A Multi-Scale, Systems Biology View of Fat Distribution and Storage Provides Novel Insight into Gene Regulation in Adipose Tissue

---

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment of the requirements for the degree

Doctor of Philosophy

by

Jordan N Reed

October 2023

ABSTRACT:

Adipose tissue is a complex organ that is capable of maintaining metabolic homeostasis. Storing fat, and specifically storing fat in the abdomen, increases the risk of cardio-metabolic diseases. Abdominal fat confers disease risk by inducing inflammation and insulin resistance. Pathways in the brain control overall obesity, while the preferential storage of fat in abdominal depots is likely controlled intrinsically by adipose tissue. Obesity and fat distribution are both complex traits, and genetic, lifestyle, and environmental factors interact to modify risk. Our understanding of the cellular and molecular mechanisms that cause obesity has allowed us to develop targeted therapeutics to treat it, but our understanding of body fat distribution is more limited.

In this dissertation, we investigate how gene expression and regulation in adipose tissue can influence fat storage at multiple biological scales. Chapter 1 summarizes abdominal and lower body adipose tissue function in health and disease states, how adipocytes contribute to tissue expansion. We discuss the genetics of complex diseases and review the genetics of obesity and fat distribution.

In Chapter 2, we predicted genes that were likely to regulate fat distribution in adipose tissue by modeling the gene-gene interactions using Bayesian networks. We first explored the parameters that influence the predictive power and biological relevance of the networks. We used optimal parameters to construct sex- and depot-specific models of adipose tissue gene regulation, and identified the putative network regulators in two independent datasets.

In Chapter 3, we narrowed this list of putative fat distribution regulators by considering publicly available data. We prioritize seven candidate genes within the Wnt signaling pathway or in mitochondria. We identify novel functions for five genes in adipogenesis or in mitochondrial function.

In Chapter 4, we study how diet composition and genetic background interact to influence body weight and other metabolic parameters in mice. We found that genetic background accounted for much of the variation, though diet was able to modify these effects. We identified genes in visceral adipose tissue that also respond to genetic background and diet interactions and found that these were partially explanatory of the observed phenotypes.

In Chapter 5, we discuss how our findings align with known biology and highlight some of our novel findings and predictions. Taking a systems biology approach, we integrate results from Chapters 2 and 3 to assess the predictive power of our models. We discuss future experimental and computational research directions.

Broadly, these studies investigated the regulation of gene expression in adipose tissue and the consequences on whole-body fat storage and metabolism. We identify novel regulators of adipocyte fat storage, and hypothesize a role for many more. We integrate predictive and experimental data at multiple biological scales to provide a holistic picture of how gene expression influences tissue function in health and disease.

AKNOWLEDGEMENTS

I want to start by acknowledging the wonderful people who have contributed to the success of this science. First, I must thank my advisor, Mete Civelek, for your incredible scientific, professional, and personal support over the past five years. It has been an honor to be your first graduate student, and I’ve enjoyed watching you grow the lab into what it is today. Thank you for teaching me to be a thoughtful systems biologist, for encouraging me to pursue difficult things, and for celebrating my successes. You’ve also helped me through the not so successful times with wisdom and grace, and I truly could not have asked for a better mentor.

I must thank the chair of my committee, Kevin Janes, for the expertise, thoughtfulness, and scientific rigor that you bring to every discussion. You have taught me to think critically and logically in all aspects of science, and I thank you for dedicating your time and careful consideration to my project. I want to thank my committee, Thurl Harris, Heather Ferris, Charles Farber, and Shayn Peirce-Cottler, for all of your help and advice. You each bring a diverse set of expertise that truly shaped this research into its current form. Thurl, thank you for being available to answer adipocyte questions and help interpret results, and Heather, thank you focusing my proposed research into something reasonable. In the time it took me to complete this research, both Charles and Shayn have become busy department heads, and I want to thank you both for continuing to make time and mental space for my project.

A big thank you goes to Susanna Keller, who has been a wonderful collaborator over the years and who lead the nutrigenomics studies. Without your thoughtfulness and expertise, this analysis would not have been possible. I also want to thank Dhanush Banka, who put countless hours into carefully collecting and analyzing the RNA-seq data for these studies, it was a pleasure working with you. Thank you to our collaborators at Novo Nordisk for your and expertise in completing the mitochondrial studies and for your thoughtful discussions of Bayesian networks. I must also thank Jacalyn Huband at UVA Research Computing, for your help implementing computational packages on UVA’s computing cluster, Marieke Jones of the UVA Health Sciences Library for teaching statistics in an approachable way and for answering all of my endless questions.

The Civelek lab members have been a source of joy, and I want to thank them all for making research fun and exciting. Noah and Yoni have been with me since the beginning, and we’ve shared the triumphs and the late nights that come with grad school. I am so glad I got to share in your scientific successes and I thank you both for your friendship and support. A huge thank you goes to Qianyi for answering every experimental question I’ve ever had and for teaching me how to ask questions. I will miss talking adipose tissue with the SubQ crew! I want to thank Nik for always finding joy in science and for leading the group game nights. I must also thank Jameson, for teaching me just about everything when I joined the lab, and Warren, for teaching me to think critically and be precise. Thank you to Lexi, Ashlyn, and Kelsey, for making the last (and toughest) part of grad school so much fun, and to Redouane and Atum, for keeping me company in cell culture. It was truly an honor to work with such amazing scientists and great friends.

I want to thank Paul Dunman, my undergraduate advisor and my first boss in science. You helped me start to think like a scientist, and encouraged me to pursue a PhD before I even thought that was possible.

Finally, I want to thank my family and friends for the love and support you’ve shown me over the last five years. You’ve housed me, fed me, sent me care packages, called me, visited me from far away, and I am forever grateful to have you in my life. I thank you all for encouraging me to pursue science, listening to me talk about my research, and being proud of my accomplishments. I could not have done it without you.

Appendix:

Chapter 5: Discussion- Systems Biology Approaches are characterize fat storage and distribution…………………….…………………………………………………………………………..….. 211

List of Figures:

List of Tables:

Chapter 1: Introduction

1.1 Adipose Tissue:

Endothermic animals require energy to conduct normal metabolism and maintain body temperature. Thus, mammals, birds, and some fish have developed specific organs, called adipose tissue, to store excess energy for times of high demand[1]. Specialized cells in adipose tissue store the excess nutrient input as triglycerides[2]. During times of increased energy demand, such as starvation or illness, adipose tissue can break down its stores and release fatty acids, which can be used by peripheral tissues. These processes are carefully orchestrated by hormonal signaling from adipose tissue to the brain[3]. Further, adipose tissue provides mechanical padding for limbs and organs and maintains body temperature via physical insulation and electrochemical heat production, called thermogenesis[2].

1.1.1 A Brief History –

Throughout human evolution, our species faced the challenges of food scarcity and under-nutrition[1]. Storing excess fat in adipose tissue during times of abundance would have been advantageous. Since the industrial revolution, large parts of the world now have cheap, immediate access to food sources, particularly calorie-dense but nutrient-lacking foods, and thus, now face challenges of over-nutrition.

Until the 1940s, our conception of adipose tissue was simplistic. The prevailing school of thought was “calories in- calories out”, or, the amount of fat stored in adipose tissue –and thus overall body weight- was directly determined by a person’s chosen diet and exercise regime[3]. Then, researchers identified that lesions in the rat hypothalamus could cause obesity[4], as could a genetic mutation on chromosome 6 of the *ob/ob* mouse[5]. For the first time, researchers showed that obesity and eating behavior were under intrinsic control by the brain and some element of the genome, and were not the result of simply choosing to eat more[6]. Further studies using *ob/ob* and *db/db*[7] mice identified leptin as a hormone secreted from adipose tissue that signaled to a receptor in the hypothalamus that controlled eating behavior[8]. Mice without leptin could not signal to the brain to stop feeding, while mice without the receptor could not respond the leptin signal. Since the discovery of leptin, researchers have identified other genes, pathways, and adipose tissue secreted factors (adipokines) that contribute to obesity and its complications, leading them to conclude that storing fat is a complex process in which adipose tissue has an active role[3]. Further, we have a better understanding of how excess adipose tissue storage dysregulates overall metabolic health and contributes to the risk of disease in some individuals[2].

1.1.2 Structure and Function-

1.1.2.1 Adipose Tissue Depots:

Adipose tissue is distributed across the human body into depots (Figure 1.1 A). Most fat is stored in white adipose tissue, named for the white appearance of the large lipid droplets in gross dissection[1,2]. Subcutaneous white adipose tissue is found directly under the skin, with large depots in the gluteo-femoral region, lower abdomen, upper arms, and face. Visceral white adipose tissue is located within the abdominal or chest cavity, with depots found among or even wrapped around the organs. While all white adipose tissue is capable of storing lipids, there are large differences the metabolic health and homeostasis maintained by certain depots. Brown adipose tissue's primary function is not lipid storage; thus, it appears brown in dissection[2]. Instead, it produces heat electrochemically and is found primarily behind the neck in infants and in adults with prolonged exposure to cold. Adipose tissue can also be found in the breast, where it has a range of additional functions[9].

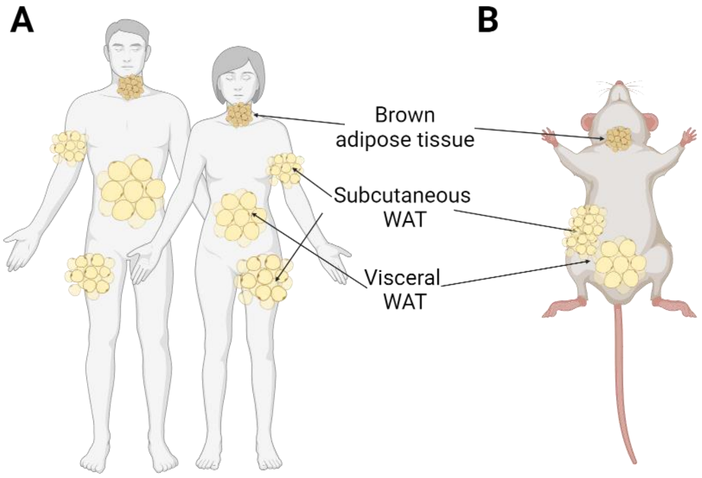

Figure 1.1: Adipose tissue depot locations in human and mouse.

(A) Humans have white adipose tissue depots in the abdomen, arms, and lower body and sometimes have intra-scapular brown fat.

(B) Mice have white adipose tissue depots in the abdomen and have intra-scapular brown fat. Figure created with BioRender.com

1.1.2.2 Cell Types:

Adipose tissue is made up of a diverse collection of cell types that are loosely organized in support of the primary functional cell, the adipocyte[2]. White adipocytes are large (70-300 µM), spherical cells whose cytoplasmic space is dominated by a single triglyceride-storing organelle, the lipid droplet. Because of this structure, large adipocytes are fragile and prone to lyse apart. Healthy adipocytes maintain communication with the brain and other organs and thus have many surface receptors to remain responsive to signaling cues[10]. Brown adipocytes, found in brown adipose tissue, have multiple smaller lipid droplets and many mitochondria[2]. These thermogenic adipocytes upregulate a gene uncoupling protein 1 (*UCP1*) whose protein is able to produce heat by dissipating the electrochemical energy produced by the electron transport chain. In white adipose tissue, a subset of adipocytes upregulate *UCP1* and produce heat, termed beige adipocytes. All adipocytes are terminally differentiated from precursor cells and cannot proliferate.

Adipocyte precursor cells called pre-adipocytes are mesenchymal, fibroblast-like cells that can migrate, proliferate, and ultimately differentiate to mature adipocytes by taking up glucose and free fatty acids (FFA) and storing them as triglycerides. These cells are derived from mesenchymal stem cells (MSC), which can ultimately giving rise to white and brown adipocytes, osteoblasts, and smooth muscle cells.

Adipose tissue and adipocytes are sensitive to hypoxia and thus require blood vessels throughout the tissue[11]. Endothelial cells, smooth muscle cells, pericytes, and some pre-adipocytes[12] are found in and around the vasculature and communicate reciprocally with adipocytes[13]. Neurons innervate the tissue and provide catecholamine signals to adipocytes through b-adrenergic receptors[14]. Adipose tissue contains many immune cells including monocytes, macrophages, T-cells, B-cells, neutrophils, etc[15]. Because of these resident and responsive cells, adipose tissue can launch a robust inflammatory response that may become dysregulated in old age[16] or obesity[15,17]. In visceral adipose tissue only, an outer layer of mesothelial cells surrounds the tissue, though their role has not been precisely determined[18].

The main function of adipose tissue is to maintain metabolic homeostasis; adipocytes in white adipose tissue accomplish this by storing excess caloric energy, and communicating with systemic organs and the brain to modify behavior[2].

1.1.2.3 Adipocyte signaling:

Adipocytes respond to many signaling cues in order to maintain metabolic homeostasis[3]. Insulin induces glucose uptake in adipocytes by turning on PI3K and AKT, which leads to GLUT4 translocation to the cell membrane[10]. Insulin signaling also leads to downstream inhibition of food intake in favor of triglyceride breakdown. JAK/STAT signaling molecules influence leptin secretion, while MAPK pathways induce tissue inflammation[10]. B-adrenergic signals from the brain can stimulate the adipocyte to break down its triglyceride stores[14]. Cyclic-AMP (cAMP) signaling mediates this lipid breakdown, and in brown adipocytes, it promotes thermogenic uncoupling[10]. AMPK signaling also helps maintain brown adipocyte identity[10,14]. Wnt signaling and TGF-β/BMP pathways regulate pre-adipocyte proliferation by turning on cell cycle genes[10,19,20]; Wnt also inhibits master adipocyte identity genes *PPARG* and *CEBPA*.

Adipocytes release a variety of signaling molecules, the most famous being leptin[3,21,22]. Leptin is a peptide hormone released into the blood stream, where it moves to the hypothalamus and binds the leptin receptor. This turns on satiety pathways in pro-opiomelanocortin (POMC) neurons, which release α-melanocyte-stimulating hormone (a-MSH) and turns off hunger pathways, which release agouti-related protein (AgRP). These signals are integrated by the melanocortin receptor (MC4R), which propagates the satiety behavior[23]. While leptin promotes glucose utilization in peripheral tissues, at sustained high doses, cells can become leptin resistant and, in turn, insulin resistant[22,24]. Other adipokines like adiponectin and omentin help promote glucose homeostasis and insulin sensitivity in surrounding tissues, while resistin and inflammatory cytokines TNF-α and IL6 lead to insulin resistance[3,24]. Finally, adipocytes release fatty acids and glycerol. While this is normal and necessary to provide peripheral tissues with energy, the uncontrolled or sustained release of FFAs can cause insulin resistance, inflammation, and lipotoxicity[25].

1.1.2.4 Mechanisms of fat storage:

Adipose tissue accommodates excess nutritional energy by increasing the number or size of adipocytes in the tissue[26–28] (Figure 1.2). To increase the total number of mature adipocytes, called hypertrophic expansion, pre-adipocytes must differentiate into mature lipid-storing adipocytes (Figure 1.2 A, C). Differentiation is initiated by transcription factors PPARɣ, CEBPα, CEBPβ, and CEBPδ, which turn on insulin signaling, glucose uptake, and triglyceride synthesis-related genes[29]. The pre-adipocyte pool is carefully regulated by Wnt and TGF-β signaling pathways to maintain a proliferative subpopulation[10,30,31] (Figure 1.2 B). Old or dysfunctional adipocytes often undergo apoptosis, reducing the total number of adipocytes and recruiting immune cells to participate in clearance[29,32].

Existing adipocytes can accommodate energy excess by increasing the size of the lipid droplet, called hyperplastic expansion (Figure 1.2 C, D). Adipocytes do this primarily by taking up free fatty acids through fatty acid transporters (FATP) and re-esterifying them to glycerol via the enzyme DGAT[33]. When adipocytes take up glucose via GLUT4, they undergo a process called *de novo* lipogenesis, in which they convert glucose to citrate in the TCA cycle, then use enzymes FASN, ACC1, ACLY, and SCD1 to convert citrate to a fatty acid[34]. This process requires mitochondria, which are bound to the surface of the lipid droplet[35]. While most of adipocyte triglyceride stores come from fatty acid uptake[33], there is evidence that *de novo* synthesized fatty acids may be used in lysosomal and autosomal membranes, involved in protein degradation[36]. Lipolysis is the process by which adipocytes break down stored triglycerides to release FFAs and glycerol. Lipases ATGL, HSL, and MGL each remove one fatty acid from the backbone, which are released into the bloodstream[37]. Brown and beige adipocytes upregulate *UCP1* and are capable of thermogenesis[38–40]. Upon cold or β-adrenergic stimulation, these cells undergo lipolysis, and the resulting fatty acids are metabolized via the TCA cycle and β-oxidation pathways. This fuels a proton gradient down the electron transport chain, where UCP1 shuttles the H^+ ions across the mitochondrial membrane without driving ATP production; instead, the energy is dissipated as heat into the surrounding cell and tissue.

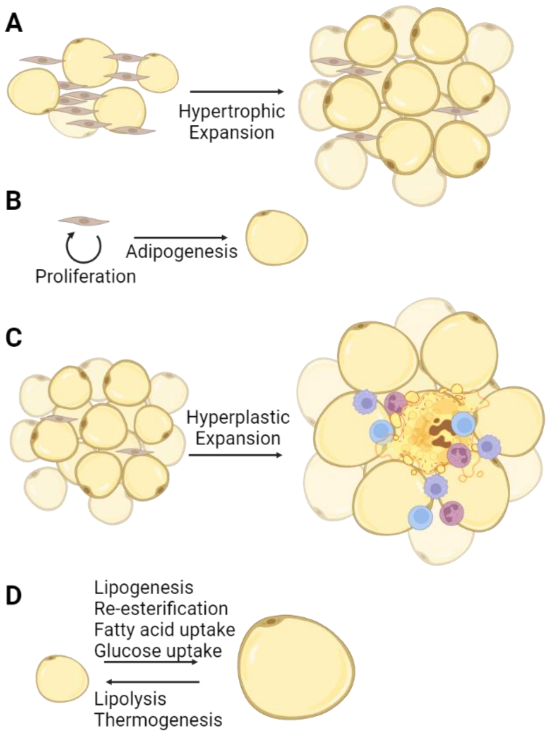

Figure 1.2: Adipose tissue expands through hypertrophic and hyperplastic adipocyte processes.

(A) In hypertrophic expansion, pre-adipocytes increase the number of mature adipocytes
(B) Pre-adipocytes proliferate and differentiate to contribute to hypertrophic processes.
(C) In hyperplastic expansion, existing adipocytes increase size by storing more lipid
(D) Adipocytes take up glucose and fatty acids, undergo lipogenesis or re-esterification to store lipids and release stored lipids via lipolysis and thermogenesis to contribute to hyperplastic processes. Figure made using BioRender.com

1.1.2.5 Maintaining Healthy adipose tissue:

Healthy adipose tissue is characterized by functional adipocytes that stably store triglycerides and can sense and respond to signaling cues[1,28]. In times of nutritional excess, adipose tissue first tries to accommodate by differentiating more pre-adipocytes into lipid-storing mature adipocytes[26]. These small adipocytes are sensitive to signals and can easily be oxygenated by the vasculature[26]. If the proliferative capacity cannot meet demand, the tissue will add more lipids to existing adipocytes. Larger adipocytes cannot be vascularized well and will become hypoxic[11]. This causes the cells to release inflammatory cytokines, which will recruit immune cells to propagate the inflammatory response[15]. If an adipocyte becomes too large, it can become apoptotic or necrotic, propagating the immune response[32]. Unresolved inflammation can contribute to insulin resistance in adipose and systemic tissues[15,24]. Eventually, the adipose tissue will reach its storage capacity, and excess lipids will be stored ectopically in other tissues[41].

1.1.2.6 Visceral adipose tissue dysfunction:

Subcutaneous adipose tissue is considered the model for healthy fat storage[43]. Subcutaneous pre-adipocytes have a large capacity to proliferate and differentiate[2,44,45]. These cells can be easily vascularized and innervated, and thus, the tissue remains normoxic and responsive to signaling cues[11,26]. Each individual has a different capacity for subcutaneous fat storage[26], and sex, hormones, and age play significant roles in fat storage[16]. Pre-menopausal women tend to store more fat in gluteo-femoral subcutaneous depots than men[43,46]. Men instead preferentially store fat in visceral abdominal depots. Visceral pre-adipocytes are less able to proliferate and differentiate to accommodate over-nutrition[44,45,47]. Excess caloric energy is stored in the lipid droplets of existing adipocytes, making them large, fragile, insulin-resistant, and hypoxic[11,26,41,49]. While visceral adipocytes are smaller than subcutaneous in healthy adults, in disease conditions, they expand more, and their expansion is correlated with diabetes phenotypes[26,48,49]. It is thought that evolutionarily, visceral fat exists to elicit an inflammatory response during gut infection or injury[50]. As such, visceral tissue is prone to inflammation. Macrophages surround dead and dying adipocytes in “crown-like structures” and release inflammatory cytokines[49,51]. Further, visceral adipocytes are more metabolically active, undergoing lipolysis at higher rates than subcutaneous cells, releasing FFAs into the bloodstream[49]. In total, the unhealthy mechanisms of visceral fat storage cause harm to adipose tissue and to systemic metabolism.

1.1.2.7 Visceral adipose contributes to systemic metabolic dysfunction:

Visceral adipose tissue is located within the abdominal cavity and is vascularized by the portal vein. Inflammatory cytokines and FFAs are released into the portal vein, where they move toward other metabolic organs, including liver and pancreas[52]. This causes lipotoxicity[53], inflammation[51,54,55], insulin resistance[56], hypertension[57], and ectopic fat deposition[32,42] in these tissues, contributing to diabetes and fatty liver disease phenotypes[32,57]. Diseased cells secrete abnormal levels of adipokines, which can further dysregulate systemic metabolism[24]. Conversely, subcutaneous adipose tissue simultaneously releases fewer inflammatory molecules and FFAs and releases them to the peripheral tissues where the detrimental effects on metabolism are lessened.

1.1.3 Phenotypes of Over-nutrition-

1.1.3.1 Obesity:

Obesity is a disease of excess adiposity. Clinicians often use the body mass index (BMI), or the ratio of weight (in lbs.) to height^2 (in ins), as an approximation of fat storage[58]. They define underweight as BMI < 18.5, normal weight as BMI 18.5-25, overweight as BMI 25 -30, obese as BMI > 30, and morbidly obese as BMI > 35. While BMI is a reliable weight indicator, it can conflate lean mass and adiposity and sometimes mis-categorize healthy individuals. Obesity is caused by various factors, including diet (caloric and nutritional content), exercise and lifestyle factors, prescription medications, stress, socioeconomic status, hormonal imbalances, and genetic variation, among others[59]. Over 40% of adults in the United States are obese, and prevalence rates are higher in Black and Hispanic populations[60]. For a majority of the population, increased BMI is associated with the prevalence of other diseases. Observational studies show that higher BMI is associated with mortality of all causes[61], cardio-vascular disease[61–64], stroke[64,65], type 2 diabetes[61,64], fatty liver disease[66], cancer[61,64,67,68], and infectious disease[61,69–71]. Interestingly, having an underweight BMI is also associated with the risk for many infectious diseases[61,70,71].

The disease risk imparted by increased BMI is due to harmful adipose tissue expansion. In vascular diseases, the increased lipid species released into the blood can build up in vessel walls[72]. These cell and lipid-filled plaques can occlude the vessel, causing heart disease, or rupture, causing downstream clots like strokes. Obesity increases insulin resistance and causes glucose dysregulation in type 2 diabetes[73]. Ectopic fat deposition in other organs, such as the liver or skeletal muscle, causes fatty liver disease and increased insulin resistance[74]. Finally, lipid species increase the inflammatory response in other tissues. As macrophages try to clear dead and dying cells, they become lipid-loaded, perpetuating aberrant inflammatory responses in vascular disease[75] and cancer progression[76].

1.1.3.2 Metabolically Healthy Obesity:

While excess adipose tissue fat storage often leads to whole-body metabolic dysfunction and a range of comorbidities, each individual has a different capacity for healthy fat storage[26,41]. Clinicians observe that a subset of obese and overweight individuals are metabolically healthy- these people remain sensitive to insulin, have normal circulating levels of glucose and triglycerides, and do not have ectopic fat deposition in other organs or dangerous plaques in vessels[77]. This condition is termed metabolically healthy obesity (MHO), and its discovery has helped some clinicians and researchers to redefine obesity as increased adiposity that presents a risk to health. Using obesity status and metabolic parameters together, researchers show that abnormal metabolic parameters have larger influence on all-cause mortality and cardiovascular and cancer-related deaths[78].

The discovery of MHO also forced clinicians to consider what types of adiposity present the greatest risk to health. Pre-menopausal women with more subcutaneous fat storage than men are at lower risk for cardiovascular disease[79,80]. The prevalence of these diseases increases in post-menopausal women, when estrogen hormone signaling is lost and fat storage shifts from lower body subcutaneous depots to visceral abdominal storage[81]. In the past several decades, more research has been devoted to understanding the differences between subcutaneous and visceral adipose tissue and the implications on metabolic health.

1.1.3.3 Body Fat Distribution:

The relative storage of fat between different depots is called body fat distribution. While there are dozens of depots throughout the body, many studies focus on the differences between disease-causing visceral abdominal fat[45,47] and healthy lower body subcutaneous fat[43]. Clinicians approximate the storage in these locations as the waist-to-hip ratio (WHR), which is the waist circumference divided by hip circumference. Using whole-body dual X-ray absorptiometry (DXA) or magnetic resonance imaging (MRI)[82], researchers find that WHR and other anthropometric traits are fairly correlated with imaging based metrics of depot-specific storage[83–89], depending on the imaging modality and study population. An individual's WHR is relatively constant throughout adult life, and researchers find that hormones, age, sex, genetics, and some HIV prescription medications can modify fat distribution[90,91]. However, our understanding of the genetic and molecular mechanisms contributing to fat distribution is limited. Thus unlike overall obesity[92,93], no targeted therapeutics or lifestyle interventions to combat abdominal obesity are known[91].

WHR is strongly associated with the prevalence of cardiovascular disease[94–96], stroke[97–100], and type 2 diabetes[101]. Because obesity also increases the risk for these diseases, researchers often regress out the effects of BMI ($\text{WHR}_{\text{adjBMI}}$). They find that $\text{WHR}_{\text{adjBMI}}$ is also highly associated with these diseases[102] and conclude that WHR likely confers disease risk through independent mechanisms than those conferred by BMI.

1.2 Complex Trait Genetics:

The chromatin sequence in a person's cells has a strong influence on the function or dysfunction of the cell, tissue, and whole body[103,104]. Most human cells have 23 pairs of chromosomes, one copy from each parent. Following the central dogma, the chromatin sequence encodes genes, which are transcribed into mRNA copies, which are then translated into proteins. Altering the sequence of a gene within the chromatin can alter the structure and function of the encoded protein, while the chromatin sequences outside of gene bodies can regulate how many copies of each gene are transcribed[103,105,106]. Thus, mutations or naturally occurring variations in the DNA sequence can have sizable effects on cellular phenotypes and disease risk[103,104].

1.2.1 Monogenic disorders-

Some diseases and phenotypes are regulated by a single gene[103,107]. Gregor Mendel first discovered this phenomenon when crossing pea plants. He found that certain traits were inherited across generations and were inherited in conserved patterns. Although forgotten for over a hundred years, his discoveries paved the way for our modern understanding of genetics[108]. In the 1900s, novel discoveries showed that DNA was the "transforming factor" that conferred variation and that the genetic sequence encoded protein amino acid sequence[108]. Since these discoveries, we have identified many single gene defects that cause disease in the same conserved patterns of inheritance described by Mendel. Specifically, most genes will have a dominant and a recessive version, called an allele; in Mendelian inheritance, any number of copies of the dominant allele causes expression of the dominant phenotype[103]. The recessive phenotype is only seen when the alleles from both parents are recessive. Often, the dominant allele encodes a functional protein, while the recessive version encodes a truncated, misfolded, or nonfunctional version. Sometimes, these proteins are rendered nonfunctional by a single base pair change in the coding sequence, called a single nucleotide polymorphism (SNP), while other protein changes are caused by large insertions or deletions of DNA sequence[103]. Inheriting at least one dominant, functional copy from either parent leads to healthy functional cells containing the protein, while having two recessive copies usually means an individual does not have any functional protein. In some cases, this does not affect the health of the individual, but if the protein is very important, having two recessive alleles will lead to disease[107]. For example, patients with phenylketonuria lack the gene *PAH*, which breaks down dietary phenylalanine. People with one or two functional copies will be phenotypically normal, while people with two recessive, non-functional copies have phenylalanine buildup in cells[109]. Disease-causing recessive alleles are often present at a very low frequency in the population, but have a large effect size on the disease or trait of interest[104].

1.2.2 Polygenic complex diseases and traits-

Most common diseases and phenotypes are not caused by a defect in a single gene; instead, they result from many different alleles and other risk factors that each confer a small amount of risk or protection[104,110]. Usually, the SNPs that confer disease risk are not found within the gene body and do not cause changes in the protein structure. Instead, the SNPs reside in regulatory regions, such as promoters or enhancers, and influence the number of copies of each gene transcribed[105,106]. Promoters are regions of the genome directly upstream from gene bodies (Figure 1.3 A). They often have conserved sequence elements where transcriptional machinery binds. Enhancer sequences also bind transcriptional machinery but are great distances upstream of the gene body; the chromatin loops around to facilitate a physical connection (Figure 1.3 A). Transcription factors bind the DNA at specific sequences and can help or hinder transcription of the gene. SNPs in enhancer or promoter regions often change the sequence of a transcription factor binding site (TFBS), which can alter the affinity of binding and, ultimately the amount of transcription (Figure 1.3 B,C). For example, height is a complex, polygenic trait. Hundreds of SNPs in the genome confer a small amount of change to overall height[111], many by altering the expression of nearby genes. Modern schools of thought believe that genetic variation's heritable effects on phenotypes are due to linear combinations of SNP effects[112,113] and non-linear interaction effects between the genetic variants and with other risk factors and confounders[110,114,115].

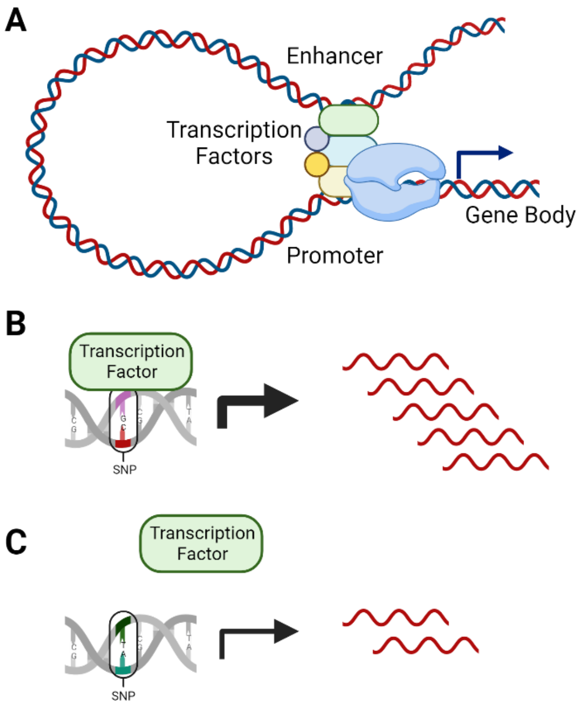

Figure 1.3: Genetic variants that affect transcription factor binding sites may regulate gene expression.

- *(A) Transcription factors bind to conserved DNA sequences to facilitate the transcription of the gene by RNA polymerases. Regulatory sequences can be proximal to the gene start site, called promoters, while enhancers are distal regulator elements that loop to facilitate connections with the promoter*
- *(B) Genetic variants that affect the binding site of important transcription factors affect the expression of those target genes. For example, one allele at a SNP may create a higher-affinity binding site and thus encourages transcription of the gene.*
- *(C) While the alternate allele at the SNP leads to less transcription factor binding and fewer copies of the target gene.*
  Figure made using BioRender.com

<u>*1.2.3 Expression Quantitative Trait Loci (eQTL)-*</u>

One way to determine if a SNP affects the expression of a nearby gene is to perform eQTL studies in tissues of interest[104,106]. In these studies, hundreds of individuals are recruited, and first, their genome is sequenced. The human genome is structured into haplotypes, or long regions on each chromosome that are consistently inherited together[116,117]. During meiotic cell division, the chromosome pairs will align, and "crossing over" events occur when a homologous segment of each chromosome breaks off and recombines with the other chromosome pair. These events happen at conserved locations in the genome and create haplotype blocks. The SNPs in these haplotypes are consistently inherited together and are in high linkage disequilibrium (LD) with each other[103]. Therefore, instead of sequencing the entire genome for every individual, we sequence a subset of SNPs unique to each haplotype block and impute, or infer, the rest of the SNPs in high LD[118].

Next, researchers take a biopsy or sample of the tissue of interest and sequence[119] the RNA species. Using the gene expression and alleles at each SNP for the same individuals, they use linear regression to determine if there is a significant association between the number of copies of the risk allele and the expression of the gene[120]. To obtain appropriate power, researchers will often consider associations between SNPs and genes that are proximal (± 1Mb) as *cis*-eQTLs, because most enhancers and promoters are on the same chromosome and are short distance away[121–123]. This allows us to correct for fewer tests and use a statistical threshold of $\sim P < 1\text{x}10^{-4}$ when determining significance.

Researchers will then consider associations between SNPs and distal genes, called *trans*-eQTLs[106]. Because of the great distance, these associations likely represent secondary effects where a SNP regulates the expression of a diffusible transcription factor, which binds DNA and regulates distal genes. These effects are usually very strong to be detected in a small sample size and over a higher multiple testing correction burden[120].

1.2.4 Genome-wide Association Studies-

We can determine which SNPs are associated with complex diseases and traits using genome-wide association studies (GWAS)[103]. In GWAS, researchers recruit hundreds of thousands of people, both with and without the disease of interest, and sequence their genomes to determine which alleles they have at each SNP of interest. Next, they perform association tests to determine which SNPs are significantly associated with having the disease (Figure 1.4 A). Because of the LD structure of the genome, often a group of proximal SNP are significantly associated[124] (Figure 1.4 B). We term this area in the genome a locus, and often refer to an associated locus instead of individual associated SNPs. The large sample size allows researchers to correct for the multiple tests performed; the genome-wide level for significance of $P < 5\text{x}10^{-8}$ corrects for the number of SNPs tested[125,126]. GWAS can be performed on continuous traits as well, and have become a powerful tool for determining the genetic regulation of a disease or trait[103,104].

However, there are significant challenges associated with identifying the mechanism by which an associated SNP affects the trait of interest, given that there can be multiple genes and signals in the locus, and the strengths of these signals can vary between populations and with biological confounders, such as sex and age[124]. Further, many GWAS have been performed primarily in European and secondarily in Asian populations[127,128]. While these have advanced our understanding of disease in these populations, other populations contain different SNPs at different allele frequencies[129]. *PCSK9* was identified as a regulator of fatty liver disease only after studying an African American population where the risk allele was present at higher frequencies[130]. Additionally, studies have found that the nearest gene to the GWAS signal is the causal gene in 70% of loci; in the other 30% of loci, this assumption does not hold, and a more distal gene is responsible for the disease risk imparted by the SNP[122]. Finally, genes not under significant genetic regulation may also contribute to the studied phenotype or disease; they may be under environmental control[131]. These considerations will be discussed further in Chapter 3 and 4.

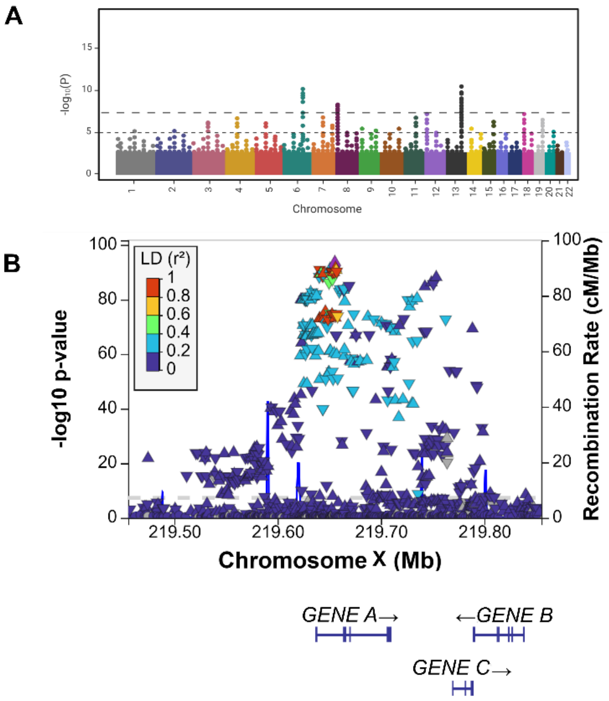

Figure 1.4: Genome-wide association studies identify genetic variants associated with disease

- *(A) Manhattan plot showing the associations between SNPs across the genome and a disease or trait of interest. Associations greater than* $5 \text{ x}10^{-8}$ *are considered genome-wide significant.*
- *(B) The associated variants in one locus within the GWAS. The color represents the linkage disequilibrium (LD) with the highest associated variant. Blue line represents the rate of crossover events at each location. Gene bodies in the locus are plotted below*
  Figure made using BioRender.com

1.2.5 Mendelian Randomization-

To determine how two polygenic traits are causally related, we can use Mendelian Randomization (MR)[132,133]. In this approach, we know that a risk factor is significantly associated with SNPs in the genome, and using central dogma, we conclude that these SNPs are regulatory of that risk factor. We want to test if the risk factor is causal for the disease of interest. Because alleles of each SNP are assigned randomly during meiosis, we employ this natural variation to create randomized groupings based on genotype. We then statistically determine if the SNPs are associated with the disease of interest only through the risk factor, not indirectly or by some unmeasured confounder. If this assumption is true, we can say the risk factor is causal for the disease of interest (Figure 1.5).

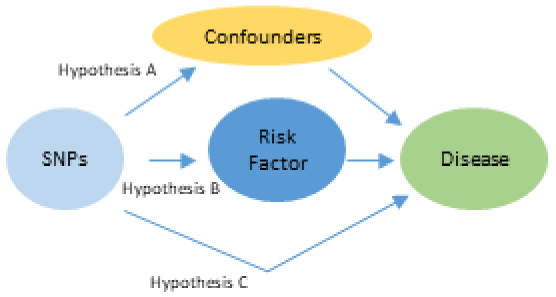

Figure 1.5: Hypothesis testing used to determine if a complex risk factor is causal for a complex disease or trait using Mendelian Randomization.

1.2.6 Genetics of Obesity-

Obesity, as measured by BMI, is a complex trait determined by genetics and other risk factors[58]. BMI is up to 55% heritable- meaning that, in some populations, up to 55% of a person's BMI can be explained by their genetic background[134]. Single gene defects in the leptin (*LEP*) and the leptin receptor (*LEPR*), or its downstream signaling partners melanocortin 4 receptor (*MC4R*) and *POMC*, can cause monogenic forms of obesity[135,136]. To date, GWAS have been performed on >1,000,000 individuals and have uncovered over 1,200 loci associated with BMI in at least one population. These studies have implicated many other genes in the leptin-melanocortin pathway, including *AGRP, PCSK1, PHIP, SH2B1, SIM1,* and *ADCY3,* and additional genes that control brain function, including *BDNF, NEGR1*, and *NTRK2*[111,136,137]. When considering the genes nearby the BMI GWAS loci, researchers found that they often contained brain specific enhancer and promoter sequences, and the genes were preferentially expressed in brain[138,139].

Using MR studies, we find that BMI is a causal risk factor that increased the odds ratio of incident disease risk for cardio-vascular disease[102,140], type 2 diabetes[102,140–143], cancer[144], and infectious disease[145]. Because of our understanding of the genetic causes of obesity, researchers have identified druggable targets and have successfully created therapeutics to modify them[10,58]. In individuals with upstream mutations in the leptin/MC4R pathway, Setmelanotide can successfully signal to MC4R to indicate satiety[146]. Further, GLP1 and GIPR agonists have shown success in recent trials[92,93].

1.2.7 Genetics of Fat Distribution-

Body fat distribution, as measured by $\mathrm{WHR}_{\mathrm{adjBMI}}$, is a complex trait determined by genetics and other risk factors; it is up to 60% heritable in some populations[147]. Monogenic forms of extreme fat distribution phenotypes are called lipodystrophies[148]. *AGPAT2, BSCL2, CAV1*, and *PTRF* mutations cause almost complete loss of all adipose tissue, while mutations in *LMNA, PPARG, PLIN1, CIDEC,* and *LIPE* all cause loss of subcutaneous fat from limbs, and *AKT2, PCYT1A, PIK3R1, MFN2, PSMB8,* and *ADRA2A* present partial lipodystrophy features. Many of these genes normally function in adipocyte fat storage pathways, and most forms of lipodystrophy are accompanied by metabolic dysfunction.

The first meta-analysis of GWAS of WHR was performed in 2009 and identified 1 locus nearby gene *LYPLAL1*[149]. Subsequent studies and meta-analysis of GWAS in WHR and $\mathrm{WHR}_{\mathrm{adjBMI}}$ predicted many more candidates[150–154] and showed that the expression of those genes differed between visceral and subcutaneous depots[154]. Using promoter and enhancer, open chromatin profiling, and gene expression data, researchers found that the genes nearby the loci were likely expressed and regulated in adipose tissue[138,154]. They focused on a group of developmentally related genes, and showed that *TBX15* regulated fat storage differently in visceral and subcutaneous adipocytes[154]. The most recent GWAS meta-analyses have uncovered ~350 loci associated with $\mathrm{WHR}_{\mathrm{adjBMI}}$[155]. These genes include *RSPO3, TBX15, VEGFA, CPEB4,* and others. Approximately one-third of the GWAS loci are associated with fat distribution in only one sex[155,156], and some have a greater magnitude of association in one sex[153].

