN

Study protocol approval was obtained from the Human Studies Committee of Institutional Review Board at the University of Minnesota, University of Utah, and Tufts Medical Center. All participants provided written informed consent. The detailed methodology and design of the GOLDN study has been described previously (Corella

2006; Shen et al. 2007). Briefly, GOLDN is part of the Program for the Genetic Interactions Network and is funded by the NIH. Participants were recruited from pedigrees from two genetically homogenous National Heart, Lung, and Blood Institute Family Heart Study field centers in Minnesota and Utah, both predominately white populations. There were 1086 subjects with complete phenotype, dietary and genotype data. Dietary intake was estimated by use of the Dietary Health Questionnaire (DHQ) which consists of 124 food items and includes both portion size and dietary supplement questions (Thompson et al. 2002).

Anthropometric and biochemical determinations for GOLDN

Blood samples were drawn after fasting overnight. Anthropometrics and blood collection, plasma separation and processing, and biochemical lipid measurements, including triglycerides and HDL cholesterol, have been described previously (Lai et al. 2004). Fasting plasma insulin was determined by the Human Insulin Specific RIA kit (Linco Research). Fasting plasma glucose was measured using a hexokinase-mediated reaction on a Hitachi 911 (Roche Diagnostics).

Statistical analysis

For rs3824369, rs3824368, rs204681, rs3760948 and rs12977684 an additive genetic model was utilized. For all other SNPs a dominant model was applied, classifying homozygotes for the major allele in one group, and carriers and homozygotes of the minor allele in another. Multivariate linear regression was used for association analyses in the GOLDN and FOS populations. To reduce variability that might obscure

potential findings, multiple covariates were incorporated into our regression model, including age, sex, alcohol and tobacco smoking status, physical activity (GOLDN only), hormone use and diabetes, cholesterol and hypertension medications. To adjust for familial relationships among subjects in both populations, the lme *kinship* procedure was used in R, which allows the specification of the full correlation structure within pedigrees into the regression equation with random effects. To determine gene by diet interactions a SNP*dietary term was introduced into the regression equation. These tests were adjusted for total energy intake by adding total energy to the model, in addition to those mentioned above. A p value < 0.05 was considered significant. Response variables that did not maintain a normal distribution were log transformed to fit the normal distribution.

Our meta-analyses were performed using the software package METAL (www.sph.umich.edu/csg/-abecasis/metal). METAL performs meta-analysis of results from two or more individual studies. It creates a single summary p-value from studies that cannot be analyzed together, for reasons such as differences in ethnicity, phenotype distribution, or gender. For analysis of PLIN2, PLIN3 and PLIN5 SNPs we used the fixed-effects inverse variance-weighted approach to combine beta coefficients and standard errors for each study into a summary z-score that combines the association evidence to determine an overall level of significance. For PLIN4 we used meta analysis to weight the effect size of each study by its sample size and combining Z statistics to determine an overall level of significance.

Cell culture & RNA isolation

HepG2, COS-7 and HEK-293T (obtained from American Type Tissue Collection) cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) containing 10% FBS and 2% penicillin-streptomycin. RNA was extracted from approximately $6x10^6$ cells using Trizol reagent. Total RNA was reverse transcribed using the RT2 miRNA First Strand kit (SABiosciences) and miRNA quantified using primers specific for human miR-522 (SABiosciences qPCR assay) and values were normalized to the housekeeping gene SNORD38b. All experiments were performed in triplicate. Trizol whole cell lysates from $3x10^6$ human pre- and mature adipocytes were purchased from Zen-Bio. Total RNA was purified using the miRNeasy kit (Qiagen) and miRNA was quantified using the protocol above.

3'UTR Luciferase Reporter Assays

The Expand High Fidelity PCR kit (Roche) was used to amplify the *PLIN4* 3'UTR sequence using gDNA from subjects homozygous for either the rs8887 G or A allele. Primers were designed to amplify a 560 nucleotide sequence including the last 460 bases of the PLIN4 3'UTR and 100 bases of downstream genomic sequence. Included in the primers were the restriction enzyme sites XhoI for the forward primer (***AACTCGAGCTGTAGGAGCCTGCAAG)*** and NotI for the reverse ***(AGCGGCCGCGACTATAAATGGTTTTTTAATGAAAAAAGAAATCACT***). These PCR products where cloned into the multiple cloning site of the PSICHeCK2 reporter vector downstream of the Renilla luciferase coding sequence.

COS7 cells, plated into 12-well plates (Costar), were co-transfected with 1μg of the pmiR-LucPLIN4-G or pmiR-LucPLIN4-A luciferase reporter vectors and 40 nM mi*RIDIAN* miR-522 mimic or an equal concentration of a non-targeting control mimic sequence (Dharmacon) using the Lipofectamine 2000 Reagent (Invitrogen). Luciferase activity was measured using the Dual-Glo Luciferase Assay System (Promega). *Renilla* luciferase activity was normalized to the corresponding firefly luciferase activity and plotted as a percentage of the control (cells co-transfected with the corresponding concentration of control mimic). This experiment was performed in triplicate wells of a 12-well plate and repeated at least three times (**Figure 4).**

Phylogenetic analysis

Nucleotide sequences for human, Neandertal, chimpanzee, gorilla, orangutan, lemur, wild boar, cow, dog and mouse were downloaded from reference assemblies available at NCBI and aligned with ClustalW. The evolutionary history of PLIN4 was inferred using the Maximum Parsimony (MP) method. The bootstrap consensus tree inferred from 10000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) are shown next to the branches. The MP tree was obtained using the Close-Neighbor-Interchange algorithm with search level 3 in which the initial trees were obtained with the random addition of sequences (100 replicates) (Tamura et al. 2007). The analysis involved 10 nucleotide sequences. There were a total of 2447 positions in the final dataset. Evolutionary analyses

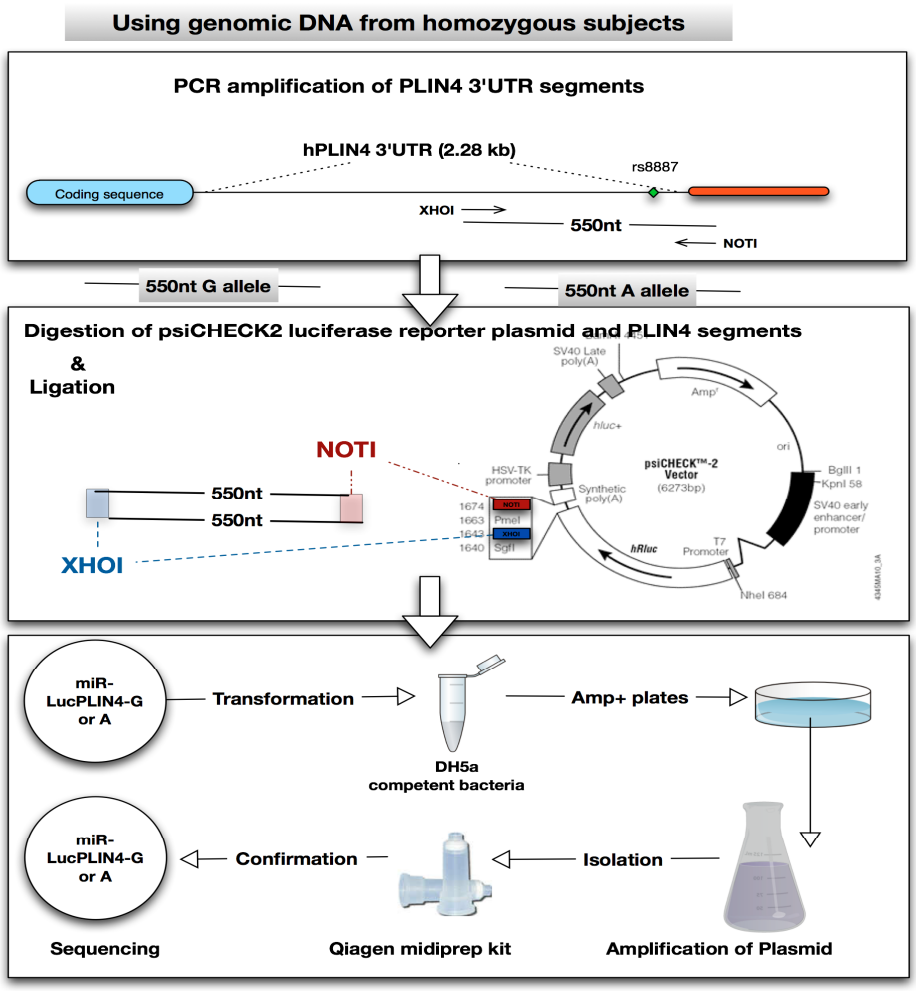


Figure 4. Creation of the rs8887 allele specific luciferase reporter vectors.
This figures outlines the procedure for creation of the allele specific PLIN4-3'UTR expression vectors. A fragment containing the rs8887 A of G allele was PCR amplified from subjects homozygous for either allele. This fragment was directionally cloned into the psiCHECK2 expression vector. The vectors where transformed, selected for and amplified. Sequencing confirmed the product contained the allele specific motif.

were conducted in MEGA4. Total and population F_{ST} statistics for HapMap populations where estimated using the SNP@evolution webtool which implements the F_{ST} calculations described in Akey et al (Akey et al. 2002a).

Retrieval and use of dbSNP information

We retrieved all dbSNP build 132 SNP (as of 11-31-10) information by downloading the vcf file available through the 1000 genomes home page (Durbin et al. 2010). To ensure we only surveyed variation in the form of SNPs and not indels and/or copy number variants we removed all SNPs not reported as bi-allelic. A subset of data containing all 3'UTR SNPs (n=210042) was extracted using Perl. This dataset was used for all subsequent analyses. To determine the percentage of SNPs that were submitted to dbSNP by the 1000 Genomes Project we used UCSC Genome Browser to identify all dbSNP build 132 SNPs where the submitter status handle was equal to only the 1000GENOMES tag. This data set was then searched against our MRESS and CNM data sets to identify those SNPs contributed by the 1000 Genomes Project. To retrieve allele frequency data on SNPs reported in MRESSs and CNMs, we utilized the Perl API variation tools accessing the latest human genome variation data, build 61. All data analysis was performed on the NUGO information network (Harttig et al. 2009).

Identifying MRE SNPs in validated targets

The database miRecords hosts a collection of validated miR-mRNA interactions built from an exhaustive literature search and the database of records was download in a tab-delimited format (Xiao et al. 2009). We next annotated each hit for target site

functionality, by checking the literature source for evidence of a loss of function experiment, which provided us with 606 validated MRE targets. To identify SNPs falling within these 606 validated target sites, a Perl script was written to search each SNP gDNA coordinate against the gDNA coordinates of each target transcript MRESS, retrieved from Ensembl.

SNP Density determination

To calculate SNP density, a Perl script was written to perform a sliding window search (6 bases corresponding to bases 2-7 of the MRESS of the 606 validated MREs for SNPs, starting 18 bases upstream of the 1st base of the MRESS and ending 18 bases downstream of the last base of the MRESS. We report the number of SNPs for each position of the window across this sequence. Values are reported as SNPs/kb.

Identification of conserved MRESS and CNM SNPs

To identify SNPs falling in conserved MRESSs, we downloaded the "good mirsrv_score conserved miRNA" datafile from the microrna.org website. This file contains all predicted mRNA target motifs for targeting microRNAs which belong to conserved microRNA families. Conservation signal is used to predict functional MREs, however it has been determined that a conservation signal above background for MREs of the most recently emerged miRNA families (non-conserved) was unlikely due to the relatively short time between the emergence of these miRs and the occurrence of new MREs within 3 'UTRs (Friedman et al. 2009a). Therefore, to eliminate false positives that would arise from this form of analysis we utilize predictions for only conserved miR

families – which are contained in the "good mirsrv_score conserved miRNA" datafile. To add an additional measure of conservation, we implemented a conservation score cutoff for predicted miR targets using a Phastcon score of >= 0.57, which roughly corresponds to conservation across all mammals (Siepel et al. 2005; Betel et al. 2008b; Betel et al. 2010). The Phastcon scores is provided in the predictions data file from microRNA.org. Additionally, predictions generated by the microRNA.org tool provide a ranking score (mirSVR score) which is calibrated to correlate linearly with the extent of down regulation of a miR on its target. Importantly, these scores may be interpreted as an empirical probability of down regulation. From these data we selected a mirsvr score cutoff <= -0.6, representing the top 12% of all predictions. A Perl script was then used to compare the gDNA coordinates of each predicted MRESS (n=197287) against the gDNA coordinates of each 3'UTR SNP in dbSNP132.

To identify CNM SNPs we utilized the Ensembl variation Perl API tools (Build 61) to retrieve the 22 bases flanking the 5' and 3' regions of every 3'UTR SNP in dbSNP 132. We generated the reverse complement for those mRNAs transcribed from the negative strand. These data were then run locally through the miRanda target prediction algorithm. To limit identification of potential false positives we implemented an arbitrary pairing score cutoff of >=150 and an energy cutoff of <=-20. We identified all predicted MRESS created by CNM SNPs by filtering hits on the position of their target prediction on the mRNA, where each SNP is located at position 23 of 45.

Retrieval of GWAS results and LD calculations

We first downloaded a catalog of human variation associating with disease phenotypes (1-25-11) (Hindorff et al. 2009). The list was then submitted to the SNAP tool (http://www.broadinstitute.org/mpg/snap/ldsearch.php) using r^2 of >0.8 for the CEU population. The resulting list was then searched against each of the CNM and MRESS SNP lists. Using a perl script, 3940 SNPs where randomly selected from the dbSNP build 132 data set and run through the SNAP tool to identify all SNPS in LD, using r^2 of >0.8 for the CEU population. This procedure was repeated 100 times. These lists were then compared against the MRESS and CNM snp lists to determine if these SNPs are in LD with GWAS SNPs more frequently that would be observed by chance.

Co-expression

We utilized the mimiRNA webtool to identify miR-mRNA predictions with co-expression evidence (Ritchie et al. 2010). The mimiRNA tool provides expression data for 564 mRNAs and 636 miRs, normalized across samples, from four large scale miR expression studies, and one mRNA expression study. We queried each miR-mRNA pair for co-expression using the tools provided on the webpage. Because not all of the miRs implicated in our work are in the mimiRNA dataset, we also searched the literature via pubmed using the search terms of the miR-name and the term "expression." To determine if the matching gene was expressed in the same tissue type, we queried the GEOprofile database.

eQTL survey

To identify association of transcript levels with MRESS and CNM SNPs we searched eQTL data from the MuTHER study using the Genevar webtool (Yang et al. 2010; Nica et al. 2011). eQTL data was generated from Fat cell biopsy (n= 160), LCL cells (n=166) and skin punch biopsy (n=160) taken from healthy adult female twins (both mono and di-zygotic). Twin pairs where separated in two to unrelated groups, thereby performing 2 independent eQTL analysis, as described in Nica, et al.

Genevar provides Spearman's rank correlation coefficient estimates for the strength of relationships between alleles and gene expression intensities for each study group. Furthermore, to test the significance of the relationship, Genevar generates a t-statistic for correlation analysis. Adjusted non-parametric permutation P-values are also provided (Yang et al. 2010).

F_{ST} calculations

Genotype characteristics of 11 HapMap Phase 3 populations were split into 4 groups of similar ancestry; Asian, African, European and American. F_{ST} values where calculated for each HapMap Phase 3 SNP between these 4 groups and reported in a downloadable file (Cheng et al. 2009a). Using a Perl script we extracted all MRESS and CNM SNPs with F_{ST} values from this dataset. We used the statistical analysis software (SAS) boxcoxar macro to transform the F_{ST} data to fit a normal distribution. We then performed an unpaired Student's t-test using the transformed values for these two groups to determine if they were significantly different. To identify MRESS and CNM F_{ST} outliers we selected those SNPs with F_{ST} > 2 SDs from the mean.

Chapter 3: Variants at the *PLIN3* and *PLIN5* loci associate with obesity and glucose related traits.

Introduction:

The PAT proteins have been identified as important regulators of lipid uptake and utilization in nearly all cell types (Brasaemle 2007). Furthermore, the aberrant expression of some PAT proteins in human tissues is associated with the occurrence of obesity (Mottagui-Tabar et al. 2003; Wolins et al. 2006). Unlike *PLIN1* and *PLIN4*, which are expressed primarily in adipose, *PLIN2* and *PLIN3* are expressed ubiquitously in human tissues, and *in vitro* and *ex vivo* experiments suggest their disregulation may contribute to lipid storage abnormalities implicated in the pathogenesis of CVD and T2D (Larigauderie 2004; Nuotio et al. 2007). In addition, *PLIN5* is expressed in adipose, and in tissues that actively oxidize fatty acids, and reduced expression of *PLIN5* is correlated with increased BMI in humans (Wolins et al. 2006).

Several studies have shown variants at the *PLIN1* locus associate with obesity related phenotypes. To date there have been no studies investigating the contribution of variants within the *PLIN2, PLIN3* or *PLIN5* loci to obesity related phenotypes in humans. To explore this, we performed a fixed-effects inverse variance-weighted meta-analysis using results from association analysis of *PLIN2, PLIN3* and *PLIN5* SNPs with anthropometric, lipid and glucose variables in two populations, the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) and the Framingham Offspring Study (FOS**)**. We also investigated the interaction of dietary PUFA, MUFA and SFA with *PLIN* loci SNPs to determine their combined potential to modulate these phenotypes.

Results:

Demographic, Biochemical, and Genotypic Characteristics of Study Participants

Demographic and biochemical characteristics of participants for the FOS and GOLDN populations are presented in **Table 3**. Genotypic characteristics are shown in **Table 4**. Genotype distributions did not deviate from Hardy-Weinberg Equilibrium in either population.

Association of PLIN3 variants with Anthropometric, Lipid and Glucose related phenotypes from Meta-Analysis.

Table 5 displays the multivariate adjusted significant p-values for associations of *PLIN3* SNPs with anthropometric and glucose phenotypes in both the FOS and GOLDN populations. The overall p-values of the combined data for meta-analysis are also provided. To identify SNPs with the strongest evidence for association, we report here only those associations replicating significant (P <=0.05) or nominally significant (P<=0.10) associations for both populations, in addition to significant association from the combined data of the meta-analysis. A significant meta-analysis association also indicates an effect in the same direction in both populations.

Only the *PLIN3* variant rs3760950 demonstrated significant or nominally significant association in both populations (FOS P=0.075, GOLDN P=0.011, meta P=0.003) with fasting insulin, with carriers of the minor allele having elevated levels. We did not observe significant associations for other SNPs at the *PLIN2, PLIN3* or *PLIN5* loci with our phenotypes of interest.

	Men		Women	
	FOS (N=1259)	**GOLDN (N=481)**	**FOS (N=1352)**	**G0LDN (N=513)**
Trait	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Age (years)	56.3(9.9)	52.9(14.4)	55.9(9.6)	52.2(14.0)
BMI (kg/m**2)	28.5(4.16)	29(4.58)	27.1(5.51)	28.4(6.19)
Waist (cm)	100.3(10.6)	102.1(11.78)	90.42(14.07)	93.21(17.9)
Weight (kg)	87.49(14.09)	82.91(14.01)	70.59(15)	69.34(15.43)
Waist/Hip Ratio	0.97(0.05)	0.96(0.09)	0.87(0.08)	0.85(0.09)
Glucose (mg.dL)	106(26.7)	106.4(20.5)	98.7(23.6)	98.6(16.3)
Homa*	8.7(7.09)	3.85(2.86)	7.26(5.93)	3.35(2.51)
Insulin (mU/L)	3.41(0.37)	14.21(8.60)	3.3(0.32)	13.2(8.08)
Triglycerides (mg/dL)	156(108)	150.7(92.9)	133(80.8)	129(80.1)
HDL Cholesterol (mg/dL)	43.2(10.9)	41.5(10)	56.9(14.6)	52.9(14)
Total PUFA N3 - g	1.43(0.58)	1.83(0.98)	1.37(0.53)	1.48(0.82)
Total PUFA N6 - g	11.03(5.12)	18.2(10.23)	9.79(4.37)	14.1(7.87)
Food Energy - kcal	1989(630)	2354(896)	1735(546)	1719(629)
Physical Activity Score	-	35.2(7.39)	-	33.1(4.73)
	%Use	**%Use**	**%Use**	**%Use**
Current Smokers (%)	16	8	18	8
Current Alcohol Users (%)	73	51	60	52
Menopausal (%)	-	-	69	-
Hormones (%)	-	-	23	20
Taking Hypertensive Medications (%)	12	23	9	17
Taking Diabetes Medications (%)	12	5	7	4
Taking Cholesterol Medications (%)	12	6	8	4

Table 3. Demographic and biochemical characteristics of FOS & GOLDN subjects.

Data are means and standard deviation (SD) for continuous variables or % usage for categorical variables. Populations are displayed by gender for all anthropometric, lipid and glucose variables investigated. The percent usage of tobacco, alcohol, hormone, hypertension, diabetes and cholesterol medication is also listed. The % of menopausal women is provided for FOS, only. * Homeostasis model assessment of insulin resistance (homa).

					FOS			GOLDN		
	alleles	Chrm	Position	Feature	Maj	Min	HWE	Maj	Min	HWE
PLIN2										
rs3802335	A/G	9	19116476	intronic	0.80	0.20	0.88	0.82	0.19	0.38
rs3824369	A/G	9	19116565	Intronic	0.52	0.48	0.02	0.53	0.47	0.80
rs3824368	G/C	9	19117902	Promoter	0.63	0.37	0.47	0.63	0.37	0.91
rs2046841	A/G	9	19122051	Promoter	0.83	0.17	0.59	0.80	0.20	0.05
PLIN3										
rs11880629	C/T	19	4788789	3' downstream	0.88	0.12	0.02	0.88	0.12	0.11
rs8289	G/A	19	4810937	Exonic	0.71	0.29	0.20	0.73	0.27	0.05
rs12977684	G/A	19	4812071	Intronic	0.78	0.22	0.56	0.83	0.17	0.564
rs3760950	G/A	19	4819689	Promoter	0.90	0.10	0.91	0.88	0.12	1
rs3760948	G/A	19	4819802	Promoter	0.64	0.36	0.86	0.63	0.37	0.76
PLIN4										
rs8887	G/A	19	4453201	3' UTR	0.55	0.45	0.92	0.56	0.45	0.81
rs11673616	A/G	19	4457915	Intronic	0.87	0.13	0.68	0.89	0.11	0.42
rs892158	G/A	19	4458716	Intronic	0.84	0.16	0.38	0.86	0.14	0.91
rs7250947	G/A	19	4461530	Exonic	0.91	0.09	0.40	0.93	0.07	0.43
rs8102428	A/G	19	4467982	Intronic	0.90	0.10	0.80	0.91	0.09	0.40
rs1609717	T/C	19	4470450	Promoter	0.95	0.05	0.05	0.94	0.06	0.80
rs884164	T/C	19	4472625	Promoter	0.92	0.08	0.61	0.93	0.07	0.03
PLIN5										
rs1610090	C/T	19	4476046	Exonic	0.87	0.13	0.53	0.90	0.10	0.11
rs11666660	A/G	19	4485279	Intronic	0.92	0.08	0.01	0.88	0.13	0.05
rs7254885	A/G	19	4490086	Promoter	0.87	0.13	0.00	0.92	0.08	0.00
rs4806985	G/A	19	4491429	Promoter	0.90	0.10	0.13	0.87	0.13	0.59

Table 4. Genotypic characteristics of *PAT* SNPs in FOS & GOLDN subjects.
dbSNP rs numbers for each SNP genotyped are given in column one. Major and minor alleles, and chromosomal position (GRCh 36.3) are provided, followed by the gene region in which the SNP falls. Allele frequencies and Hardy-Weinberg equilibrium p-values are given for each SNP in FOS and GOLDN populations.

			FOS			GOLDN			Meta-Analysis	
PLIN3										
SNP	**Phenotype**	**Gender**	**Beta**	**Se**	**P**	**Beta**	**Se**	**P**	**z-score**	**P**
rs3760950	HOMA (log)	Both	0.029	0.020	0.133	0.090	0.038	0.017	2.585	0.009
		Males	0.062	0.028	0.029	0.080	0.055	0.146	2.625	0.009
		Females	-0.001	0.027	0.981	0.086	0.050	0.084	0.947	0.343
	Insulin (log)	Both	0.030	0.017	0.075	0.086	0.034	0.011	2.909	0.003
		Males	0.059	0.025	0.019	0.072	0.050	0.147	2.749	0.006
		Females	0.005	0.022	0.837	0.093	0.045	0.040	1.320	0.187
rs11880629	BMI	Both	0.061	0.245	0.804	-0.290	0.431	0.502	-0.160	0.873
		Males	0.617	0.299	0.039	0.274	0.506	0.588	2.022	0.043
		Femlaes	-0.386	0.373	0.302	-0.917	0.679	0.177	-1.603	0.109
	WC	Both	0.141	0.245	0.565	-0.531	0.485	0.274	-0.118	0.906
		Males	0.667	0.300	0.026	0.721	0.566	0.203	2.554	0.011
		Females	-0.203	0.369	0.583	-1.818	0.758	0.017	-1.773	0.076
	Waist/Hip	Both	0.005	0.003	0.136	-0.002	0.006	0.702	1.038	0.299
		Males	0.007	0.004	0.065	0.010	0.008	0.198	2.250	0.024
		Females	0.004	0.005	0.389	-0.017	0.010	0.082	-0.238	0.812

Table 5. Significant Associations of PAT SNPs in the FOS and GOLDN populations - Main Effects.
Significant Associations of *PLIN2*, *PLIN3*, and *PLIN5* SNPs from Meta-analysis of the FOS and GOLDN populations - Main Effects. Waist Circumference(WC). Results for associations in FOS, GOLDN and Meta-analysis. Results are shown for the entire population (Both) and for males and females independently. P-values for anthropometrics were adjusted for sex, age, smoking, physical activity (GOLDN only), alcohol use, diabetes, beta-blockers, calories from fat, and estrogen and menopausal status(FOS only) in women. Lipid and Glucose p-values where also adjusted for BMI and cholesterol medications.

Meta-Analysis reveals interaction of dietary FFAs with PLIN3 and PLIN5 variants modulating anthropometric related phenotypes.

PLIN1 variants have been shown to interact with dietary and pharmacological factors modulating obesity and glucose related phenotypes (Corella et al. 2006). Given that the PLINs have been shown to respond to dietary FFAs at both the protein and transcriptional levels we hypothesized variation in the PLINs may modulate the relevant phenotypic responses of anthropometric, lipid and glucose variables. As with our main effect analyses, we report here only those associations showing significance (P<=0.05) or nominal significance (P<=0.10) in both populations, in addition to significance (P<=0.05) from the combined data of the meta-analysis.

Several interactions between *PLIN3* SNPs and dietary FFAs modulating phenotypes of interest were identified in the meta-analysis. The *PLIN3* variant rs3760950 demonstrated significant associations with body weight (FOS P=0.020, GOLDN P=0.070, meta P=0.005) and BMI (FOS P=0.10, GOLDN P=0.020, meta P=0.008), with carriers of the minor allele having reduced measurements correlating with increased MUFA intake (**Table 6**). In addition, the rs3760950 showed an interaction with SFA, modulating TAG (FOS P=0.049, GOLDN P=0.082, meta P=0.009) with carriers of the minor allele having increased levels with higher SFA intake.

Our analyses also demonstrated several sex specific GxE interactions between *PLIN3* and *PLIN5* variants and our phenotypes of interest. The *PLIN3* rs3760948 variant showed significant interaction with SFA in male subjects, modulating TAG (P=0.039, 0.036) with the minor allele having an additive effect increasing TAG levels in response to SFA. The *PLIN5* rs1610090 variant interacted with N3 PUFA, in female subjects, to

modulate body weight (FOS P=0.082, GOLDN P=0.053, meta P=0.012) with carriers of the minor allele having reduced measurements with increasing N3 PUFA intake. The *PLIN5* rs4806985 variant significantly interacted with N3 PUFA in female subjects modulating body weight (FOS P=0.024, GOLDN P=0.074, meta P=0.004), with carriers of the minor allele having reduced measurements with increased N3 PUFA intake. We observed no GxE interactions for SNPs at the *PLIN2* locus.

Functional prediction of PLIN gene variants.

We investigated if variants that showed the most consistent associations may be functional. *In-silico* analysis with the webtool MAPPER predicted the *PLIN3* rs3760950 SNP creates a binding motif for the transcription factor (TF) SP1. *In-silico* analysis predicted the minor allele of the rs3760948 SNP creates of a PPARG binding site in the promoter region of *PLIN3*. Analysis of the *PLIN5* SNP rs1610090 showed the minor allele induces a non synonymous amino acid substitution in exon 7. The rs4806895 allele falls in the promoter of the *PLIN5* gene and is predicted to alter affinity for a CEBP TF motif.

Discussion:

This study is the first to investigate *PLIN2*, *PLIN3* and *PLIN5* as candidate genes for obesity and related conditions. We report findings of two independent association analyses from white populations of European ancestry, in addition to a meta-analysis combining the summary statistics of these studies. Although we observed numerous associations with meta-analysis p-values reaching significance, we observed only one

					FOS			GOLDN			Meta	
SNP	**Phenotype**	**Interaction**	**Gender**	**Genotype**	**Beta**	**Se**	**P**	**Beta**	**Se**	**P**	**z-score**	**P**
PLIN3												
rs3760948	TAG (log)	SFA	Males	G/G(ref)			0.039			0.036		*
				G/A	0.026	0.013	0.045	0.048	0.021	0.022	2.933	0.003
				A/A	0.040	0.019	0.029	0.051	0.027	0.061	2.848	0.004
rs11880629	Insulin (log)	N3	Both		-0.138	0.083	0.096	-0.235	0.155	0.129	-2.229	***0.026***
			Males		-0.095	0.137	0.490	-0.470	0.275	0.088	-1.529	0.126
			Females		-0.056	0.104	0.591	-0.124	0.194	0.524	-0.802	0.422
	Weight	N3	Both		-10.592	8.524	0.214	-13.649	11.879	0.251	-1.667	0.095
			Males		-22.887	12.770	0.073	-18.671	18.909	0.324	-2.039	***0.041***
			Females		3.593	11.862	0.762	-8.702	16.522	0.599	-0.034	0.973
rs3760950	HDL-C	N6	Both		-0.664	0.491	0.177	-0.205	0.473	0.664	-1.368	0.171
			Males		-0.011	0.605	0.986	1.013	0.559	0.071	0.976	0.329
			Females		-1.251	0.763	0.102	-1.004	0.743	0.177	-2.109	***0.035***
	Weight	MUFA	Both		-1.467	0.675	0.030	-1.601	0.884	0.070	-2.808	***0.005***
			Males		-1.381	0.919	0.133	-1.514	1.127	0.180	-1.989	0.047
			Females		-1.245	0.998	0.213	-1.822	1.469	0.216	-1.720	0.085
	BMI	MUFA	Both		-0.173	0.105	0.100	-0.341	0.146	0.020	-2.652	***0.008***
			Males		-0.046	0.126	0.713	-0.302	0.160	0.060	-1.338	0.181
			Females		-0.232	0.167	0.166	-0.424	0.265	0.110	-2.033	0.042
	WC	MUFA	Both		-0.167	0.105	0.110	-0.269	0.170	0.113	-2.203	***0.028***
			Males		-0.085	0.126	0.504	-0.363	0.187	0.052	-1.620	0.105
			Females		-0.188	0.165	0.256	-0.217	0.301	0.471	-1.344	0.179
	HDL-C	MUFA	Both		-0.339	0.259	0.191	-0.383	0.321	0.233	-1.749	0.080
			Males		-0.127	0.315	0.687	0.495	0.350	0.158	0.437	0.662
			Females		-0.485	0.407	0.234	-1.286	0.562	0.023	-2.246	***0.025***
	TAG (log)	MUFA	Both		0.015	0.010	0.153	0.015	0.015	0.322	1.739	0.082
			Males		0.014	0.016	0.380	-0.010	0.020	0.611	0.457	0.648
			Females		0.016	0.013	0.222	0.047	0.022	0.034	2.185	***0.029***
	TAG (log)	SFA	Both		0.019	0.010	0.049	0.028	0.016	0.082	2.599	***0.009***
			Males		0.025	0.015	0.093	0.025	0.022	0.260	2.021	0.043
			Females		0.013	0.012	0.274	0.035	0.023	0.119	1.769	0.077
PLIN2												
rs2046841	ratio_w_h	N3	Both		-0.026	0.016	0.105	-0.045	0.027	0.097	2.266	0.023
rs3824369	Glucose (log)	N3	Both				0.197			0.011		
	Glucose (log)	N3	Both		0.043	0.032	0.175	0.071	0.045	0.120	-1.981	0.048
	Glucose	N3	Both		0.060	0.034	0.080	0.153	0.051	0.003	-3.080	0.002

	(log)										
	Glucose (log)	N6	Both			0.234			0.019		
	Glucose (log)	N6	Both	0.007	0.005	0.116	0.007	0.005	0.132	-2.137	0.033
	Glucose (log)	N6	Both	0.007	0.005	0.146	0.015	0.005	0.006	-2.721	0.007
PLIN5											
rs4806985	BMI	N3	Females	-5.155	1.943	0.008	-3.925	2.972	0.187	-2.939	0.003
	WGT	N3	Females	-26.441	11.710	0.024	-29.362	16.422	0.074	-2.863	0.004
rs1610090	BMI	MUFA	Both	-0.205	0.091	0.024	-0.079	0.145	0.584	-2.201	0.028
	WAIST	MUFA	Both	-0.174	0.091	0.057	-0.143	0.167	0.392	-2.067	0.039
	Glucose (log)	SFA	Females	0.009	0.003	0.001	0.003	0.004	0.443	3.155	0.002
	WAIST	SFA	Both	-0.214	0.086	0.013	-0.063	0.168	0.710	-2.284	0.022
	WGT	N3	Females	-19.995	11.482	0.082	-31.482	16.247	0.053	-2.507	0.012
	BMI	N3	Females	-3.159	1.916	0.100	-4.520	2.942	0.125	-2.215	0.027
	TAG (log)	SFA	Males	0.017	0.015	0.259	0.064	0.026	0.014	2.287	0.022

Table 6. Significant Associations from Meta-analysis of the FOS and GOLDN populations for PLIN2, 3, & 5 - Interaction Effects.

Triglycerides(TAG), Waist Circumference(WC), Saturated Fat(SAT), Omega 3 polyunsaturated fatty acids(N3 PUFA), Monounsaturated fatty acids(MUFA). Gene by diet interaction for FOS, GOLDN and meta-analysis. Results are shown for the entire population (Both) and for males and females independently. Interactions between PLIN variants and dietary N3 and N6 was included in an multivariate regression model as continuous variables. Models were adjusted for sex, age, smoking, alcohol use, diabetes, beta-blockers, N3, N6, total energy and estrogen and menopausal status in women and Lipid and Glucose p-values where also adjusted for BMI and cholesterol medications.

* The meta analysis results is not valid for the global test of the general models

association in which both populations and meta-analysis demonstrated significance, that being the *PLIN3* rs3760950 SNP associating with fasting insulin levels. Interestingly, *in-silico* analysis with the webtool MAPPER predicted the rs3760950 SNP lies in a TF binding motif for SP1. SP1 is thought to be important for the expression of genes in response to insulin and other hormones. It is tempting to speculate on a possible functional consequence of rs3760950 whereby the minor allele reduces PLIN3 expression via reduced affinity for SP1 binding. If PLIN3 facilitates the uptake of FFAs from the serum into LDs, altered levels of PLIN3 protein may contribute to an increase in circulating TAG concentrations, and an increased risk for elevated insulin resistance.

Furthermore, we have identified a series of gene-diet interactions between *PLIN3* variants and dietary FFAs modulating anthropometric and lipid traits. These findings suggest dietary habits may affect the contribution of PLIN gene variants on one's risk for obesity and related comorbidities. The rs3760950 variant showed interaction with MUFA modulating anthropometric and lipid phenotypes. Interestingly, PPARg activation has been demonstrated to alter gene expression by modulating SP1 activity in mouse 3T3L1 cells, and MUFA is a ligand of PPARg (Chung et al. 2006). In addition, several studies have demonstrated gene by MUFA interactions modulating obesity and glucose phenotypes (Memisoglu et al. 2003; Perez-Martinez et al. 2008). It may be the differential phenotypic response among rs3706950 genotypes in response to MUFA is mediated in part through SP1 activity through PPARg stimulation. However, given what little is known about *PLIN3* regulation in humans, we can only speculate on how the rs3760950 SNP may alter *PLIN3* expression or activity in response to MUFA at this time.

In addition, our results demonstrate the *PLIN3* rs3760948 SNP interacts with SFA to modulate TAG levels in male subjects. *In-silico* analysis predicted the minor allele of the rs3760948 SNP increases the likelihood of a PPARg site in the promoter region of *PLIN3*. SFA is predicted to be only a weak agonist of PPARg. It may be that SFA stimulation of PPARg may increase *PLIN3* levels in individuals with one or more copies of the rs3760948 minor allele results in a more efficient clearance of serum TAGs than those individuals not carrying the minor allele.

Our analyses demonstrated no main effect or GxE associations with variants in the human *PLIN2* locus. Like *PLIN2*, no main effect associations were observed with variants in the *PLIN5* locus. However, several *PLIN5* SNPs did show interaction with dietary FFA modulating anthropometric variables among female subjects. The variant rs1610090 showed significant interaction with MUFA modulating body weight. rs1610090 falls in exon 7 of the *PLIN5* gene causing a non-synonomous amino acid substitution. The effect of this substitution on protein folding has not been studied, but it may be that this alteration effects protein relocalization, and in combination with increased MUFA intake alters anthropometric variables between genotypes.

The *PLIN5* rs4806985 SNP also showed a gender specific association, interacting with PUFA N3 to modulate body weight among female subjects. rs4806985 is predicted to abrogate a CEBPa binding motif in the *PLIN5* promoter. CEBPa is a TF responsible for activation of many adipogenic genes. Further, lower PLIN5 levels where shown to negatively correlate with adiposity in a small sample of human subjects (Wolins et al. 2006). If PLIN5 is responsible for transport of FFAs to oxidative pathways, a reduction in this process may result in deposition of these FFAs in the LSD and result in elevated

anthropometrics. However, our data show that the rs4806985 minor allele associates with lower anthropometrics in females, in response to increasing dietary PUFA N3. It may be that PUFA N3 modulates CEPBa activity, and this in combination with the effect of the minor allele results in a differential response. Currently, the effects of N3 PUFA on CEBPa activity are not well defined in humans.

Similar to associations found in the PAT family member *PLIN1* we observed gender-specific associations with the *PLIN5* variants rs1610090 and rs4806985 with anthropometrics among female subjects. It remains to be determined if sex specific hormones play a role in the associations observed with the rs4806985 and rs1610090 variants and anthropometric traits.

A potential limitation to this study is that we did not adjust our results for multiple-tests, suggesting some of our associations may be false positives. Furthermore, there was some heterogeneity in the levels of statistical significance between the FOS and GOLDN populations. For our baseline analyses we see that, overall, the FOS population showed significant associations with many more variables than in GOLDN which may be due to the larger sample size. To identify those associations of highest confidence we have reported only those associations showing nominally significant to significant associations in both populations studied and significance when combined in a meta-analysis. Furthermore, given the direction of the effect in both populations was the same and are supported by functional hypothesis we are confident. Nevertheless, additional replication studies able to account for dietary interactions will help clarify a role for variation in *PLIN3* and obesity related traits in humans. In addition, our *in-silico* predictions have not been experimentally validated, and it is possible that one or all of the

predictions are not biologically relevant. The associations observed here could also be due to more rare variants in LD with the SNPs genotyped, but not found in public databases. To date, this locus has not been identified in GWAS analyses investigating obesity related phenotypes. This may be explained by the possible poor coverage of this region on SNPchips, or perhaps that the effect size of these variants are too small to reach the GWAS threshold of significance.

In conclusion, we have identified several novel polymorphisms at the *PLIN* loci which associate with obesity related traits. Furthermore, our meta-analysis suggest *PLIN3* and *PLIN5* SNPs interact with MUFA, PUFA N3 and SFA to modulate lipid, glucose and obesity measures in humans of European ancestry. These findings may assist in explaining the inter-individual differences in insulin sensitivity and anthropometrics in response to varying levels of MUFA in the diet. Furthermore, this information may contribute to the formulation of personalized dietary recommendations. Taken together these data, in combination with previous studies investigating PAT genes, support a relevant role for PAT family in obesity-related traits.

Chapter 4: The *PLIN4* variant rs8887 modulates obesity related phenotypes in humans through creation of a novel miR-522 seed site.

Introduction:

The promoter region of *PLIN4* contains conserved and functional peroxisome proliferator-activated receptor (PPAR) response-elements (PPREs) (Tansey et al. 2001b). PPARs are a family of nuclear-receptor transcription factors that modulate many aspects of lipid metabolism (Schoonjans et al. 1996). PUFA are known ligands for PPAR receptors suggesting PAT genes respond to dietary lipids at the transcriptional level. Interestingly, several human studies showed that genetic variation at the *PLIN1* locus associates with anthropometric phenotypes in female subjects (Qi et al. 2004a; Qi et al. 2004b; Qi et al. 2005). Moreover, other studies have demonstrated gene by environment interactions for *PLIN1* influencing weight in response to Rosiglitazone and insulin resistance levels for women consuming diets high in saturated fat (Corella 2006; Kang 2006).

Although there have been numerous functional investigations into variants located in the promoter regions of candidate genes little attention has been given to variants falling in the 3'UTR where microRNAs (miR)s may bind. miRs regulate protein output, and individual miR-to-target mRNA interactions may act to dampen mRNA translation often by 33% or less (Stark et al. 2005). In line with the common disease-common variant hypothesis it has been proposed that variants mapping within miR targets, or which create novel miR-to-target interactions, have functional consequence resulting in subtle phenotypic variation (Saunders et al. 2007a).

We hypothesize that variation in human *PLIN4* may modulate obesity related phenotypes. To explore this, we performed a sample size weighted meta-analysis using results from association analysis of seven *PLIN4* SNPs with anthropometric, lipid and glucose variables in two populations, the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) and the Framingham Offspring Study (FOS**)**. We also investigated the interaction of dietary PUFA n3 and n6 with *PLIN4* SNPs to determine their combined potential to modulate these phenotypes. *In silico* prediction for SNPs falling in *PLIN4* regulatory regions was done to assess their potential for functional consequence. Our results indicated the rs8887 SNP creates a miR-522 miR recognition element (MRE) in the *PLIN4* 3'UTR. The ability of miR-522 to regulate *PLIN4* 3'UTR was examined. We investigated how genetic drift in the *PLIN4* 3'UTR in combination with environmental exposures may act in concordance, predisposing individuals to obesity. Although our association results were not adjusted for multiple tests, the combined evidence of meta-analysis and functional *ex-vivo* data, indicates rs8887 is a modulator of anthropometrics in humans.

Results:

Meta-analysis of PLIN4 variants with anthropometric, lipid and glucose related phenotypes from the FOS and GOLDN populations

The demographic and biochemical characteristics of participants for the FOS and GOLDN populations varied slightly, and are presented in **Table 3.** Genotypic characteristics are shown in **Table 4**. Genotype distributions did not deviate from Hardy-Weinberg equilibrium.

To test for overall significance of association of *PLIN4* SNPs with phenotypes of interest, we performed a meta-analysis that revealed significant associations between rs8887 and BMI *(P*=0.002), weight *(P*=0.017) and a nominal association with waist circumference *(P*=0.056) with minor allele carriers having elevated measures in each case **(Table 7)**. We discuss here only those associations for SNPs showing consistent trends with supporting functional hypotheses. However, a complete list of our findings can be viewed in **Table 7**. The direction of these effects is in agreement with those reported for the FOS and GOLDN populations.

Meta-analysis of interaction of dietary n3 and n6 PUFA with PLIN4 variants on anthropometric, lipid and glucose related phenotypes from the FOS and GOLDN populations

We performed a meta-analysis of interaction between *PLIN4* SNPs and PUFA n3 and n6. The interaction between PUFA n3 and rs8887 showed significant association modulating BMI *(P*=0.0144), weight *(P*=0.0068) and waist circumference *(P*=0.0145) where minor allele carriers showed a reduced response to PUFA n3 compared to non-carriers. An interaction between rs884164 and PUFA n3 showed BMI *(P*=0.008), weight *(P*=0.005), waist (0.035), glucose *(P*=0.0167) and TAG *(P*=0.0144) levels are increased in carriers of the minor allele with elevated PUFA n3 intake (**Table 8**). Furthermore, rs884164 showed significant interaction with PUFA n6 with HDL (P=0.036) levels decreasing and TAG (P=0.012) increasing among minor allele carriers with elevated PUFA n6 intake (**Table 8**).

Performing the meta-analysis separately for each gender revealed several

associations in the male population. Interactions were observed between rs884164 and PUFA n3 in which BMI (*P*=0.037), weight (*P*=0.028), glucose (*P*=0.017) and TAG (*P*=0.001) levels were modulated. The minor allele subjects showed elevated levels for each trait, compared to non-carriers.

Association of PLIN4 variants with visceral and subcutaneous fat measurements from the FOS population

Visceral adipose tissue (VAT) has been shown to correlate better with obesity related phenotypes such as insulin resistance and CVD than the more traditional anthropometrics (Fox et al. 2007). We performed an association analysis with *PLIN4* SNPs and volumetric computed tomography measures from a subset of the FOS study for whom those measures were taken. rs8887 associated with VAT (*P*=0.003) with carriers of the minor allele showing increased volume compared to non-carriers. In addition carriers of the rs8887 minor allele showed increased subcutaneous adipose tissue (SAT) (*P*=0.011) compared to non-carriers. Performing our analyses by gender revealed an association with rs8887 and VAT (*P*=0.001) with only male carriers having greater volume than non-carriers (**Table 7**).

Functional prediction of PLIN4 variants

We investigated if variants showing the most consistent associations may be functional. Using CEU data from HapMap we determined that rs8887 was not in LD with other known SNPs. As rs8887 is located in the 3'UTR of the *PLIN4* mRNA, we

			FOS				GOLDN				Meta-Analysis	
SNP	**Phenotype**	**Gender**	**Beta**	**Se**	**P**	**%Var**	**Beta**	**Se**	**P**	**%Var**	**z-score**	**P**
rs8887	BMI	Both	0.614	0.221	0.005	0.396	.581	0.378	0.125	0.334	3.164	0.002
		Males	0.624	0.269	0.021		0.326	0.461	0.480		2.319	0.020
		Females	0.631	0.335	.060		0.704	0.586	0.230		2.231	0.026
	Weight	Both	3.106	1.431	0.030	0.200	2.374	2.271	0.296	0.081	2.387	0.017
		Males	2.917	1.980	0.141		0.584	3.146	0.853		1.332	0.183
		Females	3.524	2.010	0.080		3.465	3.242	0.286		2.050	0.040
	Waist	Both	0.423	0.221	0.056	0.157	0.252	0.440	0.567	0.023	1.911	0.056
		Males	0.483	0.269	0.073		-0.085	0.565	0.880		1.413	0.158
		Females	0.381	0.334	0.253		0.397	0.656	0.556		1.278	0.201
	VAT	Both	199.5	67.55	0.003							
		Males	345.4	100.5	0.001							
		Females	68.7	83.97	0.413							
	SAT	Both	229.1	90.66	0.011							
		Males	180.4	106.2	0.090							
		Females	248.9	147.2	0.054							
rs7250947	BMI	Both	-0.571	0.271	0.035	0.199	-0.425	0.489	0.385	0.095	-2.238	0.025
		Males	-0.618	0.328	0.060		-0.163	0.596	0.785		-1.723	0.085
		Females	-0.630	0.413	0.128		-0.801	0.751	0.287		-1.858	0.063
rs8102428	BMI	Both	-0.519	0.262	0.048	0.146	-0.565	0.474	0.234	0.175	-2.305	0.021
		Males	-0.558	0.315	0.077		-0.560	0.582	0.336		-2.009	0.045
		Females	-0.461	0.403	0.253		-0.647	0.729	0.375		-1.443	0.149
	TAG	Both	0.011	0.026	0.667		0.064	0.047	0.173		1.108	0.268
		Males	0.041	0.039	0.300	0.057	0.159	0.069	0.023	0.951	2.119	0.034
		Females	-0.022	0.032	0.490		-0.044	0.061	0.474		-0.970	0.332
rs1609717	HOMA	Both	0.047	0.027	0.081	0.099	0.054	0.047	0.246	0.093	2.097	0.036
		Males	0.073	0.042	0.081	0.191	0.098	0.072	0.176	0.281	2.204	0.028
		Females	0.036	0.034	0.287		0.004	0.060	0.943		0.920	0.358
	Insulin	Both	0.046	0.023	0.045	0.153	0.041	0.042	0.329	0.070	2.203	0.028
		Males	0.074	0.037	0.045	0.285	0.047	0.065	0.469	0.088	2.062	0.039
		Females	0.035	0.028	0.220		0.023	0.054	0.674		1.250	0.211

Table 7. Significant Associations of *PLIN4* SNPs in the FOS and GOLDN populations - Main Effects.
Results of meta-analysis in FOS and GOLDN. P-values for anthropometrics were adjusted for sex, age, smoking, physical activity (GOLDN only), alcohol use, diabetes, beta-blockers, calories from fat, PUFA n3 and n6, and estrogen and menopausal status (FOS only) in women. Lipid and glucose p-values were also adjusted for BMI and cholesterol medications