Mendelian Randomization studies showed that WHR and $\text{WHR}_{\text{adjBMI}}$ are causal risk factors for cardiovascular disease[102,157], stroke[102], type 2 diabetes[102,141], fatty liver disease[158], and cancer[159,160]. Measures of central obesity increase disease risks as much or more than BMI.

Despite the strong genetic component of this trait, only five genes (*KLF14, LRP5, TBX15, RSPO3,* and *SHOX2*) have been mechanistically linked to disproportionate fat storage in one depot compared to the other[161–167]. While these studies have been crucial to advancing our understanding of this phenotype, we have not yet identified any druggable targets, and therefore, no targeted therapeutics to modify WHR exist[91].

Because many of these genes putatively act in adipose tissue, fat distribution researchers take advantage of public –omics data in adipose tissue to uncover the genetic and cellular mechanisms. eQTL studies associating SNPs and adipose tissue gene expression have been performed in European men of the Metabolic Syndrome in Men (METSIM) cohort[168], primarily European-ancestry adults of the Gene Tissue Expression (GTEx)[169], and in European adults of the Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task (STARNET)[170]. Single-cell and single-nucleus RNA-sequencing approaches have been applied to adipose tissue with varying success[171–178]. Large adipocytes often lyse open when sorted into single droplets, so most sequencing will either exclude adipocytes, use nuclei instead of whole cells, or only contain small adipocytes. Other common problems persist in this data, such as lack of sequencing depth and lack of reproducibility of findings[179]. Despite these problems, single-cell experiments have illuminated sub-populations of adipocytes and pre-adipocytes poised for different types of metabolism[178].

1.3 Summary of Thesis:

In summary, adipose tissue is a complex, active organ that regulates systemic metabolism. Dysfunction in brain pathways and in communication between adipose and the brain leads to excess adiposity, while gene expression in adipose tissue determines the distribution of fat between depots. We have uncovered many mechanisms by which protein integrity and gene expression cause obesity and have designed therapeutics to target some of these. To treat the metabolic complications of excess visceral adiposity, we need to broaden our understanding of the genes and pathways that control preferential fat storage in this depot.

Chapter 2 will discuss strategies to identify and prioritize putative gene regulators of body fat distribution using Bayesian networks and publicly available data. Chapter 3 will discuss strategies to identify causal genes and report our results experimentally validating seven candidate fat distribution genes. Chapter 4 will investigate how genetic background and diet composition interact to influence metabolic parameters and adipose tissue gene expression. Chapter 5 will discuss the implications of these results.

1.4 Bibliography

1. Pond, C. M. The Evolution of Mammalian Adipose Tissues. in *Adipose Tissue Biology* (ed. Symonds, M. E.) 1–59 (Springer International Publishing, 2017). doi:10.1007/978-3-319-52031-5_1.
2. Rosen, E. D. & Spiegelman, B. M. What we talk about when we talk about fat. *Cell* **156**, 20–44 (2014).
3. Li, M.-D. Leptin and Beyond: An Odyssey to the Central Control of Body Weight. *Yale J Biol Med* **84**, 1–7 (2011).
4. Hetherington, A. W. & Ranson, S. W. The spontaneous activity and food intake of rats with hypothalamic lesions. *American Journal of Physiology-Legacy Content* **136**, 609–617 (1942).
5. Ingalls, A. M., Dickie, M. M. & Snell, G. D. Obese, a new mutation in the house mouse. *J Hered* **41**, 317–318 (1950).
6. Kennedy, G. C. The role of depot fat in the hypothalamic control of food intake in the rat. *Proc R Soc Lond B Biol Sci* **140**, 578–596 (1953).
7. Hummel, K. P., Dickie, M. M. & Coleman, D. L. Diabetes, a new mutation in the mouse. *Science* **153**, 1127–1128 (1966).
8. Coleman, D. L. Effects of parabiosis of obese with diabetes and normal mice. *Diabetologia* **9**, 294–298 (1973).
9. Kothari, C., Diorio, C. & Durocher, F. The Importance of Breast Adipose Tissue in Breast Cancer. *Int J Mol Sci* **21**, 5760 (2020).
10. Wen, X. *et al.* Signaling pathways in obesity: mechanisms and therapeutic interventions. *Signal Transduct Target Ther* **7**, 298 (2022).
11. Trayhurn, P. Hypoxia and adipose tissue function and dysfunction in obesity. *Physiol Rev* **93**, 1–21 (2013).
12. Cawthorn, W. P., Scheller, E. L. & MacDougald, O. A. Adipose tissue stem cells meet preadipocyte commitment: going back to the future. *J Lipid Res* **53**, 227–246 (2012).
13. Zhou, Y., Li, H. & Xia, N. The Interplay Between Adipose Tissue and Vasculature: Role of Oxidative Stress in Obesity. *Front Cardiovasc Med* **8**, 650214 (2021).
14. Evans, B. A., Merlin, J., Bengtsson, T. & Hutchinson, D. S. Adrenoceptors in white, brown, and brite adipocytes. *Br J Pharmacol* **176**, 2416–2432 (2019).
15. Kawai, T., Autieri, M. V. & Scalia, R. Adipose tissue inflammation and metabolic dysfunction in obesity. *Am J Physiol Cell Physiol* **320**, C375–C391 (2021).
16. Varghese, M., Song, J. & Singer, K. Age and Sex: impact on adipose tissue metabolism and inflammation. *Mech Ageing Dev* **199**, 111563 (2021).
17. Fuster, J. J. *et al.* Noncanonical Wnt signaling promotes obesity-induced adipose tissue inflammation and metabolic dysfunction independent of adipose tissue expansion. *Diabetes* **64**, 1235–1248 (2015).
18. Gupta, O. T. & Gupta, R. K. Visceral Adipose Tissue Mesothelial Cells: Living on the Edge or Just Taking Up Space? *Trends Endocrinol Metab* **26**, 515–523 (2015).
19. Ackers, I. & Malgor, R. Interrelationship of canonical and non-canonical Wnt signalling pathways in chronic metabolic diseases. *Diab Vasc Dis Res* **15**, 3–13 (2018).
20. Chen, N. & Wang, J. Wnt/β-Catenin Signaling and Obesity. *Front Physiol* **9**, 792 (2018).
21. Park, H.-K. & Ahima, R. S. Leptin signaling. *F1000Prime Rep* **6**, 73 (2014).
22. Picó, C., Palou, M., Pomar, C. A., Rodríguez, A. M. & Palou, A. Leptin as a key regulator of the adipose organ. *Rev Endocr Metab Disord* **23**, 13–30 (2022).
23. Baldini, G. & Phelan, K. D. The melanocortin pathway and control of appetite-progress and therapeutic implications. *J Endocrinol* **241**, R1–R33 (2019).
24. Kwon, H. & Pessin, J. E. Adipokines mediate inflammation and insulin resistance. *Front Endocrinol (Lausanne)* **4**, 71 (2013).
25. Stinkens, R., Goossens, G. H., Jocken, J. W. E. & Blaak, E. E. Targeting fatty acid metabolism to improve glucose metabolism. *Obes Rev* **16**, 715–757 (2015).
26. Haczeyni, F., Bell-Anderson, K. S. & Farrell, G. C. Causes and mechanisms of adipocyte enlargement and adipose expansion. *Obes Rev* **19**, 406–420 (2018).
27. Drolet, R. *et al.* Hypertrophy and hyperplasia of abdominal adipose tissues in women. *Int J Obes (Lond)* **32**, 283–291 (2008).
28. White, U. Adipose tissue expansion in obesity, health, and disease. *Front Cell Dev Biol* **11**, 1188844 (2023).

29. Lefterova, M. I., Haakonsson, A. K., Lazar, M. A. & Mandrup, S. PPARγ and the global map of adipogenesis and beyond. *Trends Endocrinol Metab* **25**, 293–302 (2014).
30. de Winter, T. J. J. & Nusse, R. Running Against the Wnt: How Wnt/β-Catenin Suppresses Adipogenesis. *Front Cell Dev Biol* **9**, 627429 (2021).
31. Li, S.-N. & Wu, J.-F. TGF-β/SMAD signaling regulation of mesenchymal stem cells in adipocyte commitment. *Stem Cell Res Ther* **11**, 41 (2020).
32. Eguchi, A. & Feldstein, A. E. Adipocyte cell death, fatty liver disease and associated metabolic disorders. *Dig Dis* **32**, 579–585 (2014).
33. Czech, M. P., Tencerova, M., Pedersen, D. J. & Aouadi, M. Insulin signalling mechanisms for triacylglycerol storage. *Diabetologia* **56**, 949–964 (2013).
34. Song, Z., Xiaoli, A. M. & Yang, F. Regulation and Metabolic Significance of De Novo Lipogenesis in Adipose Tissues. *Nutrients* **10**, 1383 (2018).
35. Veliova, M., Petcherski, A., Liesa, M. & Shirihai, O. S. The biology of lipid droplet-bound mitochondria. *Semin Cell Dev Biol* **108**, 55–64 (2020).
36. Rowland, L. A. *et al.* De novo lipogenesis fuels adipocyte autophagosome and lysosome membrane dynamics. *Nat Commun* **14**, 1362 (2023).
37. Yang, A. & Mottillo, E. P. Adipocyte lipolysis: from molecular mechanisms of regulation to disease and therapeutics. *Biochem J* **477**, 985–1008 (2020).
38. Bartesaghi, S. *et al.* Thermogenic activity of UCP1 in human white fat-derived beige adipocytes. *Mol Endocrinol* **29**, 130–139 (2015).
39. Chouchani, E. T., Kazak, L. & Spiegelman, B. M. New Advances in Adaptive Thermogenesis: UCP1 and Beyond. *Cell Metab* **29**, 27–37 (2019).
40. Bertholet, A. M. *et al.* Mitochondrial uncouplers induce proton leak by activating AAC and UCP1. *Nature* **606**, 180–187 (2022).
41. Moreno-Indias, I. & Tinahones, F. J. Impaired adipose tissue expandability and lipogenic capacities as ones of the main causes of metabolic disorders. *J Diabetes Res* **2015**, 970375 (2015).
42. Wu, H. & Ballantyne, C. M. Skeletal muscle inflammation and insulin resistance in obesity. *J Clin Invest* **127**, 43–54 (2017).
43. Williams, M. J., Hunter, G. R., Kekes-Szabo, T., Snyder, S. & Treuth, M. S. Regional fat distribution in women and risk of cardiovascular disease. *Am J Clin Nutr* **65**, 855–860 (1997).
44. Van Harmelen, V., Röhrig, K. & Hauner, H. Comparison of proliferation and differentiation capacity of human adipocyte precursor cells from the omental and subcutaneous adipose tissue depot of obese subjects. *Metabolism* **53**, 632–637 (2004).
45. Tchkonia, T. *et al.* Abundance of two human preadipocyte subtypes with distinct capacities for replication, adipogenesis, and apoptosis varies among fat depots. *Am J Physiol Endocrinol Metab* **288**, E267-277 (2005).
46. Ye, M. *et al.* Anthropometric changes and risk of diabetes: are there sex differences? A longitudinal study of Alberta's Tomorrow Project. *BMJ Open* **9**, e023829 (2019).
47. Pellegrinelli, V., Carobbio, S. & Vidal-Puig, A. Adipose tissue plasticity: how fat depots respond differently to pathophysiological cues. *Diabetologia* **59**, 1075–1088 (2016).
48. Fang, L., Guo, F., Zhou, L., Stahl, R. & Grams, J. The cell size and distribution of adipocytes from subcutaneous and visceral fat is associated with type 2 diabetes mellitus in humans. *Adipocyte* **4**, 273–279 (2015).
49. Verboven, K. *et al.* Abdominal subcutaneous and visceral adipocyte size, lipolysis and inflammation relate to insulin resistance in male obese humans. *Sci Rep* **8**, 4677 (2018).
50. West-Eberhard, M. J. Nutrition, the visceral immune system, and the evolutionary origins of pathogenic obesity. *Proc Natl Acad Sci U S A* **116**, 723–731 (2019).
51. Alvehus, M., Burén, J., Sjöström, M., Goedecke, J. & Olsson, T. The human visceral fat depot has a unique inflammatory profile. *Obesity (Silver Spring)* **18**, 879–883 (2010).
52. Item, F. & Konrad, D. Visceral fat and metabolic inflammation: the portal theory revisited. *Obes Rev* **13 Suppl 2**, 30–39 (2012).
53. Lemieux, I. *et al.* Total cholesterol/HDL cholesterol ratio vs LDL cholesterol/HDL cholesterol ratio as indices of ischemic heart disease risk in men: the Quebec Cardiovascular Study. *Arch Intern Med* **161**, 2685–2692 (2001).
54. Després, J.-P. Abdominal obesity and cardiovascular disease: is inflammation the missing link? *Can J Cardiol* **28**, 642–652 (2012).

55. Samaras, K., Botelho, N. K., Chisholm, D. J. & Lord, R. V. Subcutaneous and visceral adipose tissue gene expression of serum adipokines that predict type 2 diabetes. *Obesity (Silver Spring)* **18**, 884–889 (2010).
56. Kabir, M. *et al.* Molecular evidence supporting the portal theory: a causative link between visceral adiposity and hepatic insulin resistance. *Am J Physiol Endocrinol Metab* **288**, E454-461 (2005).
57. Mathieu, P., Poirier, P., Pibarot, P., Lemieux, I. & Després, J.-P. Visceral obesity: the link among inflammation, hypertension, and cardiovascular disease. *Hypertension* **53**, 577–584 (2009).
58. Lin, X. & Li, H. Obesity: Epidemiology, Pathophysiology, and Therapeutics. *Front Endocrinol (Lausanne)* **12**, 706978 (2021).
59. Safaei, M., Sundararajan, E. A., Driss, M., Boulila, W. & Shapi'i, A. A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity. *Comput Biol Med* **136**, 104754 (2021).
60. Hales, C. M., Carroll, M. D., Fryar, C. D. & Ogden, C. L. Prevalence of Obesity and Severe Obesity Among Adults: United States, 2017-2018. *NCHS Data Brief* 1–8 (2020).
61. Bhaskaran, K., Dos-Santos-Silva, I., Leon, D. A., Douglas, I. J. & Smeeth, L. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3·6 million adults in the UK. *Lancet Diabetes Endocrinol* **6**, 944–953 (2018).
62. Khan, S. S. *et al.* Association of Body Mass Index With Lifetime Risk of Cardiovascular Disease and Compression of Morbidity. *JAMA Cardiol* **3**, 280–287 (2018).
63. Zhao, Y. *et al.* Association of BMI with cardiovascular disease incidence and mortality in patients with type 2 diabetes mellitus: A systematic review and dose-response meta-analysis of cohort studies. *Nutr Metab Cardiovasc Dis* **31**, 1976–1984 (2021).
64. Guh, D. P. *et al.* The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC Public Health* **9**, 88 (2009).
65. Wang, X. *et al.* The relationship between body mass index and stroke: a systemic review and meta-analysis. *J Neurol* **269**, 6279–6289 (2022).
66. Jarvis, H. *et al.* Metabolic risk factors and incident advanced liver disease in non-alcoholic fatty liver disease (NAFLD): A systematic review and meta-analysis of population-based observational studies. *PLoS Med* **17**, e1003100 (2020).
67. Liu, P.-H. *et al.* Association of Obesity With Risk of Early-Onset Colorectal Cancer Among Women. *JAMA Oncol* **5**, 37–44 (2019).
68. Renehan, A. G., Tyson, M., Egger, M., Heller, R. F. & Zwahlen, M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* **371**, 569–578 (2008).
69. Mulki, R. *et al.* Body mass index greater than 35 is associated with severe Clostridium difficile infection. *Aliment Pharmacol Ther* **45**, 75–81 (2017).
70. Atamna, A., Elis, A., Gilady, E., Gitter-Azulay, L. & Bishara, J. How obesity impacts outcomes of infectious diseases. *Eur J Clin Microbiol Infect Dis* **36**, 585–591 (2017).
71. Twig, G. *et al.* Body mass index and infectious disease mortality in midlife in a cohort of 2.3 million adolescents. *Int J Obes (Lond)* **42**, 801–807 (2018).
72. Poznyak, A. *et al.* The Diabetes Mellitus-Atherosclerosis Connection: The Role of Lipid and Glucose Metabolism and Chronic Inflammation. *Int J Mol Sci* **21**, 1835 (2020).
73. Galicia-Garcia, U. *et al.* Pathophysiology of Type 2 Diabetes Mellitus. *Int J Mol Sci* **21**, 6275 (2020).
74. Pouwels, S. *et al.* Non-alcoholic fatty liver disease (NAFLD): a review of pathophysiology, clinical management and effects of weight loss. *BMC Endocr Disord* **22**, 63 (2022).
75. Wang, Z. & Nakayama, T. Inflammation, a link between obesity and cardiovascular disease. *Mediators Inflamm* **2010**, 535918 (2010).
76. Singh, A., Mayengbam, S. S., Yaduvanshi, H., Wani, M. R. & Bhat, M. K. Obesity Programs Macrophages to Support Cancer Progression. *Cancer Res* **82**, 4303–4312 (2022).
77. Blüher, M. Metabolically Healthy Obesity. *Endocr Rev* **41**, bnaa004 (2020).
78. Shi, T. H., Wang, B. & Natarajan, S. The Influence of Metabolic Syndrome in Predicting Mortality Risk Among US Adults: Importance of Metabolic Syndrome Even in Adults With Normal Weight. *Prev Chronic Dis* **17**, E36 (2020).
79. Leening, M. J. G. *et al.* Sex differences in lifetime risk and first manifestation of cardiovascular disease: prospective population based cohort study. *BMJ* **349**, g5992 (2014).

80. Huebschmann, A. G. *et al.* Sex differences in the burden of type 2 diabetes and cardiovascular risk across the life course. *Diabetologia* **62**, 1761–1772 (2019).
81. El Khoudary, S. R. *et al.* Menopause Transition and Cardiovascular Disease Risk: Implications for Timing of Early Prevention: A Scientific Statement From the American Heart Association. *Circulation* **142**, e506–e532 (2020).
82. Shah, U. A. *et al.* Imaging modalities for measuring body composition in patients with cancer: opportunities and challenges. *J Natl Cancer Inst Monogr* **2023**, 56–67 (2023).
83. Agrawal, S. *et al.* Inherited basis of visceral, abdominal subcutaneous and gluteofemoral fat depots. *Nat Commun* **13**, 3771 (2022).
84. Barreira, T. V. *et al.* Anthropometric correlates of total body fat, abdominal adiposity, and cardiovascular disease risk factors in a biracial sample of men and women. *Mayo Clin Proc* **87**, 452–460 (2012).
85. Varghese, B. *et al.* Visceral adiposity in young patients with coronary artery disease-a case control study. *Indian Heart J* **64**, 284–289 (2012).
86. Lee, V. *et al.* Estimation of visceral fat in 9- to 13-year-old girls using dual-energy X-ray absorptiometry (DXA) and anthropometry. *Obes Sci Pract* **4**, 437–447 (2018).
87. Vatanparast, H. *et al.* DXA-derived abdominal fat mass, waist circumference, and blood lipids in postmenopausal women. *Obesity (Silver Spring)* **17**, 1635–1640 (2009).
88. Froelich, M. F. *et al.* Measurement of total and visceral fat mass in young adult women: a comparison of MRI with anthropometric measurements with and without bioelectrical impedance analysis. *Br J Radiol* **93**, 20190874 (2020).
89. Yang, H. I. *et al.* A new anthropometric index to predict percent body fat in young adults. *Public Health Nutr* **23**, 1507–1514 (2020).
90. Frank, A. P., de Souza Santos, R., Palmer, B. F. & Clegg, D. J. Determinants of body fat distribution in humans may provide insight about obesity-related health risks. *J Lipid Res* **60**, 1710–1719 (2019).
91. Kesztyüs, D., Erhardt, J., Schönsteiner, D. & Kesztyüs, T. Therapeutic Treatment for Abdominal Obesity in Adults. *Dtsch Arztebl Int* **115**, 487–493 (2018).
92. Samms, R. J., Sloop, K. W., Gribble, F. M., Reimann, F. & Adriaenssens, A. E. GIPR Function in the Central Nervous System: Implications and Novel Perspectives for GIP-Based Therapies in Treating Metabolic Disorders. *Diabetes* **70**, 1938–1944 (2021).
93. Drucker, D. J. Mechanisms of Action and Therapeutic Application of Glucagon-like Peptide-1. *Cell Metab* **27**, 740–756 (2018).
94. Cao, Q. *et al.* Waist-hip ratio as a predictor of myocardial infarction risk: A systematic review and meta-analysis. *Medicine (Baltimore)* **97**, e11639 (2018).
95. de Koning, L., Merchant, A. T., Pogue, J. & Anand, S. S. Waist circumference and waist-to-hip ratio as predictors of cardiovascular events: meta-regression analysis of prospective studies. *Eur Heart J* **28**, 850–856 (2007).
96. Medina-Inojosa, J. R. *et al.* Relation of Waist-Hip Ratio to Long-Term Cardiovascular Events in Patients With Coronary Artery Disease. *Am J Cardiol* **121**, 903–909 (2018).
97. O'Donnell, M. J. *et al.* Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet* **376**, 112–123 (2010).
98. O'Donnell, M. J. *et al.* Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. *Lancet* **388**, 761–775 (2016).
99. Hu, G. *et al.* Body mass index, waist circumference, and waist-hip ratio on the risk of total and type-specific stroke. *Arch Intern Med* **167**, 1420–1427 (2007).
100. Namaganda, P., Nakibuuka, J., Kaddumukasa, M. & Katabira, E. Stroke in young adults, stroke types and risk factors: a case control study. *BMC Neurol* **22**, 335 (2022).
101. Jayedi, A. *et al.* Anthropometric and adiposity indicators and risk of type 2 diabetes: systematic review and dose-response meta-analysis of cohort studies. *BMJ* **376**, e067516 (2022).
102. Dale, C. E. *et al.* Causal Associations of Adiposity and Body Fat Distribution With Coronary Heart Disease, Stroke Subtypes, and Type 2 Diabetes Mellitus: A Mendelian Randomization Analysis. *Circulation* **135**, 2373–2388 (2017).
103. Jackson, M., Marks, L., May, G. H. W. & Wilson, J. B. The genetic basis of disease. *Essays Biochem* **62**, 643–723 (2018).

104. Lappalainen, T. & MacArthur, D. G. From variant to function in human disease genetics. *Science* **373**, 1464–1468 (2021).
105. Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45–54 (2019).
106. Signor, S. A. & Nuzhdin, S. V. The Evolution of Gene Expression in cis and trans. *Trends Genet* **34**, 532–544 (2018).
107. Murphy, R., Carroll, R. W. & Krebs, J. D. Pathogenesis of the metabolic syndrome: insights from monogenic disorders. *Mediators Inflamm* **2013**, 920214 (2013).
108. Gayon, J. From Mendel to epigenetics: History of genetics. *C R Biol* **339**, 225–230 (2016).
109. Elhawary, N. A. *et al.* Genetic etiology and clinical challenges of phenylketonuria. *Hum Genomics* **16**, 22 (2022).
110. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).
111. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. *Hum Mol Genet* **27**, 3641–3649 (2018).
112. Wright, C. F. *et al.* Assessing the Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing Variants in a Population Setting. *Am J Hum Genet* **104**, 275–286 (2019).
113. Tang, M., Wang, T. & Zhang, X. A review of SNP heritability estimation methods. *Brief Bioinform* **23**, bbac067 (2022).
114. Marian, A. J. Elements of ‘missing heritability’. *Curr Opin Cardiol* **27**, 197–201 (2012).
115. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
116. Wang, T. *et al.* The Human Pangenome Project: a global resource to map genomic diversity. *Nature* **604**, 437–446 (2022).
117. Collins, A. Allelic association: linkage disequilibrium structure and gene mapping. *Mol Biotechnol* **41**, 83–89 (2009).
118. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype Imputation. *Annu Rev Genomics Hum Genet* **10**, 387–406 (2009).
119. Hong, M. *et al.* RNA sequencing: new technologies and applications in cancer research. *J Hematol Oncol* **13**, 166 (2020).
120. Zhang, J. & Zhao, H. eQTL studies: from bulk tissues to single cells. *J Genet Genomics* S1673-8527(23)00113–3 (2023) doi:10.1016/j.jgg.2023.05.003.
121. Chepelev, I., Wei, G., Wangsa, D., Tang, Q. & Zhao, K. Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res* **22**, 490–503 (2012).
122. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
123. Laverré, A., Tannier, E. & Necsulea, A. Long-range promoter-enhancer contacts are conserved during evolution and contribute to gene expression robustness. *Genome Res* **32**, 280–296 (2022).
124. Cannon, M. E. & Mohlke, K. L. Deciphering the Emerging Complexities of Molecular Mechanisms at GWAS Loci. *Am J Hum Genet* **103**, 637–653 (2018).
125. Xu, C. *et al.* Estimating Genome-Wide Significance for Whole-Genome Sequencing Studies. *Genet Epidemiol* **38**, 281–290 (2014).
126. Chen, Z., Boehnke, M., Wen, X. & Mukherjee, B. Revisiting the genome-wide significance threshold for common variant GWAS. *G3 (Bethesda)* **11**, jkaa056 (2021).
127. Peterson, R. E. *et al.* Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell* **179**, 589–603 (2019).
128. Barroso, I. The importance of increasing population diversity in genetic studies of type 2 diabetes and related glycaemic traits. *Diabetologia* **64**, 2653–2664 (2021).
129. Fedorova, L. *et al.* Analysis of Common SNPs across Continents Reveals Major Genomic Differences between Human Populations. *Genes (Basel)* **13**, 1472 (2022).
130. Cohen, J. *et al.* Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet* **37**, 161–165 (2005).
131. Hunter, D. J. Gene-environment interactions in human diseases. *Nat Rev Genet* **6**, 287–298 (2005).
132. Bowden, J. & Holmes, M. V. Meta-analysis and Mendelian randomization: A review. *Res Synth Methods* **10**, 486–496 (2019).

133. Sanderson, E. *et al.* Mendelian randomization. *Nat Rev Methods Primers* **2**, 6 (2022).
134. Bouchard, C. Genetics of Obesity: What We Have Learned Over Decades of Research. *Obesity (Silver Spring)* **29**, 802–820 (2021).
135. Farooqi, I. S. Monogenic human obesity syndromes. *Handb Clin Neurol* **181**, 301–310 (2021).
136. Loos, R. J. F. & Yeo, G. S. H. The genetics of obesity: from discovery to biology. *Nat Rev Genet* **23**, 120–133 (2022).
137. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
138. Hansen, G. T. *et al.* Genetics of sexually dimorphic adipose distribution in humans. *Nat Genet* **55**, 461–470 (2023).
139. Ndiaye, F. K. *et al.* The expression of genes in top obesity-associated loci is enriched in insula and substantia nigra brain regions involved in addiction and reward. *Int J Obes (Lond)* **44**, 539–543 (2020).
140. Corbin, L. J. *et al.* BMI as a Modifiable Risk Factor for Type 2 Diabetes: Refining and Understanding Causal Estimates Using Mendelian Randomization. *Diabetes* **65**, 3002–3007 (2016).
141. Xu, H., Jin, C. & Guan, Q. Causal Effects of Overall and Abdominal Obesity on Insulin Resistance and the Risk of Type 2 Diabetes Mellitus: A Two-Sample Mendelian Randomization Study. *Front Genet* **11**, 603 (2020).
142. Miao, Z. *et al.* The causal effect of obesity on prediabetes and insulin resistance reveals the important role of adipose tissue in insulin resistance. *PLoS Genet* **16**, e1009018 (2020).
143. Cheng, L. *et al.* Exposing the Causal Effect of Body Mass Index on the Risk of Type 2 Diabetes Mellitus: A Mendelian Randomization Study. *Front Genet* **10**, 94 (2019).
144. Suzuki, S. *et al.* Body mass index and colorectal cancer risk: A Mendelian randomization study. *Cancer Sci* **112**, 1579–1588 (2021).
145. Butler-Laporte, G., Harroud, A., Forgetta, V. & Richards, J. B. Elevated body mass index is associated with an increased risk of infectious disease admissions and mortality: a mendelian randomization study. *Clin Microbiol Infect* S1198-743X(20)30356–6 (2020) doi:10.1016/j.cmi.2020.06.014.
146. Pressley, H., Cornelio, C. K. & Adams, E. N. Setmelanotide: A Novel Targeted Treatment for Monogenic Obesity. *J Pharm Technol* **38**, 368–373 (2022).
147. Schleinitz, D., Böttcher, Y., Blüher, M. & Kovacs, P. The genetics of fat distribution. *Diabetologia* **57**, 1276–1286 (2014).
148. Lightbourne, M. & Brown, R. J. Genetics of Lipodystrophy. *Endocrinol Metab Clin North Am* **46**, 539–554 (2017).
149. Lindgren, C. M. *et al.* Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS Genet* **5**, e1000508 (2009).
150. Heid, I. M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* **42**, 949–960 (2010).
151. Randall, J. C. *et al.* Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet* **9**, e1003500 (2013).
152. Berndt, S. I. *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* **45**, 501–512 (2013).
153. Winkler, T. W. *et al.* The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet* **11**, e1005378 (2015).
154. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
155. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* **28**, 166–174 (2019).
156. Pulit, S. L., Karaderi, T. & Lindgren, C. M. Sexual dimorphisms in genetic loci linked to body fat distribution. *Biosci Rep* **37**, BSR20160184 (2017).
157. Lv, W.-Q. *et al.* Genetically driven adiposity traits increase the risk of coronary artery disease independent of blood pressure, dyslipidaemia, glycaemic traits. *Eur J Hum Genet* **26**, 1547–1553 (2018).
158. Liu, Z. *et al.* Causal relationships between NAFLD, T2D and obesity have implications for disease subphenotyping. *J Hepatol* **73**, 263–276 (2020).

159. Lu, Y. *et al.* Assessment of causal effects of visceral adipose tissue on risk of cancers: a Mendelian randomization study. *Int J Epidemiol* **51**, 1204–1218 (2022).
160. Maina, J. G. *et al.* Abdominal obesity is a more important causal risk factor for pancreatic cancer than overall obesity. *Eur J Hum Genet* **31**, 962–966 (2023).
161. Small, K. S. *et al.* Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nat Genet* **50**, 572–580 (2018).
162. Yang, Q. *et al.* Adipocyte-Specific Modulation of KLF14 Expression in Mice Leads to Sex-Dependent Impacts on Adiposity and Lipid Metabolism. *Diabetes* **71**, 677–693 (2022).
163. Loh, N. Y. *et al.* RSPO3 impacts body fat distribution and regulates adipose cell biology in vitro. *Nat Commun* **11**, 2797 (2020).
164. Loh, N. Y. *et al.* LRP5 regulates human body fat distribution by modulating adipose progenitor biology in a dose- and depot-specific fashion. *Cell Metab* **21**, 262–273 (2015).
165. Gesta, S. *et al.* Mesodermal developmental gene Tbx15 impairs adipocyte differentiation and mitochondrial respiration. *Proc Natl Acad Sci U S A* **108**, 2771–2776 (2011).
166. Gesta, S. *et al.* Evidence for a role of developmental genes in the origin of obesity and body fat distribution. *Proc Natl Acad Sci U S A* **103**, 6676–6681 (2006).
167. Lee, K. Y. *et al.* Shox2 is a molecular determinant of depot-specific adipocyte function. *Proc Natl Acad Sci U S A* **110**, 11409–11414 (2013).
168. Civelek, M. *et al.* Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *Am J Hum Genet* **100**, 428–443 (2017).
169. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from dbGaP Accession phs000424.v8.p2on 10/01/2020.
170. Franzén, O. *et al.* Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**, 827–830 (2016).
171. Hildreth, A. D. *et al.* Single-cell sequencing of human white adipose tissue identifies new cell states in health and obesity. *Nat Immunol* **22**, 639–653 (2021).
172. Sun, W. *et al.* snRNA-seq reveals a subpopulation of adipocytes that regulates thermogenesis. *Nature* **587**, 98–102 (2020).
173. Vijay, J. *et al.* Single-cell analysis of human adipose tissue identifies depot and disease specific cell types. *Nat Metab* **2**, 97–109 (2020).
174. Sárvári, A. K. *et al.* Plasticity of Epididymal Adipose Tissue in Response to Diet-Induced Obesity at Single-Nucleus Resolution. *Cell Metab* **33**, 437-453.e5 (2021).
175. Hepler, C. *et al.* Identification of functionally distinct fibro-inflammatory and adipogenic stromal subpopulations in visceral adipose tissue of adult mice. *Elife* **7**, e39636 (2018).
176. Almanzar, N. *et al.* A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020).
177. Min, S. Y. *et al.* Diverse repertoire of human adipocyte subtypes develops from transcriptionally distinct mesenchymal progenitor cells. *Proc Natl Acad Sci U S A* **116**, 17970–17979 (2019).
178. Emont, M. P. *et al.* A single-cell atlas of human and mouse white adipose tissue. *Nature* **603**, 926–933 (2022).
179. Deutsch, A., Feng, D., Pessin, J. E. & Shinoda, K. The Impact of Single-Cell Genomics on Adipose Tissue Research. *Int J Mol Sci* **21**, 4773 (2020).

Chapter 2: Identifying and prioritizing putative causal fat distribution genes using network construction and analysis

Parts of this chapter were adapted from a manuscript, originally published on bioRxiv, (023.09.06.556534; doi: https://doi.org/10.1101/2023.09.06.556534). The authors are listed below.

Systems genetics analysis of human body fat distribution genes identifies Wnt signaling and mitochondrial activity in adipocytes

Jordan N Reed[1,2,] Jiansheng Huang[3], Yong Li[3], Lijiang Ma[4], Dhanush Banka[1],Martin Wabitsch[5], Tianfang Wang[3], Wen Ding[3], Johan L.M. Björkegren[4,6], Mete Civelek[1,2]

1. Department of Biomedical Engineering, University of Virginia. Charlottesville, Virginia.
2. Center for Public Health Genomics, University of Virginia. Charlottesville, Virginia.
3. Novo Nordisk Research Center China, Novo Nordisk A/S, Beijing, China.
4. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai. New York, New York.
5. Division of Paediatric Endocrinology and Diabetes, Department of Paediatrics and Adolescent Medicine, Ulm University Medical Centre, Ulm, Germany.
6. Department of Medicine, Karolinska Institutet, Huddinge, Stockholm, Sweden.

2.1 Introduction:

2.1.1 Biological Networks-

Much of the natural world is organized into systems of individual parts that work together in sometimes random, often organized ways[1]. From making friends and networking with colleagues, to how proteins bind and interact, natural processes usually follow similar patterns that we can describe with biological networks[2]. Networks are a graphical and mathematical way to describe the interactions between parts of the system[1]. By studying the parts as a whole, we can predict and explain phenomena that only occur when the parts work together and identify which parts are critical for the system to function[1,3].

2.1.1.1 Definitions:

Networks are made up of nodes, each representing one part of the system[1], such as a person or a protein (Figure 2.1 A). The nodes are connected by edges that describe an interaction between those nodes[1], like a link on a social media app or a protein binding event. Edges can be weighted, or have different strengths depending on some criteria, like affinity of the protein-protein interaction. These edges are bidirectional in that both nodes participate in the interaction. Edges of a network can also be directed, in which the parent node is causal to the child node. For example, phosphorylation of protein A causes degradation of protein B. We refer to the series of edges between two nodes as the path. Often, the network structure, or the arrangement of nodes and edges, can be described by some mathematical model[1,4].

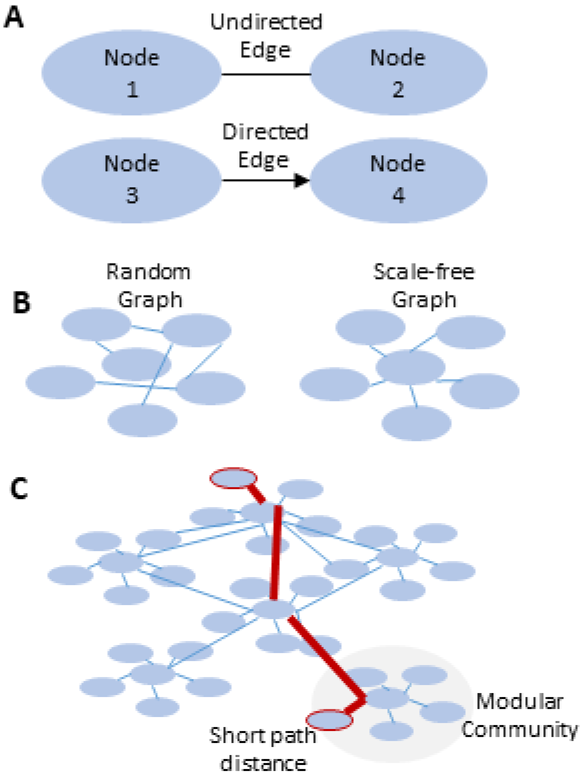

Figure 2.1: Basic definitions and properties of biological networks

- *(A) Node 1 and 2 are connected by an undirected edge, while nodes 3 and 4 are connected by a directed edge.*
- *(B) Compared to a random graph, biological networks have few nodes with many connections, and many nodes with few connections*
- *(C) This creates a modular structure, where each node is usually associated with one hub. These hubs are interconnected such that the distance between any two nodes in the network is small.*

2.1.1.2 Properties of biological networks:

Most biological networks have a similar, conserved structure. We observe that these networks are organized into modular communities, have short path distances between nodes, and have a small number of well-connected nodes[1,3].

Scale-free networks have few nodes with many edges and many nodes with few edges, often described as a 'hub and spokes' model[1,3] (Figure 2.1 B). The number of edges per node, in scale-free networks, follows the power law $P(k) \sim k^{-a}$, where k represents the number of edges and P(k) represents the fraction of nodes with k edges[5]. a is called the degree exponent, determined by fitting the line log(P(k) ~ -a log(k). Most biological networks are scale-free, with a degree exponent between 2 and 3 (Ref. 5). This observed scale-free structure is thought to be evolutionarily driven by the genomic replication of structurally similar genes, leading to redundant network connections with shared hubs[1].

Small world networks are organized into modular communities of nodes, yet have short path distances between nodes[6–8] (Figure 2.1 C). The smallest number of edges crossed to connect two nodes in the network is called the shortest path[4,9]. Biological networks often have a shorter average shortest paths than networks with the same nodes and randomly assigned edges. The clustering coefficient of a network describes the extent to which nodes in the network are interconnected, calculated as the number of three-node triplets with edges between each node divided by the number of three-node triplets without all three edges[8]. All three-node combinations in the network are considered. Biological networks often have a higher clustering coefficient than networks with the same nodes and randomly assigned edges, [6].

Biological networks are clustered into communities that have stronger connections between group members than with non-group members[1–3]. The number of edges connecting to nodes in and out of the cluster can be used to determine the cluster's connectivity[2]. In biological networks, nodes are often grouped, with a small number of 'hub' nodes connecting the peripheral nodes[10]. However, the distance between any two nodes in the full network is small, because the hubs are highly interconnected[1,3]. This makes the biological networks robust to perturbation and node removal[1,3]; we see in networks and in mechanistic biology that a small subset of all expressed genes and proteins are necessary for system function[1].

2.1.1.3 Applications of Biological Networks:

Biological networks have incredible descriptive and predictive power[11,12]. By organizing large amounts of data into interpretable mathematical models, we can observe emergent properties of the system[1,3,13], predict nodes critical to these outcomes[10], and observe how these responses lead to disease progression[11,12,14,15]. For example, by organizing the proteins in a signaling pathway into a network, we might observe feedback loops or crosstalk between two pathways. These features would not be observed when considering one protein or pathway at a time[3]. Further, we might observe groups of nodes with a shared attribute, for example, a highly connected group of genes related to cell cycle. We might hypothesize that other highly connected genes with unknown function act in the cell cycle pathway, and that the cell cycle was an important pathway in the gene set. The network structure allows one to interrogate which nodes are critical for the system to function[10,16]. We can identify nodes with many connections or remove nodes to observe how the network compensates.

Different types of biological networks are best at describing different phenomena, such as protein interactions[17], signaling pathways, gene regulation[18,19], single-cell RNA[19], metabolic reactions[20,21], and more. Protein-protein interaction networks, like STRING[22] and Reactome[23], are often undirected networks where the edge weight is related to the strength of the interaction and can be used to infer the function of proteins based on binding partners. Signaling networks, like KEGG[24], are directed networks and often have nodes with phenotypic outcomes. They are often used to predict which pathways will be turned on or off in response to shared stimuli. Gene expression networks can be directed or undirected[25,26]; both are used to predict novel gene drivers of disease. Recently, network models have been applied to single-cell RNA-seq data, where they are used to predict cell-type specific gene regulation[27] and ligand-receptor interactions between cell types[28].

The human genome encodes more than 80,000 transcripts that encode an estimated 20,000 proteins[29]. It is estimated that we do not know the function of up to 6,000 genes, and we know even fewer tissue- and disease-specific functions. As further discussed in Chapter 3, GWAS predict that hundreds of genes are potentially involved in obesity and fat distribution[30,31], yet only a subset have been mechanistically linked to disease progression. It can be costly and time-consuming to perform experiments that assess gene function, but using networks, we can organize known and novel genes and prioritize essential network genes for future study.

2.1.2 Co-expression networks:

Co-expression networks are undirected networks made using the correlation structure within gene expression data, where edges represent scaled Pearson correlations between nodes, or genes[32,33]. By design, these are scale-free, modular networks with groups of highly correlated genes emerging[32]. We can then look at the function and disease relevance of the genes within modules to make inferences about unknown genes.

2.1.2.1 Theory and Construction:

Genes are often regulated in conserved groups, potential due to a signaling pathway or by a shared transcription factor. To identify the correlated and putatively co-regulated genes within a dataset, we first find the correlations between every gene in the dataset. Since gene expression data is noisy, we can amplify the true differences by raising all correlations to a power. The power is chosen such that the network structure follows the power law. Once we have the adjusted correlations, we identify groups of highly correlated genes using hierarchical clustering. The groups are called modules; and genes that are not highly expressed, variable, or correlated with other genes are not sorted into modules. The average module gene's expression per sample has low variation and has stronger correlations with genes within the module than without. We can also calculate the module eigengene per sample, which is the first principal component through the gene expression in that module, representing the axis of highest variation. We use the R package Weighted Gene Co-Expression Network Analysis (WGCNA) to construct and analyze these networks.