				FOS				GOLDN				Meta-Analysis	
SNP	**Phenotype**	**PUFA**	**Gender**	**Beta**	**Se**	**P**	**%Var**	**Beta**	**Se**	**P**	**%Var**	**z-score**	**P**
rs8887	BMI	n3	Both	-0.469	0.391	0.230	0.48	-1.208	0.459	0.009	0.77	-2.447	0.014
			Males	-0.624	0.466	0.181		-1.158	0.542	0.033		-2.288	0.022
			Females	-0.438	0.625	0.484		-0.964	0.838	0.251		-1.216	0.224
	Weight	n3	Both	-3.867	2.522	0.125	0.30	-7.189		0.009	0.44	-2.707	0.007
			Males	-3.778	3.430	0.271		-7.080		0.057		-1.964	0.049
			Females	-4.553	3.750	0.225		-6.046		0.194		-1.728	0.084
	Waist	n3	Both	-0.461	0.391	0.238	0.23	-1.444	0.544	0.008	0.55	-2.445	0.015
			Males	-0.500	0.466	0.283		-1.230	0.672	0.068		-1.900	0.057
			Females	-0.421	0.621	0.498		-1.584	0.952	0.097		-1.483	0.138
rs884164	BMI	n3	Both	1.077	0.458	0.019	0.17	0.875	0.694	0.208	0.13	2.655	0.008
			Males	0.924	0.534	0.084		0.932	0.798	0.243		2.087	0.037
			Females	0.974	0.799	0.223		1.700	1.347	0.208		1.709	0.087
	Weight	n3	Both	6.860	2.995	0.020	0.12	6.689	4.116	0.109	0.12	2.819	0.005
			Males	6.202	3.958	0.117		8.806	5.443	0.106		2.195	0.028
			Females	5.475	4.801	0.254		7.803	7.490	0.298		1.524	0.128
	Waist	n3	Both	1.164	0.459	0.011	0.19	-0.015	0.791	0.985	0.01	2.106	0.035
			Males	0.938	0.538	0.081		-0.230	0.900	0.798		1.317	0.188
			Females	0.920	0.793	0.247		0.761	1.543	0.622		1.239	0.216
	TAG	n3	Both	0.086	0.046	0.062	0.13	0.111	0.069	0.106	0.31	2.448	0.014
			Males	0.152	0.067	0.024	0.48	0.240	0.094	0.011	1.45	3.289	0.001
			Females	0.003	0.064	0.958		-0.073	0.116	0.527		-0.303	0.762
		n6	Both	0.014	0.006	0.011	0.24	0.005	0.007	0.479	0.11	2.503	0.012
			Males	0019	0.008	0.013	0.52	0.02	0.009	0.035	1.07	3.229	0.001
			Females	0.007	0.008	0.421		-0.012	0.011	0.274		0.073	0.942
	Glucose	n3	Both	0.006	0.012	0.620	0.03	0.056	0.016	0.001	0.90	2.393	0.017
			Males	0.009	0.016	0.578		0.082	0.023	0.001		2.379	0.017
			Females	0.005	0.018	0.788		0.021	0.024	0.398		0.689	0.491
	HDL	n6	Both	-0.207	0.135	0.125	0.03	-0.215	0.145	0.138	0.31	-2.096	0.036
			Males	-0.194	0.153	0.204	0.11	-0.342	0.163	0.036	0.71	-2.215	0.027
			Females	-0.086	0.252	0.734		-0.019	0.267	0.942		0-.324	0.746
rs8102428	BMI	N3	Both	0.839	0.461	0.069	0.30	0.616	0.554	0.266	0.28	2.132	0.033
			Males	0.724	0.551	0.189		0.344	0.625	0.582		1.401	0.161
			Females	0.986	0.742	0.184		1.199	1.108	0.280		1.703	0.089
	TAG	N3	Both	0.111	0.046	0.015		0.093	0.055	0.089		2.958	0.003
			Males	0.149	0.069	0.031	0.43	0.158	0.073	0.032	2.01	2.981	0.003
			Females	0.081	0.060	0.177		-0.083	0.095	0.383		0.648	0.517
		N6	Both	0.010	0.006	0.067	0.10	0.006	0.006	0.299	0.32	2.099	0.036

rs892158	HDL	N6	Both	-0.035	0.113	0.757		0.090	0.104	0.391		0.214	0.830
			Males	0.049	0.134	0.716	0.07	0.180	0.113	0.112	0.90	1.177	0.239
			Females	-0.092	0.189	0.626		0.196	0.202	0.332		0.13	0.896

Table 8. Significant Associations from Meta-analysis of the FOS and GOLDN populations for *PLIN4* - Interaction Effects.
Gene by diet interaction for meta-analysis of FOS and GOLDN. Interactions between *PLIN4* variants and dietary PUFA n3 and n6 were included in a multivariate regression model as continuous variables. Models were adjusted as in Table 8.

searched for miRs predicted to bind the *PLIN4* mRNA using Targetscan.org and microRNA.org (Grimson et al. 2007; Betel et al. 2008a). Both programs predicted the binding of miR-522 with perfect complementarity to a seed site in the *PLIN4* mRNA when containing the minor A allele (**Figure 5**). In HapMap, rs884164 was estimated to be in LD with the downstream SNP rs8102428 (r^2=0.90), and the upstream SNPs rs11670485 (r^2=0.91) and rs892157 (r^2=0.90). Our data showed that the LD between rs884164 and rs8102428 in GOLDN (r^2=0.7) and FOS (r^2=0.54) were weaker than predicted suggesting that use of HapMap values to estimate LD at this locus may not be ideal. rs884164 falls in the promoter region of *PLIN4* and sequence analysis using the transcription factor motif prediction tool Mapper indicated that the major T allele lies in a consensus NFkB motif and the C allele abrogates this prediction (Marinescu et al. 2005a).

miR-522 targets the 3'UTR of PLIN4 containing the rs8887 minor A allele

We next investigated the functional potential of rs8887. miR-522 maps within the chromosome 19 microRNA cluster (C19MC), the largest known primate specific microRNA gene cluster (Bentwich et al. 2005; Zhang et al. 2008a). While there is evidence for miR-522 expression in placenta, testis, thymus, brain and prostate to our knowledge expression has not been demonstrated in adipose tissue (Bentwich et al. 2005; Landgraf et al. 2007). Thus to determine if miR-522 is expressed in human adipoctyes, we performed RT-PCR on total RNA samples extracted from cultured COS7, HEK293T, HepG2 cells, and primary human pre-adipocyte and mature adipocytes. This was followed by qPCR using SABiosciences miR-522 specific primers. **Figure 6a** shows the

normalized relative expression of miR-522, with highest expression in HepG2 cells, pre-adipocytes and adipocytes. *PLIN4* has been shown to be expressed in human adipose (Dalen et al. 2004).

To determine the effect of miR-522 on the *PLIN4* 3'UTR, we cloned into the siCHECK2 luciferase expression vector a 560-bp region of the *PLIN4* 3'UTR from genomic DNA of subjects homozygous for either rs8887 allele. COS7 cells were co-transfected with miR-522 mimic or control mimic, and with the A allele or the G allele *PLIN4* 3'UTR vector. The 3'UTR containing the minor A allele showed 20% reduction in luciferase signal in the presence of miR-522 compared to control mimic (**Figure 6b**). The 3'UTR containing the G allele showed a non-significant increase in luciferase signal in the presence of miR-522 compared to control mimic. These data indicate an ability of miR-522 to bind and partially repress luciferase expression via the *PLIN4* 3'UTR segment when carrying the derived A allele of rs8887. The effect of increasing concentrations of miR-522 on PLIN4 3'UTR with the A allele is saturated at 20nM of miR-522 (**Figure 6c**).

Genetic drift and the PLIN4 3'UTR

Genetic drift occurring in 3'UTRs can result in the formation of new MREs. These new MREs confer beneficial, neutral or detrimental effects on the organism, leading to conservation, neutrality or selective avoidance of the MRE, and such mechanisms have had considerable influence on 3'UTR evolution (Stark et al. 2005). We examined the *PLIN4* 3'UTR across ten mammalian species. A phylogenetic tree depicts the evolutionary distance of *PLIN4* 3'UTRs across these species (**Figure 7a**).

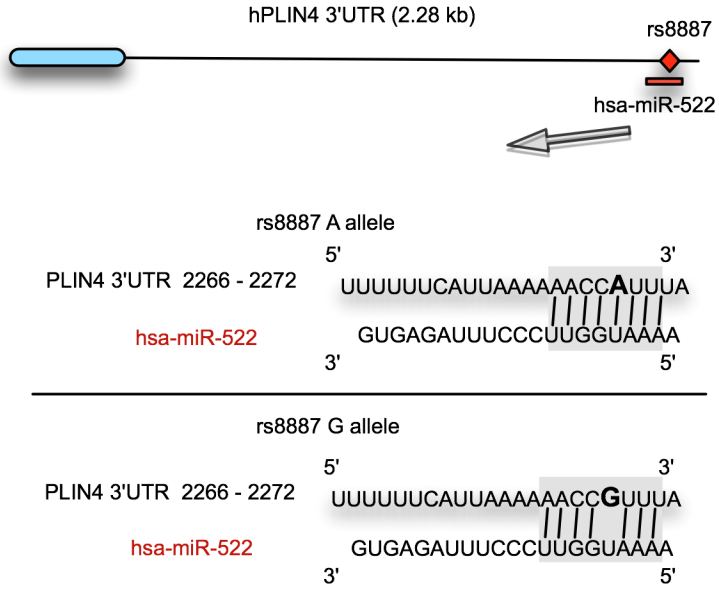


Figure 5. The rs8887 minor A allele creates a novel miR-522 MRE in the *PLIN4* 3'UTR. Diagram of the miR-522:*PLIN4* 3'UTR sequences with the A or G allele. The miR-522 seed site is highlighted in gray, and the rs8887 variants are in bold.

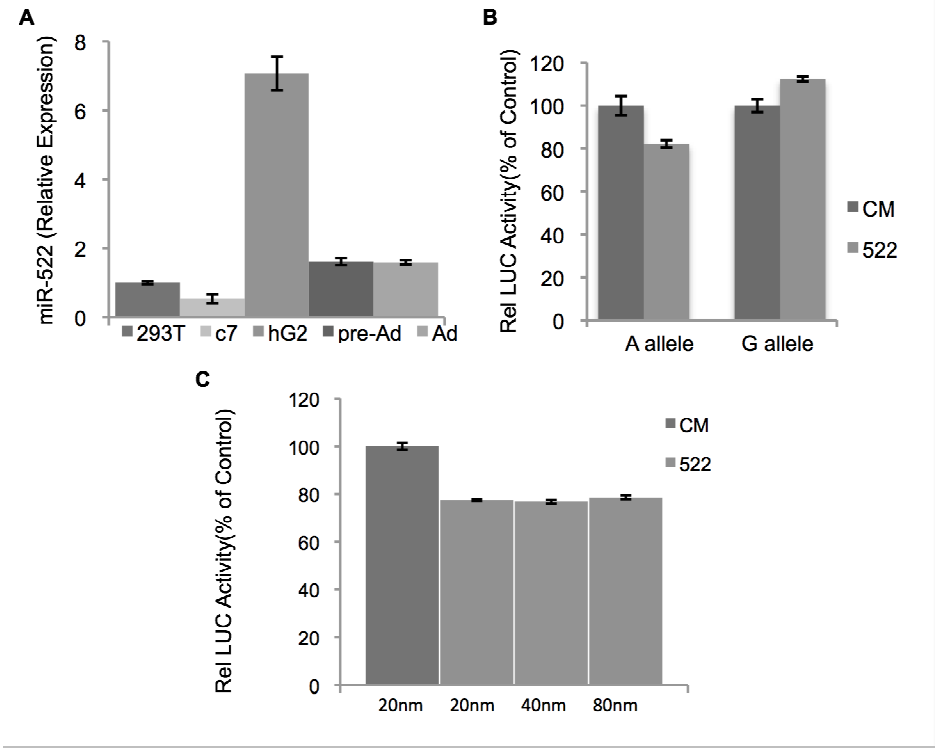


Figure 6. The *PLIN4* 3'UTR with the A allele creates a miR-522 MRE.

A) Relative miR-522 expression across indicated cell types, hek293T (293T), Cos7 (c7), hepG2 (hG2), pre-adpocytes (pre-Ad) and adipocytes(Ad). B) Luciferase expression of pmiR-LucPLIN4-G or A constructs with miR-522 (522) or control mimic (CM). Data are expressed as relative luciferase activity to control samples. C) Luciferase expression of pmiR-LucPLIN4-A constructs with increasing concentration of miR-522 compared to control mimic. All data represent experiments performed in triplicate. Statistical Analysis: P values for the difference between luciferase activity obtained for LucPLIN4-A in the presence of mir-522 or control mimic (P= .0169) or LucPLIN4-G in the presence of mir-522 or control mimic (P= .0584) were determined using the student's paired t-test.

The *PLIN4* 3'UTR has undergone significant change from mouse to primate and furthermore that change is ongoing even among recently diverged primates such as neandertal and human. **Figure 7b** shows an alignment of the last 100 bases of the *PLIN4* 3'UTR. The reference nucleotide for all sequences except mouse, neandertal and human is a C at the rs8887 position. The mouse has no orthologous sequence for this segment and neandertal and human sequence have a G or an A/G nucleotide, respectively, suggesting the recent emergence of the 522 MRE.

To investigate if the recent rs8887 SNP is undergoing selection across human populations, we utilized the fixation index (F_{ST}) statistic, which measures the divergence of alleles across populations (Akey et al. 2002a). The frequency of a particular allele in populations can vary over time and is influenced by such forces as genetic drift and natural selection and F_{ST} values can be used to approximate these influences. F_{ST} statistics among total and population subdivisions of HapMap PHASE III data were obtained with the SNP@Evolution web tool (**Table 9**) (Cheng et al. 2009b). The total F_{ST} value for rs8887 was 0.132 indicating a moderate level of differentiation between populations, suggesting rs8887 is undergoing drift.

Discussion:

We report novel associations between SNPs in human *PLIN4* and obesity related phenotypes. We also have identified a series of gene by diet interactions modulating these traits. Of particular interest rs8887 associates with a constellation of anthropometric traits which were modulated through interaction with dietary PUFA. *In*

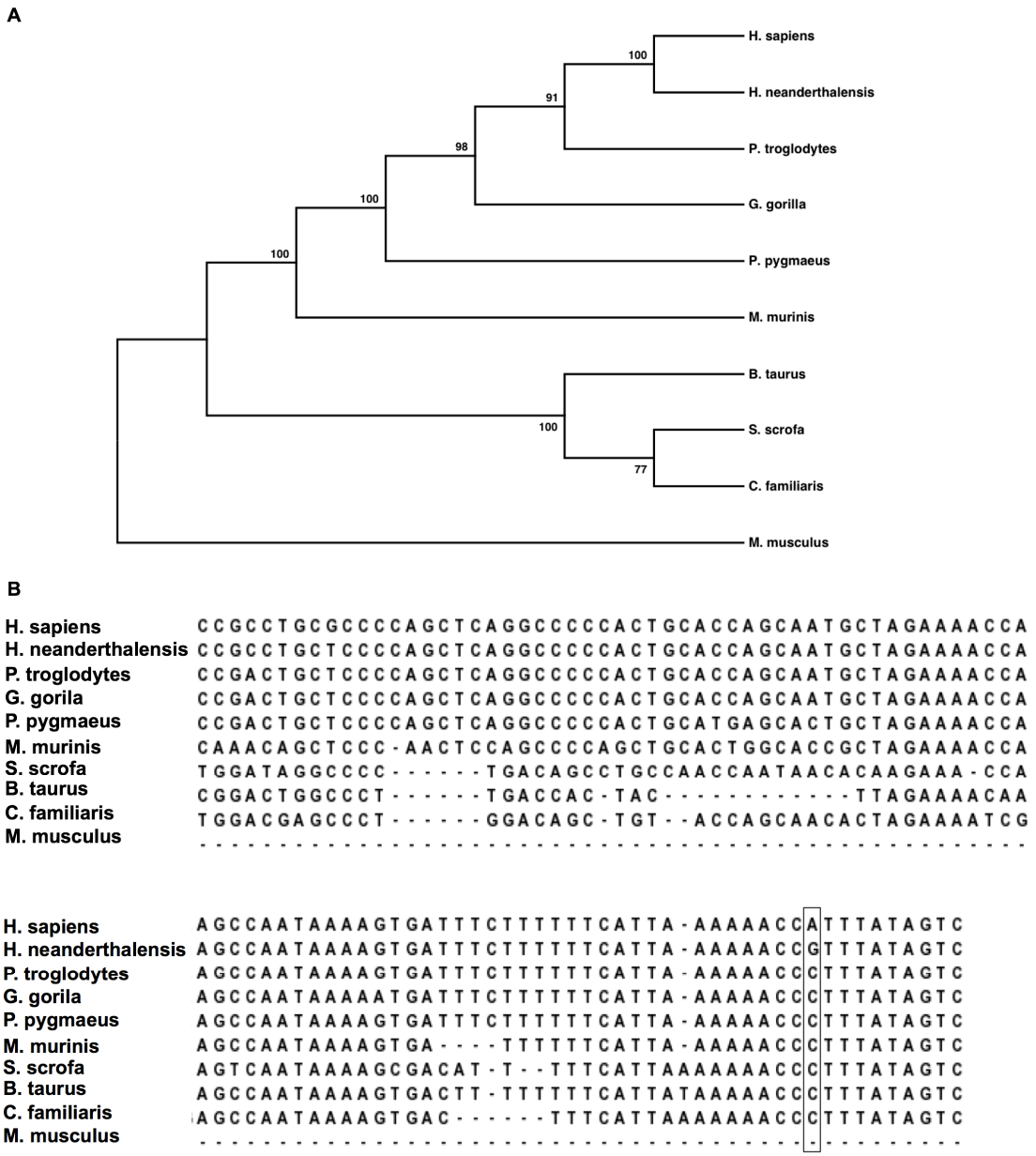


Figure 7. Evolutionary history of the *PLIN4* 3'UTR.

A) The evolutionary history of *PLIN4* was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 10000 replicates is taken to represent the evolutionary history of the taxa analyzed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches. The MP tree was obtained using the Close-Neighbor-Interchange algorithm with search level 3 in which the initial trees were obtained with the random addition of sequences (100 replicates). B) Clustal W was used to align the *PLIN4* 3'UTR sequences from H. sapiens, H. neanderthalensis, P. troglodytes, G. gorilla, P. pygmaeus, M. murinis, C. familiaris, B. Taurus, S. scrofa, and M. musculus. The last 100 bases of the alignment are show here, with the position harboring the rs8887 SNP outlined.

silico analysis of the *PLIN4* mRNA sequence predicted the minor A allele of rs8887 generates a novel seed site for miR-522 and our *ex vivo* luciferase data indicated that miR-522 reduced PLIN4 protein levels 20% via the *PLIN4* 3'UTR target site created by the rs8887 A allele. Importantly, single point mutations in MRE seed sites have shown the ability to reduce or abolish miR-mediated repression (Brennecke et al. 2005).

When 3'UTR sequences drift over the course of evolution, they are continuously exposed to potential matches with co-expressed miRs. While conservation signal is often used to predict functional MREs, it has been determined that a conservation signal above background for MREs of the most recent mammalian specific miRNA families was unlikely due to the relatively short time between the emergence of these miRs and the occurrence of new MREs within 3'UTRs (Friedman et al. 2009b). This suggests that for some of the more recent MREs to emerge, there has not been sufficient time for environment to determine which sites are beneficial, neutral or detrimental with respect to the genome. It is likely that some primate 3'UTRs have been, and are, subject to drift via the appearance of new genetic variants, resulting in loss or gain of miR-target interactions which may have potential for phenotypic modulation (Saunders et al. 2007a). However, to date, there are few known examples of genetic variation in miR-target sites contributing to phenotypic variation (Sethupathy and Collins 2008b).

C19MC is thought to be a product of an AluJ/AluS insertion into chromosome 19 during an early stage of primate evolution suggesting a role for miR-522 in higher development and phenotypic plasticity (Zhang et al. 2008a). It is likely that miR-522 is important for development given its temporal expression in placenta and fetal tissues, however its role in adipocytes is unknown. Our phylogenetic analysis indicates that

Populations	FST
ASN	0.01678
EUR	0.00258
AFR	0.03766
AME	0.08720
AEAA	0.13228

Table 9. FST values for rs8887 among populations.

ASN: samples of Asian,CHB, CHD, and JPT
- CHB: Han Chinese in Beijing, China
- CHD: Chinese in Metropolitan Denver, Colorado
- JPT: Japanese in Tokyo, Japan

EUR: samples of European, CEU and TSI
- CEU: Utah residents with Northern and Western European ancestry
- TSI: Toscans in Italy

AFR: samples of African, YRI, ASW, LWK, and MKK
- YRI: Yoruba in Ibadan, Nigeria
- ASW: African ancestry in Southwest USA
- LWK: Luhya in Webuye, Kenya
- MKK: Maasai in Kinyawa, Kenya

AME: samples of American,GIH and MEX
- GIH: Gujarati Indians in Houston, Texas
- MEX: Mexican ancestry in Los Angeles, California

AEAA: ASN, EUR, AFR and AME

variation at the rs8887 position resulting in the *PLIN4* miR-522 MRE is specific to humans and likely undergoing drift. We hypothesize that binding between rs8887 and mir-522 results in suboptimal expression of *PLIN4*, thereby contributing to the elevation in anthropometrics observed in our association analyses. A layer of complexity in controlling expression of the *PTEN* oncogene has been described as an interaction between a microRNA and pseudogene *PTENP1* (Poliseno et al. 2010). We find no evidence for a *PLIN4* pseudogene in the human genome which strengthens the implications of the miR-522-PLIN4 interaction we describe here.

If the appearance of variation at the rs8887 position is recent to human, it is tempting to speculate that the genesis of the rs8887 minor A allele may have contributed to the phenotypic diversification distinguishing humans from other primates as it is thought that the gain, or loss, of genetic regulatory mechanisms are critical for the evolutionary process (King and Wilson 1975). That this interaction may have contributed to the evolution of the human brain is food for thought (Cunnane and Crawford 2003).

Our association data indicate for rs8887 minor allele carriers that elevated intake of PUFA n3 results in decreasing anthropometrics compared to non-carriers. Due to what little is known of *PLIN4* regulation, it is difficult to propose a mechanism by which the miR-522 rs8887 interaction together with PUFA n3 could modulate anthropometrics. It is likely that PUFA n3 alters *PLIN4* expression through PPAR mediated pathways (Dalen et al. 2004). Furthermore, studies in model organisms have demonstrated anti-obesity effects of PUFA n3 which are thought to mediate their effects by modulating the activity of various transcription factors important to lipid metabolism (Sampath and Ntambi

2005). It may be that miR-522 is dysregulated in the obese and thus contributes to the dysregulation of adipogenic pathways as suggested for another C19MC member, miR-519d (Martinelli et al. 2010). Alternatively, miR-522 may be modulated by environmental factors which influence *PLIN4* through rs8887 as suggested for several other miRs (Rayner et al. 2010). If in addition to regulating *PLIN4* expression PUFA n3 down-regulates miR-522, we would expect increasing PUFA n3 intake to have a more dramatic effect in reducing weight for subjects carrying the minor allele compared to non-carriers. Specifically, if the miR-522 *PLIN4* interaction is absent in those homozygous for the G allele, reducing miR-522 activity through increasing PUFA n3 will have no additional effect on increasing *PLIN4* expression, and therefore no added contribution to weight loss. FOS subjects have on average less PUFA n3 intake than GOLDN subjects (**Table 3**). In addition, our associations for baseline anthropometrics are more significant in FOS, while p-values for interaction with PUFA n3 are less significant compared to GOLDN values (**Tables 6, 7).** The reduced levels of PUFA n3 intake in FOS subjects possibly bias associations toward significance of main effects, while biasing against significance for interaction with PUFA n3 (**Figure 8**). Identifying the function(s) of miR-522 and the conditions that induce its activation and repression will help clarify its role in mammalian development and as a potential modulator of obesity phenotypes.

We can only speculate how lower expression of PLIN4 contributes to obesity-related phenotypes. For the related PLIN1, one study demonstrated that obesity and high lipolysis rates are independently associated with lower PLIN1 protein levels in women, whereas another demonstrated reduced levels of both *PLIN1* mRNA and protein in obese

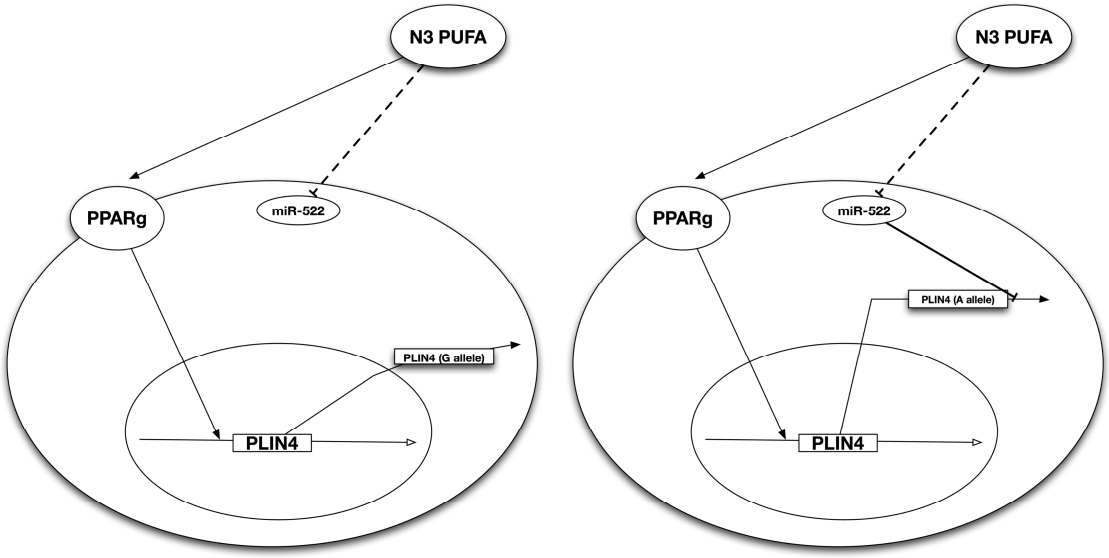


Figure 8. Model depicting hypothesized mechanism of PUFA N3 and miR-522. Our data suggest that carriers of the minor A allele of rs8887 are more sensitive to weight reduction in response to N3 PUFA intake than those homozygous for the major the G allele. One potential model that would explain this interaction is if N3 PUFA might have an inhibitory role on miR-522. If this were the case for rs8887, then carriers of the minor allele in response to N3 PUFA would have increased expression of PLIN4 through PPARG stimulation, in addition to increased PLIN4 due to inhibition of miR-522 repression on PLIN4 transcripts. This interaction may make A allele carriers more sensitive than G allele carriers. Solid lines indicate established regulatory connections and dashed lines indicated predicted connections.

compared to non-obese subjects (Mottagui-Tabar et al. 2003; Wang et al. 2003). Conversely, the *Plin1*$^{-/-}$ mouse is characterized by a lean phenotype. These data suggest the role and regulation of the PAT gene family in human obesity may be different than in model organisms.

In addition to our findings with rs8887, the rs884164 variant showed significant interaction with PUFA n3 modulating anthropometric and lipid traits. The rs884164 variant was predicted to fall in an NFkB motif. NFkB acts downstream of *TNFa* signaling and is thought to contribute to the pro-inflammatory response observed in obese individuals (Juge-Aubry et al. 2005). In addition to the activation of other pro-inflammatory cytokines, which can lead to disruptions in insulin signaling, NFkB is thought to up-regulate lipogenic factors and down-regulate adipogenic factors thereby increasing serum FFAs and further contributing to the insulin resistance and CVD associated with obesity (Ruan et al. 2002). Laurencikiene *et al* demonstrated *in vitro* that lipolysis was abolished upon the inactivation of NFkB in human adipocytes. Moreover, NFkB was shown to elevate *PLIN1* and Hormone Sensitive Lipase (*HSL)* expression during lipolytic stimulation (Laurencikiene et al. 2007b). It has been demonstrated that PUFA n3 can modulate the expression levels of NFkB target genes (Camandola et al. 1996). PUFA n3s were shown to decrease levels of NFkB target genes by limiting the translocation of NFkB subunits from the cytoplasm to the nucleus (Zhao et al. 2004). However, the extent through which NFkB affects energy storage and expenditure is not well characterized in humans. If and how rs884164 may modulate PLIN4 response to PUFA through NFkB remains to be determined.

Several variants at the *PLIN4* locus are estimated to explain a small portion of the

variance observed in anthropometric traits from main effects and by the interaction of these variants with PUFA intake. These estimates do not appear to account for a large amount of phenotypic variability. However, as with the case of *FTO* and BMI, common variants modulating anthropometric traits often explain only a small amount of the observed phenotypic variation (Frayling et al. 2007). It is likely that variation at *PLIN4* is yet another contributor to the complex nature of obesity and its associated comorbidities.

A potential limitation to this study given the hypothesis driven nature of our analyses and the correlation between traits examined, is a lack of adjusting our results for multiple-tests. Furthermore, there was some heterogeneity in the levels of statistical significance between the FOS and GOLDN populations which may be due to the larger sample size of FOS or may be explained in part by differing PUFA n3 intakes between populations. However, we are confident in our findings given that the direction of the effect in both populations was the same, that our meta-analysis demonstrated overall significance and that the rs8887 functional data support our conclusions. The type and quantity of fat in the diet are an important factor in determining risk for obesity. To this end, a variety of dietary recommendations are suggested for obesity prevention and therapy. A concern regarding these recommendations is accounting for possible inconsistencies introduced by other factors affecting the desired outcomes. The data presented here offer an example of this occurrence in that subjects carrying the rs8887 minor allele are potentially more sensitive to lowering their previously elevated anthropometrics by increasing their PUFA n3 intake. This work may help enable health

professionals to better tailor an effective weight-loss regimen based on a patients DNA profile.

Chapter 5: A genome-wide survey identifies an appreciable number of SNPs within predicted microRNA seed sites some of which are in LD with GWAS disease associating variants.

Introduction:

microRNAs (miRs) are small 20-24 nucleotide (nt) noncoding RNAs that mediate translational repression by binding to miR recognition elements (MREs) found in the 3'UTR of their mRNA targets (Bartel 2009). The most critical region for binding and repression of mRNA by a miR are positions 2-7 of the MRE, referred to as the seed site. Although there are examples of miRs targeting mRNAs without perfect Watson-Crick complementarity to the MRESS, a collection of evidence supports the MRESS as the most important feature for prediction and function. In some cases single 7mer seed sites are sufficient for a miR to repress translation, and *ex-vivo* experiments have shown single point mutations in the MRESS may reduce effectiveness or abolish miR mediated repression (Brennecke et al. 2005). Further highlighting the importance of this sequence, it has been demonstrated that a higher degree of negative selection occurs within predicted conserved MRESSs compared to conserved control sites (Chen and Rajewsky 2006). Given the importance of the MRESS, it has been proposed that single nucleotide polymorphisms (SNPs) mapping within the MRESS, or which create novel MRESS (CNM), may have functional consequences resulting in phenotypic variation (Sethupathy and Collins 2008a). Moreover, SNPs that create or abrogate MRESSs may modulate gene transcript and protein levels relevant to a phenotype of interest, generally, or under the influence of particular environmental conditions.

It has long been thought that disease causing variants act through modulation of gene expression. This has been demonstrated for many promoter SNPs in which the risk allele alters the affinity of a transcription factor to its binding motif (Botma et al. 2001; Talmud et al. 2005). Furthermore, several published examples show functional variants in MREs that modulate risk for a variety of disease states, such as breast cancer, Tourette's syndrome, and hypertension among others (Sethupathy and Collins 2008a). Two studies, including our work identifying the rs8887 minor allele creating an MRESS, have demonstrated a gene by environment interaction where a MRESS SNP modulates individual response to drug and dietary intakes. A survey of the frequency of predicted and validated MRESS SNPs using dbSNP build 126, identified an appreciable number SNPs falling within MREs across the human genome (Saunders et al. 2007b). Since this time the number of reported SNPs reported in dbSNP has nearly doubled from build 126 to the latest release, dbSNP build 132. Therefore, it remains unclear if examples like the *PLIN4* SNP rs8887 are rare occurrences or if they might be representative of a process occurring more commonly across the human genome.

Interestingly, the number of risk alleles identified with plausible mechanisms for modulation of gene expression is outweighed by SNPs falling in gene desert regions (Visel et al. 2009). It could be that these SNPs fall within distant but bona fide enhancer or suppressor elements resulting in the modulation of gene expression, as was demonstrated for the variant within the 8q24 gene desert and its effects on *TP53* expression in prostate (Wasserman et al. 2010). Alternatively, it may be these SNPs are in LD with variants not yet identified or available on GWAS chips. For example, sequencing of the *HLA-C* 3'UTR revealed a SNP modulating an MRE for the binding of

miR-148. Furthermore, this SNP was shown to be in LD with rs9264942 which is found 35 kb upstream of *HLA-C* and associates with control of HIV (Kulkarni et al. 2011). These data demonstrated that rs9264942 is a marker for a functional SNP that was not contained in commercial SNP arrays. Further underscoring this point, recent chromatin studies have identified novel non-coding gene regulatory regions, some of which contain top scoring hits for disease associating SNPs (Ernst et al. 2011).

Currently over 1000 human miR sequences are reported in the miRbase catalog (Kozomara and Griffiths-Jones 2011). Estimates suggest that over 30% of human protein-coding genes are regulated by miRs, and that each miR may potentially regulate hundreds of target transcripts (Lewis et al. 2005; Friedman et al. 2009a). Given this large number of potential miR targets in the human genome, identifying allele-specific miR-mRNA interactions may help elucidate functional roles for a portion of the many SNPs identified in genome wide association studies (GWAS) that lack functional hypotheses.

With such information in mind, one aim of the 1000 Genomes Project is to catalog over 95% of human variation in order to inform association studies of all potential causal SNPs (Durbin et al. 2010). Furthermore, initial studies in the 1000 Genomes pilot indicated that a substantial number of variants are in LD with known disease markers and that these variants are not well covered on commercial arrays. Importantly, the data currently available in the 1000 Genomes Project provides unprecedented access to millions of SNPs, some of which may elucidate functional mechanisms for the many risk alleles identified in GWAS.

Here we have performed a genome-wide survey for SNPs falling within experimentally validated and computationally predicted conserved MRESSs, by utilizing

these data (dbSNP build132) (Durbin et al. 2010). In addition to this analysis, we have surveyed this data for predicted CNM SNPs. Furthermore, we have examined all SNPs identified in GWAS for functional variants in relation to predicted MRESSs and CNM SNPs using the data from the 1000 Genomes Project. Combing with several other publically available data sources, we identified numerous MRE SNPs as possible modulators of disease relevant phenotypes. Our work demonstrates the utility of the data generated from the 1000 Genomes Project and provides insight into the frequency and relevance of MRE SNPs in human evolution and disease.

Results:

Approximately 5 % of validated MREs contain SNPs in their seed site.

To assess the frequency of SNPs falling in validated MRESSs, we first determined the genomic DNA (gDNA) coordinates of 606 validated mRNA target seed-sites for all mRNA-miR interactions, from the miRecords database (Xiao et al. 2009). For a site to be included in this list we required functional evidence for the target site (eg, loss of function experiment through a reporter assay system). We searched each reported validated site for 4 classes of "canonical" seed sites. Here we define canonical seed sites as having, at least, perfect pairing among seed site positions 2-7 (6-mer) in addition to three other classes with binding site characteristics at positions 1 or 8, demonstrated to improve likelihood of repression; 8mer (an A nt at position 1, and a complementary nt at position 8), 7mer-8m (complementary nt at position 8), and 7-mer-A1 (an A nt at position 1) (Bartel 2009). We then determined if the gDNA coordinates of all 3'UTR SNPs (from dbSNP132) fell within the gDNA coordinates of each validated MRESS from above. We

identified 31 SNPs (5%) that lie in validated MRESSs corresponding to 28 target transcripts (**Table 10**). Note that this does not include rs8887, as it is not found in the miRecords database.

No population frequency data are available for 29% of the MRE SNPs (9 of 31)**,** a value that will change as whole genome sequence data from more individuals surface and with completion of more encompassing GWAS studies. Nine SNPs have minor allele frequencies (MAF) <= 2%, and may be considered rare in the general population and therefore unlikely as common factors in complex disease. The 13 remaining SNPs have allele frequencies above 2% in at least one population listed in dbSNP (**Table 10**). Of note, 7 of the 31 (22%) SNPS identified have shown association with disease traits, emphasizing the potential importance of MRESS SNPs as modulators of disease risk.

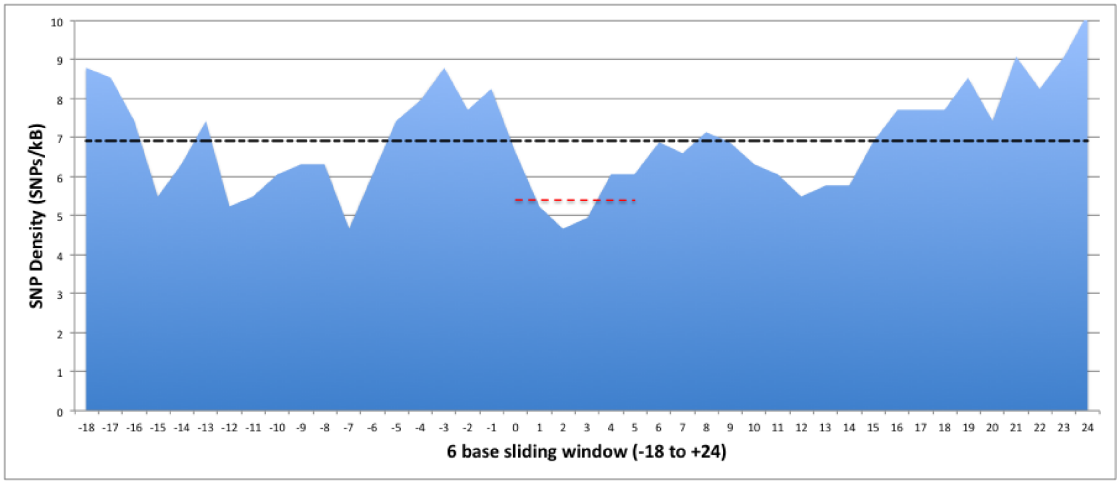
Previous studies have estimated MRESS SNP density to be lower than that observed in regions outside the MRESS, suggesting a higher rate of negative selection on MRESS (Chen and Rajewsky 2006; Saunders et al. 2007b). In light of the updated account of variation in the human genome available in dbSNP build132, we estimated the frequency of SNPs falling within MRESS and those falling outside of MRESSs using the 606 sequences from validated target sites identified from the above analysis. We performed a sliding window search of 6 bases, (the size of seed positions 2-7) starting 18 bases upstream and continuing to 24 bases downstream of each validated MRE site, sliding at a 1 base step. The 0 mark of the x-axis in **Figure 9** demarcates the second position of the MRESS (or first position of the 6mer seed site). Our data indicate that the MRESS contains the lowest amount of variation across the region, an observation in agreement with prior analyses (Chen and Rajewsky 2006; Saunders et al. 2007b).

Rs#	Coordinates	Maj	Min	MAF	Site type	Pos in MRE	miR	Gene	pubmed id
rs3783620	1: 101204463	G	A	.005 (YRI.P1)	8-mer	7th	hsa-miR-126	VCAM1	18227515
rs1059479	1: 113243892	T	G	.01 (CEPH)	8-mer	1st	hsa-miR-138	RHOC	20232393
rs12392	2: 198351529	G	A	NA	7-8mer	2nd	hsa-miR-1	HSPD1	17715156
rs5186*	3: 148459988	A	C	.306 (CEU.P1)	7-8mer	4th	hsa-miR-155	AGTR1	16675453
rs56109847*	3: 183824557	G	A	.992 (CEU.P1-LC)	8-mer	4th	hsa-miR-510	HTR3E	18614545
rs3731563	3: 48199695	T	C	.017 (GIH)	8-mer	8th	hsa-miR-21	CDC25A	19826040
rs1434536*	4: 96075965	C	T	.545 (TSC-CSHL)	7-8mer	1st	hsa-miR-125b	BMPR1B	19738052
rs6875894	5: 112179965	C	T	.027 (YRI)	7-8mer	4th	hsa-miR-135b	APC	18632633
rs6875894	5: 112179965	C	T	.027 (YRI)	7-8mer	5th	hsa-miR-135a	APC	18632633
rs79468771	6: 135539805	T	A	NA	8-mer	7th	hsa-miR-15a	MYB	18818396
rs33986155	6: 152420685	C	G	.083 (CEU.P1)	7-8mer	8th	hsa-miR-206	ESR1	17312270
rs11551509	6: 34505633	C	A	NA	8-mer	8th	hsa-miR-510	SPDEF	18922924
rs8829	7: 148504618	A	C	1 (CEPH)	8-mer	2nd	hsa-miR-101	EZH2	20478051
rs78899540	7: 27181092	A	C	.04 (YRI.P1)	7-mer-A1	2nd	hsa-miR-130a	HOXA5	17957028
rs117556949	7: 27194074	T	C	.992 (CEU.P1-LC)	8-mer	8th	hsa-miR-196a	HOXA7	15105502
rs12720208*	8: 16850399	G	A	.125 (CEU.P1)	8-mer	6th	hsa-miR-433	FGF20	18252210
rs78202059	8: 26228382	G	T	.144 (YRI.P1-LC)	7-mer-A1	8th	hsa-miR-222	PPP2R2A	20103675
rs1058153	2: 46987391	C	T	NA	8-mer	2nd	hsa-miR-21	SOCS5	17991735
rs72808106	10: 74035161	A	G	NA	7-mer-A1	4th	hsa-miR-221	DDIT4	20018759
rs16917496*	12: 123893830	C	T	.22 (CEPH)	7-mer-A1	2nd	hsa-miR-502-5p	SET8	19789321
rs111842797*	12: 123893831	A	G	NA	7-mer-A1	8th	hsa-miR-502-5p	SET8	19789321
rs76290581	12: 6760093	T	C	.051 (YRI.P1-LC)	8-mer	1st	hsa-miR-650	ING4	20381459
rs77080081	15: 40382144	A	G	.014 (CEU.P1)	7-8mer	1st	hsa-miR-125b	BMF	19471102
rs28521337*	15: 88521280	C	G	.467 (CEU.P1)	8-mer	8th	hsa-miR-485-3p	NTRK3	19370765
rs72481816	15: 88521572	G	C	NA	8-mer	8th	hsa-miR-765	NTRK3	19370765
rs72481814	15: 88522372	T	C	NA	7-8mer	4th	hsa-miR-509-3p	NTRK3	19370765
rs28574753	16: 28109760	G	A	.076 (YRI-P1)	7-8mer	2nd	hsa-miR-122	XPO6	19296470
rs75817141	17: 12044581	C	T	.02 (YRI.P1)	7-8mer	1st	hsa-miR-15b	MKK4	19861690
rs62062994	17: 48261978	G	T	NA	8-mer	3rd	hsa-miR-29c	COL1A1	18390668
rs3218074	19: 30315176	A	G	.01 (PDR90)	8-mer	8th	hsa-miR-15b	CCNE1	18701644
rs3218074	19: 30315176	A	G	.01 (PDR90)	8-mer	8th	hsa-miR-16	CCNE1	18701644
rs6094029	20: 43356176	C	T	0	7-8mer	7th	hsa-miR-449a	WISP2	19351815
rs78301106	20: 62522710	C	G	.045 (CHB+JPT.P1)	8-mer	7th	hsa-miR-122	TPD52L2	19296470

Table 10. Validated MRESSs containing SNPs.

A list comprising 31 validated MREs (with perfect complimentarity in the MRE seed site -positions 2-7) in which a SNP has been identified. The Coordinates column provides the chromosomal number and position coordinates for each SNPs - all genomic coordinates correspond to Hg19. The MAF column reports the allele frequency of the minor allele from the population where it was highest, where NA indicates a SNP with unknown allele

frequency. The Pos in MRE column refers to the position within the MRE to which the SNP maps. * SNPs reported to have shown association with disease phenotypes.



!

Figure 9. SNP Density across validated MRESSs.
A measure of SNP density (SNPs/kb) generated from the analysis of a 6 base window sliding over a 42-base region - centered on the first position of the seed site - of 606 validated MREs. The black like indicates the number of SNPs/kB across the 42 base region. The red line indicates the average SNP density across the 6 windows of seed positions 2-7.

Although there is less variation across the MRESSs there appears a considerable density of SNPs (5.5/kb) falling within what is thought to be the most important sequence for miR-mRNA interactions.

Genome wide survey of predicted MRESS and CNM SNPs

Although many variants have been associated with the modulation of phenotypes relevant to disease in GWAS studies, the challenge of determining which of them may be casual remains (2010). To identify potential causal variants mediated by MRESS creation or disruption we first performed a genome-wide survey for SNPs falling within computationally predicted conserved MRESSs. To do this, we utilized the microRNA.org portal to access a collection of predicted miR-mRNA interactions. These predictions were derived using an algorithm that incorporates an array of the most recent miR prediction guidelines, such as seed-site pairing, site context, free-energy, and conservation across multiple vertebrates (Betel et al. 2010). By comparing the gDNA coordinates of each predicted MRESS against the gDNA coordinates of dbSNP132 SNPs, we identified 2723 MRESS SNPs interrupting 5797 predicted interactions. To further prioritize these hits, we classified them by the type of seed match the MRESS SNP was predicted to interrupt; 8mer (2245), 7mer-8m (3251), 7-mer-A1 (180) or 6mer (121). Although there is overlap in the degree of efficiency of repression by these different seed type classes (likely dependent of site sequence context), there remains a hierarchy with 8mer sites being most efficient (Bartel 2009). Interestingly 38% (2245) of MRESS SNPs fall within predicted 8mer MRESSs. It has been estimated that ~50% of predicted MREs are potentially functional and it is likely that a portion of the SNPs identified here fall

within bona fide MREs (Rajewsky 2006). Overall, we estimate that 3% of high confidence predicted conserved MRESSs contain SNPs.

In addition, we estimated the number of SNPs interrupting non-conserved MRESSs. Our analysis identified 17895 SNPs interrupting 28816 predicted non-conserved MRESSs. Further prioritizing these hits, we classified them by the type of seed match the MRESS SNP was predicted to interrupt; 8mer (11801), 7mer-8m (15657), 7-mer-A1 (679) or 6mer (679). Over all we estimate that 3.5% of high confidence predicted non-conserved MRESS contain SNPs and 41% of the sites interrupted are 8mer seed types. As a positive control, we identify rs8887 as interrupting a 8mer MRESS for miR-522.

In addition to SNPs that may interrupt MRESSs, SNP alleles may also create MRESSs. To identify potential CNM SNPs we performed a genome-wide computational survey for predicted MRESSs that are created when the mRNA sequence contains the non-reference allele of hg19. Using the Ensembl variation API tools we retrieved the flanking 22 nt sequence from both upstream and downstream of the non-reference allele of every 3'UTR dbSNP132 SNP. Each sequence containing the non-reference allele was analyzed for potential miR-target sites using the miRanda software (Stark et al. 2003). This analysis provided us with 22295 CNM SNP creating 49047 miR-mRNA predictions which where also categorized by the seed-type they created; 8mer (10333), 7mer-8m (36188), and 7-mer-A1 (526) or 6mer (2000). It should be noted that there are many more predictions for CNM than MRESS SNPs. This is due to the fact that no conservation constraint was imposed on the CNM SNP predictions. Considering that many CNM SNPs presumably arise to create new regulatory sites, filtering our hits on

conservation status would be counter-intuitive. Of note, we found that approximately 28% (6946/25018) of predicted MRESS and CNM SNPs identified here were first identified by the 1000 Genomes project.

Some SNPs identified in GWAS are in LD with predicted conserved MRESS and CNM SNPs

GWAS have been a powerful approach to identify genetic variants that contribute to disease risk. However, a functional role for many of the SNPs identified has not been elucidated. It is likely some of these SNPs are in strong linkage disequilibrium (LD) with unknown functional ones, some of which could be among the predicted conserved MRESS and CNM SNPs. For MRESS SNPs we focused only on MRESSs conserved across mammals as these are more likely to be of functional significance (Friedman et al. 2009a). We searched the resulting conserved MRESS and CNM SNP data for variants in LD with SNPs showing association, of GWAS significance, with disease traits and related phenotypes. To do so, we retrieved a dataset of 4817 reported associations, collected from GWAS studies, between 3943 unique SNPs and disease traits shown to have P-values meeting a threshold of < 1.0 x 10-5 (Hindorff et al. 2009). These data were processed through SNAP (http://www.broadinstitute.org/mpg/snap/ldsearch.php) which yielded a list of SNPs (including those from the pilot 1000 Genomes Project data) in LD with those reported in the GWAS. We limited our search to an r^2 of >0.8 for the CEU population. The results of this query were searched against both the MRESS SNP and the CNM SNP predictions. This query identified 35 instances of an MRESS SNP

(some MRESS SNPs are in LD with more than one reported GWAS SNP associating with multiple phenotypes), in LD (r^2 >0.8) with a least one reported GWAS SNP or an original GWAS SNP, associating with disease phenotypes (**Table 11**). In total there were 14 MRESS SNPs in 11 genes associating with 16 traits. We also identified 124 instances of a GWAS SNP that associates with disease traits and is in LD (r^2 >0.8) with CNM SNPs (**Table 12**). There were 73 CNM SNPs in 73 genes associating with 52 traits. In total we identified 87 SNPs (14 MRESS and 73 CNM SNPs) in very strong LD with SNPs reported as associating with disease related phenotypes. These 87 SNPs represent 2.22% of the total unique SNPs reported in the GWAS data. Using the SNAP pairwise LD tool we determined that 6 of the 3943 GWAS SNPs are in LD with each other, giving us 3940 SNPs or regions associating with disease traits. To determine the possibility of this number occurring by chance we selected randomly 3940 SNPs from across the genome and determined the number of these SNPs found in our MRESS and CNM SNP data, and this analysis was repeated 100 times. We found none of the random sets to have >= 87 SNPs found in the MRESS and CNM data, suggesting this result could not have occurred by chance. These SNPs are likely the putative functional variants which represent proxy SNPs identified by GWAS.

Co-expression data identify functional candidates

Support for a prediction of a miR regulating an mRNA target is strongly lent by co-expression of both RNAs. Therefore, to further refine our list of candidate functional miRNA-mRNA interactions, we searched for evidence of co-expression of miR and mRNA for each of 87 cases where SNPs were predicted to modulate the interaction. For

each MRESS or CNM SNP we utilized the mimiR webtool which hosts miR and mRNA co-expression data, compiled from five large scale expression studies (Ritchie et al. 2010). In addition, we also searched the biomedical literature using the PubMed database with the terms of miR name and "expression." To search for mRNA expression in the cognate tissue we queried the NCBI Geoprofiles. These queries revealed miR-mRNA co-expression evidence for predictions containing 12 MRESS SNPs and 27 CNM SNPs (**Tables 13 - 14**).