The Python package iterativeWGCNA improves upon the WGCNA construction methods by iteratively reassigning genes to new modules of best fit. First, the connectivity of genes in the module to the eigengene is assessed, and genes that do not strongly fit with the eigengene are removed. Genes not in modules are re-run through the module detection pipeline iteratively until no new modules are found and similar modules are merged. iterativeWGCNA ensures that all genes in the module are strongly correlated to the module eigengene and, therefore, to each other. iterativeWGCNA assigns more genes into modules than does WGCNA, and those modules are small and more numerous.

Once we identify the communities of highly correlated genes, called modules, we next try to find the module's function and disease relevance. Using annotated pathways, such as gene ontology (GO), molecular signature database (msigDB), or KEGG[24], we can test if there is significant enrichment for any particular pathway within the module of interest. We perform a Fisher's exact test to see if the module contains more than expected pathway genes, compared to the whole gene set. Using phenotypic data collected in the same samples as the gene expression data, we can look at module gene correlations with disease indicators, biomarkers, and phenotypic traits. We calculate the Pearson correlations between the phenotypes and the module average gene expression or the module eigengene. Using Fisher's exact test, we can also test whether the module is enriched for candidate GWAS genes for the disease or trait of interest. The success of linking module gene expression to disease depends on using gene expression data from a disease-relevant cell line or tissue. Once we identify these modules of interest, we can further investigate the genes contained within.

2.1.2.2 Applications in gene expression data:

Co-expression networks are often used to predict the function of unknown genes in the tissue of interest. By annotating the module with enriched pathways, we can infer that other genes in the module may be part of those pathways[34–37], and by using unbiased construction methods, we can ensure that disease-relevant but unknown function genes will be considered[38]. We can also identify hub genes, which are the most highly connected genes in the module, and are, therefore, putatively more biologically relevant[39–44]. This approach has had success in WGCNA modules but might have limited applicability in iterativeWGCNA because the module genes have highly similar patterns of expression. By comparing gene expression in two conditions, we can make inferences about diseases[45] or identify conservation between datasets[46,47]. Finally, module genes are expressed, correlated, and putatively co-regulated in the gene set of interest[16], which may provide clues to the normal or disease function of a gene in that tissue.

2.1.3 Bayesian Networks:

Bayesian networks are directed acyclic graphs made using the underlying structure of gene expression data. According to Bayes' Theorem, the edges represent putative causal relationships between two nodes, or genes, i.e. expression of gene A → expression of gene B. Bayesian networks in biology often follow scale-free and small-world properties, but the construction is unbiased and must be assessed. These networks are powerful predictive tools; because the edges imply causality between genes, we can infer which genes may be regulatory of many others. By annotating the network with disease-relevant genes, we can identify putative regulators of disease.

2.1.3.1 Theory and Construction:

Genes expression is a tightly regulated process, where the expression of specific regulatory genes, like transcription factors and signaling molecules, controls the expression of downstream programs of genes[48]. To capture the causal relationships between the expression of many genes and identify key regulators of these processes, we must employ networks with directed edges[49].

A Bayesian network is a type of directed acyclic graph (DAG), where, based on Bayes' theorem, the probability that a node in the network has a certain expression value depends on the expression of its parent nodes, or the nodes upstream of it[26,49]. For a gene A, with parent genes B and C, the probability that A has a particular expression value depends on the expression state of B and C:

$$p(A) = p(A|B,C) = p(A|B) \cdot p(A|C),$$

where $p(A|parent)$ is the conditional probability between the two nodes[26,49,50]. These conditional probabilities are the parameters of the network, which might be known or experimentally determined, or can be learned from data. Graphs made from discrete data uses contingency tables to represent the conditional probabilities, while continuous data forms a probability distribution[50].

Each edge in the network represents one of these conditional relationships between nodes. The total structure of the graph describes all relationships between genes as a joint probability. The joint probability of a full graph, X, is the geometric sum of all individual node probabilities.

$$p(\boldsymbol{X}) = \prod p(X_i \,|D_i)$$

where X is the full graph, Xi is a node in the graph X, and Di is the set of parents for node Xi[26,49,50].

In some applications of Bayesian networks, the graph structure is known, and we want to learn a set of parameters, for example, the conditional probabilities of each node[50–52]. This can help make decisions about which outcomes will be observed in which scenarios. When using gene expression data, we know the probability distribution of expression values for each gene across all samples in the dataset. Instead, we want to learn the graph structure that best fits the observed data[26,53].

Popular score-based methods to learn the relationships within the graph generally try to maximize a likelihood function, which increases the ability of the graph's joint probability to describe the observed gene expression data[53]. Since adding more edges will almost always increases the likelihood of the new graph structure by overfitting the sample data, most scoring metrics also have a term to penalize each new edge added. The Bayesian Information Criterion is commonly used as a scoring metric:

$$BIC = k \cdot \ln(n) - 2 \cdot \ln(L)$$

where n is the number of nodes, k is the number of parameters estimated (edges), and L is the likelihood function. Maximizing this function results in the best-fit graph structure that is not overly complex[54]. Algorithms may be able to identify the global optimal solution for a network made up of a small number of nodes, but Bayesian network construction is an N-P hard problem that scales with each node added[26]. Thus, heuristic search algorithms and other computational tools must be employed to identify maxima. Greedy hill-climbing algorithms pursue the immediate optimal solution[55] and Markov blankets can reduce the parent-gene search space[49]. We also construct many iterations of the same network and merge the results to avoid local maxima[56].

Further, multiple network structures can result in the same likelihood when there is not enough evidence in the gene expression data to determine which of two related genes are parent and child. To break the equivalency, we add prior information to improve predictions, such as known direct connections or eQTL data[16,57]. In theory, genes that are regulated by SNPs on the chromatin might be more likely to be upstream of and turn on or off other genes; therefore, we denote the gene with the eQTL as the parent in these cases. Early studies in Bayesian network reproducibility showed that adding genetic information improved precision and recall[58].

Bayesian network construction is a time- and computationally-intensive process, and few popular construction tools can handle networks with more than 100 nodes. R packages *bnlearn* and *deal* construct reproducible networks on small gene expression datasets[59]. However, to investigate the relationships between thousands of genes at once, we require the Reconstructing Integrative Molecular Bayesian Network (RIMBANET) tool[56], which is able to handle up to 10,000 nodes (Table 2.1) by discretizing the gene expression data to reduce computational complexity. RIMBANET also takes in diverse prior information, including direct connections, eQTL data, and continuous gene expression data, to improve the predictive power of the network.

2.1.3.2 Applications in gene expression data:

Large-scale Bayesian networks were first applied to yeast gene expression data[60,56,58] and have been applied to gene expression data from model organisms[61–64]. In human data, Bayesian networks were used to predict novel disease regulators[47,65–71]. Many studies went on to experimentally validate that the predicted gene affected a relevant disease process in cells[16,72–79]. Bayesian networks can even be used to further interrogate the module genes identified in co-expression networks[65].

2.1.3.3 Key driver analysis:

Because of the causal structure of the network, we can more easily identify the putative regulatory genes, called key driver genes[12,80]. Depending on the goals of the analysis, we may be interested in finding key driver genes that regulate many others in the gene set[16], or we might want to find regulators of disease-related genes[81]. Studies show that, compared to other network features, key drivers are more likely to be replicated across samples and robust to sub-sampling[82].

An unbiased approach of key driver identification uses the network structure to prioritize genes that are a parent, and therefore putatively regulatory, of many others in the network[83] (Figure 2.2 A). These key drivers are putatively regulatory of normal tissue function and may be involved in disease processes. To identify these key drivers, we define a set of downstream genes as the genes within a certain shortest path distance from the tested gene. We term the distances n-hops. For a given n-hop, we determine whether the number of downstream genes is an extreme value for this dataset- whether it is more than one standard deviation from the mean of all network genes[83]. Some applications of this method use one distance, while some integrate multiple n-hops into a score for each gene. In each dataset, the top scoring or most extreme values, usually top the 10%, are called key driver genes, and are prioritized for further study[83].

Disease-related key driver genes are putatively regulatory of disease-causing genes, and are of particular interest[81] (Figure 2.2 B). Given a list of disease-related genes, such as candidate GWAS genes, we can perform a Fisher's exact test to determine if the downstream gene set of a gene in the network is enriched for the disease-related genes, compared to the entire network. The downstream gene set can be defined as the genes within a certain shortest path distance from the tested gene. Mergeomics implements this method for some path distances[81].

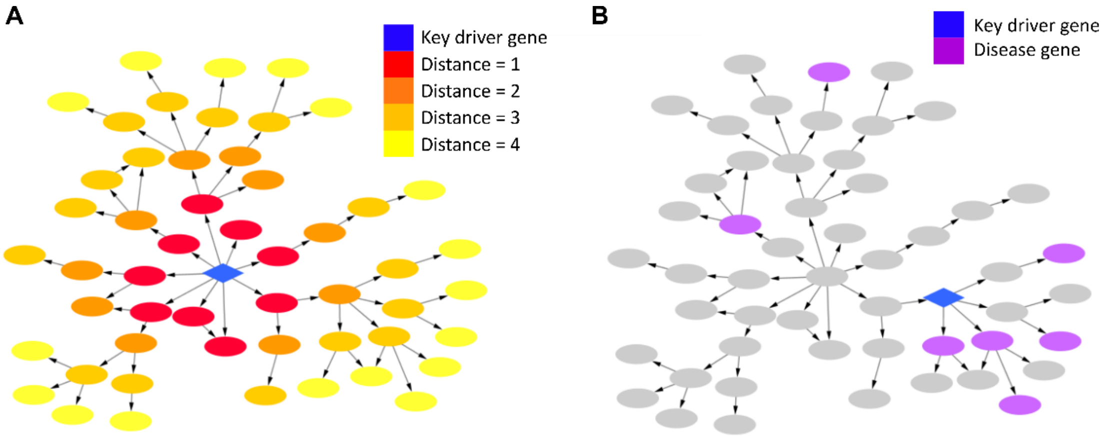

Figure 2.2: Identification of key drivers in Bayesian networks

- *(A) Using the structure of the graph, we can identify key driver nodes that regulate more than expected number of downstream nodes*
- (B) *We can identify key driver genes by annotating the nodes with disease or trait related genes. We determine which nodes are enriched for downstream disease related genes.*

This section explores the parameters that influence Bayesian network construction and key driver identification. To prioritize candidate genes involved in fat distribution, we harnessed the predictive power of the Bayesian network and key driver analysis approach to identify genes likely to drive fat storage in subcutaneous and visceral fat (Figure 2.3). Due to the differences in body fat distribution between males and females, we constructed separate Bayesian networks of each sex-depot pair to model the distinct gene regulation in each tissue. To increase the predictive power, we identified key driver genes replicated in two independent cohorts. These identified key drivers are unbiased network predictions of the gene regulation in adipose tissue.

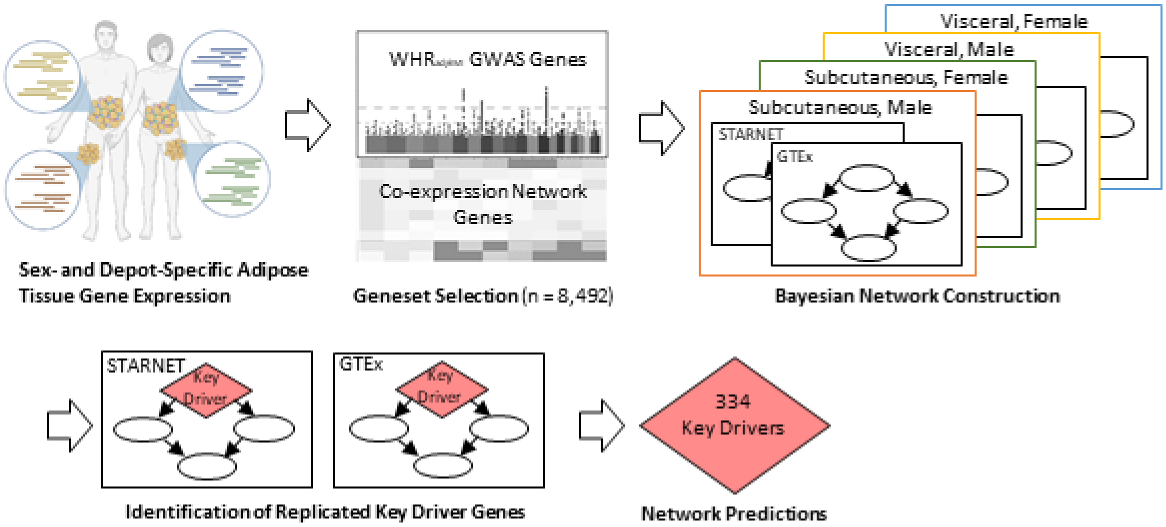

Figure 2.3: Overview of the network construction and key driver gene identification schema used in main analyses: Publicly available RNA-seq gene expression data (GTEx[84] and STARNET[85]) from subcutaneous and visceral adipose tissue were subset between males and females. Co-expression network genes and $\text{WHR}_{\text{adjBMI}}$ *GWAS genes[30] used to construct Bayesian networks representing each sex and depot using RIMBANET[56]. Key driver genes shared between STARNET and GTEx were identified in each network, and the shared key driver genes were prioritized.*
Figure made with BioRender.

2.2 Results:

2.2.1 Optimal Parameters for Bayesian Networks using RIMBANET in adipose tissue

2.2.1.1 Network maximum number of nodes is less than 10,000

To determine the utility of Bayesian network analysis transcriptome-wide, we determined the maximum number of input genes that could form a consensus network. In RIMBANET, 1,000 individual graphs are constructed, then are merged into one consensus graph that represents connections found in 70% of the iterations. The rate-limiting step for these analyses is network consensus, which must iteratively identify and remove loops in the final structure to produce an acyclic graph.

We attempted to build networks from 6,000, 8,000, and 10,000 genes that were most variably expressed in METSIM RNA-seq data from the subcutaneous adipose tissue of 434 men. We found that the time per iteration and consensus graph formation time scaled with the number of input nodes (Table 2.1). Networks with 10,000 input genes were not able to reach a consensus graph within three days of computation time, though each of the 1,000 iterations was successfully run. Future analyses showed that the complexity and resulting computation time varied by input dataset and by fat depot; subcutaneous expression data and STARNET datasets resulted in more complex networks that took longer to reach a consensus graph. We also observed that networks made with microarray data were more complex.

Table 2.1: Number of input genes effects computational success and time of Bayesian network construction

Number of Nodes	Reached consensus?	Estimated Time per iteration	Estimated Time to consensus
>1,000	yes	> 5 minutes	> 1 hr.
6,000	yes	45 minutes	3-6 hr.
8,000	yes	1-2 hr.	6-12 hr.
10,000	no	2-4 hr.	never

2.2.1.2 Selection of gene set using iterativeWGCNA module genes

We performed much of the initial testing using the METSIM subcutaneous adipose tissue RNA-seq expression data. Initially, we selected the top n most variable genes to determine some of the initial parameters. We then focused our analyses to include genes that were more likely to be regulatory of body fat distribution. We used WGCNA and iterativeWGCNA to identify genes co-expressed and likely co-regulated in modules. We infer that genes sorted into modules are expressed and regulated in adipose tissue, while genes not assigned modules are either not expressed or are not co-regulated with other adipose tissue genes. We found that iterativeWGCNA was able to sort more genes into modules than WGCNA and that those modules, while usually smaller, were more variable in size (Table 2.2).

Table 2.2: iterativeWGCNA sorts more genes into modules than WGCNA

METSIM	WGCNA	iterativeWGCNA
Genes in Modules	3024	9796
Number of Modules	28	46
Module Size Range	31-992	20-4100

Table 2.3: iterativeWGCNA applied to adipose tissue RNA-seq data

	N input genes passing QC	N genes in modules	N modules	Average module size	N GO-term enriched modules
STAR SQ	15514	10483	58	262.9	30
MET SQ	15899	9796	46	388.3	30
GTEX SQ	16466	3514	81	200.8	37
STAR V	16109	10437	80	198.9	44
GTEX V	16507	6792	98	166.7	63

While iterativeWGCNA seemed more attractive in its ability to find strong correlations between larger numbers of genes, we wanted to make sure that the genes identified represented true biology and that it could be applied to all datasets. We used STARNET, METSIM, and GTEx adipose tissue gene expression data, considering samples from both sexes together. We followed initial quality control steps described in methods and used ~16,000 expressed genes as input for each dataset. iterativeWGCNA partitioned 9,000-10,000 genes into 50-80 modules in METSIM and STARNET (Table 2.3). The other ~6,000 genes in these datasets were not strongly correlated (biweight midcorrelation) with others and were not assigned to modules. Fewer genes were assigned into modules in GTEx, which seemed to be driven by the loss of the single large module that STARNET and METSIM possessed (Figure 2.4). GTEx co-expression networks identified more modules, but they contained fewer genes (Table 2.3). GTEx RNA is lower quality and contains more noise, which may account for the information loss.

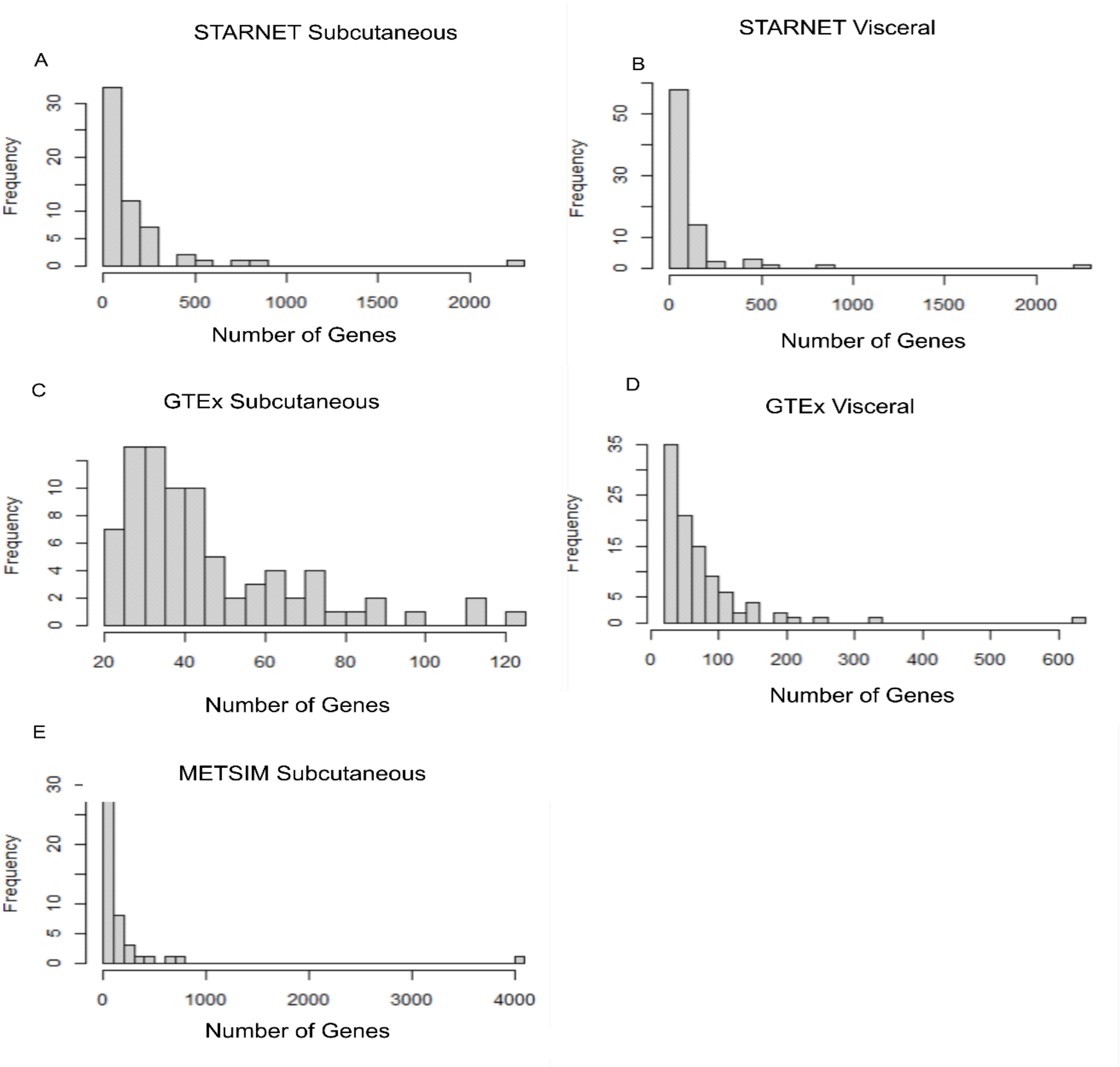

Figure 2.4: Number of Genes assigned to modules in each adipose tissue gene expression dataset

- *(A) Number of genes sorted into each iterativeWGCNA module in STARNET Subcutaneous adipose tissue gene expression data*

(B) Number of genes sorted into each iterativeWGCNA module in STARNET Visceral adipose tissue gene expression data
(C) Number of genes sorted into each iterativeWGCNA module in GTEx Subcutaneous adipose tissue gene expression data
(D) Number of genes sorted into each iterativeWGCNA module in GTEx Visceral adipose tissue gene expression data
(E) Number of genes sorted into each iterativeWGCNA module in METSIM Subcutaneous adipose tissue gene expression data

We wanted to make sure that normal and potentially disease adipocyte cell processes were represented within the modules. We tested to see whether the large modules contained genes within specific gene ontology (GO) pathways. The large module not seen in GTEx is enriched for basic cell processes like transcription and protein homeostasis (Supplemental Table 2.1). We applied GO term enrichment to the rest of the modules to see whether these processes were missing from GTEx modules. We identified modules within GTEx enriched for similar basic cell processes, though these were dispersed over a greater number of enriched modules (Supplemental Table 2.1). Further, we find a wide array of adipose tissue processes within the enriched GO terms; each dataset contains basic cell processes-related modules, lipid metabolism-related modules, and modules related to other tissue resident cell types. Despite known differences in cell-type composition, we observed a similar number of modules related to lipid processes and to immune processes in all datasets.

We conclude that, while there are differences between the gene-gene correlation structures of STARNET, GTEx, and METSIM, iterativeWGCNA is a good method for identifying groups of genes that represent shared adipose tissue biology. In future analyses, we identify the genes that are assigned to modules in both datasets as a set representative of conserved adipose tissue gene expression and use this gene set to construct meaningful adipose tissue Bayesian networks.

2.2.1.3 K-means clustering is optimal method of discretization

RIMBANET is able to create large Bayesian networks in part by discretizing the gene expression into three bins corresponding to low, medium, and high expression. Specifically, the network is built and scored using the conditional probabilities between discretized gene expression values; continuous data is only used to update prior information. Others found that the method of discretization effects the structure and confidence of the graph[26], therefore using discretized data that accurately represents the continuous data in necessary for accurate modeling and predictions.

We first investigated the distribution of expression within k-means discretized bins in METSIM RNA-seq data. We used WGCNA to identify modules within the METSIM data, then identified six modules containing ~3,000 genes that were highly correlated with $\mathrm{WHR}_{\mathrm{adjBMI}}$. We used expression from the module genes to create large Bayesian networks, each using a different set of user defined or k-means discretized expression values.

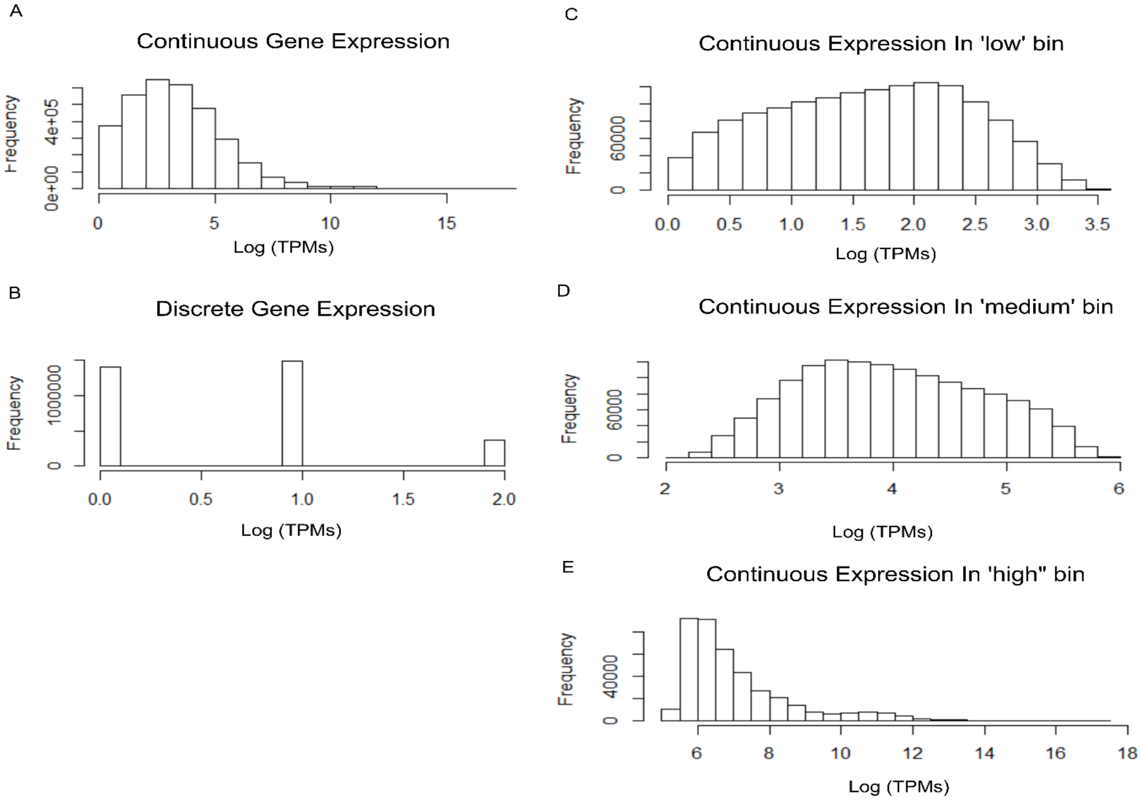

Figure 2.5: K-means cluster assigns genes to overlapping expression bins.

- *(A) Continuous gene expression in METSIM adipose tissue gene expression*
- *(B) Expression data discretized using k-means clustering*

(C) Continuous expression of values assigned to the 'low' expression bin
(D) Continuous expression of values assigned to the 'medium' expression bin
(E) Continuous expression of values assigned to the 'high' expression bin
Natural log (ln) transcripts per million + 1 displayed

We observe that the k-means discretization created bins that are overlapping, and that the low expression bin contains expression values between 0 and 3.5, in log-scaled Transcripts per Million (TPM) (Figure 2.5). Most values in the low bin had expression less than 2.5 log-scaled TPM. Equal number of values were sorted into the low and medium bins, while the high expression bin was comparatively small.

We compared k-means clustered bins to user the defined, cut-off based bins. We created schema for binning expression in five ways, increasing the size of the "low" bin (Table 2.4). The bins we defined represented a range of artificially lower and higher number of expression values sorted into the 'low' bin.

Table 2.4: User defined discretization schema are less successful at constructing predictive Bayesian networks than k-means clustering

Discretization	low range	medium range	high range	Reached consensus?	Nodes	Edges	Degree	Key Drivers
Disc 1	0-2	2-4	>4	no	6536	8785	2.20	0
Disc 2	0-3	3-6	>6	no	6051	7418	2.74	1
Disc 3	0-4	4-8	>8	no	4748	5632	2.52	0
Disc 4	0-5	5-10	>10	yes	2937	3630	2.54	0
Disc 5	0-6	6-12	>12	yes	1719	2115	2.26	6
K-Means Disc	<~2.5	~2.5 - ~6	> ~6	yes	5282	6248	2.05	17

*red = determined from one iteration, green = determined from consensus network

We found that the user defined bins most similar in size to the k-means declared bins, Disc 1, 2, and 3, could not form consensus graphs from the 1,000 iterations (Table 2.4). Artificially forcing the 'low' bin to contain more expression values, as in Disc 4 and 5, created iterations that could reach a consensus graph, but these networks contained fewer edges and nodes. We randomly selected one iteration of each of Disc 1, 2, and 3 networks, and found that these network structures were more similar to the k-means graph. The number of nodes and edges in each graph scaled with the size of the ‘low’ bin. Finally, we identified key drivers that were enriched for downstream $\text{WHR}_{\text{adjBMI}}$ key driver genes. We found that k-means discretized data produced the only network with the power to predict GWAS regulators. Therefore, we use k-means to discretize data in future studies.

2.2.1.4 Including zero values may improve the predictive power of the networks

Since we observed such differences in network structure based on the number of expression values sorted into the ‘low’ bin, we hypothesized that the size of the ‘low’ bin may have a large effect on k-means discretized networks. In our initial data quality control, we removed lowly expressed genes that had TPMs < 0.1 in more than 80% of the samples. In these analyses, we tested the effect of including genes with zero values (TPMs < 0.1) in less than 20, 40, 60, and 80% of the samples. We used gene expression from GTEx, STARNET, and METSIM, with both sexes considered together. We used iterativeWGCNA to identify module genes in each dataset, then found the union set in each depot as described above. We used the combined gene set of 6,275 genes in subcutaneous or visceral modules as the input gene set, and we included a set of 705 $\text{WHR}_{\text{adjBMI}}$ and T2D related genes to test the enrichment of potential key drivers.

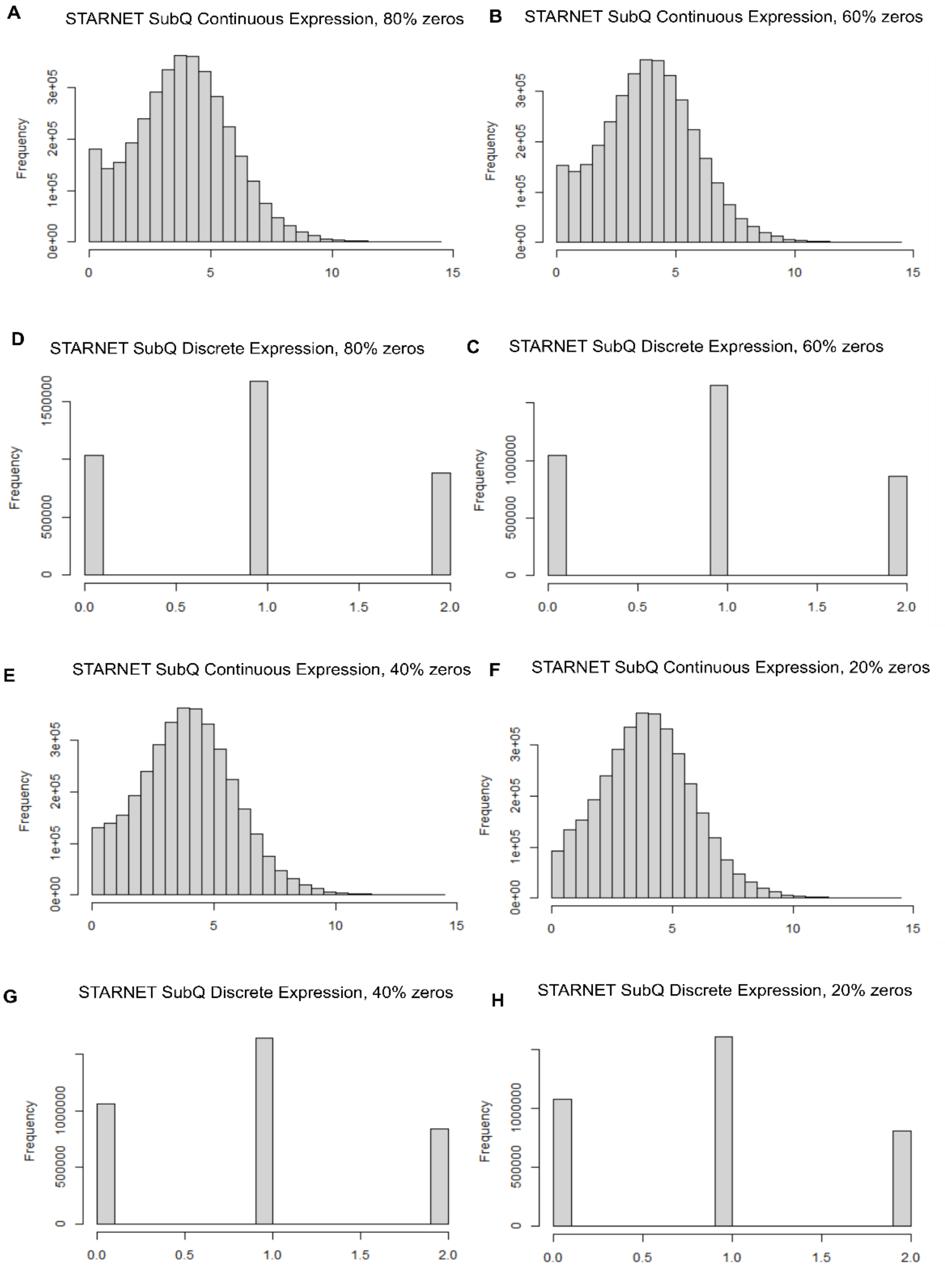

Figure 2.6: Removing genes with zero values in many samples clustered more values in the 'low' bin

(A) Continuous gene expression in STARNET subcutaneous adipose tissue data including genes with zeros < 80% of the samples.

(B) K-means discretization of gene expression in STARNET subcutaneous adipose tissue data including genes with zeros < 80% of the samples.

(C) Continuous gene expression in STARNET subcutaneous adipose tissue data including genes with zeros < 60% of the samples.

(D) K-means discretization of gene expression in STARNET subcutaneous adipose tissue data including genes with zeros < 60% of the samples.
(E) Continuous gene expression in STARNET subcutaneous adipose tissue data including genes with zeros < 40% of the samples.
(F) K-means discretization of gene expression in STARNET subcutaneous adipose tissue data including genes with zeros < 40% of the samples.
(G) Continuous gene expression in STARNET subcutaneous adipose tissue data including genes with zeros < 20% of the samples.
(H) K-means discretization of gene expression in STARNET subcutaneous adipose data including genes with zeros < 20% of the samples.

We found that increasing the number of zeros allowed in the gene set slightly decreased the amount of expression values that were sorted into the ‘low’ bin; restricting the gene set to genes with less than 20% zeros across samples resulted in more non-zero expression values sorted into the low bin (Figure 2.6). This was true for all five datasets, though only STARNET subcutaneous is shown here. The number of genes and edges scaled with the number of zero genes included, but these changes were small in magnitude (Figure 2.7). The degree distribution was between 2.6 and 2.9 for all networks, but did not scale with number of zeros included.

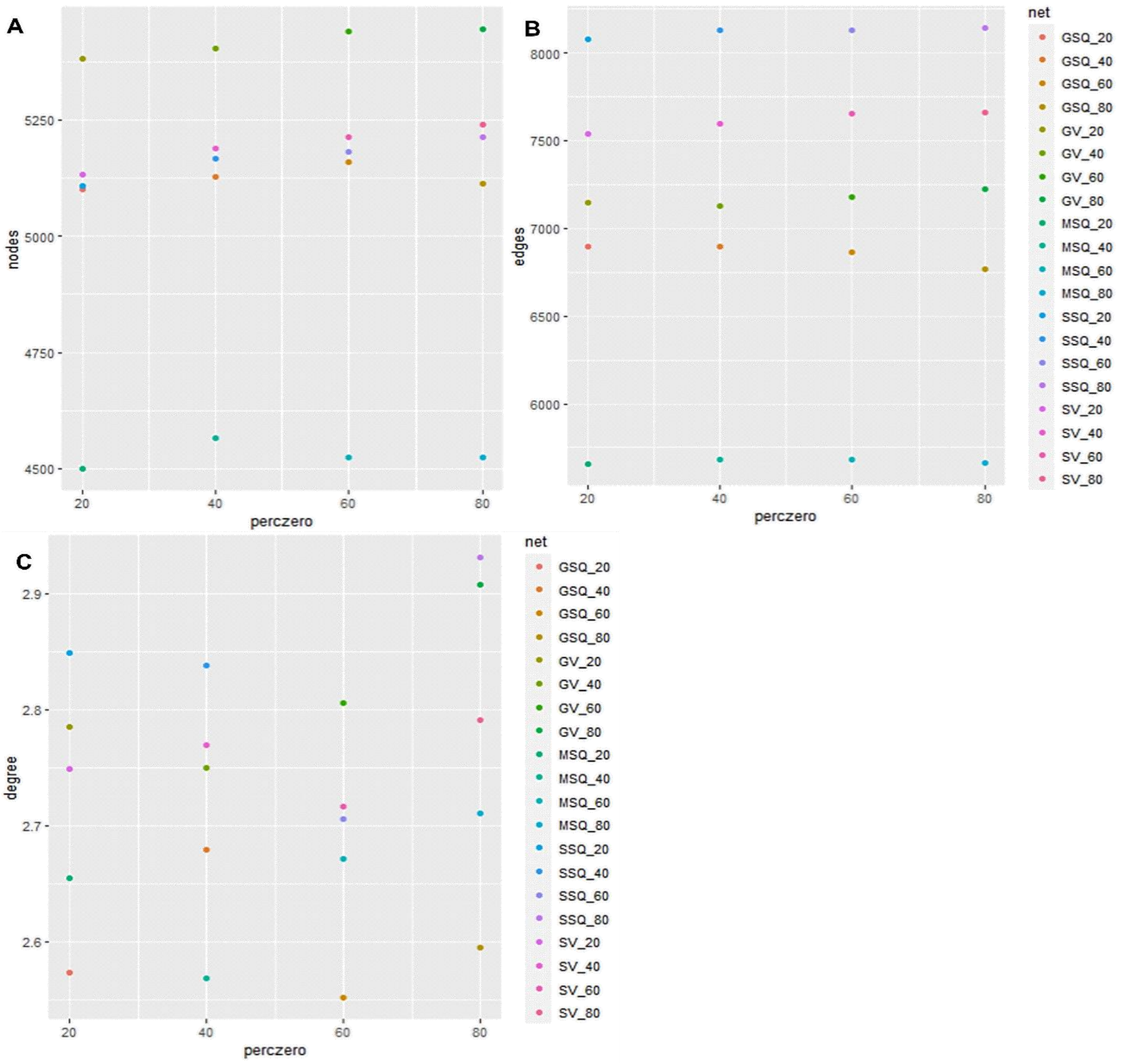

Figure 2.7: Removing genes with many zero values across samples causes no change in network structure

(A) Number of nodes included in the consensus graph versus percent zeros included
(B) Number of edges included in the consensus graph versus percent zeros included
(C) Degree of the consensus graph versus percent zeros included

The number of structural and enriched key drivers identified was similar between datasets, and the number of zero genes included had a small effect on total number of key drivers identified. All networks were scale-free and capable of making predictions (Figure 2.8-2.10). However, the key drivers identified in each network were not the same. The identity of both enriched (Figure 2.8) and structural (Figure 2.9) key driver genes were only somewhat conserved between networks including 60 and 80% zero genes, while the 20 and 40% zero gene networks were usually more dissimilar.

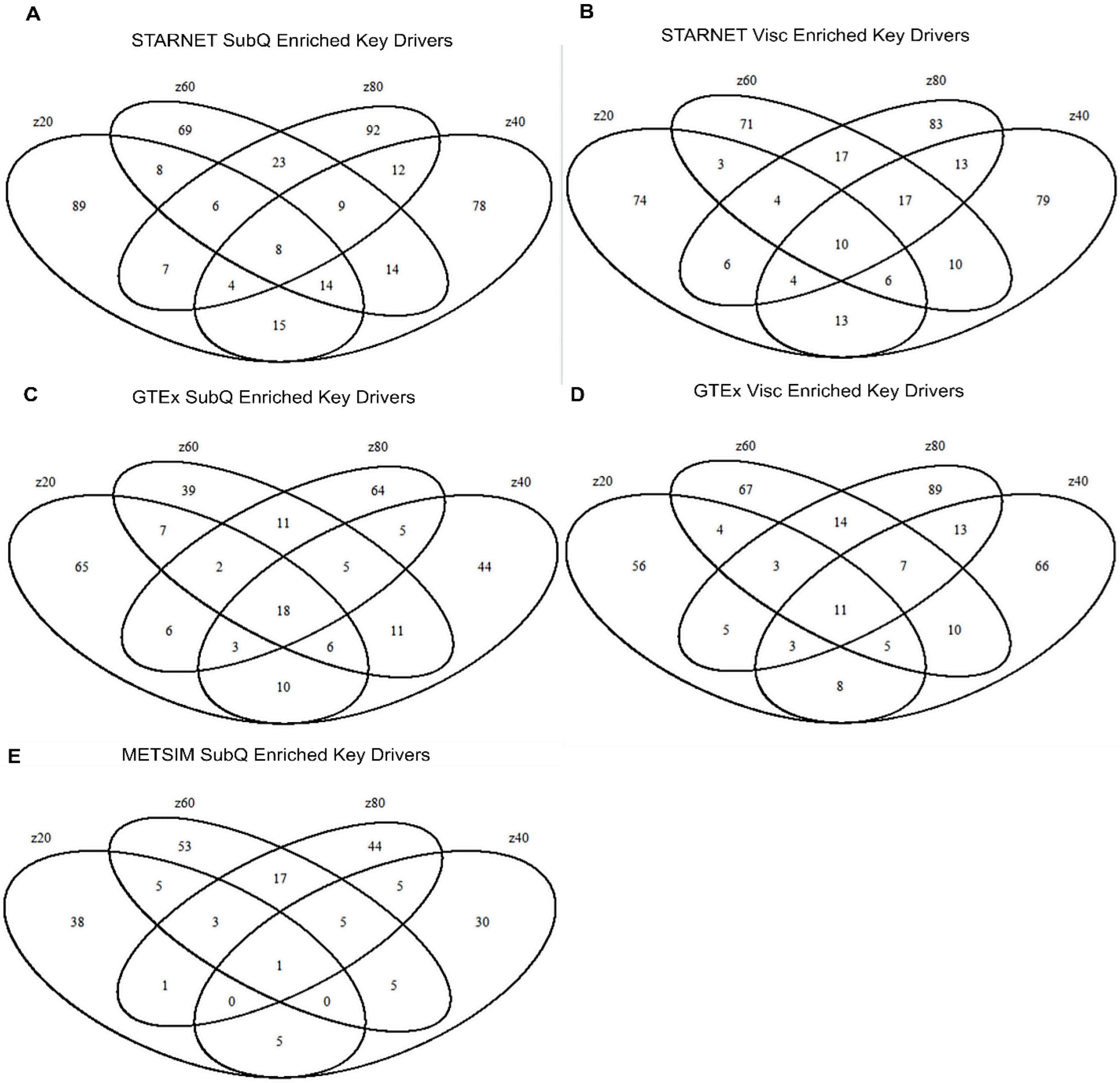

Figure 2.8: More key drivers enriched for downstream WHR_{adjBMI} genes are shared between the networks that allow 60 and 80% zeros across samples

(A) Enriched key driver genes shared in Bayesian networks made from STARNET subcutaneous adipose tissue gene expression data including genes with zeros < 20, 40, 60 and 80% of the samples.