GWAS SNP	P-value	PID	Phenotype	LD	Proxy	SNP Coord	MAF	Gene	miR	Allele	SVR	S-T
rs10089	2.00E-06	21139019	Ileal carcinoids	1	rs10089	5: 127522543	0.35	SLC12A2	hsa-miR-421	C/T	-0.6818	8mer
rs6504340	6.00E-07	20195514	Primary tooth development (number of teeth)	0.895	rs1042822	17: 46620095	0.18	HOXB2	hsa-miR-590-3p	G/T	-1.29	7mer-m8
rs9272535	9.00E-08	21131588	Chronic lymphocytic leukemia	1	rs9272934	6: 32611021	0.15 *	HLA-DQA1	hsa-miR-137	T/C	-1.0504	8mer
rs1008953	1.00E-07	20953189	Psoriasis	1	rs2245717	20: 43995880	0.86	SYS1	hsa-miR-150	T/G	-0.7042	8mer
rs8396	4.00E-24	20037589	Serum metabolites	1	rs8396	4: 159630817	0.26	PPID	hsa-miR-376b	T/C	-0.8887	8mer
rs328	9.00E-23	18193044	HDL cholesterol	0.929	rs1059611	8: 19824563	0.13	LPL	hsa-miR-136	T/C	-0.6354	7mer-m8
rs10503669	4.00E-19	18193043	HDL cholesterol	0.929	rs1059611	8: 19824563	0.13	LPL	hsa-miR-136	T/C	-0.6354	7mer-m8
rs12678919	2.00E-34	19060906	HDL cholesterol	0.929	rs1059611	8: 19824563	0.13	LPL	hsa-miR-136	T/C	-0.6354	7mer-m8
rs17482753	3.00E-11	20031538	HDL cholesterol	0.929	rs1059611	8: 19824563	0.13	LPL	hsa-miR-136	T/C	-0.6354	7mer-m8
rs325	8.00E-26	20864672	HDL cholesterol	0.929	rs1059611	8: 19824563	0.13	LPL	hsa-miR-136	T/C	-0.6354	7mer-m8
rs7016880	2.00E-07	20657596	Hypertriglyceridemia	0.929	rs1059611	8: 19824563	0.13	LPL	hsa-miR-136	T/C	-0.6354	7mer-m8
rs2083637	2.00E-10	20694148	Metabolic syndrome	0.917	rs13702	8: 19824492	0.14	LPL	hsa-miR-410	T/C	-1.1589	8mer
rs326	5.00E-12	18193046	Triglycerides	0.958	rs13702	8: 19824492	0.14	LPL	hsa-miR-410	T/C	-1.1589	8mer
rs10105606	4.00E-26	20864672	Triglycerides	0.816	rs13702	8: 19824492	0.14	LPL	hsa-miR-410	T/C	-1.1589	8mer
rs10941694	9.00E-06	20686651	Chronic kidney disease and serum creatinine concentration	1	rs12522910	5: 45259363	0.14*	HCN1	hsa-miR-653	T/C	-1.3519	8mer
rs10923931	4.00E-08	18372903	Type 2 diabetes	1	rs835576	1: 120455586	0.07	NOTCH2	hsa-miR-218	T/C	-0.7176	8mer
rs1295686	1.00E-07	20860503	Asthma	0.956	rs847	5: 131996669	0.76	IL13	hsa-miR-381	T/C	-1.1586	7mer-m8
rs20541	5.00E-15	19169254	Psoriasis	0.956	rs847	5: 131996669	0.76	IL13	hsa-miR-381	T/C	-1.1586	7mer-m8
rs12143943	5.00E-06	19734545	Cognitive performance	0.866	rs4252745	1: 204519187	0.42	MDM4	hsa-miR-542-3p	C/G	-0.6423	7mer-m8
rs6590330	2.00E-25	19838193	Systemic lupus erythematosus	1	rs1128334	11: 128328959	0.06	ETS1	hsa-miR-381	C/T	-1.166	7mer-m8
rs1128334	2.00E-11	20169177	Systemic lupus erythematosus	1	rs1128334	11: 128328959	0.06	ETS1	hsa-miR-381	C/T	-1.166	7mer-m8
rs1443512	6.00E-16	20935629	Waist-hip ratio	0.812	rs4759058	12: 54339052	0.78 *	HOXC13	hsa-miR-503	C/A	-0.7545	7mer-m8
rs504963	2.00E-08	20570966	Crohn's disease	1	rs485073	19: 49207255	0.63	FUT2	hsa-miR-186	A/G	-1.1807	7mer-m8
rs281379	7.00E-12	21102463	Crohn's disease	0.899	rs485073	19: 49207255	0.63	FUT2	hsa-miR-186	A/G	-1.1807	7mer-m8

rs602662	3.00E-20	19303062	Folate pathway vitamin levels	1	rs485073	19: 49207255	0.63	FUT2	hsa-miR-186	A/G	-1.1807	7mer-m8
rs492602	5.00E-17	18776911	Plasma level of vitamin B12	0.816	rs485073	19: 49207255	0.63	FUT2	hsa-miR-186	A/G	-1.1807	7mer-m8
rs504963	2.00E-08	20570966	Crohn's disease	1	rs603985	19: 49207257	0.63	FUT2	hsa-miR-186	T/C	-1.1807	7mer-m8
rs281379	7.00E-12	21102463	Crohn's disease	0.899	rs603985	19: 49207257	0.63	FUT2	hsa-miR-186	T/C	-1.1807	7mer-m8
rs602662	3.00E-20	19303062	Folate pathway vitamin levels	1	rs603985	19: 49207257	0.63	FUT2	hsa-miR-186	T/C	-1.1807	7mer-m8
rs492602	5.00E-17	18776911	Plasma level of vitamin B12	0.816	rs603985	19: 49207257	0.63	FUT2	hsa-miR-186	T/C	-1.1807	7mer-m8
rs6504340	6.00E-07	20195514	Primary tooth development (number of teeth)	0.895	rs1042822	17: 46620095	0.18	HOXB2	hsa-miR-186	G/T	-1.341	8mer
rs6504340	6.00E-07	20195514	Primary tooth development (number of teeth)	0.895	rs1042818	17: 46620111	0.18	HOXB2	hsa-miR-139-5p	T/C	-1.3174	8mer
rs8396	4.00E-24	20037589	Serum metabolites	1	rs8396	4: 159630817	0.26	PPID	hsa-miR-376a	T/C	-0.8887	8mer
rs12143943	5.00E-06	19734545	Cognitive performance	0.866	rs4252745	1: 204519187	0.42	MDM4	hsa-miR-494	C/G	-0.6475	7mer-m8
rs910316	1.00E-07	19343178	Height	0.967	rs10246	14: 75544470	0.50	FAM164C	hsa-miR-7	T/A	-1.1367	8mer
rs910316	1.00E-07	19343178	Height	0.967	rs10246	14: 75544470	0.50	FAM164C	hsa-miR-7	T/A	-1.1327	8mer
rs1046896	2.00E-26	20858683	Glycated hemoglobin levels	1	rs1046896	17: 80685533	0.24	FN3KRP	hsa-miR-208a	C/T	-0.9691	8mer
rs1046896	2.00E-26	20858683	Glycated hemoglobin levels	1	rs1046896	17: 80685533	0.24	FN3KRP	hsa-miR-208b	C/T	-0.9752	8mer
rs6590330	2.00E-25	19838193	Systemic lupus erythematosus	1	rs1128334	11: 128328959	0.06	ETS1	hsa-miR-300	C/T	-1.166	7mer-m8
rs1128334	2.00E-11	20169177	Systemic lupus erythematosus	1	rs1128334	11: 128328959	0.06	ETS1	hsa-miR-300	C/T	-1.166	7mer-m8
rs1295686	1.00E-07	20860503	Asthma	0.956	rs847	5: 131996669	0.76	IL13	hsa-miR-300	T/C	-1.1586	7mer-m8
rs20541	5.00E-15	19169254	Psoriasis	0.956	rs847	5: 131996669	0.76	IL13	hsa-miR-300	T/C	-1.1586	7mer-m8
rs10903129	5.00E-10	19060911	Cholesterol total	0.933	rs9689	1: 25688276	0.53	TMEM50A	hsa-miR-128	A/G	-0.818	8mer

Table 11. MRESS SNPs in LD with variants associating with disease traits.

All minor allele frequencies (MAF) reported are for the CEU pilot panel of the 1000 Genomes Project, except where indicted. * indicates MAF in low coverage 1000genomes CEU panel.

Abbreviations: PMID = PubMed accession, PS = miRanda Pairing Score, ES = miRanda energy score, S-T = seed type.

GWAS SNP	P-value	PID	Phenotype	LD	Proxy	Coord	maf	Gene	miR	Allele	PS	ES	S-T
rs3820928	5.00E-06	17903307	Pulmonary function traits (other)	0.874	rs56324594	2: 227868854	0.46*	COL4A4	hsa-miR-339-5p	C/T	179	-32.84	8mer
rs1008953	1.00E-07	20953189	Psoriasis	1	rs2245717	20: 43995880	0.86	SYS1	hsa-miR-188-3p	T/G	153	-24.72	7mer-m8
rs13098911	3.00E-17	20190752	Celiac disease	0.915	rs3136667	3: 46244284	0.1*	CCR1	hsa-miR-608	A/G	155	-29.28	6mer
rs10935120	7.00E-08	18391952	Height	0.806	rs4519744	3: 134321523	0.63*	KY	hsa-miR-663	T/C	151	-35.16	7mer-m8
rs4788084	3.00E-13	19430480	Type 1 diabetes	0.811	rs40837	16: 28510845	0.44	IL27	hsa-miR-661	A/G	155	-21.15	7mer-m8
rs7216389	9.00E-11	17611496	Asthma	0.967	rs907091	17: 37921742	0.53	IKZF3	hsa-miR-326	C/T	153	-23.56	8mer
rs2305480	1.00E-07	20860503	Asthma	0.875	rs907091	17: 37921742	0.53	IKZF3	hsa-miR-326	C/T	153	-23.56	8mer
rs907092	8.00E-06	19458352	Primary biliary cirrhosis	0.905	rs907091	17: 37921742	0.53	IKZF3	hsa-miR-326	C/T	153	-23.56	8mer
rs9303277	2.00E-09	20639880	Primary biliary cirrhosis	0.967	rs907091	17: 37921742	0.53	IKZF3	hsa-miR-326	C/T	153	-23.56	8mer
rs2872507	9.00E-07	20453842	Rheumatoid arthritis	0.875	rs907091	17: 37921742	0.53	IKZF3	hsa-miR-326	C/T	153	-23.56	8mer
rs2290400	6.00E-13	19430480	Type 1 diabetes	0.967	rs907091	17: 37921742	0.53	IKZF3	hsa-miR-326	C/T	153	-23.56	8mer
rs8067378	1.00E-07	20228799	Ulcerative colitis	0.967	rs907091	17: 37921742	0.53	IKZF3	hsa-miR-326	C/T	153	-23.56	8mer
rs2338104	1.00E-10	19060906	HDL cholesterol	0.967	rs1045255	12: 109973979	0.56	UBE3B	hsa-miR-452	G/C	154	-21.18	7mer-m8
rs2336725	1.00E-12	20881960	Height	0.967	rs891368	3: 53123273	0.58	RFT1	hsa-miR-497	A/G	162	-20.44	7mer-m8
rs10838738	5.00E-09	19079261	Body mass index	0.864	rs2293577	11: 47437202	0.63	SLC39A13	hsa-miR-665	C/T	157	-23.73	8mer
rs2814707	7.00E-09	19734901	Amyotrophic lateral sclerosis	0.904	rs700782	9: 27526047	0.27*	IFNK	hsa-miR-153	C/T	166	-21.58	8mer
rs3849942	5.00E-11	20801717	Amyotrophic lateral sclerosis	0.904	rs700782	9: 27526047	0.27*	IFNK	hsa-miR-153	C/T	166	-21.58	8mer
rs4343	3.00E-25	20066004	Angiotensin-converting enzyme activity	0.902	rs1055086	17: 61575637	0.53	ACE	hsa-miR-373*	A/G	154	-20.66	7mer-m8
rs1008953	1.00E-07	20953189	Psoriasis	1	rs2245717	20: 43995880	0.86	SYS1	hsa-miR-532-3p	T/G	169	-28.01	8mer
rs3914188	3.00E-07	21102462	Menarche (age at onset)	1	rs3914188	3: 184010048	0.72	ECE2	hsa-miR-612	G/C	159	-28.94	7mer-m8
rs3995090	4.00E-09	20010834	Pulmonary function	0.931	rs6580550	5: 147856232	0.33	HTR4	hsa-miR-148a	T/C	159	-21.6	8mer
rs11168048	1.00E-11	20010835	Pulmonary function	0.93	rs6580550	5: 147856232	0.33	HTR4	hsa-miR-148a	T/C	159	-21.6	8mer
rs7631605	1.00E-06	20932310	P-tau181p	0.845	rs1133661	3: 37028066	0.46	EPM2AIP1	hsa-miR-665	A/G	156	-25.04	8mer
rs9272535	9.00E-08	21131588	Chronic lymphocytic leukemia	0.818	rs1064991	6: 32611176	na	HLA-DQA1	hsa-miR-20b*	G/C	157	-20.53	7mer-m8
rs12449157	2.00E-07	20864672	HDL cholesterol	1	rs12449157	16: 67708897	0.13	GFOD2	hsa-miR-125a-	A/G	160	-22.52	7mer-m8

									3p				
rs4794822	6.00E-10	20172861	Neutrophil count	0.898	rs7021	17: 38153875	0.4*	PSMD3	hsa-miR-214*	T/A	156	-21.63	8mer
rs2294008	4.00E-11	20972438	Bladder cancer	1	rs1045547	8: 143763757	0.46*	PSCA	hsa-miR-597	T/G	156	-20.2	8mer
rs1007738	7.00E-07	19079262	Bone mineral density (hip)	1	rs3829940	11: 46879973	0.78	LRP4	hsa-miR-361-3p	A/G	152	-22.32	7mer-m8
rs3995090	4.00E-09	20010834	Pulmonary function	0.931	rs6580550	5: 147856232	0.33	HTR4	hsa-miR-152	T/C	156	-22.43	8mer
rs11168048	1.00E-11	20010835	Pulmonary function	0.93	rs6580550	5: 147856232	0.33	HTR4	hsa-miR-152	T/C	156	-22.43	8mer
rs2338104	1.00E-10	19060906	HDL cholesterol	0.902	rs877710	12: 109993976	0.54	MMAB	hsa-miR-564	C/G	154	-26.44	8mer
rs3995090	4.00E-09	20010834	Pulmonary function	0.931	rs6580550	5: 147856232	0.33	HTR4	hsa-miR-148b	T/C	159	-20.05	8mer
rs11168048	1.00E-11	20010835	Pulmonary function	0.93	rs6580550	5: 147856232	0.33	HTR4	hsa-miR-148b	T/C	159	-20.05	8mer
rs2271293	8.00E-16	19060911	HDL cholesterol	1	rs72556537	16: 67691477	0.13	ACD	hsa-miR-147	A/T	150	-22.2	7mer-m8
rs4929923	1.00E-08	21102462	Menarche (age at onset)	1	rs10769931	11: 8636105	0.71	TRIM66	hsa-miR-92b*	T/C	159	-26.21	7mer-m8
rs29941	3.00E-09	20935630	Body mass index	0.889	rs14810	19: 34304903	0.67*	KCTD15	hsa-miR-486-3p	C/G	150	-24.48	8mer
rs2967605	1.00E-08	19060906	HDL cholesterol	0.943	rs2241588	19: 8468485	0.17*	RAB11B	hsa-miR-92b*	C/T	154	-28.35	7mer-m8
rs887304	8.00E-07	20708005	Non-alcoholic fatty liver disease histology (lobular)	1	rs887304	12: 3757548	0.69	EFCAB4B	hsa-miR-323-5p	T/C	150	-25.81	7mer-m8
rs4952590	2.00E-06	19961619	Atopy	0.852	rs1044305	2: 42284884	0.14	PKDCC	hsa-miR-1470	T/C	155	-27.56	8mer
rs1046896	2.00E-26	20858683	Glycated hemoglobin levels	1	rs1046875	17: 80685426	0.78	FN3KRP	hsa-miR-885-3p	A/G	150	-25.57	7mer-m8
rs4809330	3.00E-15	21102463	Crohn's disease	0.93	rs1056441	20: 62370349	na	LIME1	hsa-miR-1538	T/C	153	-31.09	7mer-m8
rs2315008	9.00E-15	18758464	Inflammatory bowel disease	0.93	rs1056441	20: 62370349	na	LIME1	hsa-miR-1538	T/C	153	-31.09	7mer-m8
rs2735839	2.00E-18	18264097	Prostate cancer	0.841	rs1058205	19: 51363398	0.85	KLK3	hsa-miR-3162	C/T	164	-24.83	7mer-m8
rs10503669	4.00E-19	18193043	HDL cholesterol	0.929	rs1059611	8: 19824563	0.13	LPL	hsa-miR-1468	T/C	154	-22.31	7mer-m8
rs328	9.00E-23	18193044	HDL cholesterol	0.929	rs1059611	8: 19824563	0.13	LPL	hsa-miR-1468	T/C	154	-22.31	7mer-m8
rs12678919	2.00E-34	19060906	HDL cholesterol	0.929	rs1059611	8: 19824563	0.13	LPL	hsa-miR-1468	T/C	154	-22.31	7mer-m8
rs17482753	3.00E-11	20031538	HDL cholesterol	0.929	rs1059611	8: 19824563	0.13	LPL	hsa-miR-1468	T/C	154	-22.31	7mer-m8
rs325	8.00E-26	20864672	HDL cholesterol	0.929	rs1059611	8: 19824563	0.13	LPL	hsa-miR-1468	T/C	154	-22.31	7mer-m8
rs7016880	2.00E-07	20657596	Hypertriglyceridemia	0.929	rs1059611	8: 19824563	0.13	LPL	hsa-miR-1468	T/C	154	-22.31	7mer-m8
rs1405069	6.00E-06	20237162	Plasma chemerin levels	0.809	rs1063021	6: 36932327	0.42	PI16	hsa-miR-4313	A/G	150	-22.42	7mer-m8
rs9272535	9.00E-08	21131588	Chronic	0.818	rs1064991	6: 32611176	0.13*	HLA-	hsa-miR-1183	G/C	163	-20.08	8mer

			lymphocytic leukemia					DQA1					
rs4929923	1.00E-08	21102462	Menarche (age at onset)	1	rs10769931	11: 8636105	0.71	TRIM66	hsa-miR-3126-5p	T/C	157	-25.19	8mer
rs703842	5.00E-11	19525955	Multiple sclerosis	1	rs10783848	12: 58196528	0.39	TSFM	hsa-miR-298	G/A	151	-23.77	6mer
rs1136001	7.00E-06	20189936	Height	1	rs1121	16: 15131076	0.32	PDXDC1	hsa-miR-3652	G/A	159	-26.04	8mer
rs10483853	6.00E-06	17903303	Coronary artery calcification	0.945	rs11625196	14: 73743205	0.18	NUMB	hsa-miR-1197	G/C	167	-21.55	7mer-m8
rs572169	3.00E-18	20881960	Height	0.885	rs116657802	3: 172162394	0.27*	GHSR	hsa-miR-3666	T/C	151	-20.13	7mer-m8
rs2191566	4.00E-07	19684603	Acute lymphoblastic leukemia (childhood)	1	rs11881151	19: 44501929	0.71	ZNF155	hsa-miR-3622b-5p	G/C	153	-25.43	7mer-m8
rs3846662	3.00E-19	19060911	Cholesterol total	0.901	rs12916	5: 74656539	0.39	HMGCR	hsa-miR-1207-5p	T/C	153	-32.35	8mer
rs12654264	1.00E-20	18193044	LDL cholesterol	0.965	rs12916	5: 74656539	0.39	HMGCR	hsa-miR-1207-5p	T/C	153	-32.35	8mer
rs7703051	1.00E-08	18802019	LDL cholesterol	0.965	rs12916	5: 74656539	0.39	HMGCR	hsa-miR-1207-5p	T/C	153	-32.35	8mer
rs12916	1.00E-11	20864672	LDL cholesterol	1	rs12916	5: 74656539	0.39	HMGCR	hsa-miR-1207-5p	T/C	153	-32.35	8mer
rs3846663	6.00E-06	19197348	Quantitative traits	0.932	rs12916	5: 74656539	0.39	HMGCR	hsa-miR-1207-5p	T/C	153	-32.35	8mer
rs228769	2.00E-08	19801982	Bone mineral density (hip)	1	rs12937692	17: 42256430	0.25	C17orf65	hsa-miR-3916	G/A	169	-23.48	7mer-m8
rs1295686	1.00E-07	20860503	Asthma	1	rs1295685	5: 131996445	0.75*	IL13	hsa-miR-1202	A/G	152	-20.76	7mer-m8
rs20541	5.00E-15	19169254	Psoriasis	1	rs1295685	5: 131996445	0.75*	IL13	hsa-miR-1202	A/G	152	-20.76	7mer-m8
rs29941	3.00E-09	20935630	Body mass index	0.889	rs14810	19: 34304903	0.67*	KCTD15	hsa-miR-3151	C/G	156	-26.06	7mer-m8
rs1376877	4.00E-07	17903303	Subclinical atherosclerosis traits (other)	0.875	rs1653	2: 204163145	0.47	CYP20A1	hsa-miR-4252	G/A	158	-25.15	7mer-m8
rs2048327	1.00E-09	19198611	Coronary heart disease	0.963	rs1810126	6: 160872151	0.35	SLC22A3	hsa-miR-3617	C/T	156	-24.7	6mer
rs10782001	9.00E-10	20953189	Psoriasis	0.956	rs2305880	16: 30999462	0.78	HSD3B7	hsa-miR-3621	T/C	153	-22.38	8mer
rs1376877	4.00E-07	17903303	Subclinical atherosclerosis traits (other)	0.846	rs2469954	2: 204292517	0.40	ABI2	hsa-miR-3173	T/C	151	-22.67	7mer-m8
rs470119	4.00E-08	20139978	Hematological and biochemical traits	1	rs2782	22: 50961854	0.57	NCAPH2	hsa-miR-3177	T/C	158	-27.65	7mer-m8

rs281437	3.00E-10	18604267	Soluble ICAM-1	1	rs281437	19: 10397238	0.25	ICAM1	hsa-miR-3667-5p	C/T	151	-20.13	7mer-m8
rs2048327	1.00E-09	19198611	Coronary heart disease	0.963	rs3088442	6: 160872652	0.38	SLC22A3	hsa-miR-3911	G/A	150	-23.56	8mer
rs13098911	3.00E-17	20190752	Celiac disease	0.915	rs3136668	3: 46244280	0.1*	CCR1	hsa-miR-1321	G/A	156	-21.01	7mer-m8
rs13098911	3.00E-17	20190752	Celiac disease	0.915	rs3136668	3: 46244280	0.1*	CCR1	hsa-miR-149*	G/A	151	-23.57	7mer-m8
rs8049603	1.00E-06	19525955	Multiple sclerosis	0.945	rs3169494	16: 23076172	0.14	USP31	hsa-miR-298	A/G	152	-24.79	8mer
rs13160562	7.00E-06	19581569	Alcohol dependence	1	rs3198304	5: 96111228	0.46	ERAP1	hsa-miR-4270	A/G	152	-21.79	8mer
rs10903129	5.00E-10	19060911	Cholesterol total	1	rs34491689	1: 25826502	0.49	TMEM57	hsa-miR-1273	A/G	159	-21.5	8mer
rs37062	1.00E-06	20062063	Electrocardiographic traits	0.92	rs37036	16: 58553348	0.26	SETD6	hsa-miR-1909*	T/C	151	-23.65	7mer-m8
rs7188697	7.00E-25	19305409	QT interval	0.922	rs37036	16: 58553348	0.26	SETD6	hsa-miR-1909*	T/C	151	-23.65	7mer-m8
rs3743266	8.00E-07	21102462	Menarche (age at onset)	1	rs3743266	15: 60781513	0.33	RORA	hsa-miR-509-3-5p	T/C	159	-21.81	8mer
rs3810291	2.00E-12	20935630	Body mass index	1	rs3810291	19: 47569003	0.65*	ZC3H4	hsa-miR-502-3p	G/A	152	-21.47	7mer-m8
rs6420094	1.00E-14	20383146	Chronic kidney disease	1	rs3812035	5: 176817143	0.42	SLC34A1	hsa-miR-3198	G/T	163	-26.07	8mer
rs2014355	5.00E-96	20037589	Serum metabolites	1	rs3916	12: 121177272	0.21	ACADS	hsa-miR-3677	G/C	152	-25.7	7mer-m8
rs12053903	1.00E-07	20062063	Electrocardiographic traits	0.806	rs4073797	3: 38590850	0.3*	SCN5A	hsa-miR-1306	A/T	152	-20.6	7mer-m8
rs4788084	3.00E-13	19430480	Type 1 diabetes	0.811	rs40837	16: 28510845	0.44	IL27	hsa-miR-2861	A/G	160	-34.06	7mer-m8
rs5753037	3.00E-16	19430480	Type 1 diabetes	0.967	rs41171	22: 30422107	0.64	MTMR3	hsa-miR-3660	C/G	157	-22.69	7mer-m8
rs42041	4.00E-06	18794853	Rheumatoid arthritis	1	rs42039	7: 92244422	0.22	CDK6	hsa-miR-1205	C/T	164	-24.79	8mer
rs2191566	4.00E-07	19684603	Acute lymphoblastic leukemia (childhood)	0.959	rs442220	19: 44502168	0.26	ZNF155	hsa-miR-3647-5p	G/A	182	-20.67	8mer
rs10935120	7.00E-08	18391952	Height	0.806	rs4519744	3: 134321523	0.61*	KY	hsa-miR-1908	T/C	155	-23.13	7mer-m8
rs10935120	7.00E-08	18391952	Height	0.806	rs4519744	3: 134321523	0.61*	KY	hsa-miR-744	T/C	155	-27.51	7mer-m8
rs4679904	1.00E-06	19458352	Primary biliary cirrhosis	0.883	rs4679886	3: 160155047	0.21	TRIM59	hsa-miR-3689a-3p	C/T	151	-21.18	7mer-m8
rs4679904	1.00E-06	19458352	Primary biliary cirrhosis	0.883	rs4679886	3: 160155047	0.21	TRIM59	hsa-miR-4257	C/T	156	-22.11	7mer-m8
rs3820928	5.00E-06	17903307	Pulmonary function traits (other)	0.874	rs56324594	2: 227868854	0.50	COL4A4	hsa-miR-1274a	C/T	151	-23.68	7mer-m8
rs495337	2.00E-06	20953190	Psoriasis	1	rs6125829	20: 48568929	0.43	RNF114	hsa-miR-3919	G/T	161	-23.91	7mer-m8
rs2235617	2.00E-06	20953190	Psoriasis	1	rs6125829	20: 48568929	0.43	RNF114	hsa-miR-3919	G/T	161	-23.91	7mer-m8

rs12286037	4.00E-08	20442857	Lipoprotein-associated phospholipase A2 activity and mass	1	rs61905116	11: 116649538	0.05*	ZNF259	hsa-miR-1909	A/G	153	-24.36	7mer-m8
rs12286037	4.00E-08	20442857	Lipoprotein-associated phospholipase A2 activity and mass	1	rs61905116	11: 116649538	0.05*	ZNF259	hsa-miR-1207-5p	A/G	156	-29.73	7mer-m8
rs12272004	8.00E-10	19185284	Plasma carotenoid and tocopherol levels	0.826	rs61905116	11: 116649538	0.05*	ZNF259	hsa-miR-1909	A/G	153	-24.36	7mer-m8
rs12272004	8.00E-10	19185284	Plasma carotenoid and tocopherol levels	0.826	rs61905116	11: 116649538	0.05*	ZNF259	hsa-miR-1207-5p	A/G	156	-29.73	7mer-m8
rs28927680	2.00E-17	18193044	Triglycerides	1	rs61905116	11: 116649538	0.05*	ZNF259	hsa-miR-1909	A/G	153	-24.36	7mer-m8
rs28927680	2.00E-17	18193044	Triglycerides	1	rs61905116	11: 116649538	0.05*	ZNF259	hsa-miR-1207-5p	A/G	156	-29.73	7mer-m8
rs987237	3.00E-20	20935630	Body mass index	0.901	rs62405437	6: 50813729	0.08	TFAP2B	hsa-miR-3189	T/C	151	-20.08	8mer
rs1079596	3.00E-06	17903293	Select biomarker traits	0.932	rs6278	11: 113280724	0.06	DRD2	hsa-miR-3154	C/A	155	-26.79	7mer-m8
rs1079596	3.00E-06	17903293	Select biomarker traits	0.932	rs6278	11: 113280724	0.06	DRD2	hsa-miR-3154	C/A	155	-26.79	7mer-m8
rs4794822	6.00E-10	20172861	Neutrophil count	0.933	rs709592	17: 38175553	0.47	MED24	hsa-miR-1178	C/T	159	-20.51	8mer
rs7138803	2.00E-17	20935630	Body mass index	0.894	rs7132908	12: 50263148	0.35	FAIM2	hsa-miR-3929	G/A	169	-25.76	8mer
rs11038871	2.00E-06	20694011	Immunoglobulin A	0.894	rs7476	11: 46342834	0.94	CREB3L1	hsa-miR-4270	A/C	150	-20.87	8mer
rs2304069	7.00E-06	20041166	HIV-1 control	0.909	rs8073	5: 149362400	0.19	SLC26A2	hsa-miR-3189	T/G	162	-24.06	7mer-m8
rs2336725	1.00E-12	20881960	Height	0.967	rs891368	3: 53123273	0.58	RFT1	hsa-miR-15a	A/G	158	-20.34	7mer-m8
rs7216389	9.00E-11	17611496	Asthma	0.967	rs907091	17: 37921742	0.53	IKZF3	hsa-miR-330-5p	C/T	150	-21.21	8mer
rs2305480	1.00E-07	20860503	Asthma	0.875	rs907091	17: 37921742	0.53	IKZF3	hsa-miR-330-5p	C/T	150	-21.21	8mer
rs907092	8.00E-06	19458352	Primary biliary cirrhosis	0.905	rs907091	17: 37921742	0.53	IKZF3	hsa-miR-330-5p	C/T	150	-21.21	8mer
rs9303277	2.00E-09	20639880	Primary biliary cirrhosis	0.967	rs907091	17: 37921742	0.53	IKZF3	hsa-miR-330-5p	C/T	150	-21.21	8mer
rs2872507	9.00E-07	20453842	Rheumatoid arthritis	0.875	rs907091	17: 37921742	0.53	IKZF3	hsa-miR-330-5p	C/T	150	-21.21	8mer
rs2290400	6.00E-13	19430480	Type 1 diabetes	0.967	rs907091	17: 37921742	0.53	IKZF3	hsa-miR-330-5p	C/T	150	-21.21	8mer
rs8067378	1.00E-07	20228799	Ulcerative colitis	0.967	rs907091	17: 37921742	0.53	IKZF3	hsa-miR-330-5p	C/T	150	-21.21	8mer
rs9272346	6.00E-129	18978792	Type 1 diabetes	0.933	rs9273462	6: 32627923	0.44	HLA-DQB1	hsa-miR-3929	A/G	157	-20.13	7mer-m8
rs3180018	2.00E-13	21102463	Crohn's disease	0.851	rs932972	1: 155260096	0.35	PKLR	hsa-miR-3678-	G/A	161	-23.17	7mer-m8

									3p				
rs10760706	4.00E-07	20596022	Alopecia areata	0.962	rs9556	9: 102732237	0.67	STX17	hsa-miR-92a-1*	T/C	152	-22.35	7mer-m8
rs9818870	7.00E-13	19198612	Coronary heart disease	1	rs9818870	3: 138122122	0.24	MRAS	hsa-miR-3074	C/T	154	-22.27	7mer-m8
rs9818870	7.00E-13	19198612	Coronary heart disease	1	rs9818870	3: 138122122	0.24	MRAS	hsa-miR-3074	C/T	154	-22.27	7mer-m8
rs3761847	2.00E-07	20453842	Rheumatoid arthritis	0.967	rs9886724	9: 123665019	0.51	TRAF1	hsa-miR-3170	A/G	157	-24.29	7mer-m8

Table 12. CNM SNPs in LD with variants associating with disease traits.

All minor allele frequencies (MAF) reported are for the CEU pilot panel of the 1000 Genomes Project, except where indicted. * indicates MAF in low coverage 1000genomes CEU panel.

Abbreviations: PMID = PubMed accession, PS = miRanda Pairing Score, ES = miRanda energy score, S-T = seed type, miRlit = evidence of miR and mRNA expression collected from the literature, where numbers indicate pubmed ids, except those beginning with GDS, which indicate the Geoprofile dataset ID for which expression was demonstrated. Co = The number of cell and tissue samples in the mimiRNA database for which co-expression of miR and mRNA were found. eQTL = Reports available eQTL data in the mUTHER study, where F=-Fat, L = LCL cells, and S = skin.

GWAS SNP	P-value	Phenotype	PID	LD	Proxy	Maf	FST	Gene	miR	Allele	SVR	S-T	Co	eQTL
rs10089	2.00E-06	Ileal carcinoids	21139019	1	**rs10089**	0.35	0.08	SLC12A2	hsa-miR-421	C/T	-0.6818	8mer	23	NA
rs6504340	6.00E-07	Primary tooth dev	20195514	0.90	**rs1042822**	0.18	NA	HOXB2	hsa-miR-186	G/T	-1.341	8mer	62	NA
rs328	9.00E-23	HDL cholesterol	18193044	0.93	**rs1059611**	0.13	0.03	LPL	hsa-miR-136	T/C	-0.6354	7mer-m8	36	NA
rs10503669	4.00E-19		18193043	0.93										
rs12678919	2.00E-34		19060906	0.93										
rs17482753	3.00E-11		20031538	0.93										
rs325	8.00E-26		20864672											
rs6590330	2.00E-25	Systemic lupus	19838193	1	**rs1128334**	.06	.44	ETS1	hsa-miR-381	C/T	-1.166	7mer-m8	33	F
rs1128334	2.00E-11	erythematosus	20169177											
rs10941694	9.00E-06	Chronic kidney disease and serum creatinine concentration	20686651	1	**rs12522910**	0.14	NA	HCN1	hsa-miR-653	T/C	-1.3519	8mer	9	NA
rs326	5.00E-12	Triglycerides	18193046	0.96	**rs13702**	0.14	0.42	LPL	hsa-miR-410	T/C	-1.1589	8mer	19	NA
rs2083637	2.00E-10	Metabolic Syndrome	20694148	0.92										
rs10105606	4.00E-26	Hypertriglycerdemia	20864672	0.82										
rs1008953	1.00E-07	Psoriasis	20953189	1	**rs2245717**	0.86	0.40	SYS1	hsa-miR-150	T/G	-0.7042	8mer	49	L, F
rs12143943	5.00E-06	Cognitive performance	19734545	0.86	**rs4252745**	0.42	.52	MDM4	hsa-miR-494	C/G	-0.6475	7mer-m8	20	NA
rs12143943	5.00E-06	Cognitive performance	19734545	0.86	**rs4252745**	0.42	.52	MDM4	hsa-miR-542-3p	C/G	-0.6423	7mer-m8	19	NA
rs1443512	6.00E-16	Waist-hip ratio	20935629	0.812	**rs4759058**	0.78		HOXC13	hsa-miR-503	C/A	-0.7545	7mer-m8	29	NS
rs504963	2.00E-08	Crohn's disease	20570966	1	**rs485073**	0.63	NA	FUT2	hsa-miR-186	A/G	-1.1807	7mer-m8	63	NA
rs281379	7.00E-12		21102463	0.9										
rs504963	2.00E-08	Crohn's disease	20570966	1	**rs603985**	0.63	NA	FUT2	hsa-miR-186	T/C	-1.1807	7mer-m8	63	NA
rs281379	7.00E-12		21102463	0.9										
rs10923931	4.00E-08	Type 2 diabetes	18372903	1	**rs835576**	0.07		NOTCH2	hsa-miR-218	T/C	-0.7176	8mer	31	F, L
rs8396	4.00E-24	Serum metabolites	20037589	1	**rs8396**	0.26	0.09	PPID	hsa-miR-376b	T/C	-0.8887	8mer	18	NA
rs1295686	1.00E-07	Asthma	20860503	.956	**rs847**	0.76	0.31	IL13	hsa-miR-381	T/C	-1.1586	7mer-m8	33	F
rs20541	5.00E-15	Psoriasis	19169254	.956										
rs9272535	9.00E-08	Chronic lymphocytic leukemia	21131588	1	**rs9272934**	0.15	NA	HLA-DQA1	hsa-miR-137	T/C	-1.0504	8mer	32	NA
rs10903129	5.00E-10	Cholesterol total	19060911	0.93	**rs9689**	0.53	0.36	TMEM50A	hsa-miR-128	A/G	-0.818	8mer	43	NA

Table 13. MRESS SNPs with co-expression and in LD with variants associating with disease related traits.
All minor allele frequencies (MAF) reported are for the CEU pilot panel of the 1000 Genomes Project, except where indicted. *indicates MAF in low coverage 1000genomes CEU panel. Abbreviations: PID = PubMed accession, SVR=miRSVR score, S-T = seed type. Co = The number of cell and tissue samples in the mimiRNA database for which co-expression of miR and mRNA were found. eQTL = Reports available eQTL data in the mUTHER study, where F=-Fat, L = LCL cells, and S = skin.

GWAS SNP	P-value	Phenotype	PID	LD	Proxy	Maf	FST	Gene	miR	Allele	PS	ES	S-T	Exp lit	Co	eQTL
rs2338104	1.00E-10	HDL cholesterol	19060906	0.96	**rs1045255**	0.55	NA	UBE3B	hsa-miR-452	G/C	154	-21.18	7mer-m8		21	NA
rs2294008	4.00E-11	Bladder cancer	20972438	1	**rs1045547**	0.46	.038	PSCA	hsa-miR-597	T/G	156	-20.2	8mer		7	NA
rs4343	3.00E-25	Angiotensin-converting enzyme activity	20066004	0.90	**rs1055086**	0.48	NA	ACE	hsa-miR-373*	A/G	154	-20.66	7mer-m8		14	NA
rs9272535	9.00E-08	Chronic lymphocytic leukemia	21131588	0.81	**rs1064991**	NA	NA	HLA-DQA1	hsa-miR-20b*	G/C	157	-20.53	7mer-m8		4	NA
rs4929923	1.00E-08	Menarche (age at onset)	21102462	1	**rs10769931**	0.70	NA	TRIM66	hsa-miR-92b*	T/C	159	-26.21	7mer-m8		2	NA
rs7631605	1.00E-06	P-tau181p	20932310	0.84	**rs1133661**	0.458	NA	EPM2AIP1	hsa-miR-665	A/G	156	-25.04	8mer		4	NA
rs12449157	2.00E-07	HDL cholesterol	20864672	1	**rs12449157**	0.13	0.84	GFOD2	miR-125a-3p	A/G	160	-22.52	7mer-m8	miR: 0497147 mRNA: GDS3688	9	F, L
rs29941	3.00E-09	Body mass index	20935630	0.89	**rs14810**	0.67	NA	KCTD15	miR-486-3p	C/G	150	-24.48	8mer	miR:20651284 mRNA: GDS3615	2	NA
rs2967605	1.00E-08	HDL cholesterol	19060906	0.94	**rs2241588**	0.16		RAB11B	hsa-miR-92b*	C/T	154	-28.35	7mer-m8		2	NA
rs1008953	1.00E-07	Psoriasis	20953189	1	**rs2245717**	0.86	0.40	SYS1	hsa-miR-188-3p	T/G	153	-24.72	7mer-m8		4	L
rs1008953	1.00E-07	Psoriasis	20953189	1	**rs2245717**	0.86	0.40	SYS1	hsa-miR-532-3p	T/G	169	-28.01	8mer		6	L
rs10838738	5.00E-09	Body mass index	19079261	0.86	**rs2293577**	0.625	0.11	SLC39A13	hsa-miR-665	C/T	157	-23.73	8mer		5	NA
rs13098911	3.00E-17	Celiac disease	20190752	0.91	**rs3136667**	0.10	0.58	CCR1	hsa-miR-608	A/G	155	-29.28	6mer		5	NA
rs3743266	8.00E-07	Menarche (age at onset)	21102462	1	**rs3743266**	0.33	0.36	RORA	miR-509-3-5p	T/C	159	-21.81	8mer	miR: 21436257 mRNA: 21102462	NA	NA
rs3810291	2.00E-12	Body mass index	20935630	1	**rs3810291**	0.65*	0.67	ZC3H4	miR-502-3p	G/A	152	-21.47	7mer-m8	miR: 20497147 mRNA: 20935630	NA	F
rs1007738	7.00E-07	Bone mineral density (hip)	19079262	1	**rs3829940**	0.77	NA	LRP4	hsa-miR-361-3p	A/G	152	-22.32	7mer-m8		5	NA
rs3914188	3.00E-07	Menarche (age at onset)	21102462	1	**rs3914188**	0.72	NA	ECE2	hsa-miR-612	G/C	159	-28.94	7mer-m8		9	NA
rs4788084	3.00E-13	Type 1 diabetes	19430480	0.81	**rs40837**	0.44	.026	IL27	hsa-miR-661	A/G	155	-21.15	7mer-m8		6	NA
rs10935120	7.00E-08	Height	18391952	0.80	**rs4519744**	0.60	NA	KY	hsa-miR-663	T/C	151	-35.16	7mer-m8		14	NA
rs3820928	5.00E-06	Pulmonary function traits (other)	17903307	0.87	**rs56324594**	0.46	NA	COL4A4	hsa-miR-339-5p	C/T	179	-32.84	8mer		10	NA
rs3995090 4.00E-09 **rs11168048** 1.00E-11		Pulmonary function	20010834 20010835	0.93 0.93	**rs6580550**	0.33	0.29	HTR4	hsa-miR-148a	T/C	159	-21.6	8mer		52	

rs3995090	4.00E-09	Pulmonary function	20010834	0.93	**rs6580550**	0.33	0.29	HTR4	hsa-miR-148b	T/C	159	-20.05	8mer	45	
rs11168048	1.00E-11		20010835	0.93											
rs3995090	4.00E-09	Pulmonary function	20010834	0.93	**rs6580550**	0.33	0.29	HTR4	hsa-miR-152	T/C	156	-22.43	8mer	44	
rs11168048	1.00E-11		20010835	0.93											
rs2814707	5.00E-11	Amyotrophic lateral	20801717	0.90	**rs700782**	0.21	0.04	IFNK	hsa-miR-153	C/T	166	-21.58	8mer	41	F, L
rs3849942	3.00E-25	sclerosis	20066004												
rs4794822	6.00E-10	Neutrophil count	20172861	0.89	**rs7021**	0.4	NA	PSMD3	hsa-miR-214*	T/A	156	-21.63	8mer	4	NA
rs2271293	8.00E-16	HDL cholesterol	19060911	1	**rs72556537**	0.13	NA	ACD	miR-147	A/T	150	-22.2	7mer-m8	27	NA
rs2338104	1.00E-10	HDL cholesterol	19060906	0.90	**rs877710**	0.54	0.41	MMAB	hsa-miR-564	C/G	154	-26.44	8mer	14	NA
rs887304	8.00E-07	Non-alcoholic fatty liver disease histology (lobular)	20708005	1	**rs887304**	0.69	0.35	EFCAB4B	hsa-miR-323-5p	T/C	150	-25.81	7mer-m8	2	NA
rs2336725	1.00E-12	Height	20881960	0.96	**rs891368**	0.56	0.57	RFT1	hsa-miR-497	A/G	162	-20.44	7mer-m8	39	F, L
rs7216389	9.00E-11	Asthma	17611496	0.97	**rs907091**	0.53	0.39	IKZF3	miR-326	C/T	153	-23.56	8mer	39	F , L
rs907092	8.00E-06	Primary biliary cirrhosis	19458352	0.91											
rs9303277	2.00E-09	Primary biliary cirrhosis	20639880	0.97											
rs2872507	9.00E-07	Rheumatoid arthritis	20453842	0.88											
rs2290400	6.00E-13	Type 1 diabetes	19430480	0.97											
rs8067378	1.00E-07	Ulcerative colitis	20228799	0.97											

Table 14. CNM SNPs with co-expression and in LD with variants associating with disease related traits.

CNM SNPs in LD with variants association with disease traits. All minor allele frequencies (MAF) reported are for the CEU pilot panel of the 1000 Genomes Project, except where indicted. * indicates MAF in low coverage 1000genomes CEU panel.

Abbreviations: PMID = PubMed accession, PS = miRanda Pairing Score, ES = miRanda energy score, S-T = seed type, miRlit = evidence of miR and mRNA expression collected from the literature, where numbers indicate pubmed ids, except those beginning with GDS, which indicate the Geoprofile dataset ID for which expression was demonstrated. Co = The number of cell and tissue samples in the mimiRNA database for which co-expression of miR and mRNA were found. eQTL = Reports available eQTL data in the mUTHER study, where F=-Fat, L = LCL cells, and S = skin.

eQTL data support several MRE target predictions when a MRESS or CNM SNP is present

Variation in gene transcript levels is thought to be an important modulator of disease risk in humans and SNPs that may mediate this variation are thought to be of great functional significance (Morley et al. 2004). To investigate the contribution of SNPs associating with disease traits to transcript level variation, a number of expression Quantitative Trait Loci (eQTL) studies have been performed (Dixon et al. 2007; Spielman et al. 2007; Nica et al. 2011). These results have demonstrated a number of SNPs associating significantly with expression differences across collected tissue samples. Importantly, these studies have noted differences in the amount of transcript variation across tissue samples, suggesting SNPs may modulate regulatory mechanisms, in some cases, in a tissue specific manner (Dimas et al. 2009). Interestingly, a recent study has estimated that > 80% of miRs act to lower mRNA levels demonstrating mRNA destabilization is the primary mode of action of miRs on target mRNAs (Guo et al. 2010).

To determine if the 39 miR predictions with co-expression data are supported by eQTL data we utilized the Genevar eQTL database webtool (Yang et al. 2010). Genevar allows for querying and visualization of eQTL data for loci of interest using data from various studies. We utilized the results from the recent MuTHER study which reports eQTL data from twin pairs in 3 tissue types; ; 78 twin-pair lymphoblastoid cell line (LCLs) biopsies, 80 twin-pair skin cell biopsies and 83 twin-pair fat cell biopsies (Nica et al. 2011). Searching for eQTL data on each of the 39 MRESS and CNM SNPs we found 11 of the 39 had genotype specific transcript level data in at least one of the 3 tissue samples investigated in the MuTHER study for which there was also evidence of co-

expression in mimiR for this tissue. Four of these 11 SNPs showed marginally significant trends in the differences in transcript levels across genotypes **(Figure 10).**

SYS1 transcript levels were shown to be significantly different among rs2245717 genotypes from lymphoblastoid cell lines (LCL) in one of the two twin study groups (**Figure 10**). Although, the second twin group failed to achieve significance, the direction of the effect was in agreement with the first group. Furthermore, the lower transcript levels associate with the allele predicted to create the miR-150 binding site. Neither of the two twin adipose tissue sample groups for *GFOD2* transcript levels showed a significant difference among rs12449157 genotypes. Interestingly, both p-values are of nominal significance and the direction of the effect supports the predicted miR interaction and subsequent effect of the CNM SNP (**Figure 10**). *IKZF3* transcript levels measured in LCL cells showed significant differences among the rs907091 genotypes in both twin groups. Lower *IKZF3* levels were observed in carriers of the G allele which is predicted to create a miR-326 MRESS. However, a second probe found on the Illumina whole genome expression array, used in the study, shows conflicting data where there is no difference between transcript levels in either group. The rs3810291 SNP in the *ZC3H4* 3'UTR shows no significant difference in transcript levels between alleles in adipose samples used in the MuTHER study. However, literature mining for "eQTL," and the corresponding gene and phenotype identified an additional study showing eQTL data that supports an allelic difference, in the correct direction, and in adipose tissue for rs3810291 (Speliotes et al. 2010). Taken together, this information suggests these four SNPs may have functional significance.

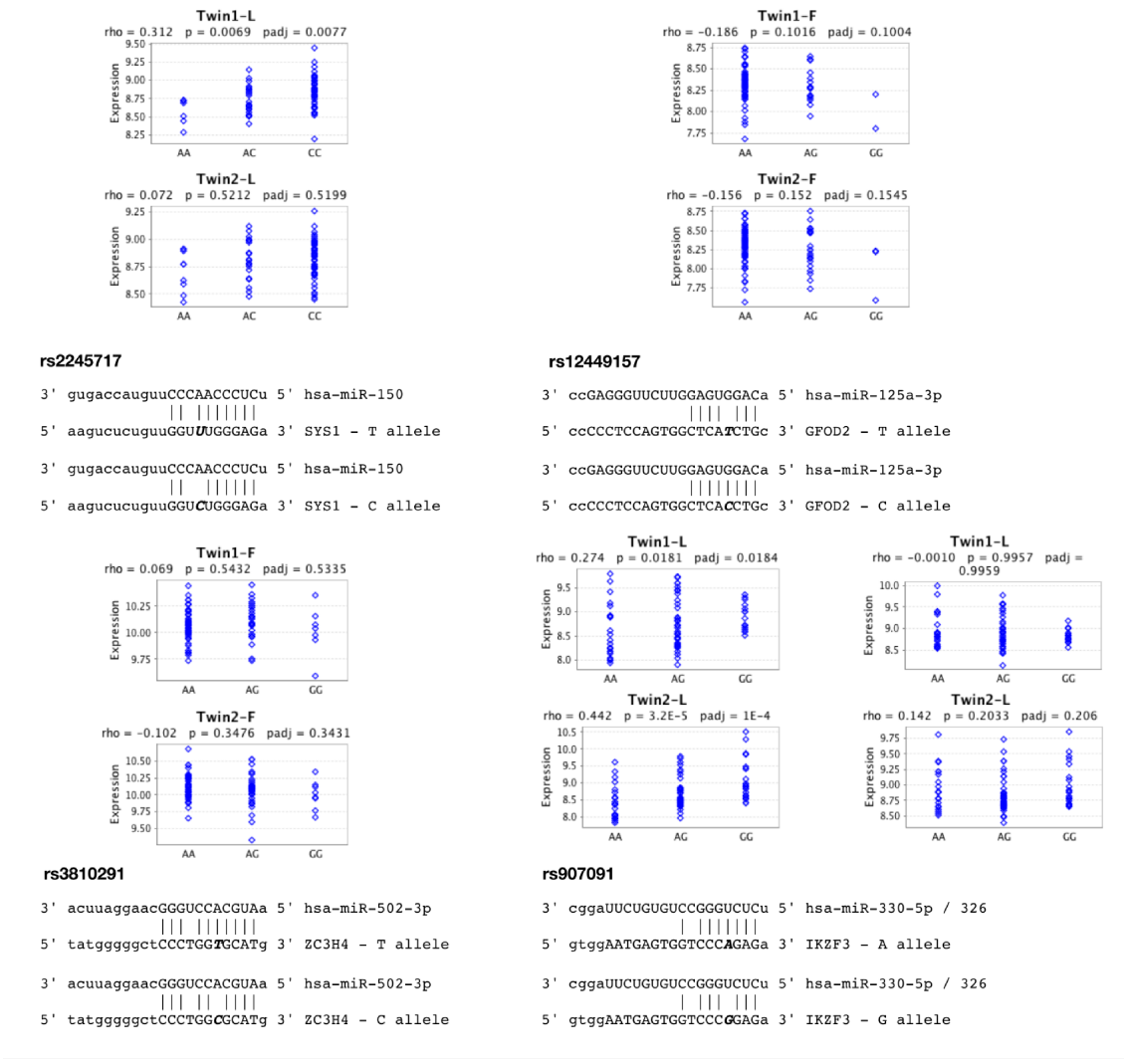


Figure 10. Four SNPs found to associate, or be in LD with a SNP that associates, with a trait(s) relevant to disease.

Each panel depicts the mRNA-miR interaction and the effect of the SNP on this interaction. Plots were generated using the Genvar webtool and published expression data from the Nica, et al.

MRESS and CNM SNPs in Positive Selection

Genetic variants that have been subject to selection are most likely the functional variants (Bustamante et al. 2005; Nielsen et al. 2005). The fixation index (F_{ST}) statistic measures population differentiation and provides a test for the influence of selective pressures, where higher F_{ST} values indicate local positive adaptation and lower values negative or neutral selection (Akey et al. 2002b). As adaptive genetic variants have been driven to higher frequencies by environmental factors (i.e, positive selection), SNPs showing high F_{ST} values may be considered high priority candidates for association studies for gene by environment studies. Such variants also may play a role in the observed variation and potentially influence disease prevalence across populations (Kullo and Ding 2007; Myles et al. 2008). To determine if the identified 3'UTR SNPs that create or disrupt predicted MREs are under selection we first downloaded genome wide F_{ST} calculations for HapMap Phase 3 data (Cheng et al. 2009a). We found a significant difference ($P = 0.0004$) between the mean transformed F_{ST} values of combined MRESS and CNM SNPs (n=2448) and the remaining (i.e. non-MRESS or non-CNM) 3'UTR SNPs (n =19906) for which F_{ST} data were available. **Figure 11** shows the number of F_{ST} values between MRESS and CNM SNPs and non-MRE SNPs across 10 F_{ST} bins. As F_{ST} values increase there is a clear increase in MRESS and CNM SNPs compared to the remainder. This observation further supports that MRESS and CNM SNPs are likely functional variants.