(B) Enriched key driver genes shared in Bayesian networks made from STARNET visceral adipose tissue gene expression data including genes with zeros < 20, 40, 60 and 80% of the samples.

(C) *Enriched key driver genes shared in Bayesian networks made from GTEx subcutaneous adipose tissue gene expression data including genes with zeros < 20, 40, 60 and 80% of the samples.*
(D) Enriched key driver genes shared in Bayesian networks made from GTEx visceral adipose tissue gene expression data including genes with zeros < 20, 40, 60 and 80% of the samples.
(E) Enriched key driver genes shared in Bayesian networks made from METSIM subcutaneous adipose tissue gene expression data including genes with zeros < 20, 40, 60 and 80% of the samples.

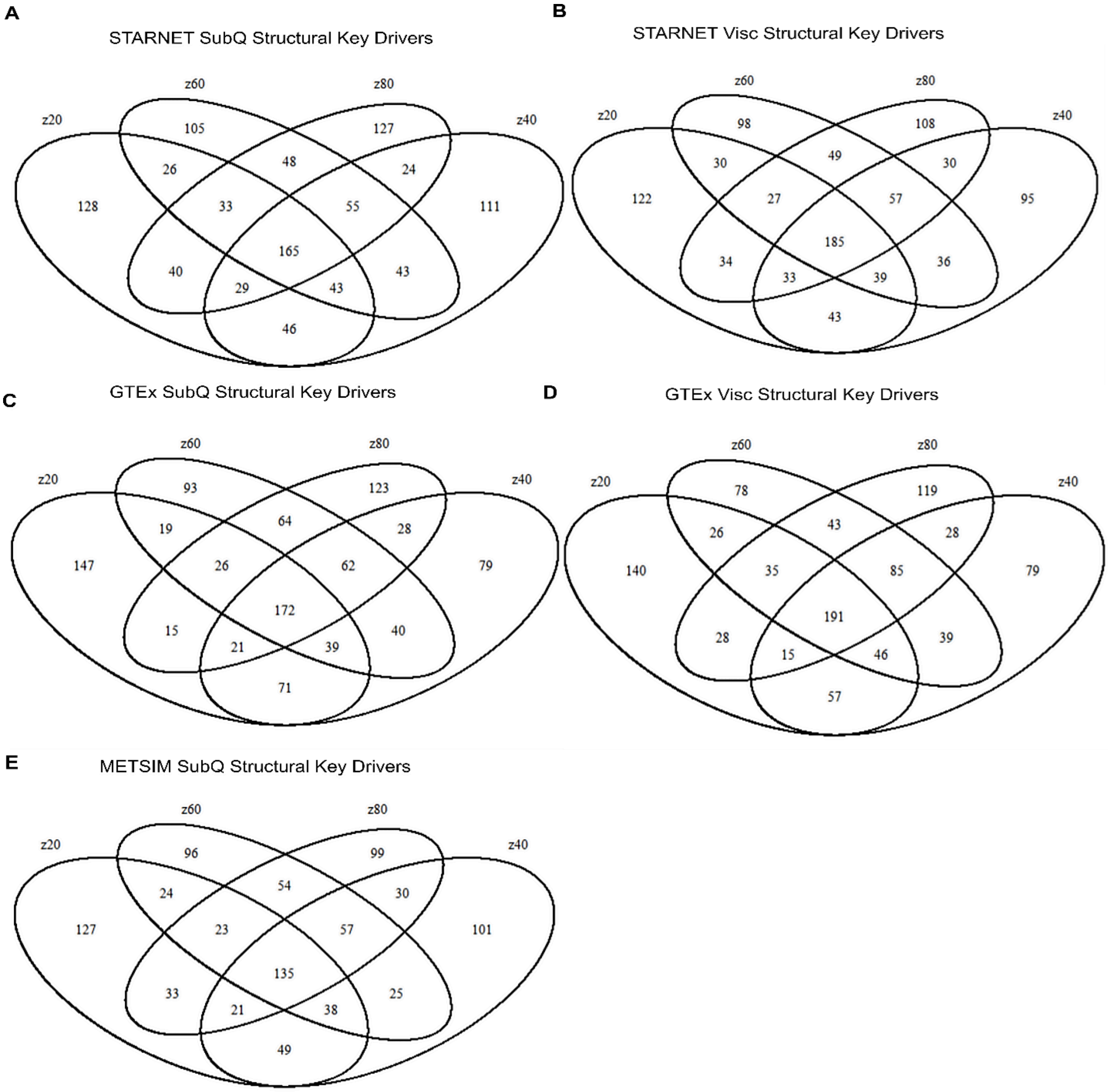

Figure 2.9: More high-scoring structural key driver genes are shared between the networks that allow 60 and 80% zeros across samples

(A) *Structural key driver genes shared in Bayesian networks made from STARNET subcutaneous adipose tissue gene expression data including genes with zeros < 20, 40, 60 and 80% of the samples.*

(B) *Structural key driver genes shared in Bayesian networks made from STARNET visceral adipose tissue gene expression data including genes with zeros < 20, 40, 60 and 80% of the samples.*
(C) *Structural key driver genes shared in Bayesian networks made from GTEx subcutaneous adipose tissue gene expression data including genes with zeros < 20, 40, 60 and 80% of the samples.*
(D) *Structural key driver genes shared in Bayesian networks made from GTEx visceral adipose tissue gene expression data including genes with zeros < 20, 40, 60 and 80% of the samples.*
(E) *Structural key driver genes shared in Bayesian networks made from METSIM subcutaneous adipose tissue gene expression data including genes with zeros < 20, 40, 60 and 80% of the samples.*

When considering both types of key driver genes, we found similar reproduction of key drivers between subcutaneous networks of including different numbers of zeros (Figure 2.10). In visceral, 60% and 80% zero networks reproduced the most key driver genes between datasets. We conclude that removing genes that are lowly expressed in some samples does not affect the structure or predictive power of the networks; however, the reproducibility of those predictions within and between datasets may be slightly improved by including genes with more zero values. Since the BIC score of the network is calculated based on the discretized data, assigning more values to the zero bin likely reduces noise in the discretized data and therefore may reduce the number of false positive connections.

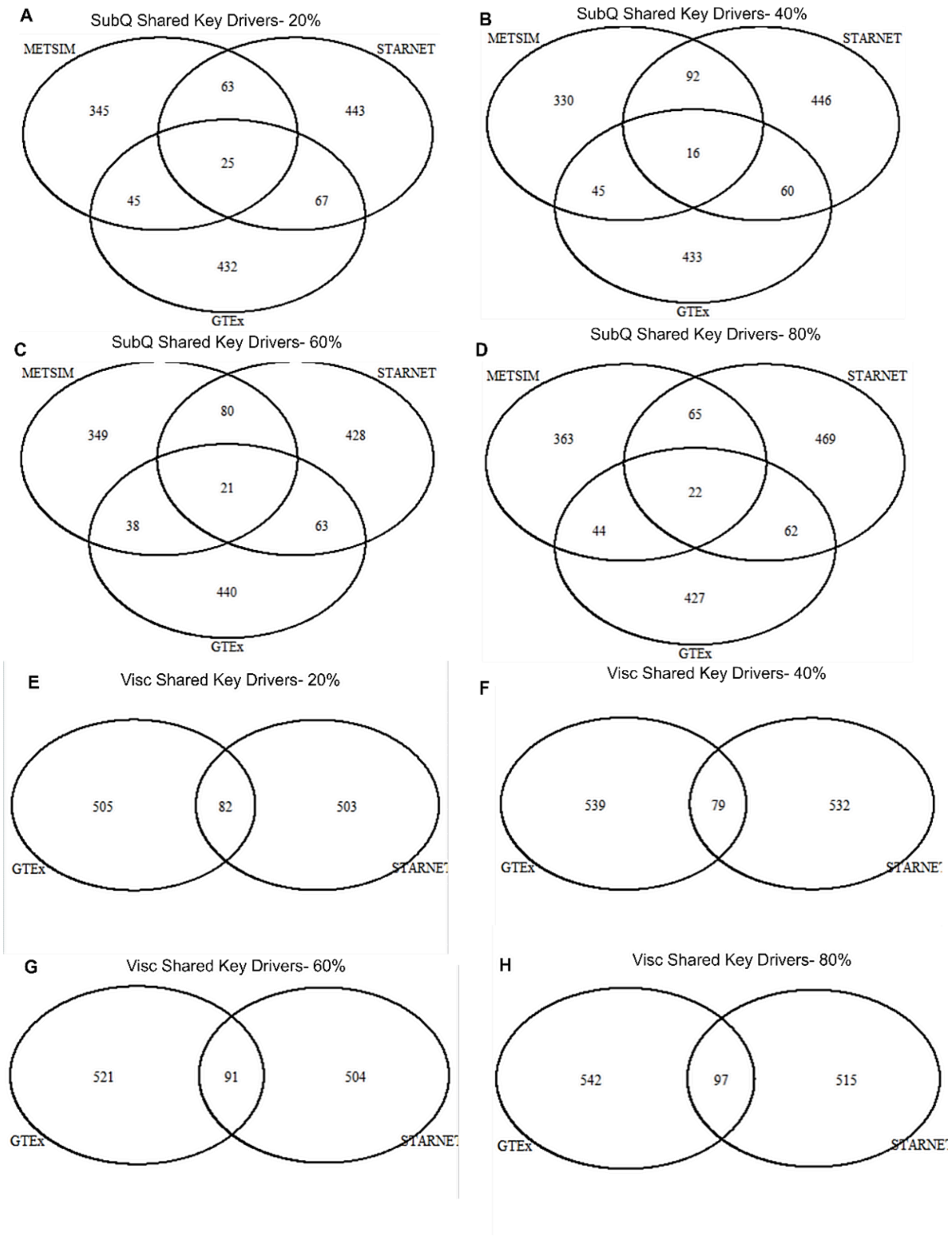

Figure 2.10: Key driver genes are similarly replicated between subcutaneous networks

- *(A) Key driver genes shared in Bayesian networks made from subcutaneous adipose tissue gene expression data including genes with zeros < 20% of the samples.*
- *(B) Key driver genes shared in Bayesian networks made from subcutaneous adipose tissue gene expression data including genes with zeros < 40% of the samples.*
- *(C) Key driver genes shared in Bayesian networks made from subcutaneous adipose tissue gene expression data including genes with zeros < 60% of the samples.*
- *(D) Key driver genes shared in Bayesian networks made from subcutaneous adipose tissue gene expression data including genes with zeros < 80% of the samples.*

(E) Key driver genes shared in Bayesian networks made from visceral adipose tissue gene expression data including genes with zeros < 20% of the samples.
(F) Key driver genes shared in Bayesian networks made from visceral adipose tissue gene expression data including genes with zeros < 40% of the samples.
(G) Key driver genes shared in Bayesian networks made from visceral adipose tissue gene expression data including genes with zeros < 60% of the samples.
(H) Key driver genes shared in Bayesian networks made from visceral adipose tissue gene expression data including genes with zeros < 80% of the samples.

2.2.1.5 Subsampled networks are able to partial replicate the predictions of the original

To test the robustness of network predictions and the utility of using different samples sizes in the same study, we subsampled all datasets to 10, 50, and 90% of the original samples. We used STARNET and GTEx, divided into males and females, and used ~7,000 input genes using Bayesian network modules. We chose genes assigned to modules in both GTEx and STARNET in at least one sex and depot. Then, randomly chose the samples included and constructed Bayesian networks in RIMBANET.

We found that the number of genes and edges included into the networks increased with the percent of total samples included (Figure 2.11 A, B). Specifically, these characteristics scaled with log number of samples (Figure 2.12). Larger networks also had higher average path distance and smaller clustering coefficient compared to random networks (Figure 2.11 C-E). These characteristics were all higher in STARNET networks compared to GTEx networks of the same size. Depot and sex did not affect network size independent of the sample size effect (Figure 2.12 I, J). The 90% networks are similar to the original networks, the 50% sub-sampled networks appear able to replicate the size of the originals but not the structure, while 10% sub-sampled networks are not similar to full networks.

We assessed the predictive power of the sub-sampled networks by identifying the reproducibility of the structural and enriched key driver genes of the full network (Figure 2.11 F, H). Since structural key drivers represent the top 10% network genes, these scaled with number of genes, and thus increased with sample size. Networks containing 90% of the original samples replicated over 50% of the original network's predictions. Key drivers enriched for downstream $\text{WHR}_{\text{adjBMI}}$ GWAS genes also increased with sample size, though did not reproduce the original results to the same extent as the structural key drivers. We found that the number of samples was correlated with the number of key drivers predicted, both structural and enriched (Figure 2.12). Though we found that STARNET networks identified more total key drivers, networks made by sub-sampling GTEx datasets were equally good at reproducing the results of the original. The reproducibility of key driver predictions differs by sex and depot (Figure 2.11 I). The key drivers shared between STARNET and GTEx were identified in the original networks and in the sub-sampled networks. The male subcutaneous and female visceral networks had the best ability to reproduce the results of the original (Table 2.5)

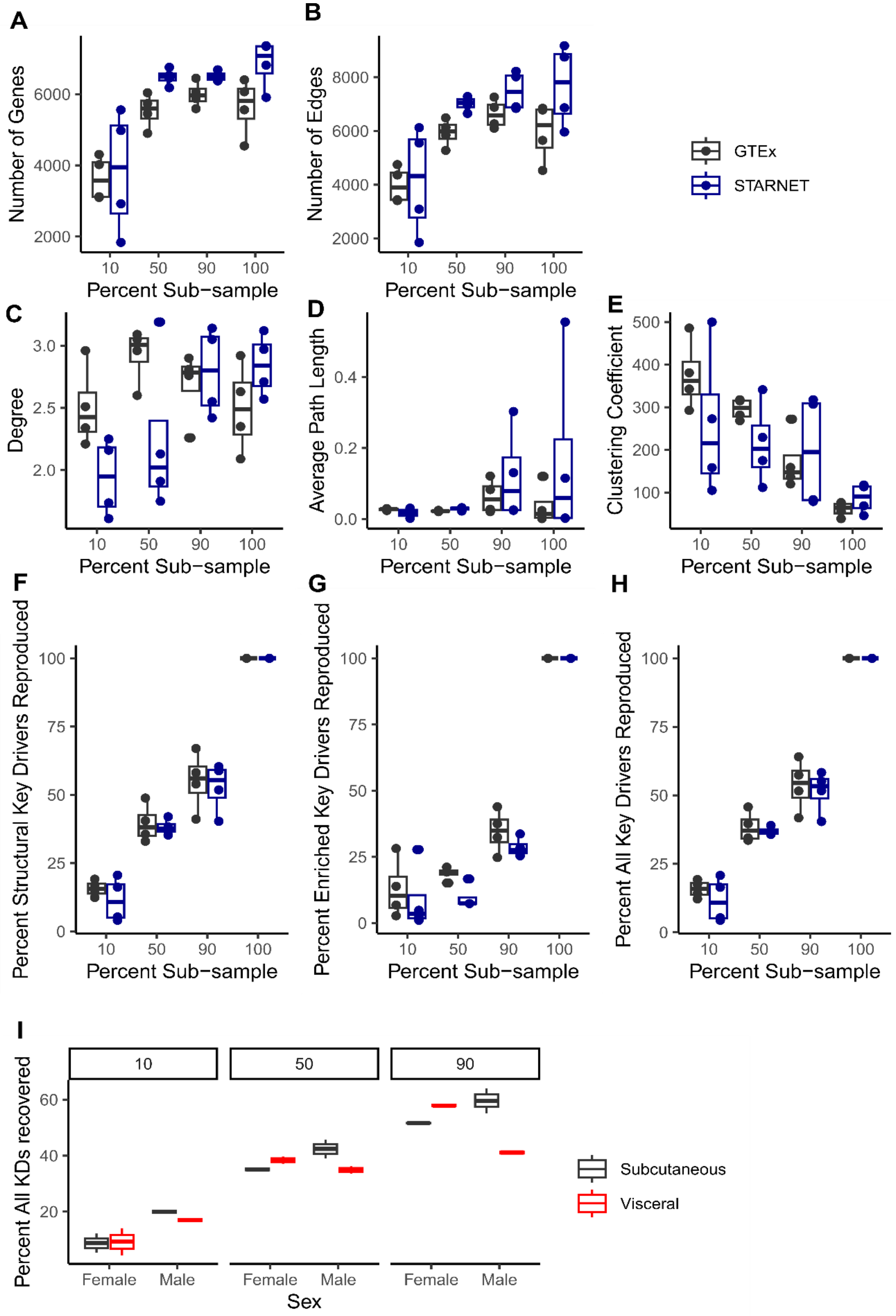
A
Number of Genes
2000
4000
6000
10
50
90
100
Percent Sub−sample
B
Number of Edges
2000
4000
6000
8000
10
50
90
100
Percent Sub−sample
GTEx
STARNET
C
Degree
2.0
2.5
3.0
10
50
90
100
Percent Sub−sample
D
Average Path Length
0.0
0.2
0.4
10
50
90
100
Percent Sub−sample
E
Clustering Coefficient
100
200
300
400
500
10
50
90
100
Percent Sub−sample
F
Percent Structural Key Drivers Reproduced
0
25
50
75
100
10
50
90
100
Percent Sub−sample
G
Percent Enriched Key Drivers Reproduced
0
25
50
75
100
10
50
90
100
Percent Sub−sample
H
Percent All Key Drivers Reproduced
0
25
50
75
100
10
50
90
100
Percent Sub−sample
I
10
50
90
Percent All KDs recovered
20
40
60
Female
Male
Female
Male
Female
Male
Sex
Subcutaneous
Visceral

Figure 2.11: Sub-sampling the networks to different percentage of input samples impact structural and predictive properties

- *(A) Number of genes in each network by the percent subsampled*
- *(B) Number of edges in each network by the percent subsampled*
- *(C) Degree exponent of each network by the percent subsampled*
- *(D) Average shortest path distance of each network compared to a random network of the same size, by the percent subsampled*
- *(E) Clustering coefficient of each network compared to a random network of the same size, by the percent subsampled*
- *(F) Percentage of structural key drivers predicted by the full network that were recovered by the sub-sampled networks*
- *(G) Percentage of enriched key drivers predicted by the full network that were recovered by the sub-sampled networks*
- *(H) Percentage of all key drivers predicted by the full network that were recovered by the sub-sampled networks*
- *(I) Percentage of all key drivers predicted by the full network that were recovered by the sub-sampled networks, grouped by sex and depot*

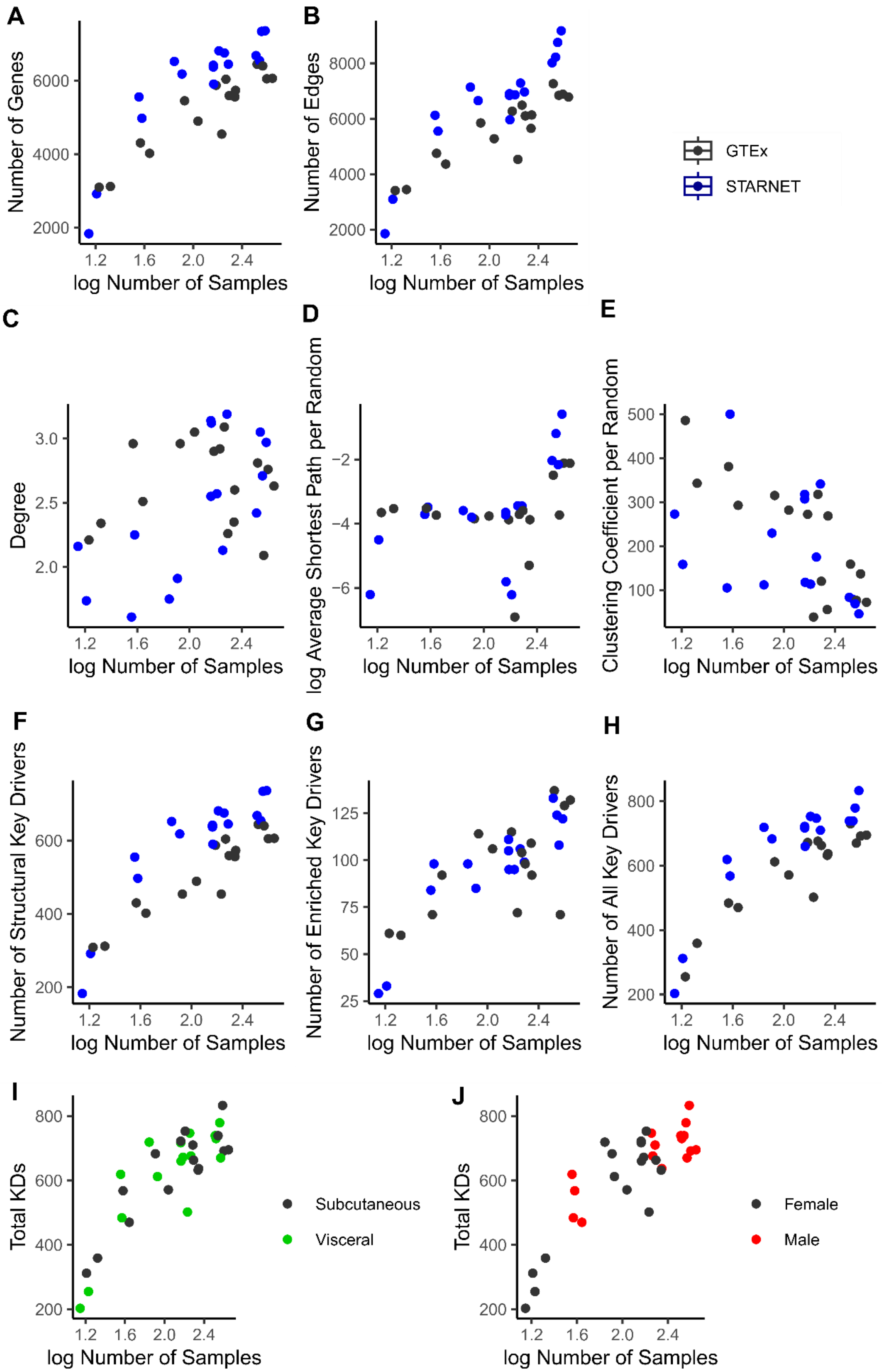
A
Number of Genes
2000
4000
6000
1.2
1.6
2.0
2.4
log Number of Samples
B
Number of Edges
2000
4000
6000
8000
log Number of Samples
GTEx
STARNET
C
Degree
2.0
2.5
3.0
log Number of Samples
D
log Average Shortest Path per Random
−6
−4
−2
log Number of Samples
E
Clustering Coefficient per Random
100
200
300
400
500
log Number of Samples
F
Number of Structural Key Drivers
200
400
600
log Number of Samples
G
Number of Enriched Key Drivers
25
50
75
100
125
log Number of Samples
H
Number of All Key Drivers
200
400
600
800
log Number of Samples
I
Total KDs
200
400
600
800
log Number of Samples
Subcutaneous
Visceral
J
Total KDs
200
400
600
800
log Number of Samples
Female
Male

Figure 2.12: Including a different number of input genes impacts structural and predictive properties

- *(A) Number of genes in each network by the log number of genes in the network*
- *(B) Number of edges in each network by the log number of genes in the network*
- *(C) Degree exponent of each network by the log number of genes in the network*
- *(D) Average shortest path distance of each network compared to a random network of the same size, by the log number of genes in the network*
- *(E) Clustering coefficient of each network compared to a random network of the same size, by the log number of genes in the network*
- *(F) Number of structural key drivers predicted by the sub-sampled networks*
- *(G) Number of enriched key drivers predicted by the sub-sampled networks*
- *(H) Number of all key drivers predicted by the sub-sampled networks, colored by dataset*
- *(I) Number of all key drivers predicted by the sub-sampled networks, colored by depot*
- *(J) Number of all key drivers predicted by the sub-sampled networks, colored by sex*

Table 2.5: 90% subsampled networks have the greatest ability to reproduce shared key drivers

Percent Sub-Sample	Number of Shared Key Drivers	Percent Overlap with Original	Sex	Depot
10	29	0	Female	Subcutaneous
10	50	5.04	Male	Subcutaneous
10	19	0	Female	Visceral
10	38	5.04	Male	Visceral
50	69	15.68	Female	Visceral
50	92	14.2	Male	Visceral
50	51	11.36	Female	Subcutaneous
50	72	17.65	Male	Subcutaneous
90	76	37.25	Female	Visceral
90	102	13.45	Male	Visceral
90	103	39.4	Male	Subcutaneous
90	79	29.5	Female	Subcutaneous

2.2.1.6 eQTL priors do not significantly improve the structure of small graphs

Because others find that including prior information improves the confidence in the identified structure[58], we tested the effects of adding or not adding eQTL prior information to construct small Bayesian networks. First, we tested the utility of adding direct connections representing genetic regulation. Using WGCNA, we created modules in METSIM RNA-seq gene expression data, then used RIMBANET to construct Bayesian networks using the genes from individual modules. Here, we defined eQTLs as genes whose expression was significantly associated with a SNP in *cis*

(<1Mb, $p < 2.46 \times 10^{-4}$) and in *trans* (>1Mb, 1×10^{-4}). A permissive *trans*-eQTL cutoff was employed purposefully, as few genome-wide significant *trans*-eQTLs are identified. We define pairs where the eSNP was associated with a gene in cis and a different gene in *trans*. We infer that the *trans* eGene levels are mediated by the *cis* eGene. We then define an edge in the network that connects the *cis* and *trans* pair if both genes are present in the input gene set.

We find that user defined edges inconsistently improve the ability of the networks to describe observed expression data, as measured by the BIC score (Table 2.6). This does not appear to be related to the size of the network or the number of *cis*-*trans* gene edges included.

Table 2.6: User defined eQTL edges do not consistently improve BIC score

Module Name	Number of eQTL pairs added	Total Genes in Module	BIC with eQTLs	BIC without eQTLs	BIC Score Improvement?
black	1	95	-78416.96	-78447.8	Yes
blue	9	251	-197577.97	-197607.86	Yes
brown	2	199	-145467.82	-145449.74	No
purple	4	174	-137225.91	-137346.39	Yes
royalblue	1	45	-38177.92	-38178.39	Yes
tan	1	72	-53779.59	-53776.9	No
turquoise	58	992	-744039.87	-743543.88	No

We statistically determined if the SNP effects on *trans* eGene expression were caused by the *cis* eGene using causal inference testing. Like Mendelian Randomization, causal inference testing identifies the likelihood of that the hypothesis of SNP->*cis* eGene->*trans* eGene is more likely that other connection structures. This test identified only eight significant SNPs that are associated with the *Trans* eGene when considering the *cis* eGene effects. In large networks, both genes in a pair were present for four pairs, on average. These were unlikely to have large effects in networks of thousands of genes and edges, and thus, were not employed in final iterations of network construction.

Next, we assessed the effect of adding cis-eQTL information to the graph. We assume that genes regulated by SNPs are more likely to be regulatory, as in the above model of SNP->cis eQTL-> effector gene. In these analyses, though we don't know the *trans*-effector gene, we still assume the *cis* eGene is a potential regulator. When determining the more likely graph of two equivalent structures, the gene with SNP associations is chosen as the parent gene. The effect of including eQTL prior information on overall graph structure was minimal. Using 500 highly expressed genes in METSIM adipose tissue, we tested the effect of artificially declaring 0%, 10%, 50%, and 90% of the network genes as eQTL-eGenes on network structure. We observed that the networks are nearly identical; over ~500 edges in the networks, between zero and three edges were different. This was untested in larger networks; the same trend of limited effect may hold, or there may be more instances of graph equivalency. Therefore, in future analyses, we included eQTL prior information, though it may not be necessary.

2.2.2 Applications of optimal Bayesian networks in adipose tissue:

2.2.2.1 Bayesian Networks model adipose tissue gene connections in a sex- and depot-specific manner:

We interrogated two independent datasets of subcutaneous and visceral adipose tissue gene expression, Genotype-Tissue Expression project (GTEx)[84] and the Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task (STARNET)[85] (Methods) and stratified each dataset by sex. There are about twice as many males as females in each resulting dataset (Table 2.7).

Table 2.7: Adipose tissue donor characteristics

	Subcutaneous				**Visceral**			
	Male		**Female**		**Male**		**Female**	
	STARNET	**GTEx**	**STARNET**	**GTEx**	**STARNET**	**GTEx**	**STARNET**	**GTEx**
Samples	387	444	162	219	362	370	147	171
Age	64.3 ± 9.1	53.0 ± 12.8	69.2 ± 7.0	52.5 ± 12.7	64.4 ± 9.0	53.2 ± 12.8	69.2 ± 6.7	51.2 ± 12.6
BMI	28.7 ± 4.3	27.7 ± 4.1	29.5 ± 4.8	26.6 ± 4.2	28.7 ± 4.2	27.5 ± 4.0	29.5 ± 4.8	26.9 ± 4.2

Ideally we would probe gene-gene interactions at a genome-wide level, but Bayesian network construction is computationally intensive, and therefore, we limited this analysis to a subset of <10,000 genes that are more likely to regulate body fat distribution (Figure 2.3). We prioritized putative regulators of body fat distribution using three strategies: (1.) genes whose expression are co-expressed with others in adipose tissue, (2.) genes proximal to body fat distribution GWAS loci[30], and (3.) genes that are putatively regulated by the transcription factor KLF14[86], see Methods. For a gene to be connected to others in a co-expression network, it must be expressed in the measured dataset, must vary between samples, and must be correlated with the expression of other genes. These properties are optimal for Bayesian network construction and can indicate gene function in the tissue of interest, therefore, we constructed adipose tissue co-expression networks for all eight datasets and identified genes connected in the corresponding STARNET and GTEx networks. The union set of replicated connected genes from co-expression networks contained 7,928 genes and made up the bulk of the input to Bayesian network construction. Genes nearest to $\mathrm{WHR}_{\mathrm{adjBMI}}$ GWAS loci likely contain a mix of 30% false positives and 70% true drivers of fat distribution[87], so we added 443 genes proximal to the 346 significant $\mathrm{WHR}_{\mathrm{adjBMI}}$ loci to the input gene set (Methods)[30,88,89]. Two previous studies identified high confidence $\mathrm{WHR}_{\mathrm{adjBMI}}$ GWAS candidate genes using colocalization methods, so we also included these 59 genes. In total, we considered this combined set of 495 genes as $\mathrm{WHR}_{\mathrm{adjBMI}}$ GWAS genes in this study. While this set does not contain all possible causal genes, it is likely enriched for them. Finally, we have previously demonstrated that *KLF14* expression regulates fat distribution in both female mice and humans[90] and is associated with the trans-regulation of 385 genes in adipose tissue specifically[86]. We hypothesized that *KLF14*'s effect on fat distribution is mediated by some of the genes it regulates transcriptionally, and we included 385 KLF14 putative target genes in the input gene set. The union set contained 8,492 genes that were used to construct all eight Bayesian networks (Supplemental Table 2.2). From this diverse gene set, we aim to prioritize putative candidate GWAS genes and genes outside of GWAS loci that may play a causal role for body fat distribution.

Table 2.8: Sex- and depot-specific adipose tissue Bayesian network construction and key driver analysis results

	Subcutaneous				Visceral			
	Male		**Female**		**Male**		**Female**	
	STARNET	**GTEx**	**STARNET**	**GTEx**	**STARNET**	**GTEx**	**STARNET**	**GTEx**
Genes	7360	6066	6814	5560	7343	6402	5908	4546
Edges	9174	6782	6865	5651	8752	6845	5962	4533
Key Drivers	833	695	753	632	779	670	660	502
Shared Key Drivers	119		88		119		51	

We chose RIMBANET to construct Bayesian networks for its ability to handle large input gene sets and its reproducibility between datasets[58,91]. We also added prior information about some genes to improve network performance, including genes with eQTLs in the corresponding adipose tissue depot (Supplemental Table 2.3). We constructed eight sex- and depot-specific networks (Methods) that contained an average of 6,250 genes connected, with an average of 6,821 directed edges between those genes (Table 2.8, Supplemental Table 2.4). Each network displayed scale-free and small-world properties consistent with known biological networks (Table 2.9)

Table 2.9: Biological properties of sex- and depot-specific adipose tissue Bayesian networks

Biological Property	Parameter	True Biological Network Values	Subcutaneous				Visceral			
			Male		Female		Male		Female	
			STARNET	GTEx	STARNET	GTEx	STARNET	GTEx	STARNET	GTEx
Scale Free	Degree Exponent	Between 2-3	2.97	2.63	2.57	2.35	2.71	2.09	3.12	2.92
Small World	Average Shortest Path: Network/Random	Less than 1	0.555	0.120	0.002	0.005	0.115	0.024	0.003	0.001
Small World	Clustering Coefficient: Network/Random	Greater than 1	45.882	72.222	113.333	55.556	68.750	76.471	117.647	38.636

In general, the male networks had more connected genes and edges than the female networks, which could be due to the difference in input sample size. To test this, we randomly subsampled the male input gene sets to include the same number of donors as the female networks, and constructed networks from these smaller sets. We identified an intermediate number of genes and edges for most subsampled networks (Table 2.10). The sparsity of the female networks is partially due to fewer female donors than male donors in the datasets.

Table 2.10: Connected genes, edges, and key drivers in full male networks (left), sub-sampled male networks (middle), female networks (right)

	Subcutaneous						Visceral					
	Male		Male Subsampled		Female		Male		Male Subsampled		Female	
	STARNET	GTEx	STARNET	GTEx	STARNET	GTEx	STARNET	GTEx	STARNET	GTEx	STARNET	GTEx
Genes	7360	6066	6586	5685	6814	5560	7343	6402	6788	5595	5908	4546
Edges	9174	6782	7106	6116	6865	5651	8752	6845	7284	6374	5962	4533
Key Drivers	833	695	751	656	753	632	779	670	749	684	660	502
Shared Key Drivers	119		75		88		119		84		51	

2.2.2.2 Bayesian Network structure identifies putative sex- and depot-specific "Key Drivers" of adipose tissue function and disease:

Using the eight constructed networks, we identified putative regulators of body fat distribution, termed key driver genes (Methods, Figure 2.3). Since we know that many body fat distribution genes are expressed and regulated in adipose tissue, a key driver gene that regulates many other genes in adipose tissue networks may also be a regulator of body fat distribution. Key driver genes have been biologically validated[16,74] for their regulatory roles in Bayesian networks.

We identified an average of 691 key driver genes per network (Table 2.8, Supplemental Table 2.5). Bayesian networks, like other models, are subject to overfitting and false positive predictions, but others observe that key driver genes are more likely than other network features to be reproduced between datasets[91] and may represent true biology. We compared the key driver predictions between our corresponding GTEx and STARNET networks and identified 334 replicated key driver genes in total (Table 2.8, Supplemental Table 2.6). Only 38 replicated key driver genes were found in multiple sex-depot groups, and only one key driver, *ARHGEF12,* was identified in all eight networks. There were more shared key driver genes found in male networks than in female networks, and in the subsampled male networks referenced above, we identified an intermediate number of shared key driver genes, showing that the number of replicated key drivers is partially a result of the input sample size (Table 2.10).

2.3 Discussion:

We conclude that Bayesian networks have incredible predictive power but are highly susceptible to spurious connections. These can occur to due to improperly defining the bin of lowest expressed genes, or by including a small sample size. Here, we explore the effect of important parameters on network structure and key driver prediction, then apply optimal Bayesian networks in adipose tissue data in a sex and depot specific manner. We identified important key drivers of each network and found which were shared between datasets.

We and others[26] find that the network structure and key driver predictions were highly responsive to discretization method and inclusion of true-lowly expressed genes in the “low” bin, while eQTL prior information had a small effect on the structure. K-means clustering was required to group expression values, even when the number of genes include in the ‘low’ bin was conserved. Removing true zeros from the data did not affect the network structure, the number of key drivers, or shared key drivers identified, but the shared key drivers were not reliably reproduced between networks. Since Bayesian networks depend on the conditional probabilities between expression of genes, accurate definition of the zero bin as true lowly expressed genes is essential to compare to low and high expression bins and score the model using likelihood estimation[26]. Other studies show that micro-array data, unlike RNA-seq, cannot accurately measure gene expression near zero[92,93]; corresponding, networks produced using microarray data were extremely densely connected and were likely overfit.

Models, including Bayesian networks, are prone to over-fitting to the construction dataset, and as such, may not be reproducible between datasets. One way to overcome overfitting is to increase the sample size. We find that increasing sample size increases the power to predict key drivers and to reproduce the key drivers identified in the original. However, more genes are included in the graph, which reduces important characteristics of small-world biological networks. There may be an optimal maximum number of genes that creates a biologically meaningful network in each dataset while conserving the predictive power.

To overcome the challenges of overfitting, we also identified key drivers in two independent cohorts and compared the shared biology. Although there are major differences between GTEx[84] and STARNET[85] that likely drive gene expression, like biopsy location, cell composition, donor ethnicity, post-mortem tissue collection, or cardiovascular disease diagnosis, the shared gene regulation likely represents conserved adipose tissue function. We found that about 10% of key driver genes were shared between the two datasets, and similar overlap with METSIM, which was used in some studies. It may be possible to batch correct and merge samples of the same tissue to create a consensus network, although we would anticipate significant information loss due to the important differences listed above. However, this may be a novel way to identify only robust key driver genes. The 90% sub-sampled networks were able to capture part of the overlap between STARNET and GTEx, and male subcutaneous and female visceral were the most able to recapitulate the predictions of the full network. In our final application, we find that the smaller sample size of the female networks results in fewer predicted key drivers and shared key driver genes. We find few genes that are both replicated between samples and robust to sub-sampling, and most key driver genes are predicted to regulate one sex and depot.

We hypothesize that key driver genes may regulate adipose tissue biology. Further prioritization is needed to identify testable key driver genes that may regulate $\mathrm{WHR}_{\mathrm{adjBMI}}$ and act in adipocytes, and further experiments are needed to determine their function. However, Bayesian networks were able to organize thousands of unknowns into one coherent model and predict the likely regulators.

2.4 Methods:

2.4.1 Input Data:

We interrogated RNA-sequencing gene expression data from subcutaneous adipose tissue and visceral abdominal adipose tissue from the Genotype-Tissue Expression project (GTEx)[84] and the Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task (STARNET)[85]. We used subcutaneous adipose tissue gene expression data from Metabolic Syndrome in Men (METSIM) cohort[89], Detailed explanations of participant inclusion, data collection, sequencing, and quantification can be found at each source. Briefly, STARNET participants are people living with coronary artery disease, from whom biopsies of abdominal subcutaneous fat and abdominal visceral fat were obtained during open thorax surgery. Samples were sequenced using the Illumina HiSeq 2000 platform. GTEx biopsies of abdominal visceral fat and leg subcutaneous fat were taken from deceased donors shortly after death. Samples were sequenced using the Illumina TruSeq platform. All datasets were obtained in transcripts per million (TPM) format.

2.4.1.1 Expression Data Processing:

For some studies, we first used annotation meta-data from each source to divide the data into males and females. We used *XIST* expression to confirm these assignments. Next, we used annotations from the R package bioMart for genome build hg38 to select only the protein coding genes within each dataset. We then removed genes with less than 0.1 TPM value in greater than 80% of the samples. Finally, we log transformed the gene expression values for subsequent analysis.

2.4.2 Bayesian Network Input Genes:

2.4.2.1 Co-expressed Genes:

We used the python package iterativeWGCNA[94] to obtain modules of co-expressed genes in each dataset. Weighted gene co-expression network analysis (WGCNA)[25] employs correlations found within the data to determine which groups of genes are highly correlated and likely co-regulated. First, we computed the correlations between all genes. We raised these correlation coefficients to an empirically determined power to increase the differences observed. Next, we performed hierarchical clustering on the correlation matrix to define modules of highly correlated genes. We then assessed the success of this clustering, and iteratively reassigned genes to the modules in which they fit best. Lowly expressed or uncorrelated genes were not assigned a module. We identified which genes were assigned to modules in each of the 8 datasets. We then compared the GTEx and STARNET module assignments for each depot and sex; we found genes assigned to modules in both datasets in the 4 depot and sex groups. We then took the union set of these 4 gene sets as the co-expressed gene set (Supplemental Table 2.2).

2.4.2.2 KLF14 trans-eQTL Network Genes:

KLF14 predicted target genes were determined previously[86]. Single nucleotide polymorphism (SNP) rs4731702 is significantly associated with KLF14 expression in *cis* in adipose tissue of multiple cohorts[86,89]. The same variant is also associated with the expression of 385 genes across the genome in *trans.*(Supplemental Table 2.2).

2.4.2.3 WHR_{adjBMI} GWAS loci-adjacent genes:

The largest WHR_{adjBMI} GWAS meta-analysis to date was performed in primarily European ancestry and discovered 346 loci associated with WHR_{adjBMI}[30]. Multiple sources have determined that the functional gene is the nearest gene to the locus in ~70% of cases[87], so we identified genes overlapping or nearest to the lead SNP (and SNPs with LD $r^2 > 0.8$) of each WHR_{adjBMI} GWAS loci using haploReg[95]. Further, we used 2 studies that identified high quality candidate genes using colocalization methods[88,89], where the SNPs that affect association with WHR_{adjBMI}

also affect the expression of the candidate gene, which is more likely to be functional (Supplemental Table 2.2).

The union set of WGCNA module genes, KLF14 targets, and putative GWAS genes made up the input to Bayesian Network construction. For each dataset, the 8,492 gene expression values were discretized into “low” “medium” and “high” bins using k-means clustering.

2.4.3 Prior Information:

Since multiple graph structures can result in the same likelihood score, we can use prior information to improve confidence in the network structure.

2.4.3.1 eQTLs:

For each dataset, we determined which genes had *cis*-eQTLs with SNPs < ± 500 kb. These eGenes were more likely to be parent nodes in the Bayesian networks. Neither STARNET nor GTEx determined *cis*-eQTLs in a sex specific manner, so these eQTL eGenes were nearly identical for male and female networks (Supplemental Table 2.3).

2.4.3.1 Continuous Data:

Although the network is built on discretized gene expression data, RIMBANET is able to use continuous gene expression data to inform network construction. First, the continuous data is used to generate Pearson correlations between all genes. Correlations with significance $p < 0.01$ are used as prior information to determine possible parents and prioritize which edges to add or remove.

2.4.4 Bayesian Network Construction:

Bayesian Networks for each dataset were constructed using RIMBANET using the discretized gene expression data, a list of eQTL eGenes, and the continuous gene expression data. The RIMBANET shell script was adapted for implementation on the University of Virginia’s high-performance computational cluster (Rivanna). RIMBANET was run with these tags: –C TRUE to specify continuous data, -w to add the continuous dataset, -d to add the discretized dataset, -e to add eQTL eGenes. RIMBANET scores networks using the BIC score and uses the 1,000 network iterations. It creates a consensus network by retaining edges present in 30% of the iterations. Finally, RIMBANET produces a directed acyclic graph by removing complete cycles.

2.4.5 Properties of Biological Networks:

Scale-free networks are have edge probability distributions that follow the power law[5]. We used the igraph() package in R to calculate the degree exponent for each network. Small world networks are highly clustered, yet have a short average distance between nodes[8]. We used the qgraph() package in R to calculate the clustering coefficient and average path length between nodes for each network. Since these properties scale with the number of nodes in the network, we compared these metrics to a random graph of the same size.

2.4.6 Key Driver Analysis:

Key driver genes of each network were identified with two methods: 1. by the number of downstream genes regulated by each potential key driver gene and 2. by the enrichment of disease genes in the set of downstream genes regulated by each potential key driver gene.

To identify type 1 key driver gene testing, first, every gene in the network was profiled to determine its number of downstream genes at distances 1-10 edges away using the shortest path. Then, the mean and standard deviation in the number of downstream genes at each edge distance was calculated for the network.

For each potential key driver gene, we calculated ten score functions:

$$Score_{dist=n} = (G_{dist=n} - NM_{dist=n})/NS_{dist=n}$$

where n is the distance in edges, G is the number of downstream genes the potential key driver gene has at distance n, NM is the network mean number of downstream genes at distance n, NS is the network standard deviation in the number of downstream genes at distance n. This is a metric of the extremeness of G, effectively a z-score.

Finally, we calculated a total score for each potential key driver gene by summing the ‘z-scores’, weighting smaller edge distances away from the potential key driver higher than large edge distances:

$$GeneScore = Score_{dist=1} \cdot 1 + Score_{dist=2} \cdot 1 + Score_{dist=3} \cdot 1 + Score_{dist=4} \cdot 0.75 + Score_{dist=5} \cdot 0.5 + Score_{dist=6} \cdot 0.25 + Score_{dist=7} \cdot 0.125 + Score_{dist=8} \cdot 0.0625 + Score_{dist=9} \cdot 0.03125$$

The top 10% highest scoring genes in the network were declared type 1 key driver genes, similar to previous studies[16].

To identify type 2 key driver genes, a ‘neighborhood’ of downstream genes was declared for every gene in the network as those genes within 4 edges away. Next, $\mathrm{WHR}_{\mathrm{adjBMI}}$ GWAS genes[30], defined above, were identified within the downstream neighborhood. Finally, using Fisher’s exact test, we determined whether the number of GWAS genes in the downstream neighborhood was significantly more than expected by chance, compared to the whole network. Genes with significant downstream enrichments for $\mathrm{WHR}_{\mathrm{adjBMI}}$ GWAS genes were declared type 2 key driver genes.

Key driver genes for each network were the union set of type 1 and type 2 key driver genes. Shared key driver genes were genes identified as either type 1 or type 2 key driver genes in both STARNET and GTEx networks of the same type.

2.4.7 Data and code availability:

Gene expression and eQTL data from GTEx can be found at dbGaP Accession phs000424.v8.p2on 10/01/2020. Gene expression and eQTL data STARNET can be found at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001203.v3.p1, data available on request. Code used in these analyses can be found at https://github.com/jnr3hh/Reed_Civelek_2023_manuscript.

Acknowledgements:

This work was supported by NIH/NIDDK grant R01 DK118287 and NIH training grant T32 HL007284.

2.5 Bibliography:

1. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**, 101–113 (2004).
2. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc Natl Acad Sci U S A* **99**, 7821–7826 (2002).
3. Dančík, V., Basu, A. & Clemons, P. Properties of Biological Networks. in *Systems Biology: Integrative Biology and Simulation Tools* (eds. Prokop, A. & Csukás, B.) 129–178 (Springer Netherlands, 2013). doi:10.1007/978-94-007-6803-1_5.
4. Patra, S. & Mohapatra, A. Review of tools and algorithms for network motif discovery in biological networks. *IET Syst Biol* **14**, 171–189 (2020).
5. Barabasi, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
6. Zhang, Z. & Zhang, J. A Big World Inside Small-World Networks. *PLOS ONE* **4**, e5686 (2009).
7. Amaral, L. A. N., Scala, A., Barthélémy, M. & Stanley, H. E. Classes of small-world networks. *Proceedings of the National Academy of Sciences* **97**, 11149–11152 (2000).
8. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
9. Ren, Y., Ay, A. & Kahveci, T. Shortest path counting in probabilistic biological networks. *BMC Bioinformatics* **19**, 465 (2018).
10. Liu, X. *et al.* Computational methods for identifying the critical nodes in biological networks. *Brief Bioinform* **21**, 486–497 (2020).
11. Chagoyen, M., Ranea, J. A. G. & Pazos, F. Applications of molecular networks in biomedicine. *Biol Methods Protoc* **4**, bpz012 (2019).
12. Civelek, M. & Lusis, A. J. Systems genetics approaches to understand complex traits. *Nat Rev Genet* **15**, 34–48 (2014).
13. Bhalla, U. S. & Iyengar, R. Emergent properties of networks of biological signaling pathways. *Science* **283**, 381–387 (1999).
14. Dipple, K. M., Phelan, J. K. & McCabe, E. R. Consequences of complexity within biological networks: robustness and health, or vulnerability and disease. *Mol Genet Metab* **74**, 45–50 (2001).
15. Schadt, E. E. & Lum, P. Y. Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. Reverse engineering gene networks to identify key drivers of complex disease phenotypes. *J Lipid Res* **47**, 2601–2613 (2006).
16. Zhang, B. *et al.* Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707–720 (2013).
17. Slater, O., Miller, B. & Kontoyianni, M. Decoding Protein-protein Interactions: An Overview. *Curr Top Med Chem* **20**, 855–882 (2020).
18. Badia-I-Mompel, P. *et al.* Gene regulatory network inference in the era of single-cell multi-omics. *Nat Rev Genet* (2023) doi:10.1038/s41576-023-00618-5.
19. Dai, H., Jin, Q.-Q., Li, L. & Chen, L.-N. Reconstructing gene regulatory networks in single-cell transcriptomic data analysis. *Zool Res* **41**, 599–604 (2020).
20. Kim, H. U., Sohn, S. B. & Lee, S. Y. Metabolic network modeling and simulation for drug targeting and discovery. *Biotechnol J* **7**, 330–342 (2012).
21. Yilmaz, L. S. & Walhout, A. J. Metabolic network modeling with model organisms. *Curr Opin Chem Biol* **36**, 32–39 (2017).
22. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607–D613 (2019).
23. Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res* **50**, D687–D692 (2022).
24. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353–D361 (2017).
25. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
26. Friedman, N., Linial, M., Nachman, I. & Pe'er, D. Using Bayesian networks to analyze expression data. *J Comput Biol* **7**, 601–620 (2000).

27. Han, X. *et al.* Construction of a human cell landscape at single-cell level. *Nature* **581**, 303–309 (2020).
28. Hu, Y., Peng, T., Gao, L. & Tan, K. CytoTalk: De novo construction of signal transduction networks using single-cell transcriptomic data. *Sci Adv* **7**, eabf1356 (2021).
29. de Klerk, E. & 't Hoen, P. A. C. Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. *Trends Genet* **31**, 128–139 (2015).
30. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* **28**, 166–174 (2019).
31. Bouchard, C. Genetics of Obesity: What We Have Learned Over Decades of Research. *Obesity (Silver Spring)* **29**, 802–820 (2021).
32. van Dam, S., Võsa, U., van der Graaf, A., Franke, L. & de Magalhães, J. P. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform* **19**, 575–592 (2018).
33. Chowdhury, H. A., Bhattacharyya, D. K. & Kalita, J. K. (Differential) Co-Expression Analysis of Gene Expression: A Survey of Best Practices. *IEEE/ACM Trans Comput Biol Bioinform* **17**, 1154–1173 (2020).
34. Dawson, H. D. *et al.* Structural and functional annotation of the porcine immunome. *BMC Genomics* **14**, 332 (2013).
35. H, B. *et al.* Large-scale gene co-expression network as a source of functional annotation for cattle genes. *BMC genomics* **17**, (2016).
36. Liu, W. *et al.* Functional Annotation of Caenorhabditis elegans Genes by Analysis of Gene Co-Expression Networks. *Biomolecules* **8**, 70 (2018).
37. Childs, K. L., Davidson, R. M. & Buell, C. R. Gene coexpression network analysis as a source of functional annotation for rice genes. *PLoS One* **6**, e22196 (2011).
38. Rocha, J. J. *et al.* Functional unknomics: Systematic screening of conserved genes of unknown function. *PLOS Biology* **21**, e3002222 (2023).
39. Wong, D. *et al.* FHL5 Controls Vascular Disease-Associated Gene Programs in Smooth Muscle Cells. *Circ Res* **132**, 1144–1161 (2023).
40. Filteau, M., Pavey, S. A., St-Cyr, J. & Bernatchez, L. Gene coexpression networks reveal key drivers of phenotypic divergence in lake whitefish. *Mol Biol Evol* **30**, 1384–1396 (2013).
41. Udyavar, A. R. *et al.* Co-expression network analysis identifies Spleen Tyrosine Kinase (SYK) as a candidate oncogenic driver in a subset of small-cell lung cancer. *BMC Syst Biol* **7 Suppl 5**, S1 (2013).
42. Zhang, H., Zhao, X., Wang, M. & Ji, W. Key modules and hub genes identified by coexpression network analysis for revealing novel biomarkers for larynx squamous cell carcinoma. *J Cell Biochem* **120**, 19832–19840 (2019).
43. Wang, Q., Roy, B. & Dwivedi, Y. Co-expression network modeling identifies key long non-coding RNA and mRNA modules in altering molecular phenotype to develop stress-induced depression in rats. *Transl Psychiatry* **9**, 125 (2019).
44. Bagot, R. C. *et al.* Circuit-wide Transcriptional Profiling Reveals Brain Region-Specific Gene Networks Regulating Depression Susceptibility. *Neuron* **90**, 969–983 (2016).
45. Haas, B. E. *et al.* Adipose co-expression networks across Finns and Mexicans identify novel triglyceride-associated genes. *BMC Med Genomics* **5**, 61 (2012).
46. Oldham, M. C., Horvath, S. & Geschwind, D. H. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci U S A* **103**, 17973–17978 (2006).
47. Talukdar, H. A. *et al.* Cross-Tissue Regulatory Gene Networks in Coronary Artery Disease. *Cell Syst* **2**, 196–208 (2016).
48. Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45–54 (2019).
49. White, A. & Vignes, M. Causal Queries from Observational Data in Biological Systems via Bayesian Networks: An Empirical Study in Small Networks. *Methods Mol Biol* **1883**, 111–142 (2019).
50. Arora, P. *et al.* Bayesian Networks for Risk Prediction Using Real-World Data: A Tool for Precision Medicine. *Value Health* **22**, 439–445 (2019).
51. Sinha, S. Reproducibility of parameter learning with missing observations in naive Wnt Bayesian network trained on colorectal cancer samples and doxycycline-treated cell lines. *Mol Biosyst* **11**, 1802–1819 (2015).

52. Seixas, F. L., Zadrozny, B., Laks, J., Conci, A. & Muchaluat Saade, D. C. A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer׳s disease and mild cognitive impairment. *Comput Biol Med* **51**, 140–158 (2014).
53. Beretta, S. *et al.* Learning the Structure of Bayesian Networks: A Quantitative Assessment of the Effect of Different Algorithmic Schemes. *Complex.* **2018**, (2018).
54. de Campos, C. P., Zeng, Z. & Ji, Q. Structure learning of Bayesian networks using constraints. in *Proceedings of the 26th Annual International Conference on Machine Learning* 113–120 (Association for Computing Machinery, 2009). doi:10.1145/1553374.1553389.
55. Tsamardinos, I., Brown, L. E. & Aliferis, C. F. The max-min hill-climbing Bayesian network structure learning algorithm. *Mach Learn* **65**, 31–78 (2006).
56. Zhu, J. *et al.* Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet* **40**, 854–861 (2008).
57. Lin, L. *et al.* Temporal genetic association and temporal genetic causality methods for dissecting complex networks. *Nat Commun* **9**, 3980 (2018).
58. Zhu, J. *et al.* Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. *PLoS Comput Biol* **3**, e69 (2007).
59. Zhang, L., Rodrigues, L. O., Narain, N. R. & Akmaev, V. R. bAIcis: A Novel Bayesian Network Structural Learning Algorithm and Its Comprehensive Performance Evaluation Against Open-Source Software. *J Comput Biol* **27**, 698–708 (2020).
60. Zhu, J. *et al.* An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet Genome Res* **105**, 363–374 (2004).
61. Li, H., Wu, G., Zhang, J. & Yang, N. Identification of the heart-type fatty acid-binding protein as a major gene for chicken fatty acid metabolism by Bayesian network analysis. *Poult Sci* **89**, 1825–1833 (2010).
62. Zhang, G. *et al.* Differential metabolic and multi-tissue transcriptomic responses to fructose consumption among genetically diverse mice. *Biochim Biophys Acta Mol Basis Dis* **1866**, 165569 (2020).
63. Chella Krishnan, K. *et al.* Integration of Multi-omics Data from Mouse Diversity Panel Highlights Mitochondrial Dysfunction in Non-alcoholic Fatty Liver Disease. *Cell Syst* **6**, 103-115.e7 (2018).
64. Shu, L. *et al.* Prenatal Bisphenol A Exposure in Mice Induces Multitissue Multiomics Disruptions Linking to Cardiometabolic Disorders. *Endocrinology* **160**, 409–429 (2019).
65. Perry, R. N., Albarracin, D., Aherrahrou, R. & Civelek, M. Network Preservation Analysis Reveals Dysregulated Metabolic Pathways in Human Vascular Smooth Muscle Cell Phenotypic Switching. *Circ Genom Precis Med* **16**, 372–381 (2023).
66. Huan, T. *et al.* A systems biology framework identifies molecular underpinnings of coronary heart disease. *Arterioscler Thromb Vasc Biol* **33**, 1427–1434 (2013).
67. Preciados, M., Yoo, C. & Roy, D. Estrogenic Endocrine Disrupting Chemicals Influencing NRF1 Regulated Gene Networks in the Development of Complex Human Brain Diseases. *Int J Mol Sci* **17**, 2086 (2016).
68. Yi, Z. *et al.* Key driver genes as potential therapeutic targets in renal allograft rejection. *JCI Insight* **5**, e136220, 136220 (2020).
69. Ogami, K. *et al.* Computational gene network analysis reveals TNF-induced angiogenesis. *BMC Syst Biol* **6 Suppl 2**, S12 (2012).
70. Li, Y. *et al.* A Bayesian gene network reveals insight into the JAK-STAT pathway in systemic lupus erythematosus. *PLoS One* **14**, e0225651 (2019).
71. Watson, C. T. *et al.* Integrative transcriptomic analysis reveals key drivers of acute peanut allergic reactions. *Nat Commun* **8**, 1943 (2017).
72. Carcamo-Orive, I. *et al.* Predictive network modeling in human induced pluripotent stem cells identifies key driver genes for insulin responsiveness. *PLoS Comput Biol* **16**, e1008491 (2020).
73. Shu, L. *et al.* Shared genetic regulatory networks for cardiovascular disease and type 2 diabetes in multiple populations of diverse ethnicities in the United States. *PLoS Genet* **13**, e1007040 (2017).
74. Mäkinen, V.-P. *et al.* Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease. *PLoS Genet* **10**, e1004502 (2014).
75. Hartman, R. J. G. *et al.* Sex-Stratified Gene Regulatory Networks Reveal Female Key Driver Genes of Atherosclerosis Involved in Smooth Muscle Cell Phenotype Switching. *Circulation* **143**, 713–726 (2021).

76. Huan, T. *et al.* Integrative network analysis reveals molecular mechanisms of blood pressure regulation. *Mol Syst Biol* **11**, 799 (2015).
77. McKenzie, A. T. *et al.* Multiscale network modeling of oligodendrocytes reveals molecular components of myelin dysregulation in Alzheimer's disease. *Mol Neurodegener* **12**, 82 (2017).
78. Hamed, M., Spaniol, C., Zapp, A. & Helms, V. Integrative network-based approach identifies key genetic elements in breast invasive carcinoma. *BMC Genomics* **16 Suppl 5**, S2 (2015).
79. Acerbi, E., Zelante, T., Narang, V. & Stella, F. Gene network inference using continuous time Bayesian networks: a comparative study and application to Th17 cell differentiation. *BMC Bioinformatics* **15**, 387 (2014).
80. Sieberts, S. K. & Schadt, E. E. Moving toward a system genetics view of disease. *Mamm Genome* **18**, 389–401 (2007).
81. Shu, L. *et al.* Mergeomics: multidimensional data integration to identify pathogenic perturbations to biological systems. *BMC Genomics* **17**, 874 (2016).
82. Cohain, A. *et al.* Exploring the reproducibility of probabilistic causal molecular network models. in *Biocomputing 2017* 120–131 (WORLD SCIENTIFIC, 2016). doi:10.1142/9789813207813_0013.
83. Zhang, B. & Zhu, J. Identification of Key Causal Regulators in Gene Networks. in (2013).
84. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from dbGaP Accession phs000424.v8.p2on 10/01/2020.
85. Franzén, O. *et al.* Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**, 827–830 (2016).
86. Small, K. S. *et al.* Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nat Genet* **50**, 572–580 (2018).
87. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
88. Raulerson, C. K. *et al.* Adipose Tissue Gene Expression Associations Reveal Hundreds of Candidate Genes for Cardiometabolic Traits. *Am J Hum Genet* **105**, 773–787 (2019).
89. Civelek, M. *et al.* Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *Am J Hum Genet* **100**, 428–443 (2017).
90. Yang, Q. *et al.* Adipocyte-Specific Modulation of KLF14 Expression in Mice Leads to Sex-Dependent Impacts on Adiposity and Lipid Metabolism. *Diabetes* **71**, 677–693 (2022).
91. Cohain, A. *et al.* EXPLORING THE REPRODUCIBILITY OF PROBABILISTIC CAUSAL MOLECULAR NETWORK MODELS. *Pac Symp Biocomput* **22**, 120–131 (2017).
92. Malone, J. H. & Oliver, B. Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biology* **9**, 34 (2011).
93. Xu, X. *et al.* Parallel comparison of Illumina RNA-Seq and Affymetrix microarray platforms on transcriptomic profiles generated from 5-aza-deoxy-cytidine treated HT-29 colon cancer cells and simulated datasets. *BMC Bioinformatics* **14 Suppl 9**, S1 (2013).
94. Greenfest-Allen, E., Cartailler, J.-P., Magnuson, M. A. & Stoeckert, C. J. iterativeWGCNA: iterative refinement to improve module detection from WGCNA co-expression networks. 234062 Preprint at https://doi.org/10.1101/234062 (2017).
95. Ward, L. D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* **44**, D877-881 (2016).

Chapter 3: Identifying and experimentally validating genes that cause fat storage in adipocytes

Parts of this chapter were adapted from a manuscript, originally published on bioRxiv, (023.09.06.556534; doi: https://doi.org/10.1101/2023.09.06.556534). The authors are listed below.

Systems genetics analysis of human body fat distribution genes identifies Wnt signaling and mitochondrial activity in adipocytes

Jordan N Reed[1,2,] Jiansheng Huang[3], Yong Li[3], Lijiang Ma[4], Dhanush Banka[1],Martin Wabitsch[5], Tianfang Wang[3], Wen Ding[3], Johan L.M. Björkegren[4,6], Mete Civelek[1,2]

1. Department of Biomedical Engineering, University of Virginia. Charlottesville, Virginia.
2. Center for Public Health Genomics, University of Virginia. Charlottesville, Virginia.
3. Novo Nordisk Research Center China, Novo Nordisk A/S, Beijing, China.
4. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai. New York, New York.
5. Division of Paediatric Endocrinology and Diabetes, Department of Paediatrics and Adolescent Medicine, Ulm University Medical Centre, Ulm, Germany.
6. Department of Medicine, Karolinska Institutet, Huddinge, Stockholm, Sweden.

3.1 Introduction:

Around the completion of the sequencing of the human genome in 2010[1], researchers undertook the first genome-wide association studies (GWAS) to identify single nucleotide variation (SNP)s associated with disease[2,3]. Some were surprised to find that the identification of mechanistically causal genes and even causal SNPs was less than trivial[4]. Knowing the mechanism of action of the causal gene can facilitate drug development, while knowing the causal SNP can help identify upstream regulators of the gene and can be used to identify individuals at higher risk for disease.

3.1.1 Problems associated with GWAS follow-up-

3.1.1.1 Haplotype Structure:

The human genome is fragmented into haplotype blocks causing groups of SNPs in high linkage disequilibrium (LD) with each other[5–7]. If one SNP is significantly associated with the trait of interest, likely other SNPs in high LD will also be strongly associated[4]. Before using experimental validation, researchers employ a variety of computational 'fine-mapping' approaches to determine which of the many SNPs in the locus is likely causal[8]. First, since different ancestral groups have different haplotype structures, we can see which SNPs are significantly associated in multiple populations[9]. Specifically, the African genome is evolutionarily the oldest and has had more time to fragment into smaller haplotype blocks, meaning fewer SNPs are in LD with each other[10,11]. This requires recruiting more diverse populations into GWA studies and more participants to be powered to test more SNP-trait associations[8]. Other computational approaches use summary statistics or information about the chromatin structure to predict which SNPs are more likely to be causal[8]. Both approaches can help narrow tens or hundreds of SNPs in a locus to a handful that can be tested experimentally.

3.1.1.2 Gene-dense loci:

Most loci contain multiple gene bodies, as well as non-coding RNAs and anti-sense transcripts[4]. While the majority of SNPs identified in GWAS regulate the nearest gene, a subset of SNPs regulate more distal genes. The *FTO* locus contains SNPs nearby the gene body *FTO*, but the SNPs regulate the expression of distal genes *IRX3* and *IRX5*[12]. Further, some genes in the locus have a clear mechanistic link to the disease or trait of interest, while some require phenotyping to determine how they contribute to the disease in the cell type or tissue of interest. eQTL data can help determine the gene of interest at the locus[4]. If the same SNPs are associated with the trait of interest and with the expression of a nearby gene, one could hypothesize that the gene might be regulatory of the trait. Colocalization methods statistically determine if the same SNPs are associated with both[13–15]. This requires robust eQTL data, which can be hard to come by. We need to know the tissue in which the genes likely act and have access to a biopsy of that tissue in living or recently deceased people. In diabetes, it can be hard to determine in which of the many metabolic organs that gene could act, while in cardiovascular disease, it is difficult to obtain major arteries of living people.

3.1.1.3 European study populations:

As mentioned before, much of the genotyping, gene expression, and multi-omics data in most diseases are primarily collected in European ancestry populations[16]. As such, conclusions about which SNPs and genes are causal for disease may only apply to patients of European ancestry[17]. Due to genetic drift, the European population contains a subset of the genetic variation present in older African genomes[18]. We know that SNPs have different allele frequencies in different populations[19], and that different populations have different LD structures that predict different sets of associated SNPs[5,11]. By extending our GWAS and eQTL cohorts to include individuals of diverse ancestry, we can identify more SNPs and genes associated with disease. Further, the findings in these studies will be easier to fine-map and extensible to more individuals.

3.1.1.4 Non-linear effects:

After completing the first few GWAS, researchers tried to use the associated SNPs to account for the heritability of each disease[20], but found that SNPs could only explain a portion of the heritability. Recruiting more participants and increasing the power could not fully explain the "missing heritability"[21,22]. It is possible that non-linear effects between SNPs, genes, and confounding variables could explain the heritability. For example, we often consider an additive model of SNP effects, however, there are often cases where a certain gene is only deleterious in a certain context, be that another SNP's allele, another gene's expression, a risk factor's presence, etc[23]. In the case of phenylketonuria, the disease is only deleterious when the patient eats a high protein diet[24]. When the diet is modified to exclude excess phenylalanine, these patients function normally. Other genes may not be regulated directly by SNPs, but still contribute to disease processes[23].

3.1.2 Experimental Validation-

3.1.2.1 SNP-to-Gene:

Early methods to test whether the mechanistic link between SNP alleles and the expression of a gene of interest co-opted the luminescent protein luciferin[25]. The luciferin gene is added to a plasmid construct under regulation of the endogenous enhancer or promoter containing the SNP of interest. Both alleles are tested to see if one version can produce more luciferin and therefore has stronger regulatory activity. While these studies were illuminating, they are time consuming to repeat for multiple SNPs or genes of interest. Further, they use sequences expressed from an artificial plasmid, which can't replicate the complex, nonlinear effects of the 3-D chromatin structure.

To edit the endogenous SNP, researchers use the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system[26]. A guide-RNA targets the location of the SNP and a

Cas9 enzyme makes the edit. Both alleles are tested to see if expression of the target gene is affected by the SNP. Although this system uses the endogenous chromatin structure, it is still very time consuming to edit and test multiple SNPs and genes.

To decrease the time associated with identifying the causal SNP, researchers developed massively parallel reporter assays (MPRA)[27]. This assay relies on plasmid constructs to insert a putative enhancer sequence with one allele for one SNP of interest, as well as a minimal promoter and reporter gene, and barcode sequences to label the cells. The plasmids are inserted into the cells of interest, and we can measure the expression of the barcoded reporter genes to infer the activity of each allele. While this system does not use the endogenous chromatin structure, it can profile hundreds or thousands of SNPs at once, making it an attractive option to narrow the candidate SNP list.

3.1.2.2 Gene-to-Phenotype:

To test whether there is a mechanistic link between gene expression and the phenotype or disease of interest, researchers employ *in vitro* and *in vivo* models. Usually, these studies involve modulating the expression of the gene of interest, and observing the resulting phenotypic changes[3,26].

Using a cell type of interest, researchers can use pharmacologic or chemical agents to lower the expression of the gene of interest. If none exist, small interfering RNA molecules (siRNA) can be used to target and deplete the gene of interest[28–30]. Some cells easily take up plasmid DNA[31,32], while some cells require lenti-virus to introduce the new constructs[33,34]. Conversely, we can add extra copies of the gene, usually on a plasmid under a constitutive promoter. The success of these studies depends on using a cell type relevant to the disease of interest and on measuring phenotypes that contribute to the disease process.

In fat distribution studies, the genes of interest usually act in adipose tissue and adipocytes[35–37]. Unfortunately, endlessly replicating human pre-adipocyte cell lines do not exist[38]. Instead, we use pre-adipocytes isolated from the subcutaneous fat of a male with Simpson-Golabi-Behmel Syndrome (SGBS cells)[39]. SGBS cells can proliferate and differentiate into mature adipocytes, up to passage ~45-50. After, these cells lose their ability to store lipids and differentiate into mature adipocytes, and can no longer be used as pre-adipocytes. Some researchers instead choose the mouse fibroblast cell line NIH 3T3-L1, which can take up plasmid DNA and can differentiate into mature adipocytes at any passage number[40]. While there are differences between mouse and human biology, these cells share similar features to human adipocytes that make them an useful resource for high throughput experiments[41,42]. Finally, other researchers choose to use primary pre-adipocytes isolated from human adipose tissues[43]. These cells can be hard to obtain, rapidly lose the ability to differentiate when cultured in a dish, and can exhibit large differences between individuals that introduce noise into the data. Subcutaneous adipose can be biopsied easily, while obtaining visceral adipose requires opening the abdominal cavity and is collected during surgery. However, these cells are necessary when studying females or visceral cells, since to date, no female pre-adipocyte or visceral pre-adipocyte cell lines exist[38].

Using cell lines, one can make conclusions about cell behavior and inferences about how that behavior contributes to disease. In obesity and fat distribution studies, the cellular phenotype is often increased or decreased fat storage, which clearly contributes to disease processes[38]. We often look at which fat storage processes (adipogenesis, lipogenesis, lipolysis, etc) are altered by the gene of interest. For other diseases, it may be less clear how the cellular phenotype contributes. For example, a gene that increases smooth muscle cell migration may help or hinder atherosclerosis progression, and further studies are required to link the cell behavior to disease processes[44].

In vitro models, though more time consuming to produce, allow us to look holistically at the effect of the gene on disease phenotypes[45]. Mouse models are commonly used, as mice grow quickly and recapitulate many of the phenotypes seen in humans. Although there are differences in fat storage between mice and humans, there are many conserved pathways that point to shared genetic mechanisms and similar biological outcomes[41,42]. In mice, the subcutaneous fat depot is supra-abdominal, but studies confirm it plays the same physiological role in healthy fat storage as human subcutaneous fat. Mice also have permanent brown fat depots behind the neck, while most adult humans do not, though most humans do have the same thermogenic capacity as mice. Finally, female mice store less fat mass and more lean mass than male mice, though the opposite is true in humans[42]. Despite these differences, there are large similarities in fat storage processes and outcomes between humans and mice, making them an effective model of fat storage[42].

Mouse models were used to identify leptin (*ob/ob*)[46] and the leptin receptor (*db/db*)[47], and these models are still used to study genetic obesity. High fat, high sugar, or high cholesterol diets are also used to induce obesity in mice[45], and are often used to promote atherosclerosis in mice[48]. To test the effects of a gene in a mouse model, that gene is knocked out or extra copies are added. Mice are usually bred in Mendelian ratios, such that wild type litter mates can be compared directly to the tested mice. Using these mice, we can look at body weight, fat and lean mass, glucose and insulin tolerance, fat pad size, adipocyte size, and other phenotypes[45]. Further, we can isolate the pre-adipocytes from these mice and perform cellular phenotyping, as described above.

Other animal models of fat storage exist, notably zebrafish, flies, and worms, which can be used to rapidly profile phenotypes[49]. For example, zebrafish were used to confirm that the deletion of gene *RSPO3* shifted fat storage from abdominal to peripheral fat storage. While these fat pads do not fully correspond with human anatomy, the results aligned well with the human data and provided additional evidence that *RSPO3* affects fat distribution[50]*. C. elegans* worms only store fat in one location, but can be used for rapid screening of potentially obesogenic genes[51]. Ideally, all studies in model organisms would be followed up or supported by human data.

3.1.2.3 Challenges of studying fat distribution:

Fat distribution is a complex, full body phenotype that is difficult to recapitulate with *in vitro* and *in vivo* models. Because it depends on differences between two tissues, fully characterizing a fat distribution gene requires testing its effects in both depots. Due to the lack of visceral cell lines, one must either use a model organism with multiple fat depots or primary human pre-adipocytes[38,45,49]. The time and money required, the difficulty acquiring or producing the cells or mouse, and the lack of available cell lines all hinder the progress of fat distribution research.

3.1.2.4 Previous experimental validation in fat distribution and fat storage:

All of the previously validated fat distribution genes (*RSPO3, LRP5, KLF14, SHOX2*, and *TBX15*) effect pathways crucial to the expansion of subcutaneous and visceral adipocyte populations[50,52–57]. *RSPO3* and *LRP5* studies used primary human pre-adipocytes isolated from both depots, while *KLF14, SHOX2*, and *TBX15* relied on mouse cells isolated from both depots. *LRP5* and *RSPO3* affect adipocyte differentiation by controlling Wnt signaling, while *TBX15* controls adipocyte differentiation and mitochondrial function. While these are the only examples of genes with proven effects on depot-specific fat storage, researchers have shown that many other genes control adipocyte fat storage in one depot or cell type.

The Wnt signaling pathway is a well-established driver of cell fate, differentiation, and proliferation in many cell types; Wnt inhibits adipogenic differentiation by transcriptionally upregulating osteogenic genes while downregulating *PPARG* and *CEBPA.* In many contexts, the non-canonical Ca^{2+} form of Wnt signaling is inhibitory of the canonical Wnt pathway. Wnt signaling activity is positively associated with visceral adiposity, and many Wnt pathway genes, especially ligands and receptors used in Ca^{2+} non-canonical Wnt signaling, are differentially expressed between fat depots.

Mitochondrial function correlates strongly with cardio-metabolic diseases and can alter adipogenic differentiation. In mature adipocytes, mitochondria can facilitate physical connections with the lipid droplet, dissipate excess energy via thermogenesis through UCP1, and promote lipid homeostasis. Human visceral fat has increased mitochondrial activity compared to subcutaneous fat, and in multiple metabolic disease states, only visceral mitochondria become dysregulated. We hypothesize that other putative drivers of fat distribution affect Wnt signaling or mitochondrial function in adipocytes, with different outcomes in each depot.

In this chapter, we applied some of these *in vitro* models to study select key driver gene function in adipocyte processes (Figure 3.1). First, we used curated, publicly available data to prioritize testable, likely functional genes within the set of key drivers identified in Chapter 2. Then, we modulated the expression of the key driver genes in pre-adipocyte cells, and measured the effects on adipogenesis and mitochondrial activity.

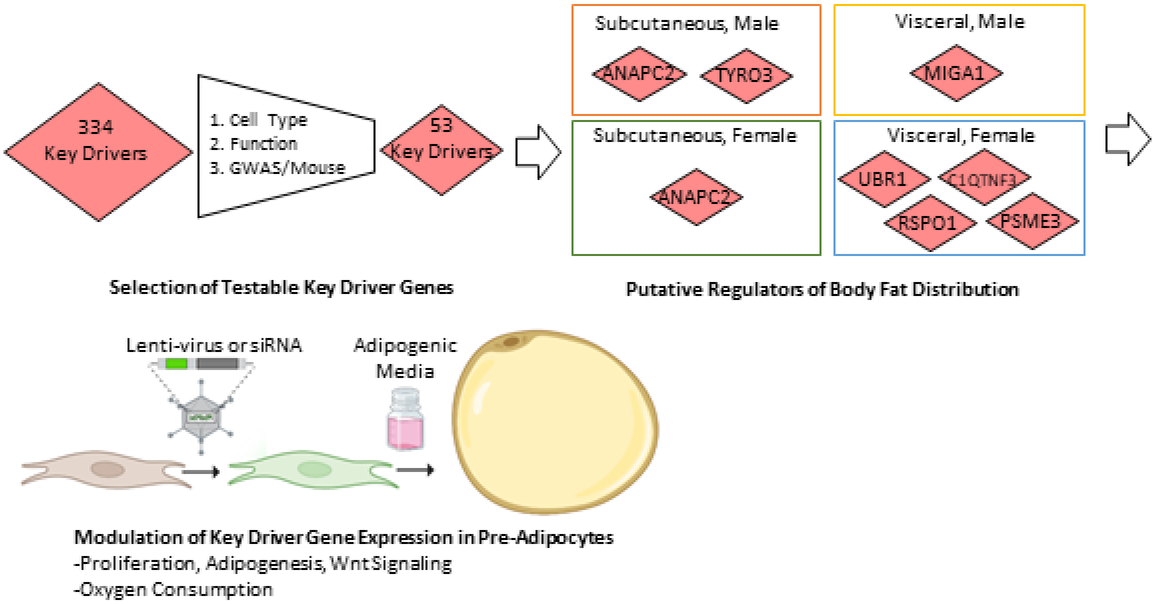

Figure 3.1: Overview of key driver gene functional validation: Key driver genes shared between STARNET and GTEx were identified in each network in chapter 2. 53 key driver genes expressed in (pre-)adipocytes but unstudied in adipose tissue were prioritized for further study. Seven selected key driver genes identified were perturbed in human pre-adipocyte cells, and functional readouts of adipogenesis, Wnt signaling, proliferation, and mitochondrial oxygen consumption were collected.

Figure created with BioRender

3.2 Results:

3.2.1 Prioritization pipeline identifies 53 novel and putatively functional adipocyte and pre-adipocyte key driver genes:

While the set of 334 replicated key driver genes (Supplemental Table 2.6) likely contains many novel mechanistic drivers of fat distribution or fat storage, it also likely contains false positives and well-characterized genes. Further, we hypothesize that fat distribution is driven, mainly, by adipocyte expansion, and that other adipose tissue processes, such as tissue structure, immune function, vascularization, etc, might contribute to fat distribution or its comorbidities by maintaining adipocyte health. We modeled these inter-cell interactions in networks, however we focused our experimentation only on genes likely acting in adipocytes. We employed three steps to narrow this list to likely functional, testable adipocyte key driver genes (Figure 3.2A, Methods). First, using a curated set of six publicly available single-cell and single-nucleus RNA-sequencing studies from human and mouse adipose tissue, we identified the cell types in which each key driver gene was expressed[58–64] and removed 110 genes that were primarily expressed in non (pre-)adipocyte cell types (Figure 3.2B). These genes were primarily expressed in immune cells, smooth muscle cells, and endothelial cells (Figure 3.3). Second, we identified many well-studied genes in the key driver analysis, such as *FGF1, DPP4, LRP6*, and *RXRA*. While this points to the fact that our approach can identify well-known regulators of adipocyte function, we were interested in adding to the body of literature by validating genes unstudied in (pre-)adipocytes. We performed a comprehensive search of the existing literature to identify genes with known function in adipocytes and we removed 45 key driver genes involved in adipocyte processes (Figure 3.2C, Supplemental Bibliography 3.1). Third, we used multiple lines of genetic evidence to prioritize a subset of the remaining genes. We reasoned that the set of 495 $\mathrm{WHR}_{\mathrm{adjBMI}}$ GWAS genes likely contains many genes that play a mechanistic role in fat distribution, and we prioritized 41 $\mathrm{WHR}_{\mathrm{adjBMI}}$ GWAS genes within the 179 remaining key driver genes (Figure 3.2D, Supplemental Table 2.6). Their identity as a regulator of genes within the network and their location nearby a significant GWAS locus is

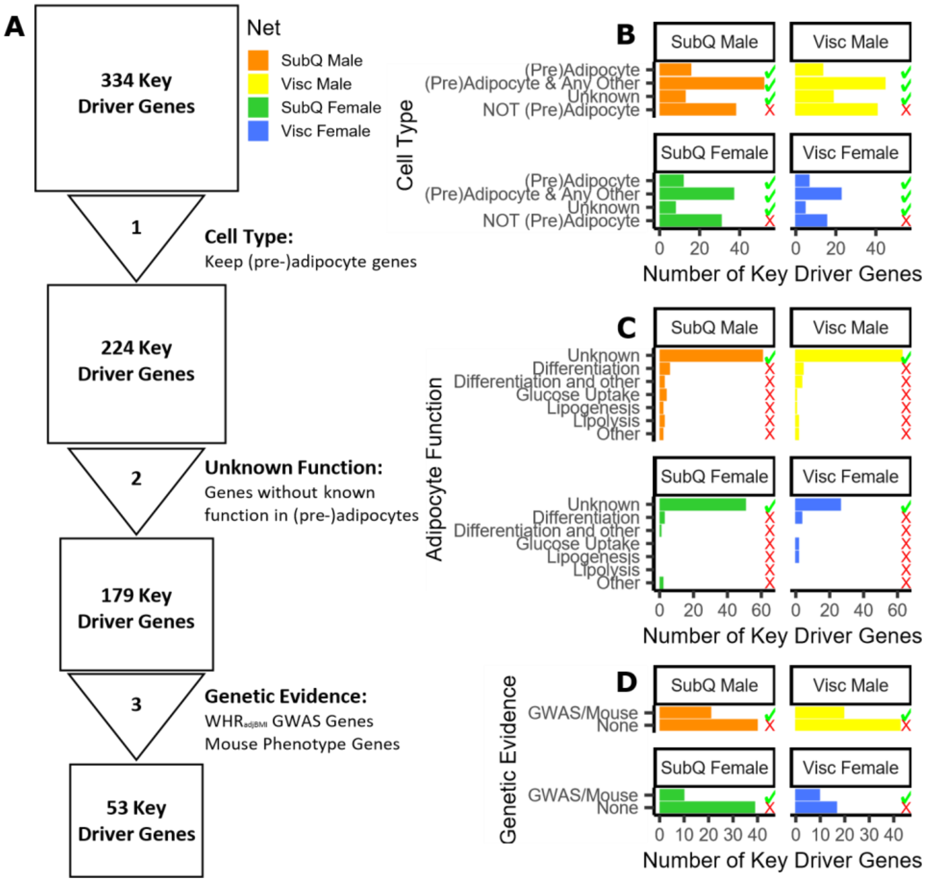

Figure 3.2: Prioritization of key driver genes for functional testing.