To identify F_{ST} outliers we selected all SNPs falling 2 standard deviations (SDs) or more from the mean (**Table 17**). In total, 24 MRESS or CNM SNPs were identified falling 2 SDs from the mean. Among these is the *GFOD2* SNP, rs12449157 (F_{ST} = 0.8399) for which we show evidence of co-expression and eQTL effect.

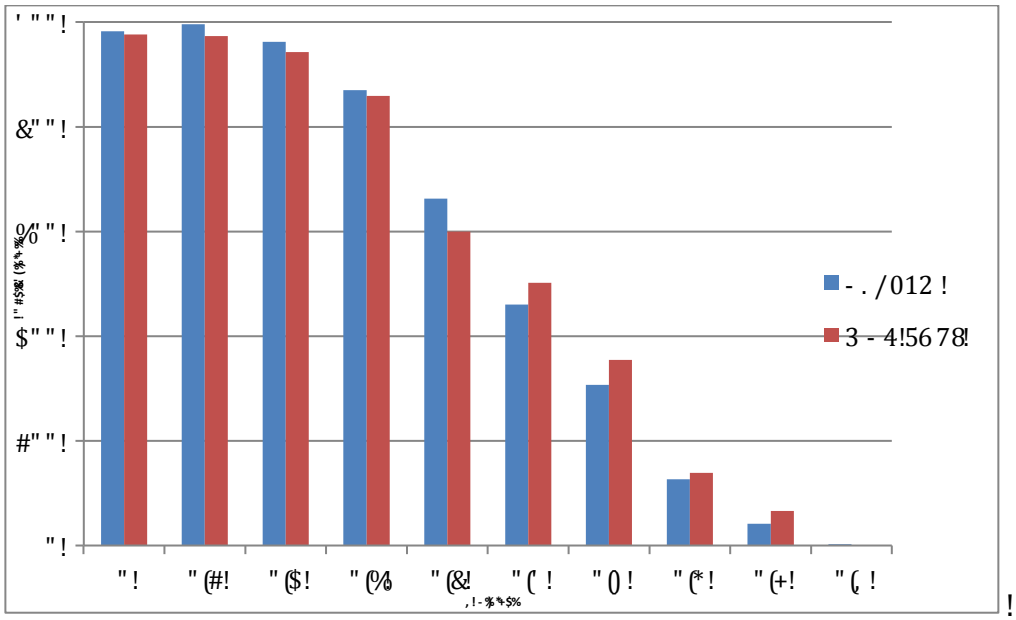


Figure 11. Fst of predicted MRESS SNPs compared to non MRESS SNPs.
A plot showing the number of combined MRESS and CNM SNPs or non MRESS and non CNM 3'UTR SNPs across 10 F_{ST} bins. Data plotted compares 2704 MRESS and CNM SNPs with F_{ST} data and a random sample of 2448 F_{ST} values from the remainder of 3'UTR SNPs. A significant difference between mean F_{ST} values for combined MRESS and CNM SNPs and F_{ST} values for the remaining 3'UTR SNPs was observed (P=0.0004).

MRESS SNPs	F_{ST}	Gene	miR	CNM SNPs	F_{ST}	Gene	miR
rs3822506	0.8743	TCERG1	miR-590	rs7665492	0.8942	ENAM	miR-3916
rs1217382	0.8469	BCL2L15	miR-17	rs1043809	0.8900	EPN2	miR-3616-3p
rs3087542	0.8428	EMCN	miR-197	rs2470102	0.8859	MYEF2	miR-1180
rs6778889	0.8426	C3orf43	miR-34a	rs7290134	0.8695	TNFRSF13C	miR-1205
rs3742988	0.8385	CDAN1	miR-378	rs8057598	0.8596	NOL3	miR-769-3p
rs11629508	0.8381	HEXA	miR-15b	rs1969589	0.8545	RGMA	miR-593*
rs1071738	0.8298	PALLD	miR-182	rs1246014	0.8476	COPS7B	miR-1273d
rs1050755	0.8222	SMNDC1	miR-329	rs12449157	0.8399	***GFOD2***	miR-125a-3p
rs7513934	0.8195	CC2D1B	miR-284	rs16990309	0.8398	SLC23A2	miR-760
				rs3742988	0.8385	CDAN1	miR-326
				rs2292549	0.8361	GPBAR1	miR-936
				rs1995939	0.8338	STARD9	miR-3943
				rs3199486	0.8321	STARD9	miR-2278
				rs873258	0.8312	TSPAN14	miR-873
				rs1801449	0.8178	CAPN3	miR-3664

Table 15. Fst outliers among MRESS and CNM SNPs.
MRESS and CNM SNPs showing highest levels of population differentiation among HapMap phase 3 data- All SNPs falling 2 SDs from the mean.

Conclusions:

In the work presented here, we utilized the 1000 Genomes Project data to perform a genome-wide scan of human variation within validated and predicted miR binding sites, our hypothesis being that genetic variants at miR binding sites are functional, and important contributors to phenotypic variation and disease susceptibility. We have taken careful measures to assign SNPs as creating or altering miR-mRNA interactions. We identified 5797 instances of a SNP falling within a conserved predicted MRESS based on stringent filtering of interaction scores predicted by Betel et al (Betel et al. 2010). Interestingly, 38% of these predicted disruptions were identified in 8mer target predictions. 8mer target sites have been shown to have the highest efficacy of target repression and therefore are considered higher priority predictions than those with lesser complimentarity (Bartel 2009). Overall, we estimate that 3% of predicted conserved MRESSs contain SNPs. Our analysis also identified 49407 instances of a SNP creating an MRESS. We also determined that 92 of the MRESS and CNM SNPs identified are in LD with SNPs identified in GWAS. We demonstrate that 2.2% of GWAS SNPs are, or are in LD with, MRESS or CNM SNPs. However, this may be a conservative estimate, given that 1) we limited our SNP selection based on conservation and other strict cutoffs, 2) the catalog of GWAS SNPs investigated is not all encompassing, 3) that these GWAS studies do not consider gene by environment interactions and 4) LD estimates only cover SNPs up to the 1000 Genomes Project pilot study 1 data.

Recently, it has been demonstrated that SNPs previously identified in GWAS are in LD with SNPs found in enhancer motifs regulating gene expression (Ernst et al. 2011). Furthermore, other studies have linked SNPs falling in gene regulatory motifs, and not

found on commercial SNP arrays, to be in LD with top scoring GWAS hits (Wasserman et al. 2010; Kulkarni et al. 2011). In a similar fashion, we suggest that many of the SNPs found in this study to be in LD with GWAS SNPs may have functional significance. To further explore this possibility we utilized several publicly available data sets and tools and showed 39 of these 87 variants found to have evidence of co-expression of target mRNA and the predicted miR. We found that four SNPs from this list have supporting eQTL data demonstrating variation in transcripts between alleles. It is possible that other candidate SNPs identified here are capable of inducing allele specific expression changes. However, these alterations may be specific to a particular cell type not analyzed in the MuTHER study, or may be observed only in response to a particular environmental or experimental condition not utilized in the MuTHER Study.

Our analyses have identified four SNPs predicted to modulate allele-specific miR-mRNA interactions which are supported by expression and eQTL data. The rs907091 SNP falls in the *IZKF3* transcript and is in LD (r2 >0.90) with eight SNPs associating with increased risk for a variety of autoimmune diseases. IZKF3 is a transcription factor important for B-cell activation, and mice lacking this gene develop a lupus like syndrome, suggesting a role for IZKF3 in autoimmunity (Sun et al. 2003). The rs907091 minor T allele is predicted to create a CNM for mir-326. There is evidence for expression of miR-326 and *IZKF3* in human B-lymphocytes. Interestingly, miR-326 is important for T-cell differentiation and has been implicated in the pathogenesis of autoimmune multiple sclerosis (Du et al. 2009). A study investigating transcript levels between the T and C alleles of rs907091 in a lymphoblastoid cell line (LCL) demonstrate significantly lower levels of *IZKF3* in subjects carrying the T allele (Nica et al. 2011).

These data suggest that carriers of the T allele may have reduced levels of IZKF3, in part through miR-326.

In addition the minor allele of rs3810291 is predicted to create an MRE for mir-502-3p within the *ZC3H4* transcript and associates with BMI. ZC3H4 is a poorly characterized zinc finger protein. There is eQTL evidence supporting this prediction where minor allele carriers have reduced *ZC3H4* expression compared to non-carriers, in adipose tissue (Speliotes et al. 2010). Both mir-502-3p and *ZC3H4* are expressed in adipose (Estep et al. 2010). The rs2245717 SNP, predicted to create an MRE for miR-155 in the *SYS1* transcript, is in perfect LD with rs1008953 a SNP associating with psoriasis (Stuart et al. 2010). The MRE-creating allele of *SYS1* is also associated with lower *SYS1* transcript levels in LCL cells (Nica et al. 2011). While a role for SYS1 in immune function could not be found in the literature, it is known that miR-155 is involved in the immune response (Tsitsiou and Lindsay 2009). The rs12449157 SNP is found in the poorly characterized glucose-fructose oxidoreductase domain containing 2 (*GFOD2*) transcript showing association with HDL-C (Waterworth et al. 2010). Our analysis predicts that the minor allele of rs12449157 creates a CNM for mir-125a-3p and that it is associated with reduced *GFOD2* levels. Interestingly, both RNAs are expressed in adipose tissue (Estep et al. 2010). Further, we identify rs12449157 as an F_{ST} outlier suggesting this SNP may be undergoing population specific selection.

In addition to these four SNPs, we identified 39 others with data indicating co-expression with the predicted target mRNA and these should be considered as candidates for functional studies. Of these 39 candidates, SNPs within the *TMEM50A and HOXB2* loci each have shown eQTL peaks identified from lymphoblastoid cell lines (Dixon et al.

2007; Zhang et al. 2008b). While our analysis has generated many MRESS and CNM SNP predictions for which no miR expression data are available, it is likely that as more miR expression and eQTL data become accessible, particularly for different cell types and specific conditions, many of these SNPs could be seen as functionally relevant. Recent data indicate some miRs may act intracellularly, carried by HDL particles to recipient cells (Vickers et al. 2011). Therefore, it may be that co-expression is not essential for all predicted miR-mRNA interactions.

As new variants arise in a population and are exposed to different environmental conditions, those variants may be subject to forces of selection. Moreover, if these SNPs alter gene expression they may modulate the individual's response to the environment and potentially the risk for particular disease state. Based on this, we hypothesized that allele-specific miR-mRNA interactions would show a greater level of selection than SNPs not classified as MRESS SNPs. We show that, as a group, predicted MRESS and CNM SNPs have a significantly higher mean F_{ST} than do those SNPs which do not create or disrupt a predicted MRESS. We identify those MRESS and CNM SNPs showing the highest degree of population subdivision and suggest these SNPs and the interactions they are predicted to modulate, as high priority functional candidates.

We show that the frequency of MRESS SNPs in validated MREs (5.5 SNPs/kb) is less than in surrounding regions and this supports prior work showing a higher degree of negative selection on MRESSs (Chen and Rajewsky 2006; Saunders et al. 2007b). Although the level of variation within this region is lower, we do show that the occurrence of variation across validated MRESSs is not rare (~5%). Supporting the notion that miR SNPs are high priority candidates for functional consequence we show

that 22% of SNPs falling within validated MRESSs have reported associations related to a disease phenotype or risk. Of note, our results differ somewhat from the MRESS SNPs reported in Saunders, et al (Saunders et al. 2007b). This is most likely due to the fact that we utilized a more current database of validated MRE targets, and also that we required functional evidence of MRESS for inclusion.

In Summary, we have surveyed the most current human SNP data and identified variants that provide functional hypotheses for observed GWAS associations. Our work also suggests that a considerable number of SNPs create or abrogate MREs in the human genome. Our results further suggest MRE SNPs that modulate gene expression are likely to be under selective pressure. With relevance to human disease we show that publicly available resources can be used to identify high priority candidate SNPs for functional studies.

Chapter 6: Summary and Conclusions.

In this thesis work we have achieved three main objectives. First, we have identified numerous tag and potential functional SNPs within the PAT gene family and genotyped them in two populations of European ancestry. Our association analyses identified the *PLIN3* SNP rs3760950 modulating fasting insulin levels and the *PLIN4* SNP rs8887 modulating a constellation of anthropometric traits in two populations of European ancestry.

Second, we demonstrated that the *PLIN3* SNP rs3760950 and the *PLIN4* SNPs rs8887 and rs884164 modulate obesity related traits in response to dietary FFAs. Our data also demonstrated several gender specific interactions where the *PLIN3* SNP rs3760948 modulates TAG in response to SFA in males, and the *PLIN5* SNPs rs1610090 and rs4806985 modulate weight in response to N3 PUFA in females.

Third, we established a functional role for the *PLIN4* SNP rs8887, which showed the most consistent associations in traits across both populations. Our work revealed a novel regulatory mechanism where the minor A allele of rs8887 generates a seed site for the previously uncharacterized miR-522. Further, our data suggest that this allele specific miR-mRNA interaction modulates anthropometrics in humans in response to dietary N3 PUFA. Although our association results where not corrected for multiple testing we are confident in the associations reported given their replication of significance in both populations, over all significance from meta-analysis and for rs8887 supporting functional data.

In addition to these goals met, the finding that rs8887 creates a MRESS lead us to question the frequency of SNPs creating or destroying MRESSs. Our bioinformatic

analyses demonstrated that the number of MRESS and CNM SNPs are appreciable in both validated and conserved predicted MREs in the human genome. Furthermore, this work uncovered four SNPs as high priority candidates for functional and association studies.

Along with the increase in obesity prevalence has come a rise in health care costs in addition to increases in chronic disease states affecting quality of life and overall mortality. This is thought to be a result of a disruption of the intricate metabolic balance developed over human evolution for optimal energy balance, brought on by environmental changes of the industrialized world. However, it has been well documented that the individual susceptibility for obesity varies tremendously. Therefore, identifying and understanding the genetic factors that contribute to variability in metabolic rate are considered key factors in assessing the individual's risk.

Important, and previously over looked, LDs of the adipocytes and other cell types are key players in the maintenance of energy metabolism. Numerous *in vivo* and *in vitro* studies have demonstrated that proteins surrounding these metabolically active organelles are required for the proper uptake and turnover of TAGs. The PAT gene family has been implicated as key regulators of LD homeostasis and their dysregulation implicated in disease susceptibility. In combination with previous findings of *PLIN1* SNPs, our identification of SNPs in *PLIN3* and *PLIN4* as modulators of obesity related traits further demonstrates the importance of this gene family in human obesity.

Given the ineffectiveness of current strategies for obesity control it is reasonable to assume that dietary recommendations for obesity control could be improved if predictive information for the individual's response to particular diets were available.

The work presented here makes a contribution to this field in that subjects carrying the rs8887 minor allele are potentially more sensitive to lowering their previously elevated anthropometrics by increasing their PUFA N3 intake through a novel biological mechanism. This work provides a link between a previously uncharacterized miR with adipocyte biology and obesity. Importantly, our work is the first to implicate a SNP in a candidate obesity gene as having a functional consequence whereby it modulates a mode of transcriptional regulation in response to exposure of a particular dietary factor, a contribution realized in the literature (Richardson et al. 2011). Furthermore, we demonstrate several other instances of PAT gene SNPs modulating obesity related traits in response to diet.

It is hypothesized that common complex diseases, such as obesity, occur as a consequence of common genetic variation - the common disease, common variant hypothesis (Reich and Lander 2001). In this scenario risk for disease is dependent on the collective contribution of genetic variants, with small to moderate effect size, which an individual may carry. We demonstrate an instance of this where rs8887 modulates PLIN4 expression by 20% and shows a significant interaction with PUFA altering obesity related traits. Importantly, our analysis on the frequency of MRESS SNPs suggests that a considerable number of SNPs create or abrogate MRESS in the human genome and that many of these SNPs are in LD with identified GWAS SNPs. We conclude that MRESS SNPs, such as the one created by rs8887, constitute an appreciable number of functional variants across the human genome. Further, our data suggest MRESS SNPs are more likely to be under selective pressure indicating that these SNPs may modulate phenotypic response to a variety of environmental conditions.

Although the combined data presented here is a compelling argument for the involvement of PAT gene variation and miR-522 with human obesity, there are several limitations that we must recognize. Associations identified through the candidate gene approach are of immediate interest because of their potential to modulate the levels and/or activity of proteins involved in relevant biological pathways. However, if a large number of SNPs in candidate genes with traits of interest are investigated an association may be observed in the form of a false positive. Furthermore, this SNP may be falsely argued as biologically relevant due to its location within the candidate locus. This potential pitfall may be overcome by a multiple comparison correction or by replication of association results in additional population based studies. In addition, functional studies to demonstrate a biological effect of the SNP in the pathway/trait of interest to support the association may further support the association.

In our analyses we investigated more than 20 SNPs and 10 traits. However, we report only those findings that reached a significance level of $P < 0.10$ in both the GOLDN and FOS populations, thus emphasizing SNPs showing replication. Our functional analysis of the *PLIN4* SNP rs8887 also lends support to the validity of the association.

In addition to the limitations inherent in the candidate gene approach, several limitations come with 3'UTR luciferase assays that must be acknowledged. A main limitation of this type of analysis, for which our work was no exception, is the fact that endogenous expression of miR-522 was low in the cell line used for transfection and as a result our experiments were based on miRNA over-expression, rather than inhibition of endogenous miRNAs. An ideal experiment would have been to analyze the effect of

miR-522 inhibition on PLIN4 levels in a cell line endogenously expressing miR-522. Also, analyzing cell lines where *PLIN4* is not present and where the inhibition of miR-522 might induce *PLIN4* expression would be of interest. However, the fact that miR-522 is both primate specific and tissue specific severely limits the number of experimental cell systems available for the exploration of the miR-522:*PLIN4* interaction. An additional limitation to testing miR-522's effect on *PLIN4* comes from the fact there is no readily available antibody for human PLIN4. As a result, it is not possible for us to determine PLIN4 protein levels in response to miR-522 over expression or inhibition.

Importantly, similar experiments in the literature have shown luciferase data comparable to what we report here (Bartel 2009). Individual miRs have been shown to act as subtle regulators of gene expression whereby several miRs may target a single mRNA to fine-tune protein output. Therefore, the observation of a 20% reduction in PLIN4 3'UTR luciferase activity in the presence of the miR-522 is not unusual for an individual miR acting on a single target site (Bartel 2009).

To overcome these limitations in future investigations we propose first to perform similar tests of association and interaction with the rs8887 SNP, and other PAT variants, using cohorts of similar background with larger sample sizes and more comprehensive dietary information, thus allowing for greater power to detect true association. In addition, the screening of easily transfectable primate cell lines for endogenous expression of miR-522 would facilitate future experiments for the validation of miR-522's effect on *PLIN4* transcript levels.

In summary, we have identified several SNPs in the candidate PAT gene family modulating obesity related traits in humans. We demonstrate these SNPs further

modulate obesity traits in response to dietary FFAs. Functional investigations reveal the rs8887 SNP creates a regulatory motif for the PLIN4 transcript and suggest a functional mechanism. We investigated the frequency of this mechanism across the human genome and conclude the occurrence appreciable. This information may be further utilized to select high priority SNP candidates for epidemiological and functional studies and may provide predictive, preventative or diagnostic information to health care professionals.

References:

2010. On beyond GWAS. *Nat Genet* **42**(7): 551.

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. 2002a. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* **12**(12): 1805-1814.

-. 2002b. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* **12**(12): 1805-1814.

Arner P. 2000. Obesity—a genetic disease of adipose tissue. *Br J Nutr*.

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**(2): 263-265.

Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136**(2): 215-233.

BECKMAN M. 2006 Science. 3;311(5765):1232-4.

Bell CG, Walley AJ, Froguel P. 2005. The genetics of human obesity. *Nat Rev Genet* **6**(3): 221-234.

Bell M, Wang H, Chen H, McLenithan JC, Gong DW, Yang RZ, Yu D, Fried SK, Quon MJ, Londos C et al. 2008. Consequences of lipid droplet coat protein downregulation in liver cells: abnormal lipid droplet metabolism and induction of insulin resistance. *Diabetes* **57**(8): 2037-2045.

Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E et al. 2005. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* **37**(7): 766-770.

Betel D, Koppal A, Agius P, Sander C, Leslie C. 2010. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol* **11**(8): R90.

Betel D, Wilson M, Gabow A, Marks DS, Sander C. 2008a. The microRNA.org resource: targets and expression. *Nucleic Acids Res* **36**(Database issue): D149-153.

-. 2008b. The microRNA.org resource: targets and expression. *Nucleic Acids Res* **36**(Database issue): D149-153.

Botma GJ, Verhoeven AJ, Jansen H. 2001. Hepatic lipase promoter activity is reduced by the C-480T and G-216A substitutions present in the common LIPC gene variant, and is increased by Upstream Stimulatory Factor. *Atherosclerosis* **154**(3): 625-632.

Brasaemle DL. 2007. Thematic review series: Adipocyte Biology. The perilipin family of structural lipid droplet proteins: stabilization of lipid droplets and control of lipolysis. *J Lipid Res* **48**(12): 2547-2559.

Brasaemle DL, Barber T, Wolins NE, Serrero G, al. e. 1997. Adipose differentiation-related protein is an ubiquitously expressed lipid storage droplet- …. *The Journal of Lipid Research*.

Brasaemle DL, Rubin B, Harten IA, Gruia-Gray J, al. e. 2000. Perilipin A Increases Triacylglycerol Storage by Decreasing the Rate of Triacylglycerol Hydrolysis. *Journal of Biological Chemistry*.

Brennecke J, Stark A, Russell RB, Cohen SM. 2005. Principles of microRNA-target recognition. *PLoS Biol* **3**(3): e85.

Buers I, Robenek H, Lorkowski S, Nitschke Y, Severs NJ, Hofnagel O. 2009. TIP47, a lipid cargo protein involved in macrophage triglyceride metabolism. *Arterioscler Thromb Vasc Biol* **29**(5): 767-773.

Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD et al. 2005. Natural selection on protein-coding genes in the human genome. *Nature* **437**(7062): 1153-1157.

Camandola S, Leonarduzzi G, Musso T, Varesio L, Carini R, Scavazza A, Chiarpotto E, Baeuerle PA, Poli G. 1996. Nuclear factor kB is activated by arachidonic acid but not by eicosapentaenoic acid. *Biochem Biophys Res Commun* **229**(2): 643-647.

Chang BHJ, Li L, Paul A, Taniguchi S, Nannegari V, al. e. 2006. Protection against Fatty Liver but Normal Adipogenesis in Mice Lacking Adipose Differentiation- …. *Molecular and Cellular Biology*.

Chen K, Rajewsky N. 2006. Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet* **38**(12): 1452-1456.

Cheng F, Chen W, Richards E, Deng L, Zeng C. 2009a. SNP@Evolution: a hierarchical database of positive selection on the human genome. *BMC evolutionary biology* **9**: 221.

-. 2009b. SNP@Evolution: a hierarchical database of positive selection on the human genome. *BMC Evol Biol* **9**: 221.

Chung SS, Choi HH, Cho YM, Lee HK, Park KS. 2006. Sp1 mediates repression of the resistin gene by PPARgamma agonists in 3T3-L1 adipocytes. *Biochem Biophys Res Commun* **348**(1): 253-258.

Corella D. 2006. Perilipin Gene Variation Determines Higher Susceptibility to Insulin Resistance in Asian Women When Consuming a High-Saturated Fat, Low-Carbohydrate Diet. *Diabetes Care* **29**(6): 1313-1319.

Corella D, Lai CQ, Demissie S, Cupples LA, Manning AK, Tucker KL, Ordovas JM. 2007. APOA5 gene variation modulates the effects of dietary fat intake on body mass index and obesity risk in the Framingham Heart Study. *J Mol Med* **85**(2): 119-128.

Corella D, Ordovas J. 2005. SINGLE NUCLEOTIDE POLYMORPHISMS THAT INFLUENCE LIPID METABOLISM: Interaction with Dietary Factors. *Annu Rev Nutr* **25**(1): 341-390.

Corella D, Qi L, Tai ES, Deurenberg-Yap M, Tan CE, Chew SK, Ordovas JM. 2006. Perilipin gene variation determines higher susceptibility to insulin resistance in Asian women when consuming a high-saturated fat, low-carbohydrate diet. *Diabetes Care* **29**(6): 1313-1319.

Cunnane SC, Crawford MA. 2003. Survival of the fattest: fat babies were the key to evolution of the large human brain. *Comp Biochem Physiol A Mol Integr Physiol* **136**(1): 17-26.

Dalen KT, Dahl T, Holter E, Arntsen B, Londos C, Sztalryd C, Nebb HI. 2007. LSDP5 is a PAT protein specifically expressed in fatty acid oxidizing tissues. *Biochim Biophys Acta* **1771**(2): 210-227.

Dalen KT, Schoonjans K, Ulven SM, Weedon-Fekjaer MS, Bentzen TG, Koutnikova H, Auwerx J, Nebb HI. 2004. Adipose tissue expression of the lipid droplet-

associating proteins S3-12 and perilipin is controlled by peroxisome proliferator-activated receptor-gamma. *Diabetes* **53**(5): 1243-1252.