(A) 334 key driver genes prioritized to 53 putative candidate regulators of WHR$_{adjBMI}$ and adipocyte function using publicly available data.
(B) Cell types in adipose tissue single cell- and single nucleus- RNA-seq data[58–64] in which each key driver is expressed.
(C) Key driver gene with known function in pre-adipocyte and adipocyte fat storage pathways.
(D) Genetic evidence (status as WHR$_{adjBMI}$ GWAS gene or causes mouse fat storage phenotype) for key driver genes.
Green checkmark indicates genes kept in analysis pipeline, red X indicates genes removed.

strong evidence that they likely have a functional role in fat distribution. Additionally, we queried the functional role these genes play when knocked out in mouse models. Although there are differences in fat storage between mice and humans, there are many conserved pathways that point to shared genetic mechanisms and similar biological outcomes[41,42]. We found that 15 genes, when knocked out in mice, affect fat storage phenotypes, and we hypothesized that they play a similar role in human fat storage (Figure 3.2D). In total, we prioritize 53 key driver genes with putative roles in human fat distribution via altered adipocyte fat storage, that currently have unknown function in adipose tissue (Supplemental Table 2.6).

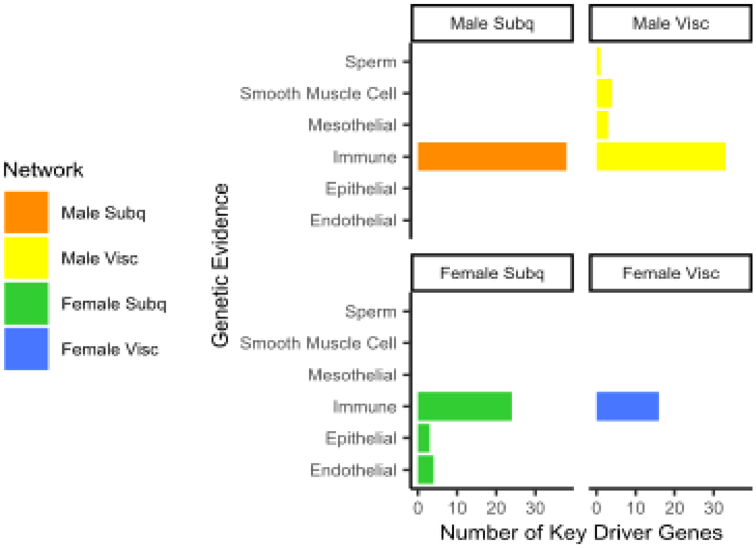

Figure 3.3: The 110 key driver genes removed from analyses due to primary expression in other cell types, in Figure 3.2, step 1.

We applied the same genetic evidence criteria (GWAS gene, mouse phenotype) to the 45 well-studied key driver genes that were removed in step 2 (Figure 3.2C). We found that almost half of these genes were identified without additional evidence from human GWAS or mouse phenotyping (Figure 3.4), showing the strength of this approach as an orthogonal method with which to identify candidate functional genes.

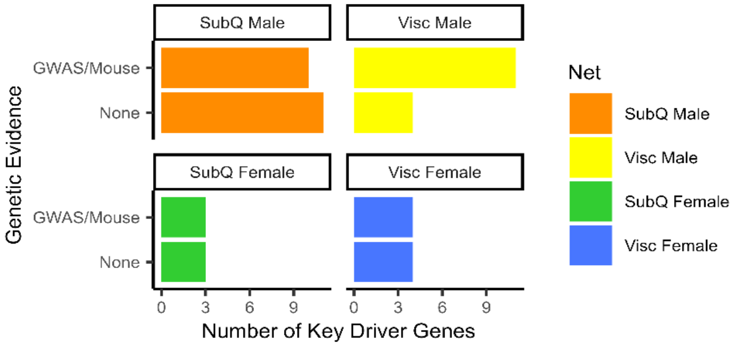

Figure 3.4: The 45 key driver genes with known function in adipose tissue that were removed in Figure 3.2, step 2 were subjected to the Genetic Evidence criteria used in Figure 3.2, step 3.

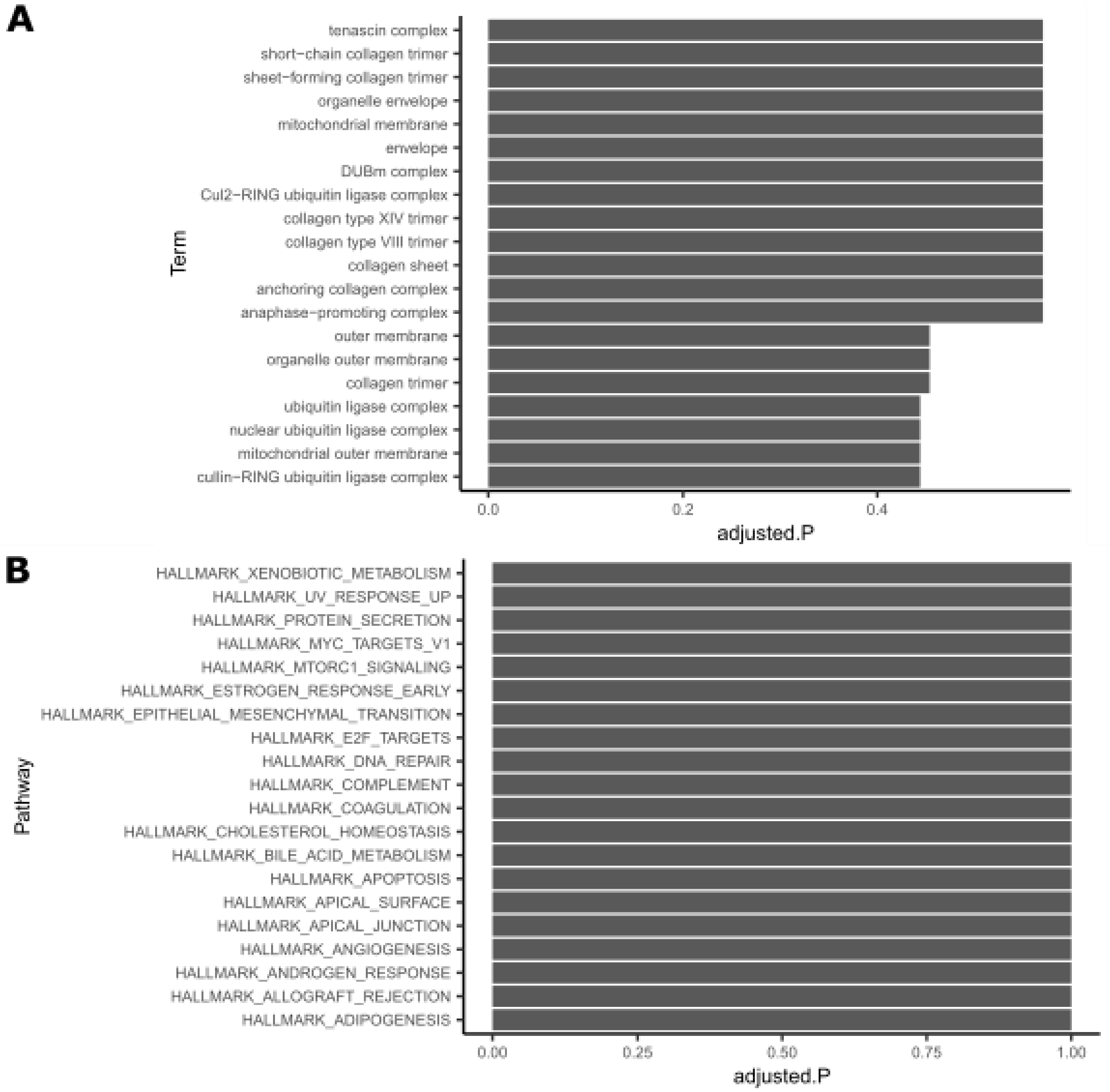

Figure 3.5: The 53 key driver genes prioritized for further study were not enriched for any specific pathways.
(A) The top 20 Gene Ontology (GO) biological processes pathways, tanked by adjusted p-value.
(B) The top 20 msigDB Hallmark pathways, ranked by p-value, adjusted p-value shown. Fisher's exact test used to test enrichments, FDR correction used to adjust p-values.

3.2.2 Four key driver genes in the Wnt signaling pathway are highly correlated with WHR_{adjBMI}:

To further characterize the 53 key driver genes, we attempted to find enrichment of specific pathway genes, but did not identify any enriched gene ontology (GO) terms or msigDB Hallmark pathways, likely due to the removal of well-characterized genes (Figure 3.5). We performed a second literature search to determine the primary function of the 53 genes in other cell types and found that 13 affect the activity of the Wnt signaling pathway in other cell types (Table 3.1, Supplemental Bibliography 3.2). These 13 genes were identified as key driver genes in all four sex-depot networks, and ten are WHR$_{\text{adjBMI}}$ GWAS candidate genes.

Table 3.1: Key driver genes that act in Wnt signaling in other cell types

Key Driver Gene	Network		Public Data	
	Depot	Sex	Cell Type	Evidence
BAZ1B	Subcutaneous	Male	Multiple	GWAS
HELZ	Subcutaneous	Male	Multiple	GWAS
MTMR9	Subcutaneous	Male	Unknown	GWAS
TYRO3	Subcutaneous	Male	Unknown	GWAS
ANTXR1	Visceral	Male	Multiple	GWAS
ARFGEF2	Visceral	Male	Multiple	GWAS
ARMCX3	Visceral	Male	Multiple	Mouse
ANAPC2	Subcutaneous	Both	Adipocyte	GWAS
BNIP2	Subcutaneous	Both	Multiple	Mouse
KIAA1522	Visceral	Female	Unknown	GWAS
PSME3	Visceral	Female	Adipocyte	GWAS
RSPO1	Visceral	Female	Multiple	Mouse
ZNF148	Visceral	Female	Multiple	GWAS

In canonical Wnt signaling, β-catenin and transcription factors TCF/LEF repress *PPARG* and *CEBPA* expression, which are necessary to initiate adipogenesis[65,66] (Figure 3.6A). In other cell types, the 13 gene's proteins interact with the canonical and non-canonical Wnt signaling pathway in a variety of ways (Figure 3.6A, Supplemental Bibliography 3.2). For example, RSPO1 (r-spondin 1) is a known member of the Wnt signaling pathway; like other r-spondins, it prevents LRP degradation at the cell membrane[67,68]. TYRO3 (TYRO3 protein tyrosine kinase) activates Wnt signaling by upregulating AKT[69,70]. PSME3 (proteasome activator subunit 3, REG-ɣ, PA28-ɣ) is known to bind and target GSK3β for degradation, which releases sequestered β-catenin[71,72]. ANAPC2 (anaphase promoting complex 2), interacts with Disheveled to inhibit Wnt signaling[73]. Their role in adipose remains unknown.

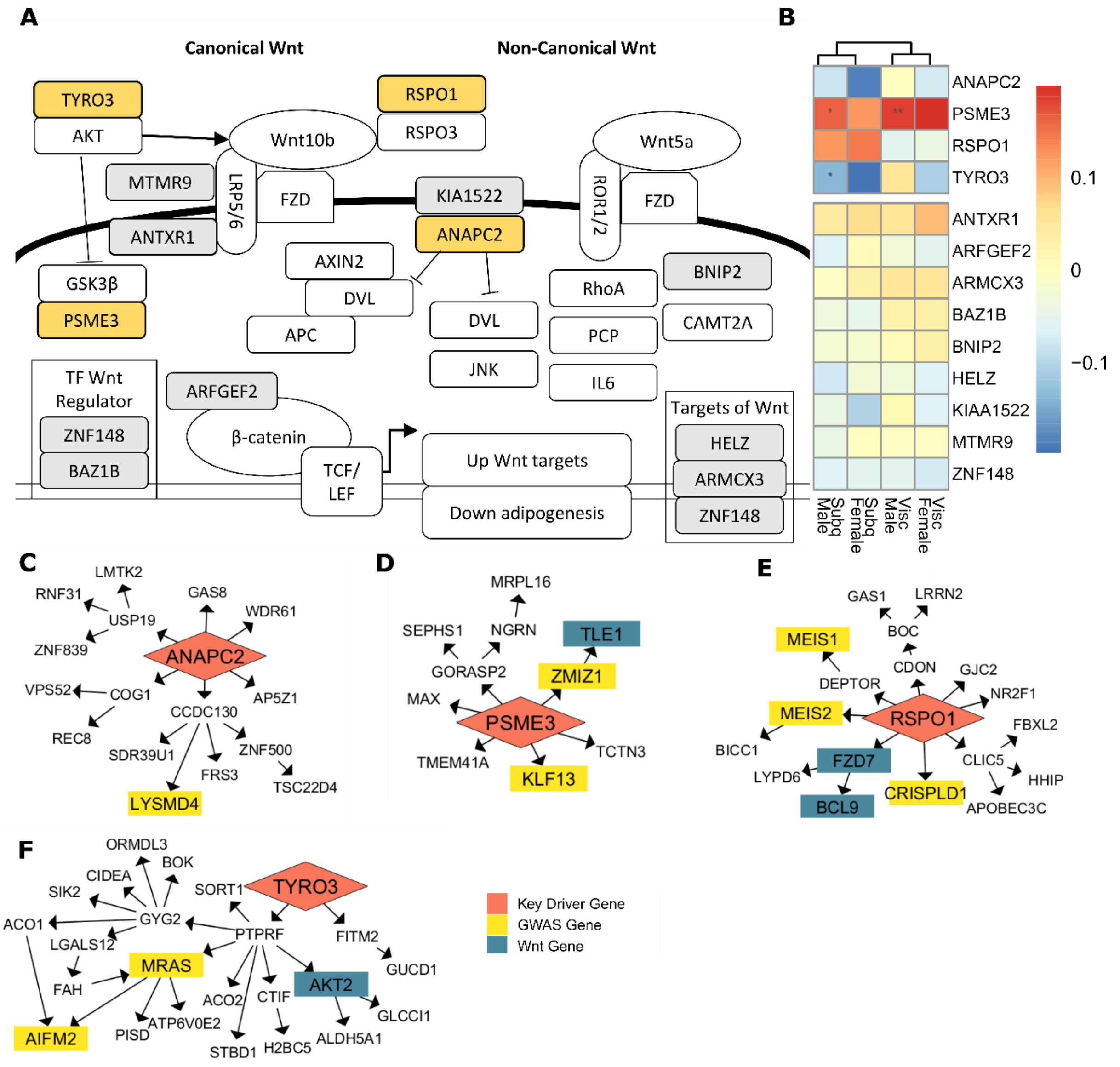

Figure 3.6: Thirteen prioritized key driver genes may affect fat storage in adipocytes through the Wnt signaling pathway.
(A) The Wnt signaling pathway consists of canonical β-catenin signaling and non-canonical pathways. Key driver genes interact with Wnt pathways in other cell types (gray and yellow).
*(B) Key driver gene expression in adipose tissue correlations with WHR_{adjBMI} in STARNET. Pearson correlations are shown by color, p-values adjusted using FDR correction shown with * (*** = adj.P < 0.001, * = adj.P < 0.05).*
Four selected key driver genes (red) regulate both WHR_{adjBMI} downstream genes[74] (yellow) and Wnt signaling downstream genes (blue, GO term "Wnt signaling pathway") in GTEx and STARNET:
(C) ANAPC2 in the GTEX Subcutaneous Female network,
(D) PSME3 in the STARNET Visceral Female network,
(E) RSPO1 in the GTEx Visceral Female network, and
(F) TYRO3 in the GTEx Subcutaneous Male network.

We prioritized genes related to fat distribution by calculating the correlation of each gene's expression in each depot with overall $\text{WHR}_{\text{adjBMI}}$, measured in STARNET, since GTEx did not measure WHR. We found that, while not all are significant, four genes have Pearson correlations with $\text{WHR}_{\text{adjBMI}} > 0.12$ in at least one sex-depot (Figure 3.6B) and these correlations differ by sex or by depot. While these correlations are not numerically large because they incorporate the noise of human data and are adjusted for BMI, the correlations are extreme values for this data (Figure 3.7).

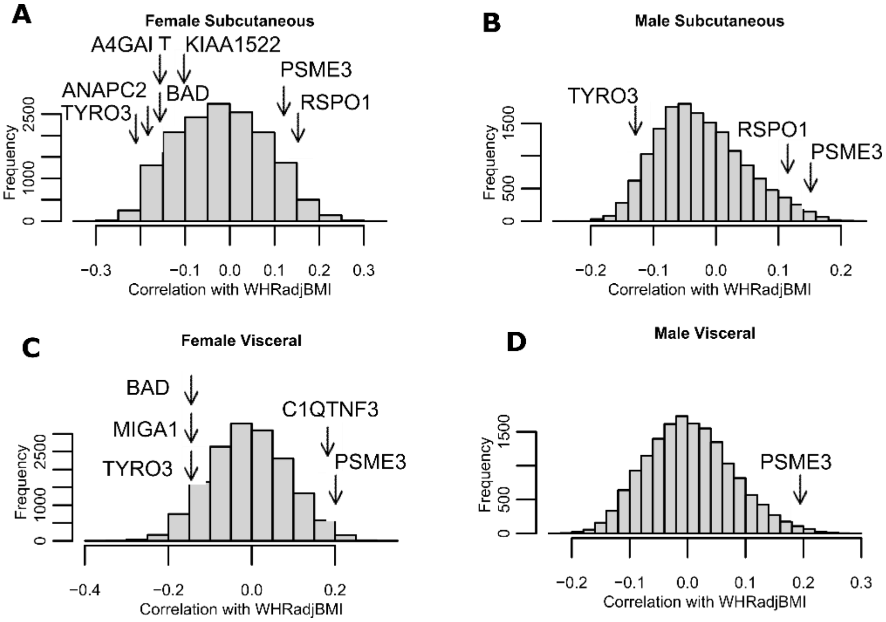

Figure 3.7: Gene expression correlations with $\text{WHR}_{\text{adjBMI}}$ in STARNET in
(A) female subcutaneous,
(B) male subcutaneous,
(C) female visceral, and
(D) male visceral samples.
Arrows indicate prioritized Wnt-related genes (Table 3.1) and mitochondrial-related genes (Table 3.2) whose correlation with $\text{WHR}_{\text{adjBMI}}$ is greater than 0.1.

The four strongly correlated genes, *ANAPC2, PSME3, RSPO1*, and *TYRO3,* regulate a large number of downstream genes in their STARNET and GTEx networks (Figure 3.6 C-F). Those downstream genes contain $\text{WHR}_{\text{adjBMI}}$ GWAS genes (Supplemental Table 2.6), and often contain genes that are part of the Wnt signaling pathway. Additional evidence of involvement in fat distribution, such as expression differences between depots, $\mathrm{WHR}_{\mathrm{adjBMI}}$ GWAS signal strength, and enrichment of relevant GO terms in downstream gene sets further prioritize these four genes (Figures 3.8-3.11). *ANAPC2, PSME3,* and *TYRO3* are all found within gene dense $\mathrm{WHR}_{\mathrm{adjBMI}}$ GWAS loci, and may be the causal gene at the locus, while *RSPO1* is a putative regulator $\mathrm{WHR}_{\mathrm{adjBMI}}$ without genetic regulation. The lead SNP in the *ANAPC2* locus, rs144926207, is located in the intronic region of *ANAPC2*, and reference allele T is associated with higher $\mathrm{WHR}_{\mathrm{adjBMI}}$[74] and higher excision of an intronic region of *ANAPC2* in subcutaneous and visceral depots in GTEx[75] (Figure 3.12).

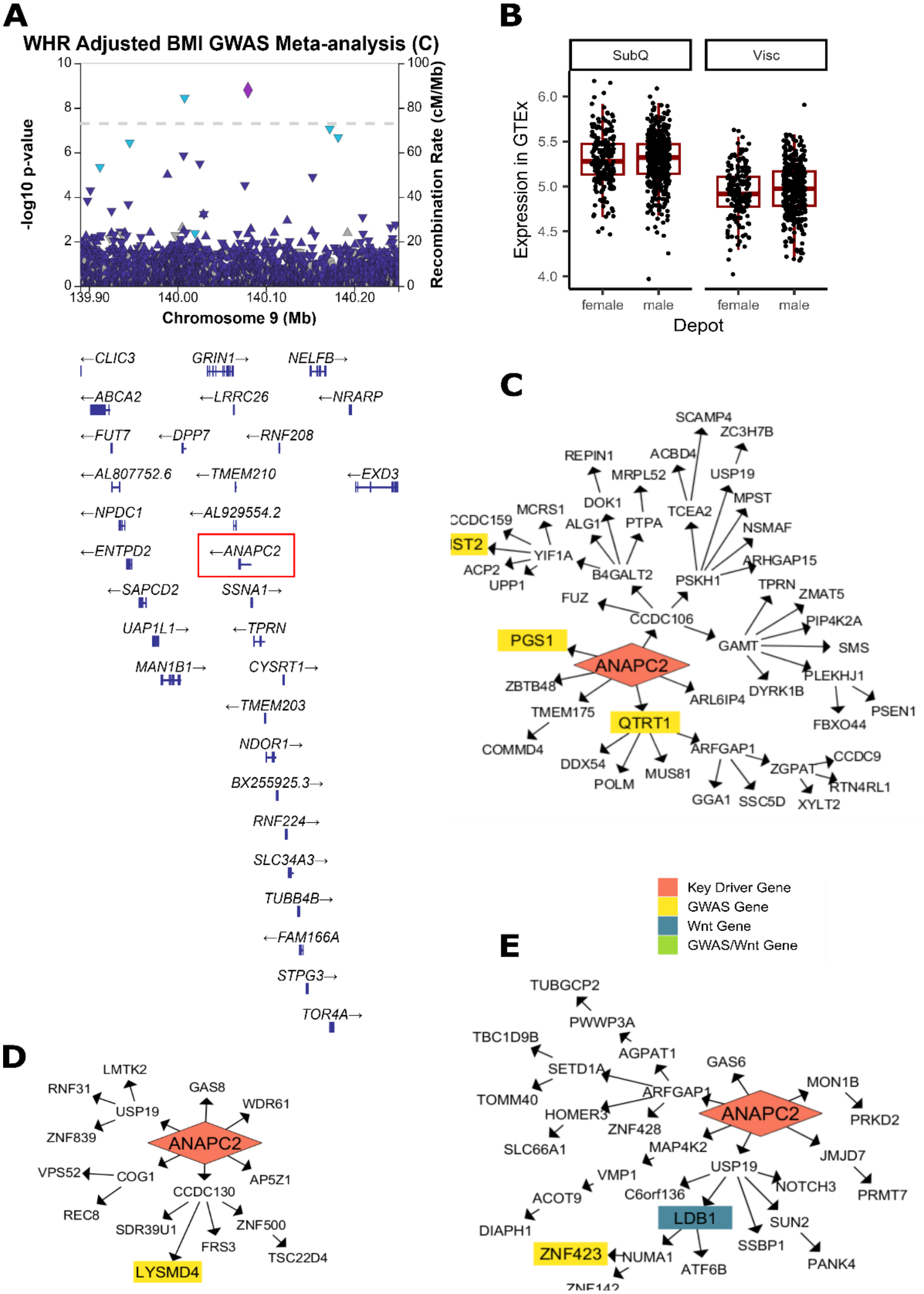

Figure 3.8- Additional evidence of ANAPC2 involvement in WHR_{adjBMI} and Wnt signaling. (A) A significant GWAS signal is detected near ANAPC2 in WHR_{adjBMI} GWAS meta-analysis[1].

(B) In GTEx, both males and females show higher expression of ANAPC2 in subcutaneous fat depots over visceral depots, though the change is non-significant after p-value adjustment.

(C) ANAPC2 is a key driver that regulates downstream WHR_{adjBMI} GWAS genes in the STARNET subcutaneous male network,

(D) the GTEx subcutaneous male network, and

(E) regulates both WHR_{adjBMI} GWAS genes and Wnt genes in STARNET female subcutaneous.

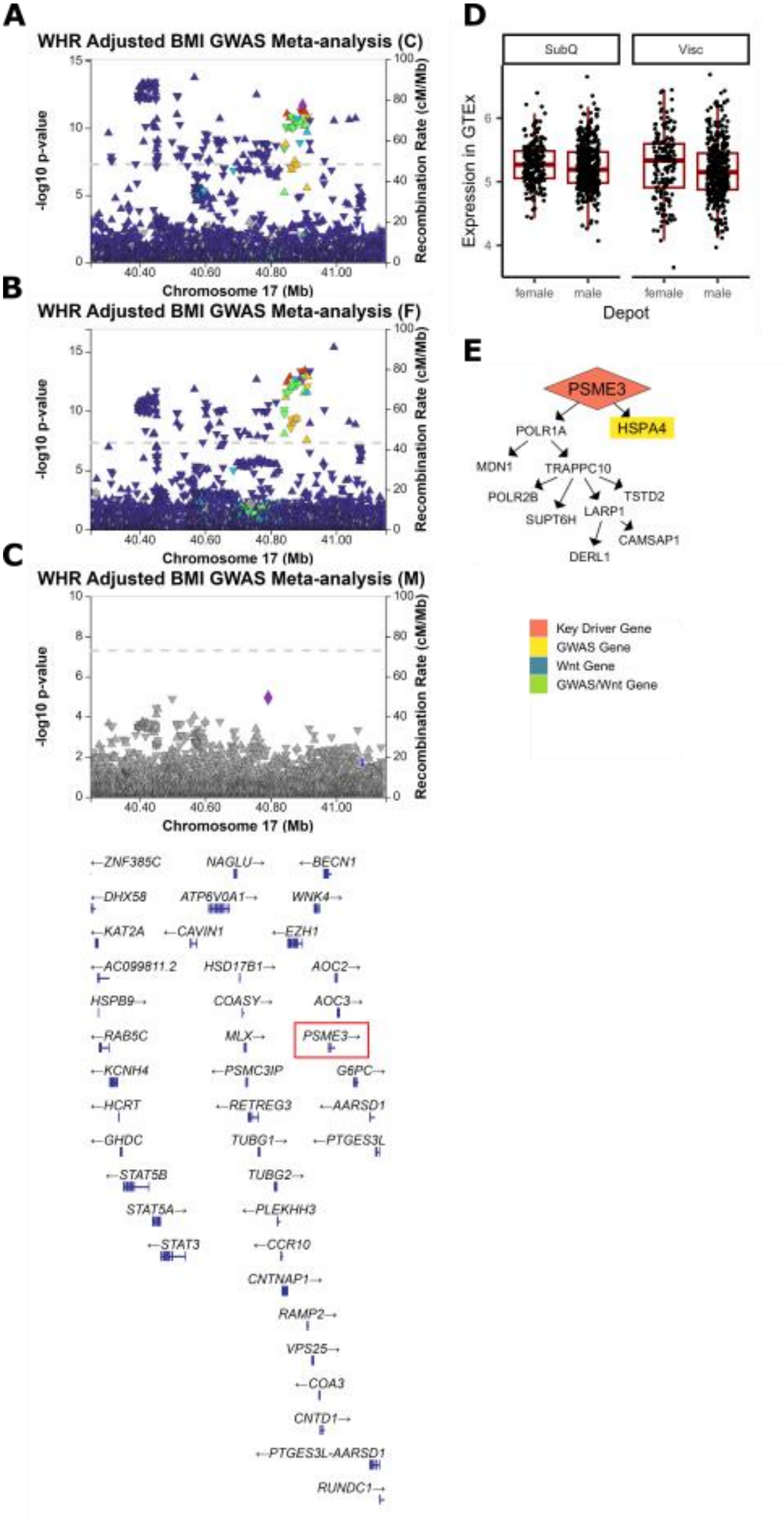
A
WHR Adjusted BMI GWAS Meta-analysis (C)
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 17 (Mb)
B
WHR Adjusted BMI GWAS Meta-analysis (F)
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 17 (Mb)
C
WHR Adjusted BMI GWAS Meta-analysis (M)
-log10 p-value
Recombination Rate (cM/Mb)
Chromosome 17 (Mb)
40.40
40.60
40.80
41.00
←ZNF385C
NAGLU→
←BECN1
←DHX58
ATP6V0A1→
WNK4→
←KAT2A
←CAVIN1
←EZH1
←AC099811.2
HSD17B1→
AOC2→
HSPB9→
COASY→
AOC3→
←RAB5C
MLX→
PSME3→
←KCNH4
←PSMC3IP
G6PC→
←HCRT
←RETREG3
←AARSD1
←GHDC
TUBG1→
←PTGES3L
←STAT5B
TUBG2→
STAT5A→
←PLEKHH3
←STAT3
←CCR10
CNTNAP1→
RAMP2→
VPS25→
←COA3
CNTD1→
←PTGES3L-AARSD1
RUNDC1→
D
SubQ
Visc
Expression in GTEx
female
male
female
male
Depot
E
PSME3
POLR1A
HSPA4
MDN1
TRAPPC10
POLR2B
SUPT6H
LARP1
TSTD2
CAMSAP1
DERL1
Key Driver Gene
GWAS Gene
Wnt Gene
GWAS/Wnt Gene

Figure 3.9- Additional evidence of PSME3 involvement in WHR$_{adjBMI}$.

(A) A significant GWAS signal is detected near PSME3 in WHR$_{adjBMI}$ GWAS meta-analysis[1]*.*
(B) This same signal is strongly associated with WHR$_{adjBMI}$ in the female-specific GWAS meta-analysis,
(C) but is not present in the male-specific GWAS.
(D) In GTEx, there are no significant changes in PSME3 gene expression between depots or sexes.
(E) PSME3 regulates WHR$_{adjBMI}$ GWAS gene in the GTEx visceral female network.

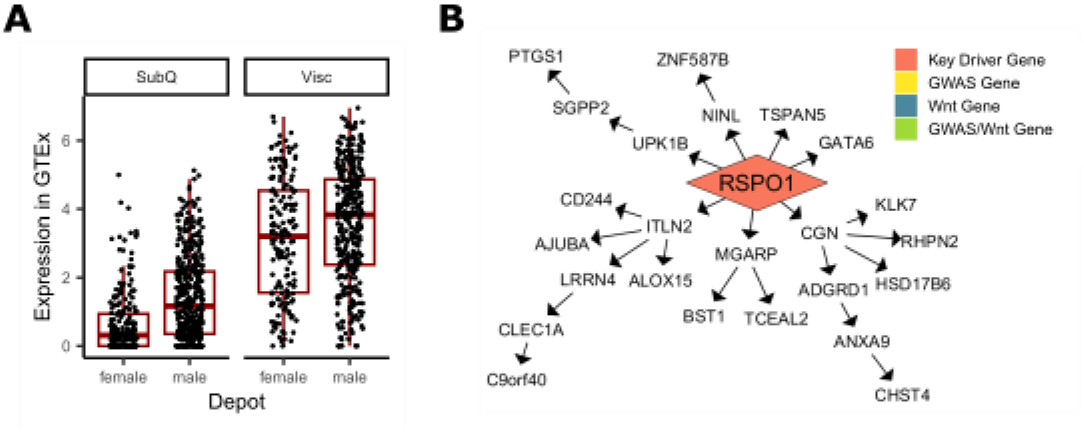

Figure 3.10- Additional evidence of RSPO1 involvement in WHR$_{adjBMI}$.
(A) In GTEx, both sexes show higher expression of RSPO1 in visceral fat depots over subcutaneous depots (adjusted.P = 5.1e-21), and males show higher expression than females in both depots (adjusted.P = 4.9e-57).
(B) RSPO1 is a key driver that regulates many downstream genes in the STARNET female visceral network.

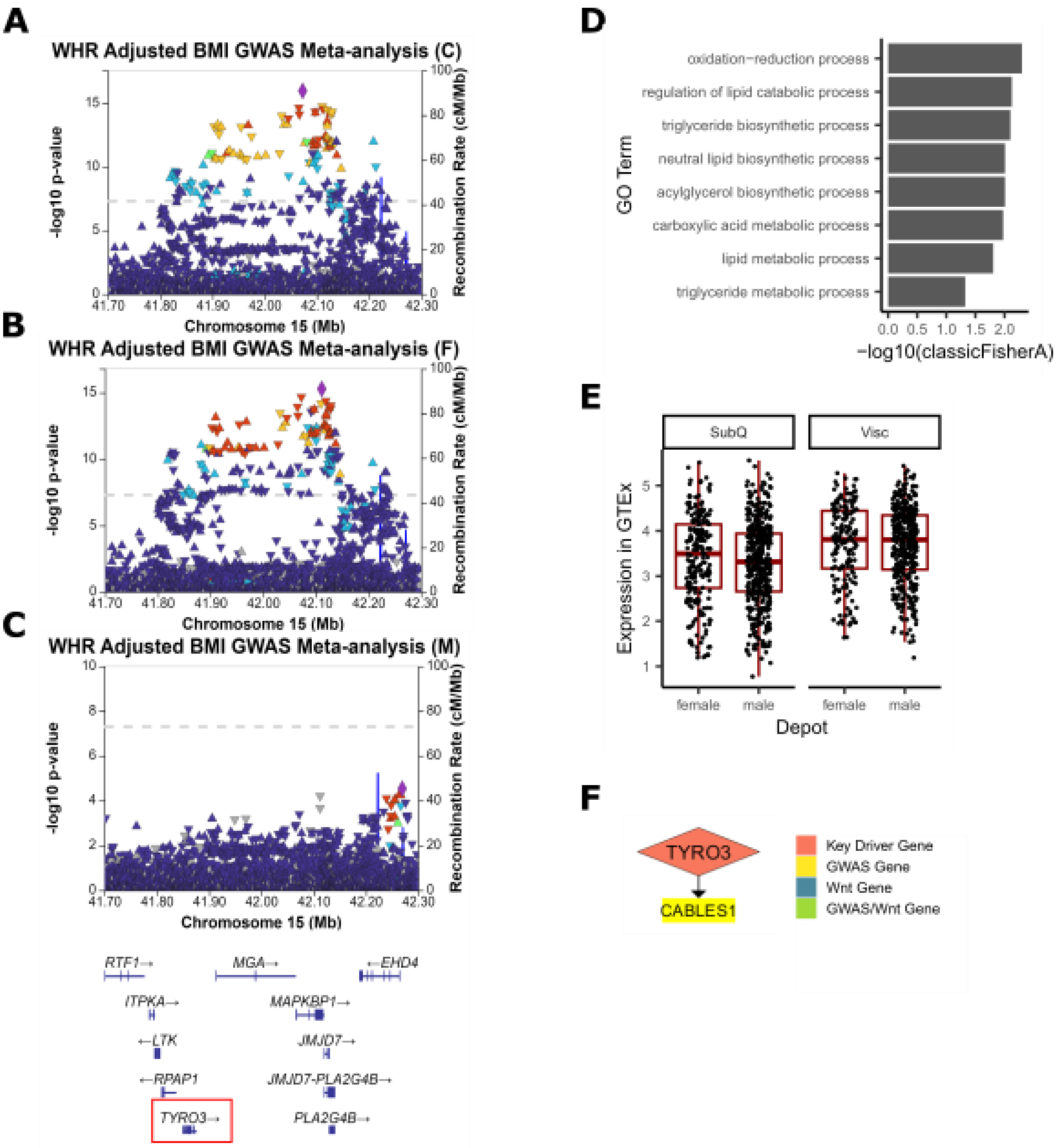

Figure 3.11- Additional evidence of TYRO3 involvement in WHR_{adjBMI}.
(A) A significant GWAS signal is detected near TYRO3 in WHR_{adjBMI} GWAS meta-analysis[1].
(B) This same signal is strongly associated with WHR_{adjBMI} in the female-specific GWAS meta-analysis,
(C) but is not present in the male-specific GWAS.
(D) In GTEx subcutaneous male networks, the genes downstream of TYRO3 are significantly enriched for lipid-related biological processes.
(E) In GTEx, both males and females show higher expression of TYRO3 in visceral fat depots over subcutaneous depots, though the change is non-significant after p-value adjustment.
(F) TYRO3 regulates one WHR_{adjBMI} GWAS gene in the STARNET subcutaneous male network.

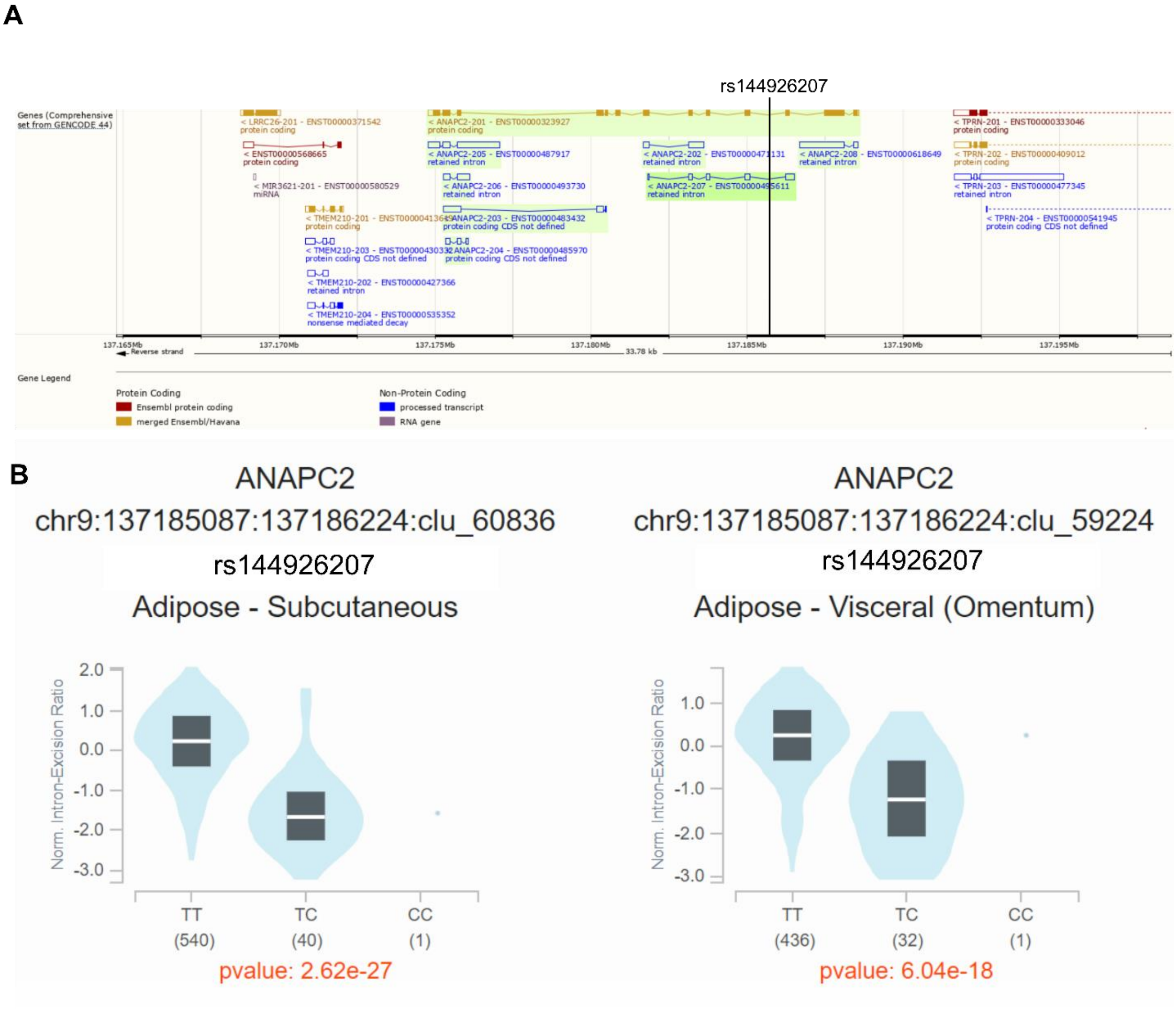

Figure 3.12- Genetic regulation of ANAPC2 intronic excision by WHR$_{adjBMI}$ SNP rs144926297 in GTEx

- *(A) Alternate transcripts of ANAPC2 in GENCODE 44, the lead SNP in the ANAPC2 WHR$_{adjBMI}$, rs144926297 is located within the intron*
- *(B) Normalized intronic excision by genotype in subcutaneous and visceral fat in GTEx*

Both subcutaneous male sub-sampled networks (Table 2.10) recovered *ANAPC2* and *TYRO3* as key driver genes, showing that the subsampled networks are able to replicate the predictions of the original.

By virtue of their status as prioritized network key driver genes and their strong correlation with WHR$_{adjBMI}$ in humans, we hypothesize that these four genes affect fat storage in adipocytes. If the four genes also affect Wnt signaling activity in adipocytes, their effects on overall fat storage may be due to increases or decreases in Wnt signaling activity.

3.2.3 ANAPC2, PSME3, and RSPO1 overexpression alter adipogenesis, not proliferation:

To test these hypotheses, we first overexpressed each of the four Wnt key driver genes or a GFP control plasmid in human male pre-adipocyte cell line[39,76] using lenti-virus (Methods, Figure 3.1). We were not able to perform similar experiments in female and visceral cells since no such cell lines exist at this time. We confirmed the overexpression of each gene compared to GFP controls using qPCR (Figure 3.13 A). We also overexpressed gene *RSPO3* as a positive control, because previous literature shows that *RSPO3* has no effect on subcutaneous pre-adipocyte proliferation, but impairs adipogenesis[50].

We assessed the ability of these cell lines to proliferate by seeding each at the same density and counting representative wells every 24h (Methods). We found no differences in the rate of increase in cell number or the mean doubling time during exponential growth between any lines and GFP controls (Figure 3.13 B, C).

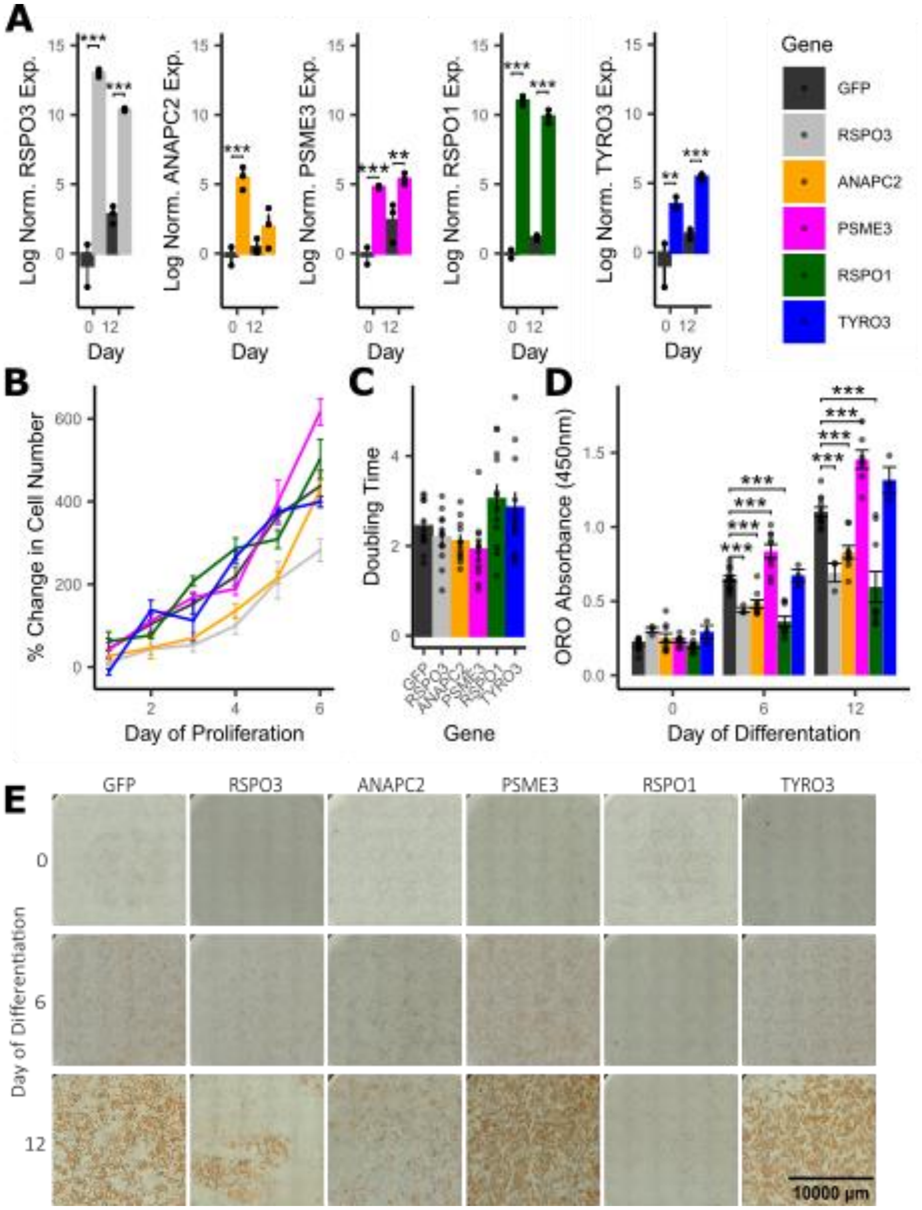

Figure 3.13: RSPO1, PSME3, and ANAPC2 affect fat storage in a pre-adipocyte cell line.
(A) Expression of key driver genes compared to GFP controls at day 0 and 12 after onset of differentiation (n = 3).
(B) Percent change in cell number over 6 days (n = 4).
(C) Calculated doubling time in days (n = 12).
(D, E) Oil Red O staining of cells was performed for each gene of interest and GFP controls at day 0, 6 and 12 after beginning differentiation,
(E) representative images of one well of a 12-well plate are shown (n = 3-15).
*RSPO3 serves as a positive control. All plots show mean ± standard error of the mean. Differences between groups determined using 2-way ANOVA by day and gene (Gene of Interest vs GFP controls), post-hoc tests were performed using pooled t-test with Dunnett's adjustment. Adjusted p-values shown with * (*** = adj.P < 0.001, ** = adj.P < 0.01, * = adj.P < 0.05).*

We quantified the alterations in adipogenesis due to key driver overexpression by staining cells with the neutral lipid specific dye, Oil Red O (ORO), at 0, 6, and 12 days after the onset of differentiation. As expected, we observe an increase in lipid accumulation at day 6 and 12 compared to day 0 in all cell lines, confirming successful induction of adipogenesis, with a 4.25-fold increase in ORO absorbance in GFP controls at day 12 compared to day 0 (Figure 3.13 D,E). Compared to GFP controls, *ANAPC2* and *RSPO1* overexpressing cells were deficient in adipogenesis, with 24.6% and 45.4% less lipid accumulation, respectively, than GFP controls at day 12. We observed that *PSME3* overexpressing cells show a 31.6% increase in lipid accumulation compared to controls at day 12. *TYRO3* overexpression caused no significant differences in lipid accumulation from controls at any time point. Positive control *RSPO3* overexpressing cells were also deficient in lipid accumulation by 37.3% compared to controls at day 12, consistent with previous studies[50]. Expression of adipocyte markers *CEBPA, PPARG,* and *ADIPOQ* also increased over the 12 days, and showed significant differences between *ANAPC2* and *RSPO1* overexpressing cells and controls, in agreement with adipogenesis (Figure 3.14)

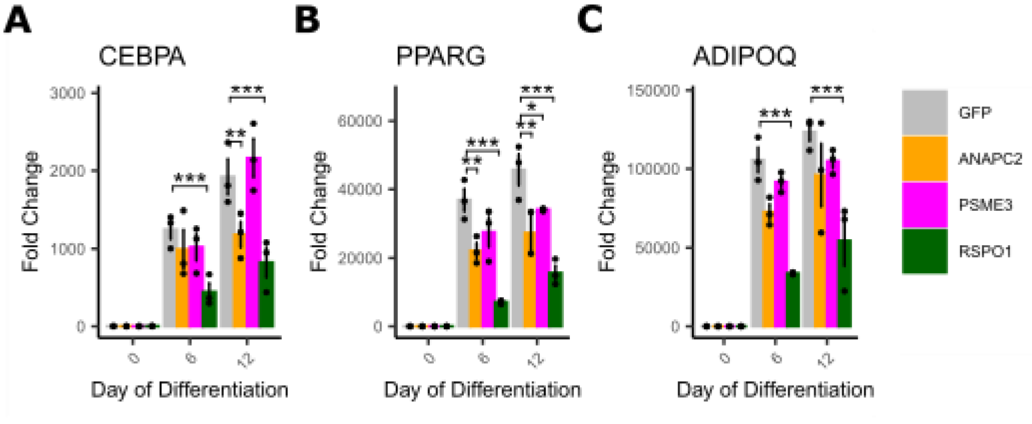

Figure 3.14- Markers of differentiation increase over time. Gene expression of
(A) CEBPA,
(B) PPARG,
*(C) ADIPOQ increase over time, with some significant differences between cells overexpressing genes of interest. n = 3 replicates used in all assays. Differences between groups were determined using 1-way ANOVA within each timepoint by gene (Gene of Interest vs NTCcontrols). All post-hoc tests were performed using pooled t-test with Dunnett's adjustment. Adjusted p-values shown with * (*** = adj.P < 0.001,* = adj.P < 0.05, # = adj.P < 0.1). Error bars show mean ± SEM*

To fully characterize the effects of these genes, we would need to perform the same functional studies in both subcutaneous and visceral cells to determine the magnitude of adipogenic alterations in each cell type and infer effects on body fat distribution. However, pre-adipocyte cell lines from female subcutaneous tissue or any visceral tissue do not exist; therefore, in the absence of this data, we used the correlations with $\text{WHR}_{\text{adjBMI}}$ in STARNET (Figure 3.6 B) to determine if our data could explain relationship between gene expression and body fat distribution. *RSPO1* findings are well aligned with human data – e.g., if RSPO1 inhibits fat storage in both subcutaneous (Figure 3.13 D) and visceral adipocytes, then expression of *RSPO1* in visceral adipose should decrease visceral fat storage and hence decrease $\text{WHR}_{\text{adjBMI}}$, and expression in subcutaneous adipose should decrease subcutaneous fat storage to increase $\text{WHR}_{\text{adjBMI}}$. This is perfectly replicated in STARNET correlation data (Figure 3.6 B) - positive correlations between subcutaneous *RSPO1* expression and $\text{WHR}_{\text{adjBMI}}$, but negative correlations between visceral *RSPO1* expression and $\text{WHR}_{\text{adjBMI}}$. Although *ANAPC2* and *PSME3* findings are not immediately explained by the correlation data, phenotypic effects in visceral adipocytes of different magnitude or direction could explain the observed changes.

<u>*3.2.4 RSPO1 activates Wnt signaling to inhibit adipogenesis*</u>:

We assessed the activity of the canonical Wnt signaling pathway using cells expressing both an overexpression plasmid and the 7TFC-luciferase construct, which measures the output of β-catenin-TCF/LEF transcription via luminescence (Methods). *PSME3* and *RSPO1* overexpressing cells were able to activate canonical Wnt signaling significantly more than GFP controls; luminescence increased by 0.74-fold and 1.34-fold in *PSME3* and *RSPO1* overexpressing cells, respectively (Figure 3.15 A).

Next, we looked at individual molecules of the canonical Wnt signaling pathway[66]. We quantified the ratio of active (non-phosphorylated) β-catenin to total β-catenin using immunoblotting (Figure 3.15 B,C), and we observed that *PSME3* and *RSPO1* overexpressing cells had a 25.4% and

33.1% increase in active/total β-catenin species compared to controls, respectively, consistent with the luciferase reporter assay. We quantified *AXIN2* mRNA expression, a target of canonical Wnt signaling, using qPCR (Figure 3.15 D). We observed that, compared to GFP control cells, cells overexpressing *ANAPC2* and *RSPO1* increased *AXIN2* expression by 1.9-fold and 2.2-fold, respectively. Finally, using ELISAs, we measured the ratio of active to total GSK3β, an inhibitor of canonical Wnt signaling pathway (Figure 3.15 E). Compared to GFP controls, we observed a 28.2% decrease in the ratio of active to total GSK3β in *PSME3* overexpressing cells.

We also assessed the consequences of gene overexpression on non-canonical Wnt signaling[77,78]. We observed a significant 21.7% decrease in the mRNA expression of *IL6*, a target of multiple types of non-canonical Wnt signaling, in *PSME3* overexpressing cells (Figure 3.15 F). *RSPO1* overexpressing cells also show a non-significant 15.8 % decrease in *IL6* expression. Using ELISAs, we measured the ratio of active to total CAMK2A, a member of the Ca^{2+} non-canonical Wnt pathway, and we observed a significant 70.2% and 85.0% decrease in active to total CAMK2A ratio in *PSME3* and *RSPO1* overexpressing cells, respectively (Figure 3.15 G). Finally, we measured the ratio of active to total JNK, a member of the planar cell polarity (PCP) non-canonical Wnt pathway and observed no differences between cell lines (Figure 3.15 H).

RSPO1 appears to play a straightforward role in inhibiting adipogenesis by turning on canonical Wnt signaling. *PSME3* likely has a different mechanism by which promotes adipogenesis, although these cells may experience some of the protective metabolic effects of canonical Wnt signaling. We did not uncover the mechanism by which *ANAPC2* inhibits adipogenesis, but in subcutaneous male cells, we now know that it does not affect proliferation rates or Wnt signaling activity.

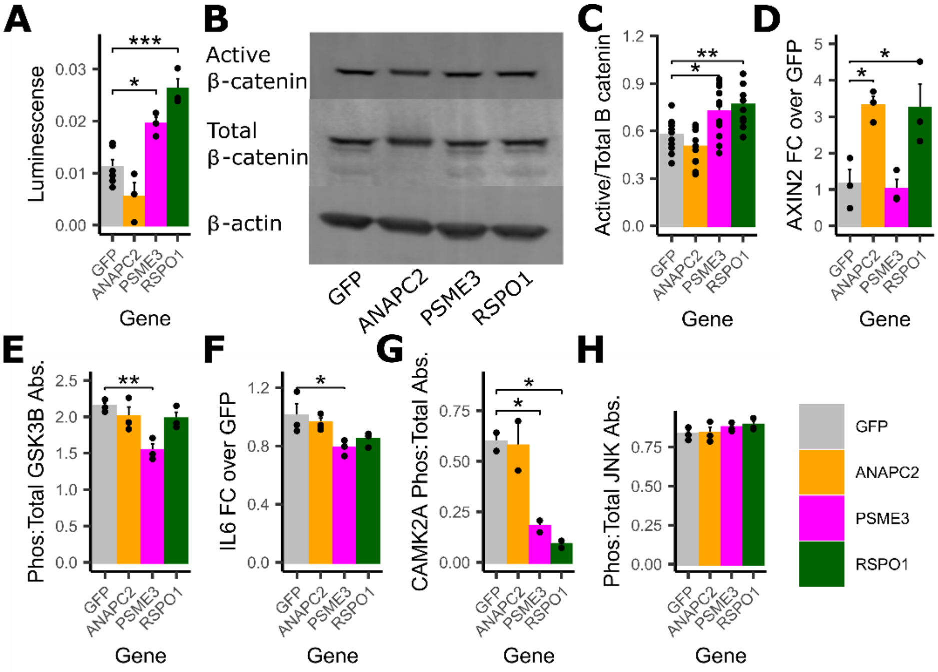

Figure 3.15: RSPO1 and PSME3 activate canonical Wnt signaling while inhibiting the Ca^{2+} non-canonical Wnt pathway.
(A) Wnt transcriptional activity measured by luminescence of luciferase reporter (n = 3-6).
(B) Representative images and
(C) Quantification of active (non-phosphorylated) and total β-catenin by immunoblotting (n = 12).
(D) Gene expression of AXIN2 measured by qPCR (n = 3).
(E) Ratio of active (phosphorylated): total GSK3β measured by ELISA (n = 3).
(F) Gene expression of IL6 measured by qPCR (n = 3).
(G) Ratio of active (phosphorylated): total CAMK2A measured by ELISA (n = 2).
(H) Ratio of active (phosphorylated): total JNK measured by ELISA (n = 3).
*All plots show mean ± standard error of the mean. Differences between groups determined using 1-way ANOVA by gene (Gene of Interest vs GFP controls), post-hoc tests were performed using pooled t-test with Dunnett's adjustment. Adjusted p-values shown with * (*** = adj.P < 0.001, ** = adj.P < 0.01, * = adj.P < 0.05).*

3.2.5 Eight key driver genes related to mitochondrial function are correlated with WHR_{adjBMI} or UCP1 expression:

Among the 53 prioritized key driver genes, we identified a group of 13 genes that alter the function of the mitochondria in other cell types (Table 3.2). These genes were of particular interest, due to the variety of ways mitochondria impact adipocyte function and lipid storage, and the 13 genes affect a variety of mitochondrial functions in other cell types (Supplemental Bibliography 3.3).

Table 3.2: Key drivers that act in mitochondria in other cell types

Key Driver Gene	Network		Public Data	
	Depot	Sex	Cell Type	Evidence
INO80D	Subcutaneous	Male	Adipocyte	Both
SPART	Subcutaneous	Male	Multiple	Mouse
TRIP12	Subcutaneous	Male	Multiple	GWAS
A4GALT	Visceral	Male	Adipocyte	GWAS
ARMCX3	Visceral	Male	Multiple	Mouse
BAD	Visceral	Male	Multiple	GWAS
MIGA1	Visceral	Male	Adipocyte	Mouse
NMT1	Visceral	Male	Multiple	GWAS
YME1L1	Both	Both	Multiple	Mouse
C1QTNF3	Visceral	Female	Adipocyte	GWAS
PSME3	Visceral	Female	Adipocyte	GWAS
UBR1	Visceral	Female	Multiple	Mouse
ZNF148	Visceral	Female	Multiple	GWAS

For example, in vascular smooth muscle cells, cardiomyocytes, and hippocampal neurons, C1QTNF3 (Complement C1q Tumor Necrosis Factor-Related Protein 3) signals through PGC-1 to increase biogenesis, oxygen consumption, and ATP synthesis[79–81]. *MIGA1's* protein (Mitoguardin 1) is localized to the outer mitochondrial membrane (OMM) and promotes mitochondrial fusion[82,83]. *Psme3* knockout mice have larger mitochondria that are structurally dysregulated[84] and fibroblasts overexpressing *Psme3* have elevated Bax, Cytochrome C, with concomitant anti-apoptotic effects[85]. *UBR1* (Ubiquitin Protein Ligase E3 Component N-Recognin 1) encodes an E3-ligase that degrades proteins using the N-end rule[86]. In yeast, UBR1 is localized to the OMM and targets misfolded proteins[87] and in mouse embryonic fibroblasts, UBR1, with UBR2 and UBR4, target PINK1 in the cytosol and prevent mitophagy[88].

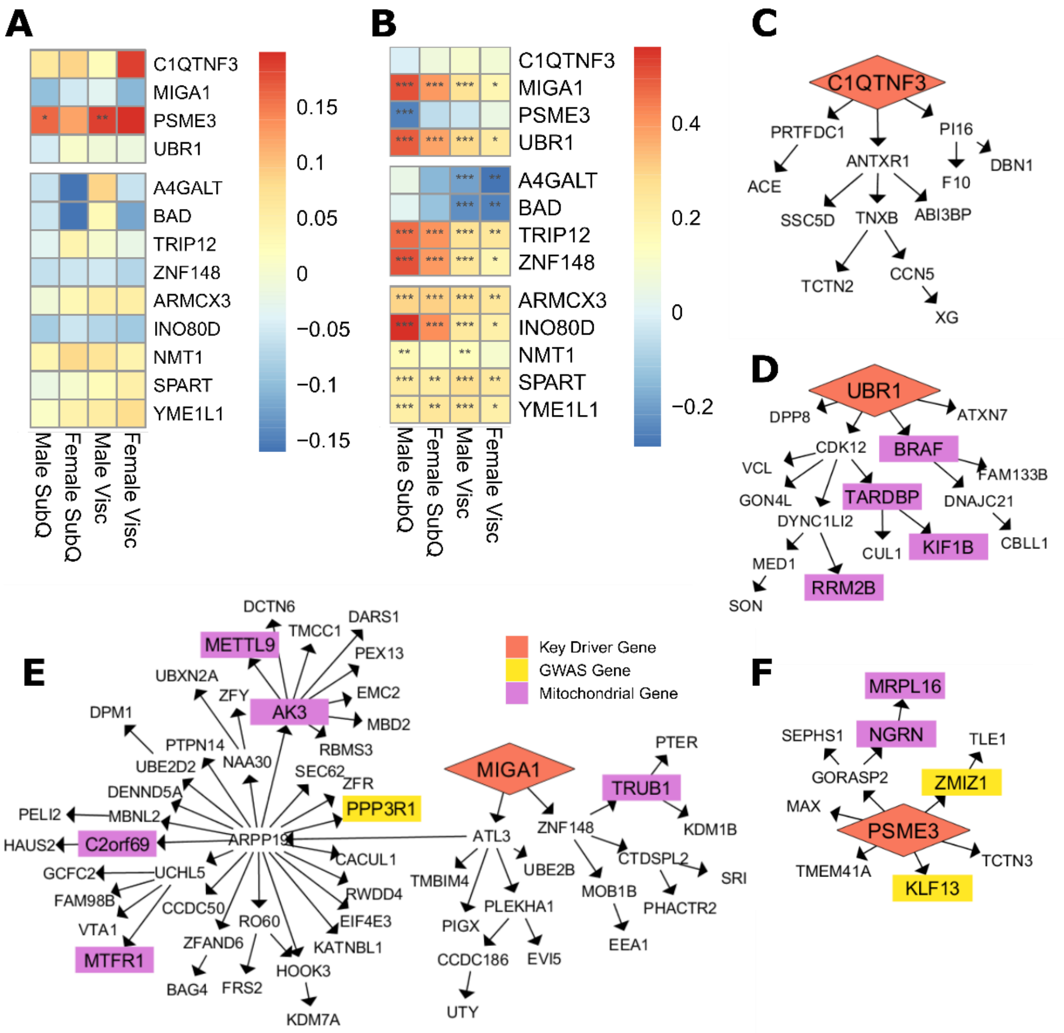

Figure 3.16: Thirteen prioritized key driver genes may affect mitochondrial function in adipocytes.
(A) Key driver gene expression in adipose tissue correlations with WHR_{adjBMI} *in STARNET.*
*(B) Key driver gene expression in adipose tissue correlations with UCP1 expression in STARNET. Pearson correlations are shown by color, p-values adjusted using FDR correction shown with * (*** = adj.P < 0.001, * = adj.P < 0.05).*
(C-F) Four selected key driver genes are regulate both WHR_{adjBMI} *downstream genes (yellow, Pulit et al) and mitochondrial downstream genes (purple, GO term “Mitochondrion”) in GTEx and STARNET.*
(C) C1QTNF3 in the GTEX Visceral Female network,
(D) UBR1 in the GTEx Visceral Female,
(E) MIGA1 in the GTEx Visceral Male network, and
(F) PSME3 in the STARNET Visceral Female network.

We used depot-specific expression correlations with WHR_{adjBMI} in STARNET to prioritize genes for further study (Figure 3.16). We identified four genes with Pearson correlation > 0.13 (Figure 3.7). We also considered correlation to *UCP1* expression in STARNET, since this gene is a driver of thermogenesis in adipocytes. We found nine correlated genes, eight of which were expressed at appreciable levels (TPM ≥ 5, *INO80D* removed) in the pre-adipocyte cells used for testing (Figure 3.17).

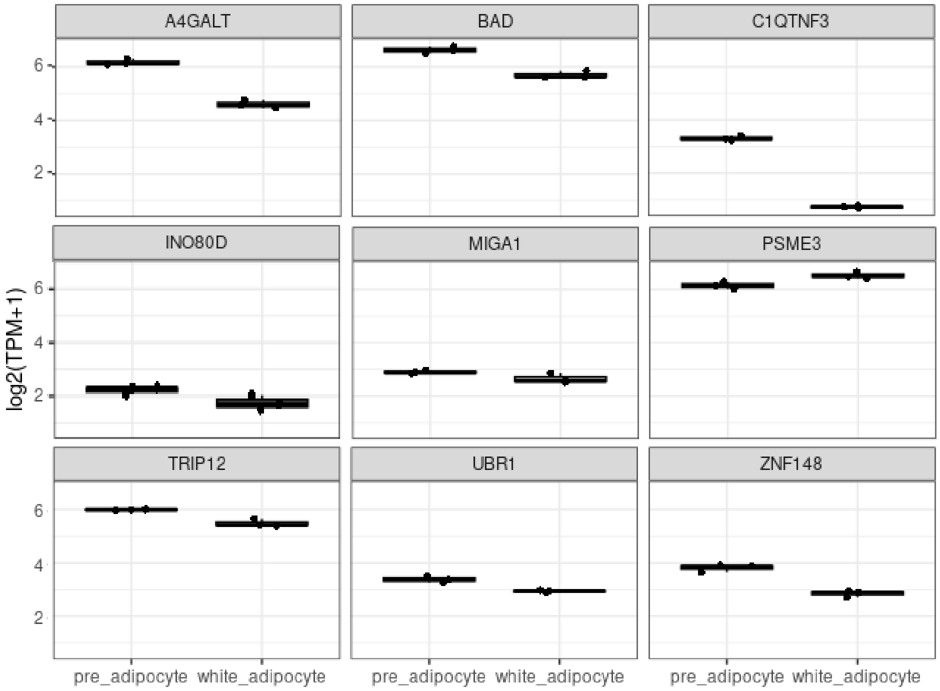

Figure 3.17- Mitochondrial candidate key driver expression in primary adipocyte cells.

The eight mitochondrial key driver genes, including *C1QTNF3, MIGA1, UBR1,* and *PSME3,* regulate a large number of downstream genes in the corresponding STARNET and GTEx networks (Figure 3.16 C-F). Those downstream genes contain body fat distribution GWAS genes (Table 2.6) and genes related to mitochondrial function. Additional evidence of involvement in fat distribution, such as expression differences between sexes and depots, and $\text{WHR}_{\text{adjBMI}}$ GWAS signal strength, further prioritized some of these genes (Figures 3.9, 3.18-3.20). *C1QTNF3* and *PSME3* are found in $\text{WHR}_{\text{adjBMI}}$ GWAS loci along with other genes, and we hypothesize they may be the causal gene in these loci, while *MIGA1* and *UBR1* are putative mechanistic genes that are not in GWAS loci.

By virtue of their status as prioritized network key driver genes and their strong correlation with $\text{WHR}_{\text{adjBMI}}$ in humans, we hypothesize that these eight genes affect mitochondrial function in adipocytes (Figure 3.1).

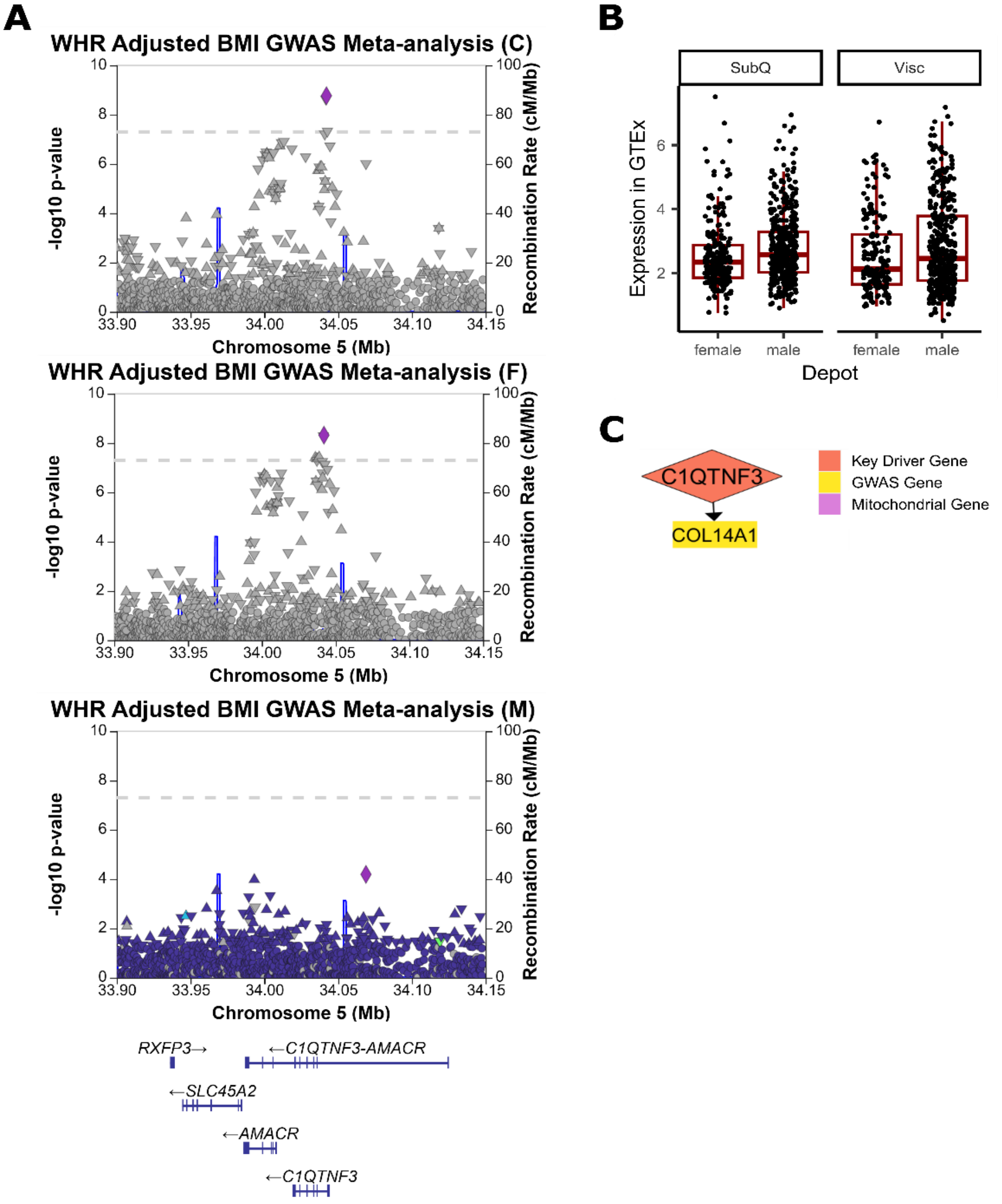

Figure 3.18- Additional evidence of C1QTNF3 involvement in $\text{WHR}_{\text{adjBMI}}$.
(A) A significant GWAS signal is detected near C1QTNF3 in $\text{WHR}_{\text{adjBMI}}$ GWAS meta-analysis[1]. This same signal is strongly associated with $\text{WHR}_{\text{adjBMI}}$ in the female-specific GWAS meta-analysis, but is not present in the male-specific GWAS.
(B) In GTEx, both males and females show higher expression of C1QTNF3 in visceral fat depots over subcutaneous depots, though the change is non-significant after p-value adjustment.
(C) C1QTNF3 regulates one $\text{WHR}_{\text{adjBMI}}$ GWAS gene in the STARNET subcutaneous male network.

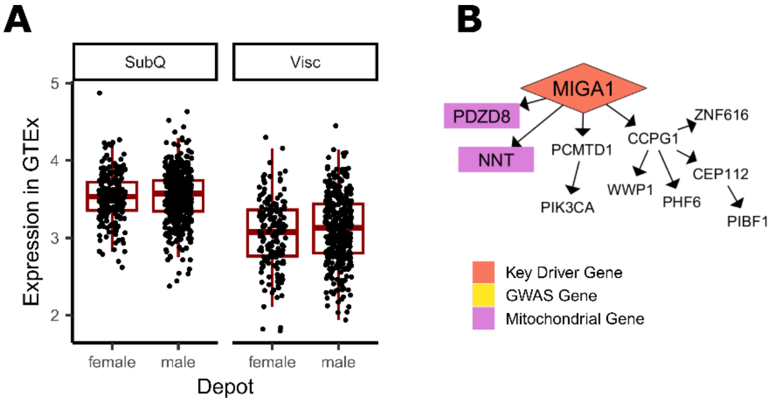

Figure 3.19- Additional evidence of MIGA1 involvement in WHR$_{adjBMI}$.
(A) In GTEx, both males and females show higher expression of MIGA1 in visceral fat depots over subcutaneous depots, though the change is non-significant after p-value adjustment.
(B) MIGA1 regulates mitochondrial genes in the STARNET subcutaneous male network.

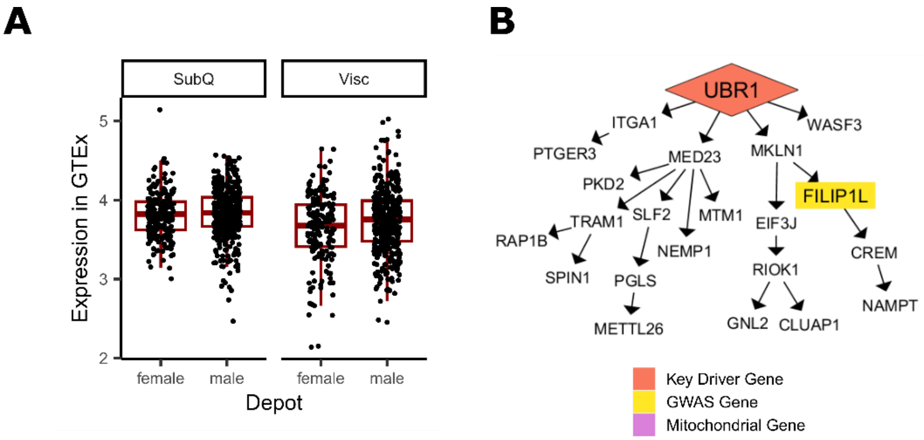

Figure 3.20- Additional evidence of UBR1 involvement in WHR$_{adjBMI}$.
(A) In GTEx, both males and females show higher expression of UBR1 in visceral fat depots over subcutaneous depots, though the change is non-significant after p-value adjustment.
(B) UBR1 regulates one WHR$_{adjBMI}$ GWAS genes in the STARNET subcutaneous male network.

3.2.6 Knockdown of MIGA1 and UBR1 inhibits oxygen consumption in differentiated adipocytes:

To test these hypotheses, we downregulated each of the eight genes using siRNA in primary human female pre-adipocyte cells, with non-targeting siRNA as a control (Methods). We obtained mature adipocytes by differentiating these cells for 18 days, then measured the mRNA expression of each gene to confirm that the knockdown efficiency was still more than 50% compared to controls (Figure 3.21 A). *A4GALT, ZNF148*, and *TRIP12* did not meet these criteria and *BAD* was not able to be detected, thus we removed these genes from subsequent analyses.

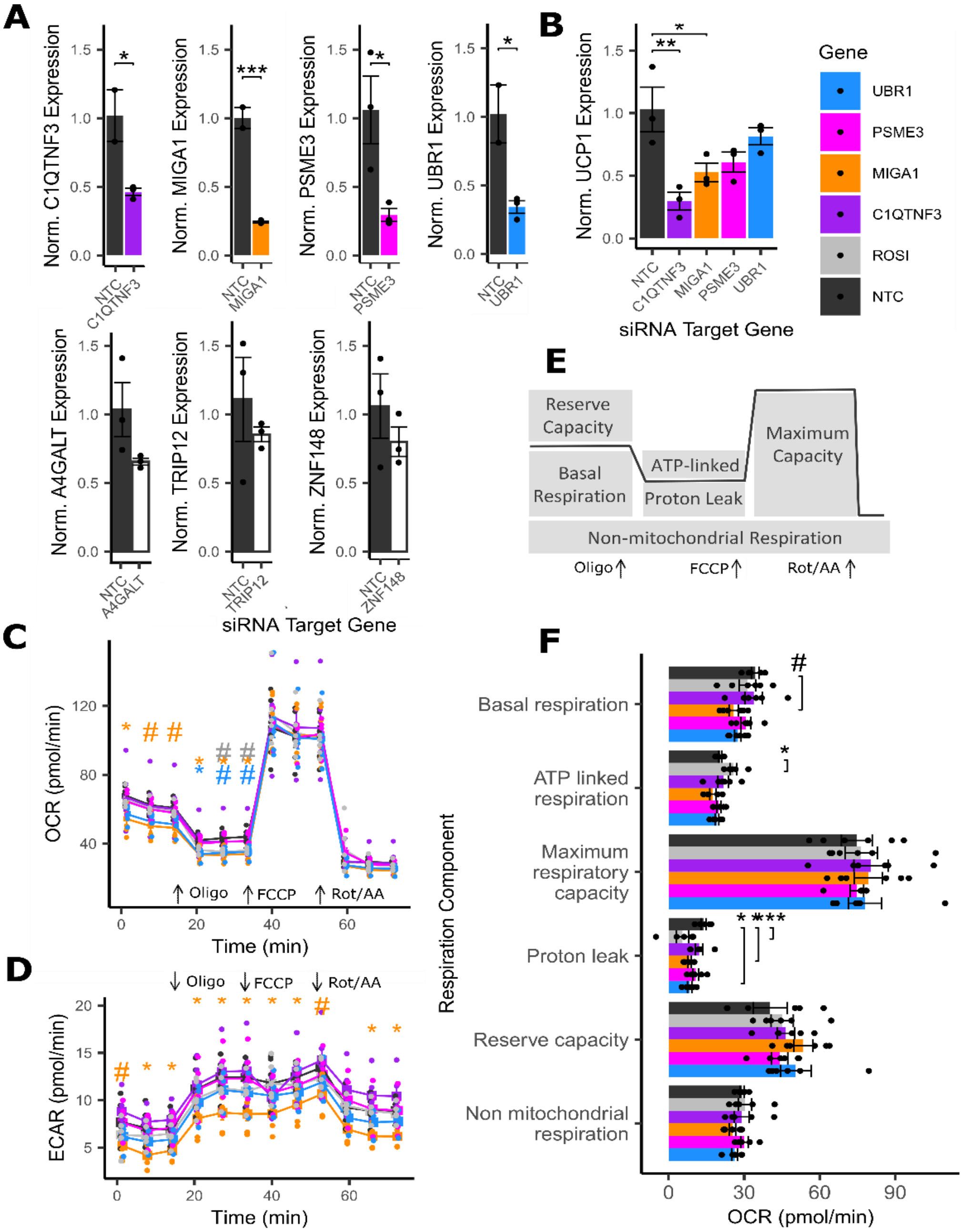

Figure 3.21: UBR1 and MIGA1 affect mitochondrial function in adipocytes.
(A)Gene expression of key driver genes in non-targeting control cells and in siRNA knockdown lines.
(B)Expression of UCP1 in siRNA knockdown lines and controls.
(C)Oxygen consumption rates genes in non-targeting control cells and in siRNA knockdown lines.
(D)Extracellular acidification rates genes in non-targeting control cells and in siRNA knockdown lines.
(E)Phenotypes calculated from oxygen consumption rates under various stimulation.
(F)Analysis of mitochondrial phenotypes upon siRNA perturbation under stimulations.

n = 6 replicates used in all assays. ROSI cells were treated with 2µM rosiglitazone *for 24 h prior to assay. Differences between groups in A, B, and E were determined using 1-way ANOVA by gene (Gene of Interest vs NTC controls). Differences between groups in C and D were determined using 1-way ANOVA within each timepoint by gene (Gene of Interest vs NTCcontrols). All post-hoc tests were performed using pooled t-test with Dunnett's adjustment. Adjusted p-values shown with * (*** = adj.P < 0.001,* = adj.P < 0.05, # = adj.P < 0.1).*

We examined the effect of the knockdown of the remaining four genes on *UCP1* expression in differentiated adipocytes. We found that *C1QTNF3* and *MIGA1* knockdown significantly reduced *UCP1* expression, 71.1% and 48.7% respectively, compared to controls (Figure 3.21 B). We also measured the expression of mature adipocyte markers. *PPARG*, *CEPBA*, and *FAPB4*. *UBR1* knockdown resulted in a significant 40.6% increase in *FABP4* expression (Figure 3.22).

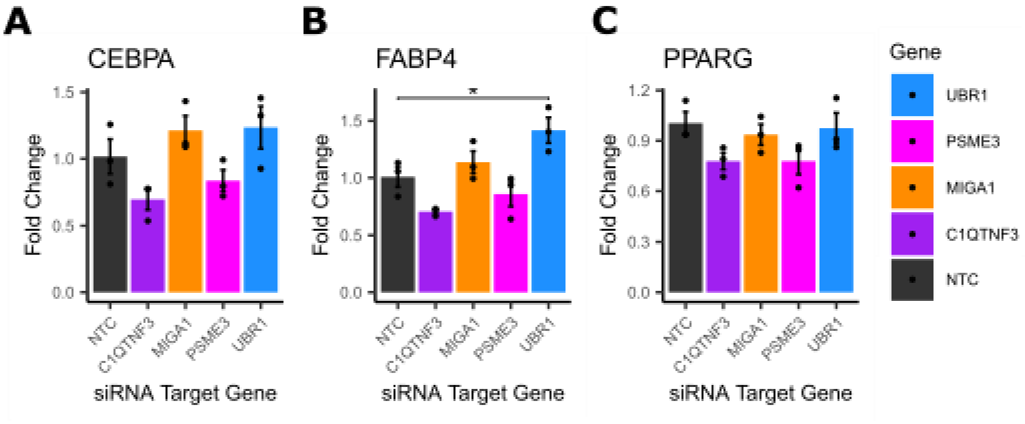

Figure 3.22- Marker of mature adipocytes FABP4 increases when mitochondrial key driver gene UBR1 is knocked down. Gene expression of
(A) CEBPα,
(B) PPARɣ,
(C) FAPB4 in mature adipocytes treated with siRNA for the indicated gene.
*n = 3-6 replicates used in all assays. Differences between groups were determined using 1-way ANOVA by gene (Gene of Interest vs NTCcontrols). Post-hoc tests were performed using pooled t-test with Dunnett's adjustment. Adjusted p-values shown with * (*** = adj.P < 0.001,* = adj.P < 0.05, # = adj.P < 0.1).*

We then determined the effect of knockdown of each gene on cellular oxygen consumption rate (OCR) using the Seahorse assay[89] (Methods). Importantly, we found that *MIGA1* knockdown significantly reduced the basal OCR and OCR after adding ATP-synthase inhibitor compared to controls (Figure 3.21 C). Additionally, we found that *UBR1* knockdown significantly reduced the OCR after adding ATP synthase inhibitor compared to controls. Only *MIGA1* knockdown resulted in significantly reduced extracellular acidification rate (ECAR) in differentiated adipocytes compared to controls (Figure 3.21 D).

Since each stimulation used in the OCR assay inhibits specific parts of the respiratory chain, we derived deeper mitochondrial phenotypes[90,91] (Figure 3.21 E, Methods). For example, we calculated the amount of oxygen consumed by mitochondrial proton leak, which refers to the futile H^+ shuttling across the OMM that does not produce ATP; thermogenic uncoupling through UCP1 contributes to proton leak[92]. Both *MIGA1* and *UBR1* knockdown cells showed significantly less proton leak compared to controls; proton leak was decreased by 40.4% and 39.9% respectively (Figure 3.21 F).

Of the genes tested, *MIGA1* had the strongest effect on cellular respiration. Its effects on ATP-independent oxygen consumption, calculated proton leak, and *UCP1* expression indicate that *MIGA1* may promote thermogenesis. To say definitively that MIGA1 promotes thermogenesis, we would measure the contributing processes lipolysis, glucose uptake, and beta-oxidation, and we could observe the response in oxygen consumption to cold exposure or beta-adrenergic stimulation. We could also measure the expression of other thermogenic genes such as *PGC1A*, *CIDEA,* and *PRDM16*. *UBR1* had a smaller effect on oxygen consumption, only under baseline conditions, and does not seem to be involved in adipocyte thermogenesis.

3.3 Discussion:

We present the first large-scale investigation into gene regulators of body fat distribution in adipose tissue in a sex- and depot-specific manner. We constructed large Bayesian networks using male and female adipose tissue gene expression from subcutaneous and visceral samples and identified over 300 putative regulators spanning both sexes and depots. Using additional evidence, we prioritized 53 unstudied key driver genes that may affect adipocyte function, and putatively regulate body fat distribution. Because of the unbiased nature of our initial key driver selection, we were able to prioritize putative candidate GWAS genes, as well as putative causal genes that are not in GWAS loci.

Identifying the causal GWAS gene at a given locus is a difficult task[93], given that there can be multiple genes and signals in the locus, and the strengths of these signals can vary between populations and with biological confounders, such as sex and age. Further, while studies have found that the nearest gene to the GWAS signal is the causal gene in 70% of loci[94], in the other 30% of loci, this assumption does not hold. For example, while *RSPO3* and *KLF14* are the only gene in the locus, the locus containing *TBX15* also contains *WARS2*, and therefore mechanistic studies had to be performed to show that *TBX15* was the causal gene[55,74]. Finally, genes not in significant GWAS loci may also contribute to the studied phenotype; in fact in the latest $\text{WHR}_{\text{adjBMI}}$ GWAS, SHOX2 and LRP5 are not located in significant loci, and still alter fat distribution in humans[54,57,74].