Dalen KT, Ulven SM, Arntsen BM, Solaas K, Nebb HI. 2006. PPARalpha activators and fasting induce the expression of adipose differentiation-related protein in liver. *J Lipid Res* **47**(5): 931-943.

Delgado-Lista J, Perez-Martinez P, Perez-Jimenez F, Garcia-Rios A, Fuentes F, Marin C, Gomez-Luna P, Camargo A, Parnell LD, Ordovas JM et al. 2010. ABCA1 gene variants regulate postprandial lipid metabolism in healthy men. *Arteriosclerosis, thrombosis, and vascular biology* **30**(5): 1051-1057.

Dieudonne MN, Leneveu MC, Giudicelli Y, Pecquery R. 2004. Evidence for functional estrogen receptors alpha and beta in human adipose cells: regional specificities and regulation by estrogens. *Am J Physiol Cell Physiol* **286**(3): C655-661.

Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M et al. 2009. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325**(5945): 1246-1250.

Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, Taylor J, Burnett E, Gut I, Farrall M et al. 2007. A genome-wide association study of global gene expression. *Nat Genet* **39**(10): 1202-1207.

Du C, Liu C, Kang J, Zhao G, Ye Z, Huang S, Li Z, Wu Z, Pei G. 2009. MicroRNA miR-326 regulates TH-17 differentiation and is associated with the pathogenesis of multiple sclerosis. *Nat Immunol* **10**(12): 1252-1259.

Ducharme NA, Bickel PE. 2008. Minireview: Lipid Droplets in Lipogenesis and Lipolysis. *Endocrinology*.

Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, Hurles ME, McVean GA. 2010. A map of human genome variation from population-scale sequencing. *Nature* **467**(7319): 1061-1073.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*.

Estep M, Armistead D, Hossain N, Elarainy H, Goodman Z, Baranova A, Chandhoke V, Younossi ZM. 2010. Differential expression of miRNAs in the visceral adipose tissue of patients with non-alcoholic fatty liver disease. *Alimentary pharmacology & therapeutics* **32**(3): 487-497.

Fan W, Yanase T, Nomura M, Okabe T, Goto K, Sato T, Kawano H, Kato S, Nawata H. 2005. Androgen receptor null male mice develop late-onset obesity caused by decreased energy expenditure and lipolytic activity but show normal insulin sensitivity with high adiponectin secretion. *Diabetes* **54**(4): 1000-1008.

Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. 1975. The Framingham Offspring Study. Design and preliminary data. *Prev Med* **4**(4): 518-525.

Fontaine KR, Redden DT, Wang C, Westfall AO, Allison DB. 2003. Years of life lost due to obesity. *JAMA* **289**(2): 187-193.

Fox C, Massaro J, Hoffmann U, Pou K, Maurovich-Horvat P, Liu C, Vasan R, Murabito J, Meigs J, Cupples L et al. 2007. Abdominal Visceral and Subcutaneous Adipose Tissue Compartments: Association With Metabolic Risk Factors in the Framingham Heart Study. *Circulation* **116**(1): 39-48.

Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW et al. 2007. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**(5826): 889-894.

French SA, Story M, Jeffery RW. 2001. Environmental influences on eating and physical activity. *Annu Rev Public Health* **22**: 309-335.

Friedman RC, Farh KK, Burge CB, Bartel DP. 2009a. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**(1): 92-105.

-. 2009b. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**(1): 92-105.

Gao J, Serrero G. 1999. Adipose Differentiation Related Protein (ADRP) Expressed in Transfected COS-7 Cells Selectively …. *Journal of Biological Chemistry*.

Gavrilova O, Marcus-Samuels B, Graham D, Kim JK, Shulman GI, Castle AL, Vinson C, Eckhaus M, Reitman ML. 2000. Surgical implantation of adipose tissue reverses diabetes in lipoatrophic mice. *The Journal of clinical investigation* **105**(3): 271-278.

Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**(1): 91-105.

Guo H, Ingolia NT, Weissman JS, Bartel DP. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**(7308): 835-840.

Harttig U, Travis AJ, Rocca-Serra P, Renkema M, van Ommen B, Boeing H. 2009. Owner controlled data exchange in nutrigenomic collaborations: the NuGO information network. *Genes Nutr* **4**(2): 113-122.

Haslam D, James W. 2005. Obesity. *The Lancet* **366**(9492): 1197-1209.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America* **106**(23): 9362-9367.

Imamura M, Inoguchi T, Ikuyama S, Taniguchi S, Kobayashi K, Nakashima N, Nawata H. 2002. ADRP stimulates lipid accumulation and lipid droplet formation in murine fibroblasts. *Am J Physiol Endocrinol Metab* **283**(4): E775-783.

Juge-Aubry CE, Henrichot E, Meier CA. 2005. Adipose tissue: a regulator of inflammation. *Best Pract Res Clin Endocrinol Metab* **19**(4): 547-566.

Kang E. 2006. The 11482G>A Polymorphism in the Perilipin Gene Is Associated With Weight Gain With Rosiglitazone Treatment in Type 2 Diabetes. *Diabetes Care* **29**(6): 1320-1324.

Kaput J, Ordovas JM, Ferguson L, van Ommen B, Rodriguez RL, Allen L, Ames BN, Dawson K, German B, Krauss R et al. 2005. The case for strategic international alliances to harness nutritional genomics for public and personal health. *The British journal of nutrition* **94**(5): 623-632.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* **188**(4184): 107-116.

Koutsari C. 2006. Thematic review series: Patient-Oriented Research. Free fatty acid metabolism in human obesity. *The Journal of Lipid Research* **47**(8): 1643-1650.

Kozomara A, Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**(Database issue): D152-157.

Kulkarni S, Savan R, Qi Y, Gao X, Yuki Y, Bass SE, Martin MP, Hunt P, Deeks SG, Telenti A et al. 2011. Differential microRNA regulation of HLA-C expression and its association with HIV control. *Nature*.

Kullo IJ, Ding K. 2007. Patterns of population differentiation of candidate genes for cardiovascular disease. *BMC Genet* **8**: 48.

Laclaustra M, Corella D, Ordovas J. 2007. Metabolic syndrome pathophysiology: The role of adipose tissue. *Nutrition, Metabolism and Cardiovascular Diseases* **17**(2): 125-139.

Lai CQ, Demissie S, Cupples LA, Zhu Y, Adiconis X, Parnell LD, Corella D, Ordovas J. 2004. Influence of the APOA5 locus on plasma triglyceride, lipoprotein subclasses, and CVD risk in the Framingham Heart Study. *J Lipid Res* **45**(11): 2096-2105.

Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M et al. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**(7): 1401-1414.

Larigauderie G. 2004. Adipophilin Enhances Lipid Accumulation and Prevents Lipid Efflux From THP-1 Macrophages: Potential Role in Atherogenesis. *Arterioscler Thromb Vasc Biol* **24**(3): 504-510.

Laurencikiene J, van Harmelen V, Arvidsson Nordstrom E, Dicker A, Blomqvist L, Naslund E, Langin D, Arner P, Ryden M. 2007a. NF-kappa B is important for TNF-alpha -induced lipolysis in human adipocytes. *J Lipid Res*.

Laurencikiene J, van Harmelen V, Arvidsson Nordström E, Dicker A, Blomqvist L, Näslund E, Langin D, Arner P, Rydén M. 2007b. NF-kappaB is important for TNF-alpha-induced lipolysis in human adipocytes. *J Lipid Res* **48**(5): 1069-1077.

Lee B, Zhu J, Wolins NE, Cheng JX, Buhman KK. 2009. Differential association of adipophilin and TIP47 proteins with cytoplasmic lipid droplets in mouse enterocytes during dietary fat absorption. *Biochim Biophys Acta* **1791**(12): 1173-1180.

Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**(1): 15-20.

Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. 2003. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature genetics* **33**(2): 177-182.

Londos C, Sztalryd C, Tansey JT, Kimmel AR. 2005. Role of PAT proteins in lipid metabolism. *Biochimie* **87**(1): 45-49.

Lyon HN, Hirschhorn JN. 2005. Genetics of common forms of obesity: a brief overview. *The American journal of clinical nutrition* **82**(1 Suppl): 215S-217S.

Marinescu VD, Kohane IS, Riva A. 2005a. The MAPPER database: a multi-genome catalog of putative transcription factor binding sites. *Nucleic Acids Res* **33**(Database issue): D91-97.

-. 2005b. MAPPER: a search engine for the computational identification of putative transcription factor binding sites in multiple genomes. *BMC Bioinformatics* **6**: 79.

Martinelli R, Nardelli C, Pilone V, Buonomo T, Liguori R, Castano I, Buono P, Masone S, Persico G, Forestieri P et al. 2010. miR-519d overexpression is associated with human obesity. *Obesity* **18**(11): 2170-2176.

Martinez-Botas J, Anderson JB, Tessier D, al. e. 2000. Absence of perilipin results in leanness and reverses obesity in Lepr db/db mice. *Nature Genetics*.

Memisoglu A, Hu FB, Hankinson SE, Manson JE, De Vivo I, Willett WC, Hunter DJ. 2003. Interaction between a peroxisome proliferator-activated receptor gamma gene polymorphism and dietary fat intake in relation to body mass. *Hum Mol Genet* **12**(22): 2923-2929.

Miura S, Gan JW, Brzostowski J, Parisi MJ, Schultz CJ, Londos C, Oliver B, Kimmel AR. 2002. Functional conservation for lipid storage droplet association among Perilipin, ADRP, and TIP47 (PAT)-related proteins in mammals, Drosophila, and Dictyostelium. *J Biol Chem* **277**(35): 32253-32257.

Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. 2004. Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**(7001): 743-747.

Motomura W, Inoue M, Ohtake T, Takahashi N, Nagamine M, Tanno S, Kohgo Y, Okumura T. 2006. Up-regulation of ADRP in fatty liver in human and liver steatosis in mice fed with high fat diet. *Biochem Biophys Res Commun* **340**(4): 1111-1118.

Mottagui-Tabar S, Ryden M, Lofgren P, Faulds G, Hoffstedt J, Brookes AJ, Andersson I, Arner P. 2003. Evidence for an important role of perilipin in the regulation of human adipocyte lipolysis. *Diabetologia* **46**(6): 789-797.

Myles S, Davison D, Barrett J, Stoneking M, Timpson N. 2008. Worldwide population differentiation at disease-associated SNPs. *BMC Med Genomics* **1**: 22.

Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, Travers M, Potter S, Grundberg E, Small K et al. 2011. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet* **7**(2): e1002003.

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol* **3**(6): e170.

Nuotio K, Isoviita PM, Saksi J, Ijäs P, Pitkäniemi J, Sonninen R, Soinne L, Saimanen E, Salonen O, Kovanen PT et al. 2007. Adipophilin expression is increased in symptomatic carotid atherosclerosis: correlation with red blood cells and cholesterol crystals. *Stroke* **38**(6): 1791-1798.

Ogden C. 2006. Prevalence of Overweight and Obesity in the United States, 1999-2004. *JAMA: The Journal of the American Medical Association* **295**(13): 1549-1555.

Ohsaki Y, Maeda T, Maeda M, Tauchi-Sato K, Fujimoto T. 2006. Recruitment of TIP47 to lipid droplets is controlled by the putative hydrophobic cleft. *Biochem Biophys Res Commun* **347**(1): 279-287.

Ordovas J, Corella D. 2004. NUTRITIONAL GENOMICS. *Annu Rev Genom Human Genet* **5**(1): 71-118.

Paul A, Chang BH, Li L, Yechoor VK, Chan L. 2008. Deficiency of adipose differentiation-related protein impairs foam cell formation and protects against atherosclerosis. *Circ Res* **102**(12): 1492-1501.

Perez-Martinez P, Lopez-Miranda J, Cruz-Teno C, Delgado-Lista J, Jimenez-Gomez Y, Fernandez JM, Gomez MJ, Marin C, Perez-Jimenez F, Ordovas JM. 2008. Adiponectin gene variants are associated with insulin sensitivity in response to dietary fat consumption in Caucasian men. *J Nutr* **138**(9): 1609-1614.

Persson J, Degerman E, Nilsson J, Lindholm M. 2007. Perilipin and adipophilin expression in lipid loaded macrophages. *Biochemical and Biophysical Research Communications* **363**(4): 1020-1026.

Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. 2010. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**(7301): 1033-1038.

Prusty D, Park BH, Davis KE, Farmer SR. 2002. Activation of MEK/ERK signaling promotes adipogenesis by enhancing peroxisome proliferator-activated receptor gamma (PPARgamma) and C/EBPalpha gene expression during the differentiation of 3T3-L1 preadipocytes. *J Biol Chem* **277**(48): 46226-46232.

Qi L, Cho YA. 2008. Gene-environment interaction and obesity. *Nutr Rev* **66**(12): 684-694.

Qi L, Corella D, Sorlí JV, Portolés O, Shen H, Coltell O, Godoy D, Greenberg AS, Ordovas JM. 2004a. Genetic variation at the perilipin (PLIN) locus is associated with obesity-related phenotypes in White women. *Clin Genet* **66**(4): 299-310.

Qi L, Shen H, Larson I, Schaefer EJ, Greenberg AS, Tregouet DA, Corella D, Ordovas JM. 2004b. Gender-specific association of a perilipin gene haplotype with obesity risk in a white population. *Obes Res* **12**(11): 1758-1765.

Qi L, Tai E, Tan C, Shen H, Chew S, Greenberg A, Corella D, Ordovas J. 2005. Intragenic linkage disequilibrium structure of the human perilipin gene (PLIN) and haplotype association with increased obesity risk in a multiethnic Asian population. *J Mol Med* **83**(6): 448-456.

Rajewsky N. 2006. microRNA target predictions in animals. *Nat Genet* **38 Suppl**: S8-13.

Rayner KJ, Suarez Y, Davalos A, Parathath S, Fitzgerald ML, Tamehiro N, Fisher EA, Moore KJ, Fernandez-Hernando C. 2010. MiR-33 contributes to the regulation of cholesterol homeostasis. *Science* **328**(5985): 1570-1573.

Reich DE, Lander ES. 2001. On the allelic spectrum of human disease. *Trends Genet* **17**(9): 502-510.

Richardson K, Louie-Gao Q, Arnett DK, Parnell LD, Lai CQ, Davalos A, Fox CS, Demissie S, Cupples LA, Fernandez-Hernando C et al. 2011. The PLIN4 Variant rs8887 Modulates Obesity Related Phenotypes in Humans through Creation of a Novel miR-522 Seed Site. *PloS one* **6**(4): e17944.

Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. 1992. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *Am J Epidemiol* **135**(10): 1114-1126; discussion 1127-1136.

Ritchie W, Flamant S, Rasko JE. 2010. mimiRNA: a microRNA expression profiler and classification resource designed to identify functional correlations between microRNAs and their targets. *Bioinformatics* **26**(2): 223-227.

Robenek H, Robenek MJ, Buers I, Lorkowski S, Hofnagel O, Troyer D, Severs NJ. 2005. Lipid droplets gain PAT family proteins by interaction with specialized plasma membrane domains. *J Biol Chem* **280**(28): 26330-26338.

Ruan H, Hacohen N, Golub TR, Van Parijs L, Lodish HF. 2002. Tumor necrosis factor-alpha suppresses adipocyte-specific genes and activates expression of preadipocyte genes in 3T3-L1 adipocytes: nuclear factor-kappaB activation by TNF-alpha is obligatory. *Diabetes* **51**(5): 1319-1336.

Sampath H, Ntambi J. 2005. POLYUNSATURATED FATTY ACID REGULATION OF GENES OF LIPID METABOLISM. *Annu Rev Nutr* **25**(1): 317-340.

Saunders MA, Liang H, Li WH. 2007a. Human polymorphism at microRNAs and microRNA target sites. *Proc Natl Acad Sci U S A* **104**(9): 3300-3305.

-. 2007b. Human polymorphism at microRNAs and microRNA target sites. *Proceedings of the National Academy of Sciences of the United States of America* **104**(9): 3300-3305.

Schoonjans K, Staels B, Auwerx J. 1996. Role of the peroxisome proliferator-activated receptor (PPAR) in mediating the effects of fibrates and fatty acids on gene expression. *Journal of lipid research* **37**(5): 907-925.

Sethupathy P, Collins FS. 2008a. MicroRNA target site polymorphisms and human disease. *Trends in genetics : TIG* **24**(10): 489-497.

-. 2008b. MicroRNA target site polymorphisms and human disease. *Trends Genet* **24**(10): 489-497.

Shen J, Arnett DK, Peacock JM, Parnell LD, Kraja A, Hixson JE, Tsai MY, Lai CQ, Kabagambe EK, Straka RJ et al. 2007. Interleukin1beta genetic polymorphisms interact with polyunsaturated fatty acids to modulate risk of the metabolic syndrome. *J Nutr* **137**(8): 1846-1851.

Shimizu M, Takeshita A, Tsukamoto T, Gonzalez FJ, Osumi T. 2004. Tissue-selective, bidirectional regulation of PEX11 alpha and perilipin genes through a common peroxisome proliferator response element. *Mol Cell Biol* **24**(3): 1313-1323.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**(8): 1034-1050.

Speliotes EK Willer CJ Berndt SI Monda KL Thorleifsson G Jackson AU Allen HL Lindgren CM Luan J Magi R et al. 2010. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics* **42**(11): 937-948.

Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG. 2007. Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet* **39**(2): 226-231.

Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM. 2005. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell* **123**(6): 1133-1146.

Stark A, Brennecke J, Russell RB, Cohen SM. 2003. Identification of Drosophila MicroRNA targets. *PLoS Biol* **1**(3): E60.

Stuart PE, Nair RP, Ellinghaus E, Ding J, Tejasvi T, Gudjonsson JE, Li Y, Weidinger S, Eberlein B, Gieger C et al. 2010. Genome-wide association analysis identifies three psoriasis susceptibility loci. *Nat Genet* **42**(11): 1000-1004.

Stunkard AJ, Harris JR, Pedersen NL, McClearn GE. 1990. The body-mass index of twins who have been reared apart. *N Engl J Med* **322**(21): 1483-1487.

Stunkard AJ, Sorensen TI, Hanis C, Teasdale TW, Chakraborty R, Schull WJ, Schulsinger F. 1986. An adoption study of human obesity. *N Engl J Med* **314**(4): 193-198.

Sun J, Matthias G, Mihatsch MJ, Georgopoulos K, Matthias P. 2003. Lack of the transcriptional coactivator OBF-1 prevents the development of systemic lupus erythematosus-like phenotypes in Aiolos mutant mice. *J Immunol* **170**(4): 1699-1706.

Sztalryd C, Bell M, Lu X, Mertz P, Hickenbottom S, Chang BH, Chan L, Kimmel AR, Londos C. 2006. Functional compensation for adipose differentiation-related protein (ADFP) by Tip47 in an ADFP null embryonic cell line. *J Biol Chem* **281**(45): 34341-34348.

Talmud PJ, Palmen J, Putt W, Lins L, Humphries SE. 2005. Determination of the functionality of common APOA5 polymorphisms. *J Biol Chem* **280**(31): 28215-28220.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**(8): 1596-1599.

Tansey J. 2003. Functional Studies on Native and Mutated Forms of Perilipins. A ROLE IN PROTEIN KINASE A-MEDIATED LIPOLYSIS OF TRIACYLGLYCEROLS IN CHINESE HAMSTER OVARY CELLS. *Journal of Biological Chemistry* **278**(10): 8401-8406.

Tansey JT, Sztalryd C, Gruia-Gray J, Roush DL. 2001a. Perilipin ablation results in a lean mouse with aberrant adipocyte lipolysis, enhanced leptin …. *Proceedings of the National Academy of Sciences*.

Tansey JT, Sztalryd C, Gruia-Gray J, Roush DL, Zee JV, Gavrilova O, Reitman ML, Deng CX, Li C, Kimmel AR et al. 2001b. Perilipin ablation results in a lean mouse with aberrant adipocyte lipolysis, enhanced leptin production, and resistance to diet-induced obesity. *Proceedings of the National Academy of Sciences of the United States of America* **98**(11): 6494-6499.

Targett-Adams P, McElwee MJ, Ehrenborg E, al. e. 2005a. A PPAR response element regulates transcription of the gene for human adipose differentiation- …. *Biochimica et Biophysica Acta (BBA)-Gene Structure and …*.

Targett-Adams P, McElwee MJ, Ehrenborg E, Gustafsson MC, Palmer CN, McLauchlan J. 2005b. A PPAR response element regulates transcription of the gene for

human adipose differentiation-related protein. *Biochim Biophys Acta* **1728**(1-2): 95-104.
Thompson FE, Subar AF, Brown CC, Smith AF, Sharbaugh CO, Jobe JB, Mittl B, Gibson JT, Ziegler RG. 2002. Cognitive research enhances accuracy of food frequency questionnaire reports: results of an experimental validation study. *J Am Diet Assoc* **102**(2): 212-225.
Tsitsiou E, Lindsay MA. 2009. microRNAs and the immune response. *Curr Opin Pharmacol* **9**(4): 514-520.
Unger RH. 2003. Lipid overload and overflow: metabolic trauma and the metabolic syndrome. *Trends Endocrinol Metab* **14**(9): 398-403.
van Herpen NA, Schrauwen-Hinderling VB. 2008. Lipid accumulation in non-adipose tissue and lipotoxicity. *Physiology & behavior* **94**(2): 231-241.
Vanherpen N, Schrauwenhinderling V. 2008. Lipid accumulation in non-adipose tissue and lipotoxicity. *Physiol Behav* **94**(2): 231-241.
Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT. 2011. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat Cell Biol* **13**(4): 423-433.
Visel A, Rubin EM, Pennacchio LA. 2009. Genomic views of distant-acting enhancers. *Nature* **461**(7261): 199-205.
Wang X, Reape TJ, Li X, Rayner K, Webb CL, Burnand KG, Lysko PG. 1999. Induced expression of adipophilin mRNA in human macrophages stimulated with oxidized low-density lipoprotein and in atherosclerotic lesions. *FEBS Lett* **462**(1-2): 145-150.
Wang Y, Sullivan S, Trujillo M, Lee MJ, Schneider SH, Brolin RE, Kang YH, Werber Y, Greenberg A, Fried SK. 2003. Perilipin expression in human adipose tissues: effects of severe obesity, gender, and depot. *Obes Res* **11**(8): 930-936.
Wasserman NF, Aneas I, Nobrega MA. 2010. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. *Genome Res* **20**(9): 1191-1197.
Waterworth DM, Ricketts SL, Song K, Chen L, Zhao JH, Ripatti S, Aulchenko YS, Zhang W, Yuan X, Lim N et al. 2010. Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arteriosclerosis, thrombosis, and vascular biology* **30**(11): 2264-2276.
Wei P, Taniguchi S, Sakai Y, Imamura M, Inoguchi T, Nawata H, Oda S, Nakabeppu Y, Nishimura J, Ikuyama S. 2005. Expression of adipose differentiation-related protein (ADRP) is conjointly regulated by PU.1 and AP-1 in macrophages. *J Biochem (Tokyo)* **138**(4): 399-412.
Weiss EC, Galuska DA, Kettel Khan L, Gillespie C, Serdula MK. 2007. Weight regain in U.S. adults who experienced substantial weight loss, 1999-2002. *Am J Prev Med* **33**(1): 34-40.
Wolins NE, Quaynor BK, Skinner JR, Schoenfish MJ, Tzekov A, Bickel PE. 2005. S3-12, Adipophilin, and TIP47 package lipid in adipocytes. *J Biol Chem* **280**(19): 19146-19155.
Wolins NE, Quaynor BK, Skinner JR, Tzekov A, Croce MA, Gropler MC, Varma V, Yao-Borengasser A, Rasouli N, Kern PA et al. 2006. OXPAT/PAT-1 is a PPAR-

induced lipid droplet protein that promotes fatty acid utilization. *Diabetes* **55**(12): 3418-3428.

Wolins NE, Rubin B, Brasaemle DL. 2001. TIP47 associates with lipid droplets. *J Biol Chem* **276**(7): 5101-5108.

Wolins NE, Skinner JR, Schoenfish MJ, Tzekov A, al. e. 2003. Adipocyte Protein S3-12 Coats Nascent Lipid Droplets*. *Journal of Biological Chemistry*.

Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. 2009. miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res* **37**(Database issue): D105-110.

Yamaguchi T, Matsushita S, Motojima K, Hirose F, Osumi T. 2006. MLDP, a novel PAT family protein localized to lipid droplets and enriched in the heart, is regulated by peroxisome proliferator-activated receptor alpha. *J Biol Chem* **281**(20): 14232-14240.

Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, Deloukas P, Dermitzakis ET. 2010. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* **26**(19): 2474-2476.

Zhang R, Wang YQ, Su B. 2008a. Molecular evolution of a primate-specific microRNA family. *Mol Biol Evol* **25**(7): 1493-1502.

Zhang W, Duan S, Kistner EO, Bleibel WK, Huang RS, Clark TA, Chen TX, Schweitzer AC, Blume JE, Cox NJ et al. 2008b. Evaluation of genetic variation contributing to differences in gene expression between populations. *Am J Hum Genet* **82**(3): 631-640.

Zhao Y, Joshi-Barve S, Barve S, Chen LH. 2004. Eicosapentaenoic acid prevents LPS-induced TNF-alpha expression by preventing NF-kappaB activation. *J Am Coll Nutr* **23**(1): 71-78.

Acknowledgments:

I thank Jose Ordovas for his outstanding mentorship and my committee, Gordon Huggins, Alan Kopin and Donna Slonim for their discussion and guidance on this work. I thank Larry Parnell and Chao-Qiang Lai for their enthusiasm and expertise. I thank the BU department of Biostatistics for their discussion and expertise on genetic epidemiology. I thank the laboratory of Carlos Fernandez-Hernando for the guidance and opportunity in performing functional assays.