Status as a replicated key driver provided independent evidence that allowed us to prioritize 41 of the 495 genes in $\text{WHR}_{\text{adjBMI}}$ GWAS loci, including *COL8A1*, *PSME3,* and *ANAPC2*. We showed here that *PSME3* and *ANAPC2* have novel functional effects on adipogenesis; *COL8A1* was also prioritized in the colocalization studies by Raulerson et al[95] and is likely a functional regulator of fat distribution as well.

Our network analyses were able to capture known biology, as well as make novel predictions. We identified 45 key driver genes that were already well-studied in adipocytes. Many of these genes were tested in the SGBS or NIH 3T3-L1 lines, and many effect classical fat storage pathways such as adipogenesis, glucose uptake, lipolysis, etc (Supplemental Bibliography 3.1). These well-studied genes included 12 genes in $\mathrm{WHR}_{\mathrm{adjBMI}}$ GWAS loci and 33 non-GWAS genes. Genes *ACO1, ACAT1,* and *SLC25A1* have well-characterized effects on adipogenesis and lipogenesis (Supplemental Bibliography 3.1) and are also regulated in *trans* by KLF14; these three key driver genes may mediate KLF14's effects on female fat distribution.

Our analyses also highlighted 23 novel genes in two well-established pathways, Wnt signaling and mitochondrial function, as putative drivers of adipocyte function; we demonstrated a functional role for five of the genes in these pathways. Two genes, *PSME3* and *RSPO1,* showed antagonistic effects on Ca^{2+} non-canonical and canonical Wnt signaling, consistent with literature[78]. Although the implications in adipocytes are not fully known and warrant further study, non-canonical Wnt ligands cause inflammation and vascular disease[78] and are released from visceral fat more than subcutaneous fat[96]. While these two pathways were the focus of this paper, other pathways that contribute to fat distribution are likely represented within the 53 genes. Three validated key driver genes, *UBR1, ANAPC2*, and *PSME3,* have known roles in protein degradation; protein quality control may also be an important pathway regulating adipocyte function. Because of the bulk nature of the adipose tissue gene expression datasets, we also identified key drivers that likely act in immune cells, endothelial cells, and smooth muscle cells (Figure 3.3). Although these were not the focus of our *in vitro* validation, these cell types and genes may contribute to the overall body fat distribution phenotype. Of course, key driver genes and other predictions made by networks are putative, and must be experimentally validated.

We showed that key driver gene *RSPO1* inhibited lipid accumulation by upregulating the canonical Wnt signaling pathway. This is similar to the role played by *RSPO3*, which has been shown to have different effects in visceral and subcutaneous cells that explain its effects on fat distribution[50]. In addition, *RSPO2* was recently shown to inhibit adipogenesis[84] and *RSPO1* is a novel serum marker of obesity[97]. A recent study shows it may have a role in adipocyte beiging as well[98]. In single-cell and single-nucleus studies, *RSPO1* is expressed in pre-adipocytes and in mesothelium, the cells lining the outer wall of visceral adipose, and in GTEx, *RSPO1* expression is higher in visceral compared to subcutaneous samples (Figure 3.10). RSPO1 may be released primarily from visceral mesothelial cells, upregulating Wnt to a greater degree in visceral adipose than in subcutaneous depots, which could contribute to overall differences in WHR.

We showed that key driver gene *ANAPC2* inhibited adipogenesis. Further, we saw that *ANAPC2* overexpression led to elevated *AXIN2* expression but no increases in other read-outs of Wnt signaling activity. This may explain the smaller decrease in adipogenesis compared to *RSPO1* overexpressing cells, which had a large effect on Wnt activity, although the loss of overexpression at day 12 may also account for the partial impairment of adipogenesis. *ANAPC2* encodes the cullin-like member of the anaphase promoting complex/cyclosome (APC/C) E3-ligase[99], which degrades cell cycle machinery to promote progression to the next cell cycle phase. Interestingly, we saw no differences in proliferation due to *ANAPC2* overexpression. Studies show the necessity of exiting the cell cycle before beginning adipogenesis[100]. Others have shown that manipulating the expression of APC/C interactors *Bubr1* and *Cdc20* in cells has directionally consistent effects on adipogenesis[101,102]. GWAS and QTL studies show that the same variant rs144926207 is associated with higher $\mathrm{WHR}_{\mathrm{adjBMI}}$ and lower retention of intron 4 in subcutaneous and visceral adipose tissue (Figure 3.12). *ANAPC2* intronic retention causes cell cycle delays in other cell types[103]; *ANAPC2* may also delay and inhibit adipogenesis by continuing cell cycle mechanisms. Further, evidence of genetic regulation and functionality in adipocytes prioritizes this gene as candidate regulator of $\mathrm{WHR}_{\mathrm{adjBMI}}$. Future experimentation is required to test these hypotheses.

We showed that key driver gene *PSME3* promoted adipogenesis, promoted Wnt signaling activity, and had no effect on mitochondrial function. Its protein product functions as a cap for the 20S proteasome[104], and in this context REG-ɣ degrades GSK3β[72]. We confirmed that *PSME3* overexpressing cells had lower levels of active:total GSK3β than controls, although Wnt signaling is unlikely the mechanism by which *PSME3* increases adipogenesis, due to the directionally inconsistent effects. 20S proteasome member Psmb4 promotes adipogenesis in brown adipocytes by maintaining proteostasis[105] and mutations in *PSMB8* cause lipodystrophy in humans[106]. *PSME3* has diverse roles in many cell types, with and without the 20S proteasome, including in development, fertility, cancer, and metabolism[104]; aging[107], ribogenesis and autophagy[104], and in lipid accumulation in hepatocytes and mouse models via sirtuin degradation[108]. Therefore, further experimentation is required to determine the mechanism by which *PSME3* increases adipogenesis.

We showed that key driver gene *MIGA1* knockdown decreased multiple components of mitochondrial function in differentiated adipocytes, suggesting that *MIGA1* contributes to normal mitochondrial function and may promote thermogenesis. Other outer membrane proteins that regulate mitochondrial fusion, MIGA2 and MFN2*,* also control the membrane's interactions with lipid droplets, which affects lipid storage[109–111]. MIGA1 could affect fat storage in adipocytes through a dual role in controlling the OMM's interactions with other mitochondria and with the lipid droplet, but further studies are required to test these hypotheses.

We showed that key driver gene *UBR1* knockdown lowered mitochondrial function in differentiated adipocytes. It had no effect on *UCP1* expression, so its role in regulating mitochondrial function is likely non-thermogenic. *UBR1* could function by maintaining protein quality and decreasing PINK1-induced mitophagy, although future studies are required to determine its precise role in mitochondria and adipocyte fat storage.

Fat distribution is a complex, full-body phenotype that is difficult to recapitulate in *in vitro* and *in vivo* models. We used subcutaneous and visceral gene expression data to make predictions, but we tested those predictions in a subcutaneous pre-adipocyte cell line that was available to us. Only *ANAPC2* was tested in the same sex-and depot-derived cell as the network in which it was a key driver. Remarkably, we observed that many predicted visceral fat key drivers had function in subcutaneous adipocyte cells, but due to these limitations, we have not fully characterized the effects of these genes on adipocyte function or fat distribution. All five genes require further studies via comparative experiments in adipocytes from subcutaneous and visceral depots. Further, the genes that were found to have no effect on adipogenesis or mitochondrial function in our studies may impact these processes in visceral adipocyte cells.

There is strong evidence that the validated key driver genes act in a sex-biased manner; therefore, follow-up studies must include cells of both sexes. Of the 53 prioritized key drivers, only five genes are replicated key drivers in both male and female networks; highlighting the unique *in silico* gene regulatory structure in each sex. Four of the validated key driver genes, *MIGA1*, *ANAPC2, PSME3,* and *RSPO1,* have a role in fertility and the development of gonads or gametes[82,112–114]. Further, RSPO1 upregulates estrogen receptor *ESR1 in vitro*[115]*,* and *PSME3* is located in a gene-dense female-specific $\mathrm{WHR}_{\mathrm{adjBMI}}$ GWAS locus (Figure 3.9). Two of the validated genes, *RSPO1* and *PSME3*, were discovered in female visceral networks and tested in male cells, while *MIGA1* was identified in male visceral networks and tested in female cells; these may have distinct roles in the opposite sex that were not investigated in this study.

We showed the validity and strength of Bayesian network modeling to predict known and novel gene regulators. We provided additional evidence of the role of Wnt signaling and mitochondria in adipocyte function, and putatively in body fat distribution. Finally, we hypothesize a broader role for five genes in regulating fat distributions in humans.

3.4 Methods:

3.4.1 Gene Expression and Phenotypic Data:

We interrogated RNA-sequencing gene expression data from subcutaneous adipose tissue and visceral abdominal adipose tissue from the Genotype-Tissue Expression project (GTEx)[75] and the Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task (STARNET)[93]. Detailed explanations of participant inclusion, data collection, sequencing, and quantification can be found at each source. Briefly, STARNET participants are people living with coronary artery disease, from whom biopsies of abdominal subcutaneous fat and abdominal visceral fat were obtained during open thorax surgery. Samples were sequenced using the Illumina HiSeq 2000 platform. GTEx biopsies of abdominal visceral fat and leg subcutaneous fat were taken from deceased donors shortly after death. Samples were sequenced using the Illumina TruSeq platform. Both datasets were obtained in transcripts per million (TPM) format. The participants of STARNET were deeply metabolically phenotyped, while the GTEx subjects were assayed for more basic anthropometric traits.

Expression Data Processing: We first used annotation meta-data from each source to divide the data into males and females. We used *XIST* expression to confirm these assignments. Next, we used annotations from the R package bioMart for genome build hg38 to select only the protein coding genes within each dataset. We then removed genes with less than 0.1 TPM value in greater than 80% of the samples. Finally, we log transformed the gene expression values for subsequent analysis.

3.4.2 Testable Key Driver Gene Selection:

3.4.2.1 Identification of cell type:

We used 7 publicly available single cell- or single nucleus- adipose tissue or adipose tissue derived- stromal vascular fraction RNA sequencing datasets from both human and mouse[58–64].

Since there was some disagreement between studies, all cell types in which the gene was expressed in any study are reported in Supplemental Table 2.6. We removed genes that were only expressed in non-adipocyte cell types.

3.4.2.2 Identification of Function in Adipocytes:

We identified well-studied key driver genes using a comprehensive literature search (Supplemental Bibliography 3.1). For each gene, we used GeneCards to identify alternate names for each gene or corresponding protein. We then searched PubMed and Google Scholar for functional studies in cells demonstrating a role for that gene in pre-adipocytes or adipocytes. Terms searched include “adipocyte”, “adipogenesis”, “differentiation”, “lipogenesis”, “lipolysis”, “glucose uptake”, “browning”, and “thermogenesis”.

3.4.2.3 Additional Genetic Evidence:

We prioritized key driver genes based on two types of genetic evidence. First, we identified $\text{WHR}_{\text{adjBMI}}$ GWAS genes[74] within the set of key driver genes. Second, we prioritized genes involved in fat storage and distribution in mouse models. We queried the Mouse Genome Informatics database[116] and the International Mouse Phenotyping Consortium[117] to determine if the gene knockout in mice results in significant differences in fat pad size, total body fat mass, lean mass, or related phenotypes. We did not consider overall body size differences, as these may be indicative of BMI related phenotypes.

3.4.3 Lentivirus Construction:

Overexpression plasmids were constructed by VectorBuilder (VectorBuilder Inc, Chicago, Illinois, USA) using the mammalian gene expression lenti-viral vector backbone with 1 open reading frame. This backbone contains 3rd generation lenti-viral integration sites and ampicillin resistance. Using GTEx, we identified the most abundant isoform in adipose tissue for each gene. This isoform was added to the plasmid, followed by a P2A linker, then the GFP reporter sequence.

This construct was under the CMV promoter. Control plasmids contain the GFP reporter gene under the CMV promoter with no P2A linker.

7TFC was a gift from Roel Nusse, purchased from Addgene (Addgene, Watertown, Massachusetts, USA; Addgene plasmid # 24307). The 7TFC plasmid contains 3rd generation lenti-virus integration sites, the Firefly Luciferase gene under 7 repeats of the TCF promoter and an mCherry marker under the SV40 promoter and ampicillin bacterial resistance gene.

We obtained the plasmids in *E. coli* swabs in agar. We cultured the *E. coli* on Luria-Bertani broth agar plates containing 100 µg/mL ampicillin, and sub-cultured single colony forming units in 50 mL Luria-Bertani broth containing 100 µg/mL ampicillin. Plasmids were isolated from *E. coli* using the Nucleobond Xtra Midi prep kit (Takara Bio, San Jose, California, USA) following the manufacturer's protocol (Cat# 740422.50). Plasmids were packaged into 3rd generation replication deficient lenti-virus in HEK-293T cells using the Lenti-Pac HIV Expression Packaging Kit (Genecopoeia, Rockville, Maryland, USA) following manufacturer's protocol (Cat# LT001).

3.4.4 Transduction and Sorting of Human Pre-adipocyte Overexpression Cell Lines:

We obtained human male pre-adipocyte Simpson-Golabi-Behmel syndrome (SGBS) cells from Dr. Martin Wabitsch[39] at passage number 35-40. All pre-adipocyte cells were grown and maintained as described previously[76] in DMEM:F12 media (ThermoFisher Scientific, Waltham, Massachusetts, USA) containing 10% Fetal Bovine Serum, 1% Penicillin/Streptomycin, 8.1 ng/mL biotin and 3.5 ng/mL pantothenate. During transduction with lenti-virus, the fetal bovine serum was first heat-inactivated at 65 C for 30 minutes, and 8 µg/mL polybrene was added to improve transduction efficiency. We plated cells in 6 well plates and grew them to 70% confluence before transducing the cells with lenti-viral-particles containing each of the plasmids listed above. Cells containing high levels of GFP or mCherry were sorted using the FACS Aria Fusion Cell Sorter

(BD Biosciences, Franklin Lakes, New Jersey, USA). We used cells with passage number less than 46 for subsequent assays.

3.4.5 siRNA of Target Genes

Table 3.3: siRNA sequences

Gene Symbol	**Sequence**
A4GALT	GGACACGGACUUCAUUGUU
A4GALT	GCACUCAUGUGGAAGUUCG
A4GALT	AGAAAGGGCAGCUCUAUAA
A4GALT	UGAAAGGGCUUCCGGGUGG
BAD	GAUCGGAACUUGGGCAGGG
BAD	CAGAGUUUGAGCCGAGUGA
BAD	GAGCUCCGGAGGAUGAGUG
BAD	UUGUGGACUCCUUUAAGAA
C1QTNF3	CCGCAAAUUCUAAAUCUUA
C1QTNF3	UCAACCUAGUAGAGGACAA
C1QTNF3	AGGUGAGAAGGGCGACAAA
C1QTNF3	CAGUAUCAGGUGUGUAUUU
MIGA1	GCUCUGACCUUUCGCAAUA
MIGA1	ACACAGAGAAGUACGGCAU
MIGA1	GGAAAUAUCUCUUUAUCGU
MIGA1	CUUGAAGACAGCAGCGCUA
PSME3	GAAUCAAUAUGUCACUCUA
PSME3	UCUGAAGGAACCAAUCUUA
PSME3	GCUAAGAACUGUUGAGAGU
PSME3	GACCAGAUUUCUAGAUAUU
UBR1	GGAAAUCAGCGCGGAGUUA
UBR1	GUACAAUCGUGUGGACAUA
UBR1	GCGAAGAAAUGGACUGUCU
UBR1	GAUCAGCAAACCCACAAUA
ZNF148	UGGAAUAGCUACUCAAUUU
ZNF148	UUGAAUAGCCCGAGCCUUA
ZNF148	GGAUCAAGCUCCCAAGCAU
ZNF148	CUAAGAACAACUCCAGAUA
TRIP12	GAACACAGAUGGUGCGAUA
TRIP12	GACAAAGACUCAUACAAUA
TRIP12	GCUCAUAUCGCAAAGGUUA
TRIP12	GGUAGUGACUCCACCCAUU

3.4.6 Transfection and Differentiation of Human Primary Pre-adipocytes

Human female subcutaneous primary pre-adipocytes were purchased from Zenbio (Zenbio, Cat# SP-F-SL; Lot# SL0061) and were differentiated according to the standard Zenbio white adipocyte differentiation protocol. Briefly, human primary subcutaneous pre-adipocytes were seeded on collagen-coated 96-well plate (20,000cell/well, Corning, 354650) with 200µl of PM-1 medium (Zenbio, #PM-1) and the cells were established overnight, then transfected with siRNAs of target genes or scramble controls for 3 days using Lipofectamine RNAiMAX (Invitrogen, cat# 13778-150). The culture medium was replaced with 150µl of differentiation medium DM-2 (Zenbio, #DM-2), and cells were cultured for 7 days. Media was replaced with maintenance medium AM-1 and cells were cultured for additional 7 days for the following qPCR and Seahorse experiments.

3.4.7 Quantification of Gene Expression in lenti-viral treated Human Pre-Adipocytes and Differentiating Cells:

We grew cells in 12 well plates. Once they reached confluency, they were washed with Phosphate Buffered-Saline (PBS) and incubated in 400 µL Trizol, then scraped and harvested. We extracted RNA using the RNeasy Micro Kit (Qiagen, Velno, Netherlands), following manufacturer's protocol (Cat# 74004). We digested the DNA species on the Qiagen spin column using the RNAse-free DNAse kit (Qiagen, Velno, Netherlands) following manufacturer's protocol (Cat# 79254). We quantified the isolated RNA using the Qubit with RNA Broad Range assay kit (ThermoFisher Scientific, Waltham, Massachusetts, USA), following manufacturer's protocol (Cat# Q10210). We reverse transcribed cDNA from the RNA templates using SuperScript IV Reverse Transcriptase Kit (ThermoFisher Scientific, Waltham, Massachusetts, USA) with Oligo(dT)20 primers, following manufacturer's protocol (Cat# 18090010). We quantified cDNA abundance using quantitative-polymerase chain reaction (qPCR). Samples and standard curves were prepared using GoTaq qPCR Master Mix (Promega, Madison, Wisconsin, USA) and gene specific primers (Integrated DNA Technologies, Coralville, Iowa, USA) (Table 3.4). Samples were measured using the

QuantStudio 5 Real Time PCR system (ThermoFisher Scientific, Waltham, Massachusetts, USA), and were analyzed using Thermo Fisher Connect qPCR Standard Curve analysis software.

3.4.8 Quantification of Gene Expression in siRNA treated Human Differentiated Adipocytes

The total RNA samples were isolated from human differentiated adipocytes using KingFisher™ Flex Magnetic Particle Processor according to the MagMAX™ mirVana™ Total RNA isolation protocol. We obtained the cDNA samples from the RNA templates using SuperScript IV Reverse Transcriptase Kit (ThermoFisher Scientific, Waltham, Massachusetts, USA) with Oligo(dT)20 primers, following manufacturer's protocol (Cat# 18090010). We quantified cDNA abundance using quantitative-polymerase chain reaction (qPCR) using the QuantStudio 5 Real-Time PCR system (ThermoFisher Scientific, Waltham, Massachusetts, USA). The conditions were: 42 °C for 5 min, a 10s denaturation step at 95 °C, followed by 40 cycles of 95 °C for 5s and 58 °C for 40s.

Table 3.4: qPCR Primer Sequences

Primer name	Sequence
Primers used in adipogenesis/Wnt signaling assays	
ADIPOQ F	GGAGATCCAGGTCTTATTGGTCC
ADIPOQ R	GCACCTTCTCCAGGTTCTCC
ANAPC2 F	GCGAGAAGAAGTCCACACTATG
ANAPC2 R	GACTCTCAAGAAGCACCCATAC
AXIN2 F	GACCAAGTCCTTACACTCCTTATT
AXIN2 R	TCTAAGGTATCCACGCATTTCTC
B2M F	AGATGAGTATGCCTGCCGTGT
B2M R	TGCTGCTTACATGTCTCGATC
CEBPA F	TATAGGCTGGGCTTCCCCTT
CEBPA R	AGCTTTCTGGTGTGACTCGG
IL6 F	CCAGGAGAAGATTCCAAAGATGTA
IL6 R	CGTCGAGGATGTACCGAATTT
PPARG F	ACCCAGAAAGCGATTCCTTCA
PPARG R	TCCACTTTGATTGCACTTTGGT
PSME3 F	CTGAGATCCGGCTGTTGATT
PSME3 R	CAGGAGCTGTACCCACATTT
RSPO1 F	GTGAAATGAGCGAGTGGTCT
RSPO1 R	GTAGCACCCTGCGTGTC
RSPO3 F	GAAACACGGGTCCGAGAAATA
RSPO3 R	CCCTTCTGACACTTCTTCCTTT

TYRO3 F	GAGTGTATGGAGGACGTGTATG
TYRO3 R	GTTCCATTCGCAGACAAGTAAAG
Primers used in mitochondrial function assays	
A4GALT-F	GCATCTACCTGGACACGGACTT
A4GALT-R	ATGCACAGCGCCATGAACTCGT
AKAP9-F	GGCGTCATTGATGGCTATGCAG
AKAP9-R	GCTGTTGCTCTGCCTCCAATTC
ARAP1-F	GTGTGGACTACATCACGCAGTG
ARAP1-R	CCGAGGAAACATCATCCACGTG
BAD-F	CCAACCTCTGGGCAGCACAGC
BAD-R	TTTGCCGCATCTGCGTTGCTGT
C1QTNF3-F	GGAGACTACAGCTTTCGAGGCT
C1QTNF3-R	TTGTCGCCCTTCTCACCTTTGG
LDHD-F	GATGGATGCCTGCAACAGGTAC
LDHD-R	TCCGTTCTGCTGGACTATCTCC
MIGA1-F	GTGCAGGAGATGCCATTGCTGA
MIGA1-R	ACTCCTCTTGGAGACGATAGGC
PSME3-F	ATGAATCTCCCAGTCCCTGACC
PSME3-R	GGGCATCACAAACACCTTGGTTC
QTRT1-F	GTAGTCTGCGTGGCTCTTGGAT
QTRT1-R	GCCGAAGTCCTTCTCAAACACC
UBR1-F	GTAGCAACCACATCAGGATCGG
UBR1-R	CTGTAAGGCAGACATCTGAGCC
ZNF148-F	CACGTTTGTGAGCACTGCAATGC
ZNF148-R	GCAGGTACTTCTGTATGAAACGC
TRIP12-F	GGCTGCCTCAAAGGATACCATC
TRIP12-R	GCAGCACAAAGTCTCTGAAGGAC
GAPDH-F	GTCTCCTCTGACTTCAACAGCG
GAPDH-R	ACCACCCTGTTGCTGTAGCCAA
hCEPBA-F	ACAAGAACAGCAACGAGTACCG
hCEPBA-R	CATTGTCACTGGTCAGCTCCA
hFABP4-F	ACGAGAGGATGATAAACTGGTGG
hFABP4-R	GCGAACTTCAGTCCAGGTCAAC
hUCP1-F	AGTTCCTCACCGCAGGGAAAGA
hUCP1-R	GTAGCGAGGTTTGATTCCGTGG
hPPARG-F	AGCCTGCGAAAGCCTTTTGGTG
hPPARG-R	GGCTTCACATTCAGCAAACCTGG

3.4.9 Cellular Phenotyping:

3.4.9.1 Proliferation Assay:

We plated 20,000 cells/well in 24 well plates. After 24 hours, 4 wells were trypsinized and cells were counted using a hemocytometer. Wells were washed with PBS and growth media was replaced. Every 24 hours, 4 more wells were counted, for a total of 6 days. Growth media was replaced every 2 days. Most cells reached exponential growth by day 4 (Figure 3.13 B). We then calculated the doubling time of each cell line using the formula:

$$T_d = (T_2 - T_1) \cdot [\frac{\ln(2)}{\ln\left(\frac{N_2}{N_1}\right)}]$$

where T_d is doubling time, T_1 and T_2 are initial and final time measurements, and N_1 and N_2 are the initial and final quantity of cells.

3.4.9.2 Adipogenesis Assay:

We differentiated cells into lipid-containing adipocytes as detailed previously[76]. Briefly, we plated 40,000 cells/well in 12 well plates. Cells were incubated for 2-5 days until they reached 100% confluency, then incubated for 48 hours post-confluency. Adipogenic media (DMEM:F12, 1% Penicillin/Streptomycin, , 8.1 ng/mL biotin, 3.5 ng/mL pantothenate, 0.01 mg/ml transferrin, 20 nM insulin, 100 nM cortisol, 0.2 nM triiodothyronine, 25 nM dexamethasone, 250 µM 3-isobutyl-1-methylxanthine, and 2 µM rosiglitazone) was added to each well to initiate differentiation. After 4 days, we changed the media to DMEM:F12, 1% Penicillin/Streptomycin, 8.1 ng/mL biotin, 3.5 ng/mL pantothenate, 0.01 mg/ml transferrin, 20 nM insulin, 100 nM cortisol, and 0.2 nM triiodothyronine. Every 4 days, this media was replaced.

3.4.9.3 Quantification of Adipogenesis:

We quantified the amount of lipid stored in cells using Oil Red O (ORO) dye. 0.25 grams of dye was suspended in 48 mL of 98% isopropanol and 32 mL of DI water. Unsuspended dye was removed from the ORO solution using 0.045 µM vacuum filtration. We repeated filtration (~3X) every 24 hours until no precipitate was observed. Cells were washed with PBS, then fixed in 300 µL 4% paraformaldehyde for 15 minutes. Cells were washed with 60% isopropanol, then dried completely. 250 µL of ORO solution was added to the cells for 5 minutes. Cells were washed with DI water twice, then dried completely. We imaged the full wells using the EVOS microscope (below). Oil Red O dye was then eluted from cells in 200 µl of 100% isopropanol for 2 minutes. Eluted ORO was quantified by measuring absorbance at 450 nm.

3.4.9.4 Imaging:

We took images using the EVOS M7000 imaging system (ThermoFisher Scientific, Waltham, Massachusetts, USA) at 10x magnification, using phase contrast and color. We constructed full well composite images by taking 30 adjacent images in a 5x6 grid that covers most of the well. Composite images were stitched together using imageJ:Fiji plugin Grid/Collection Stitching[118].

3.4.10 Quantification of Wnt Signaling:

3.4.10.1 Quantification of Wnt Signaling Transcriptional Activation:

We performed luciferase assays using the SGBS:7TFC reporter line. SGBS:7TFC cells were transduced with lenti-virus containing the gene of interest or GFP control plasmids. Images were taken to ensure a high percentage of dual mCherry and GFP expressing cells. 10,000 cells/well were plated in clear bottom, white-walled 96 well plates, with 6 replicates of each gene or control. After 24 hours of incubation, luciferase activity was measured using the Luciferase Assay System (Promega, Madison, Wisconsin, USA) following manufacturers protocol (Cat# E1500). Briefly, the cells were lysed in 20 µl lysis buffer, 100 µl of luciferin-containing reagent was added, then emitted light was measured for 10 seconds using a luminescence plate reader. Luminescence readouts in each well were normalized to mCherry fluorescence to account for total luciferase insertions by the 7TFC cassette.

3.4.10.2 Quantification of Protein Activation:

Active (Ser33/Ser37/Thr41 non-phosphorylated) β-catenin and total β-catenin species were measured using western blotting. Total proteins were isolated in RIPA buffer containing 1% protease and 1% phosphatase inhibitors (ThermoFisher Scientific, Waltham, Massachusetts, USA, Cat# 89901, Cat# 78429, Cat# 78426). We quantified total protein species using the bicinchoninic acid (BCA) assay (ThermoFisher Scientific, Waltham, Massachusetts, USA) following the manufacturer's protocol (Cat# 23225). We denatured samples at 70°C for ten minutes, then ran 20 µg total protein on a NuPAGE 4-10% BisTris Gel at 240V for 40 minutes (ThermoFisher Scientific, Waltham, Massachusetts, USA, Cat# NP0336BOX). We transferred the protein to an Immobilon-FL PVDF membrane at 80V for 60 minutes (MilliporeSigma, Burlington, Massachusetts, USA, Cat# IPFL00010). We labeled active and total β-catenin and β-actin control bands using primary antibodies (Cell Signaling Technologies, Danvers, Massachusetts, USA; Non-phospho (Active) β-Catenin (Ser33/37/Thr41) (D13A1) Rabbit mAb Cat#8814, dilution 1:500; β-Catenin (15B8) Mouse mAb Cat#37447, dilution 1:1000). Bands were labelled with fluorescently conjugated secondary antibodies (ThermoFisher Scientific, Waltham, Massachusetts, USA; Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Cyanine3, Cat# A10521, dilution 1:20,000; Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Cyanine3, Cat# A10520, dilution 1:20,000). We imaged the labelled protein on Amersham Imager 600 (Global Life Sciences Solutions, Marlborough, Massachussets, USA) using RGB fluorescence settings. Densitometry calculations were performed using imageJ.

Amount of active and total GSK3β, JNK, and CAMK2A were quantified using Enzyme-linked immunoassays (ELISA). Cells were harvested and lysed according to each manufacturer's protocol. Active GSK3β (Ser9 phosphorylated) and total GSK3β were measured using an ELISA kit (RayBiotech, Peachtree Corners, Georgia, USA) using manufacturer's protocols (Cat# PEL-GSK3b-S9-T). Active JNK (Thr183/Tyr185 phosphorylated) and total JNK were measured using an ELISA kit (RayBiotech, Peachtree Corners, Georgia, USA) using manufacturer's protocols

(Cat# PEL-JNK-T183-T-1). Active CAMK2A (Thr286 phosphorylated) and total CAMK2A were measured using an ELISA kit (Assay BioTechnology, Fremont California, USA) using manufacturer's protocols (Cat# FLUO-CBP1509 and CB5092).

3.4.11 Quantification of oxygen-consumption rate and extracellular acidification rate:

Oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) was determined using a Seahorse XF96 analyzer in combination with the Seahorse mitochondrial stress test kit according to a standard protocol[89]. In brief, human primary pre-adipocytes were plated and differentiated as described above. Differentiated cells were washed with DPBS twice and incubated with Seahorse XF assay medium supplemented with 2 mM glutamax, 10 mM glucose, 1 mM sodium pyruvate (PH 7.4) for 45 min at 37 °C in a non-CO2 environment. Both OCR and ECAR were subsequently measured in real time using XF96 extracellular flux analyzer (Seahorse Bioscience). The optimized concentration of compounds for mito-stress assay were 1.5 μM of oligomycin, 1.5 μM of carbonyl cyanide-p-trifluoromethoxyphenylhydrazone (FCCP), and 0.5 μM of Rotenone&antimycin A. Following the extracellular flux analysis, the OCR and ECAR were normalized by cell number, quantified using Hoechst staining.

3.4.11.1 Quantification of deep OCR phenotypes:

We calculated basal mitochondrial respiration, ATP-linked respiration, proton leak, maximal respiratory capacity, reserve capacity, and non-mitochondrial respiration from the OCR assay as described previously[90,91]. We defined condition 'A' as timepoints 1, 2, and 3 under basal stimulation; condition 'B' as timepoints 4,5,6 under oligomycin stimulation; condition 'C' as timepoints 7,8,9 under FCCP stimulation; and condition 'D' as timepoints 10,11 and 12 under Rot/AA stimulation. We considered each of the three timepoints within each condition as technical replicates, and the six samples as biological replicates. We averaged the three timepoints per condition into one value per biological replicate. We then defined non-mitochondrial respiration as D; basal mitochondrial respiration as A - D; ATP-linked respiration as A - B, proton leak as B - D; maximal respiratory capacity as C - D; and reserve capacity as C-A.

3.4.12 Statistical Methods:

Differences in proliferation and differentiation assays using the GFP-expressing control cells and cells expressing the gene of interest were assessed using 2-way ANOVA by gene and time (day). Post-hoc tests were performed between GFP controls and genes of interest within each timepoint using pooled t-tests with p-value adjustment using Dunnett's adjustment. Differences in Wnt signaling using the GFP-expressing control cells and cells expressing the gene of interest and mitochondrial assays using the non-targeting control-expressing cells and cells with siRNA for the gene of interest were assessed using 1-way ANOVA by gene. Post-hoc tests were performed between controls and genes of interest using pooled t-tests with p-value adjustment using Dunnett's adjustment. Analyses were performed using base R's anova() function and the emmeans package's emmeans() and contrasts(). We reported only significant p-values, with the exception of Figure 3.21 where indicated. All bar plots display the mean, with error bars displaying the standard error of the mean (S.E.M).

Acknowledgements:

This work was supported by NIH/NIDDK grant R01 DK118287 and NIH training grant T32 HL007284.

3.5 Bibliography:

1. Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
2. Jackson, M., Marks, L., May, G. H. W. & Wilson, J. B. The genetic basis of disease. *Essays Biochem* **62**, 643–723 (2018).
3. Lappalainen, T. & MacArthur, D. G. From variant to function in human disease genetics. *Science* **373**, 1464–1468 (2021).
4. Cannon, M. E. & Mohlke, K. L. Deciphering the Emerging Complexities of Molecular Mechanisms at GWAS Loci. *Am J Hum Genet* **103**, 637–653 (2018).
5. Collins, A. Allelic association: linkage disequilibrium structure and gene mapping. *Mol Biotechnol* **41**, 83–89 (2009).
6. Ward, L. D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* **44**, D877-881 (2016).
7. Wang, T. *et al.* The Human Pangenome Project: a global resource to map genomic diversity. *Nature* **604**, 437–446 (2022).
8. Hutchinson, A., Asimit, J. & Wallace, C. Fine-mapping genetic associations. *Hum Mol Genet* **29**, R81–R88 (2020).
9. Mahajan, A. *et al.* Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet* **54**, 560–572 (2022).
10. Pengelly, R. J. *et al.* Whole genome sequences are required to fully resolve the linkage disequilibrium structure of human populations. *BMC Genomics* **16**, 666 (2015).
11. Vergara-Lope, A. *et al.* Linkage disequilibrium maps for European and African populations constructed from whole genome sequence data. *Sci Data* **6**, 208 (2019).
12. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med* **373**, 895–907 (2015).
13. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am J Hum Genet* **99**, 1245–1260 (2016).
14. Foley, C. N. *et al.* A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat Commun* **12**, 764 (2021).
15. Wu, Y. *et al.* Colocalization of GWAS and eQTL signals at loci with multiple signals identifies additional candidate genes for body fat distribution. *Hum Mol Genet* **28**, 4161–4172 (2019).
16. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005–D1012 (2019).
17. Lewis, C. M. & Vassos, E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med* **12**, 44 (2020).
18. Ragsdale, A. P. *et al.* A weakly structured stem for human origins in Africa. *Nature* **617**, 755–763 (2023).
19. Fedorova, L. *et al.* Analysis of Common SNPs across Continents Reveals Major Genomic Differences between Human Populations. *Genes (Basel)* **13**, 1472 (2022).
20. Tang, M., Wang, T. & Zhang, X. A review of SNP heritability estimation methods. *Brief Bioinform* **23**, bbac067 (2022).
21. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
22. Marian, A. J. Elements of ‘missing heritability’. *Curr Opin Cardiol* **27**, 197–201 (2012).
23. Hunter, D. J. Gene-environment interactions in human diseases. *Nat Rev Genet* **6**, 287–298 (2005).
24. Elhawary, N. A. *et al.* Genetic etiology and clinical challenges of phenylketonuria. *Hum Genomics* **16**, 22 (2022).
25. Nair, A. K. & Baier, L. J. Using Luciferase Reporter Assays to Identify Functional Variants at Disease-Associated Loci. *Methods Mol Biol* **1706**, 303–319 (2018).
26. Rao, S., Yao, Y. & Bauer, D. E. Editing GWAS: experimental approaches to dissect and exploit disease-associated genetic variation. *Genome Med* **13**, 41 (2021).
27. Abell, N. S. *et al.* Multiple causal variants underlie genetic associations in humans. *Science* **375**, 1247–1254 (2022).
28. Sharma, S. & Rao, A. RNAi screening: tips and techniques. *Nat Immunol* **10**, 799–804 (2009).

29. Perwitasari, O., Bakre, A., Tompkins, S. M. & Tripp, R. A. siRNA Genome Screening Approaches to Therapeutic Drug Repositioning. *Pharmaceuticals (Basel)* **6**, 124–160 (2013).
30. Rao, D. D., Vorhies, J. S., Senzer, N. & Nemunaitis, J. siRNA vs. shRNA: similarities and differences. *Adv Drug Deliv Rev* **61**, 746–759 (2009).
31. Bai, H., Lester, G. M. S., Petishnok, L. C. & Dean, D. A. Cytoplasmic transport and nuclear import of plasmid DNA. *Biosci Rep* **37**, BSR20160616 (2017).
32. Dean, D. A. Cell-specific targeting strategies for electroporation-mediated gene delivery in cells and animals. *J Membr Biol* **246**, 737–744 (2013).
33. Page, A., Fusil, F. & Cosset, F.-L. Toward Tightly Tuned Gene Expression Following Lentiviral Vector Transduction. *Viruses* **12**, 1427 (2020).
34. Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *J Virol* **72**, 8463–8471 (1998).
35. Schleinitz, D. *et al.* Fat depot-specific mRNA expression of novel loci associated with waist-hip ratio. *Int J Obes (Lond)* **38**, 120–125 (2014).
36. Karastergiou, K. *et al.* Distinct developmental signatures of human abdominal and gluteal subcutaneous adipose tissue depots. *J Clin Endocrinol Metab* **98**, 362–371 (2013).
37. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
38. Ruiz-Ojeda, F. J., Rupérez, A. I., Gomez-Llorente, C., Gil, A. & Aguilera, C. M. Cell Models and Their Application for Studying Adipogenic Differentiation in Relation to Obesity: A Review. *Int J Mol Sci* **17**, 1040 (2016).
39. Wabitsch, M. *et al.* Characterization of a human preadipocyte cell strain with high capacity for adipose differentiation. *Int J Obes Relat Metab Disord* **25**, 8–15 (2001).
40. Cave, E. & Crowther, N. J. The Use of 3T3-L1 Murine Preadipocytes as a Model of Adipogenesis. *Methods Mol Biol* **1916**, 263–272 (2019).
41. Chusyd, D. E., Wang, D., Huffman, D. M. & Nagy, T. R. Relationships between Rodent White Adipose Fat Pads and Human White Adipose Fat Depots. *Front Nutr* **3**, 10 (2016).
42. Börgeson, E., Boucher, J. & Hagberg, C. E. Of mice and men: Pinpointing species differences in adipose tissue biology. *Front Cell Dev Biol* **10**, 1003118 (2022).
43. Palumbo, P. *et al.* Methods of Isolation, Characterization and Expansion of Human Adipose-Derived Stem Cells (ASCs): An Overview. *Int J Mol Sci* **19**, 1897 (2018).
44. Jebari-Benslaiman, S. *et al.* Pathophysiology of Atherosclerosis. *Int J Mol Sci* **23**, 3346 (2022).
45. Doulberis, M. *et al.* Rodent models of obesity. *Minerva Endocrinol* **45**, 243–263 (2020).
46. Ingalls, A. M., Dickie, M. M. & Snell, G. D. Obese, a new mutation in the house mouse. *J Hered* **41**, 317–318 (1950).
47. Hummel, K. P., Dickie, M. M. & Coleman, D. L. Diabetes, a new mutation in the mouse. *Science* **153**, 1127–1128 (1966).
48. Emini Veseli, B. *et al.* Animal models of atherosclerosis. *Eur J Pharmacol* **816**, 3–13 (2017).
49. Kleinert, M. *et al.* Animal models of obesity and diabetes mellitus. *Nat Rev Endocrinol* **14**, 140–162 (2018).
50. Loh, N. Y. *et al.* RSPO3 impacts body fat distribution and regulates adipose cell biology in vitro. *Nat Commun* **11**, 2797 (2020).
51. Ke, W. *et al.* Genes in human obesity loci are causal obesity genes in C. elegans. *PLoS Genet* **17**, e1009736 (2021).
52. Small, K. S. *et al.* Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nat Genet* **50**, 572–580 (2018).
53. Yang, Q. *et al.* Adipocyte-Specific Modulation of KLF14 Expression in Mice Leads to Sex-Dependent Impacts on Adiposity and Lipid Metabolism. *Diabetes* **71**, 677–693 (2022).
54. Loh, N. Y. *et al.* LRP5 regulates human body fat distribution by modulating adipose progenitor biology in a dose- and depot-specific fashion. *Cell Metab* **21**, 262–273 (2015).
55. Gesta, S. *et al.* Mesodermal developmental gene Tbx15 impairs adipocyte differentiation and mitochondrial respiration. *Proc Natl Acad Sci U S A* **108**, 2771–2776 (2011).
56. Gesta, S. *et al.* Evidence for a role of developmental genes in the origin of obesity and body fat distribution. *Proc Natl Acad Sci U S A* **103**, 6676–6681 (2006).
57. Lee, K. Y. *et al.* Shox2 is a molecular determinant of depot-specific adipocyte function. *Proc Natl Acad Sci U S A* **110**, 11409–11414 (2013).

58. Hildreth, A. D. *et al.* Single-cell sequencing of human white adipose tissue identifies new cell states in health and obesity. *Nat Immunol* **22**, 639–653 (2021).
59. Sun, W. *et al.* snRNA-seq reveals a subpopulation of adipocytes that regulates thermogenesis. *Nature* **587**, 98–102 (2020).
60. Vijay, J. *et al.* Single-cell analysis of human adipose tissue identifies depot and disease specific cell types. *Nat Metab* **2**, 97–109 (2020).
61. Sárvári, A. K. *et al.* Plasticity of Epididymal Adipose Tissue in Response to Diet-Induced Obesity at Single-Nucleus Resolution. *Cell Metab* **33**, 437-453.e5 (2021).
62. Hepler, C. *et al.* Identification of functionally distinct fibro-inflammatory and adipogenic stromal subpopulations in visceral adipose tissue of adult mice. *Elife* **7**, e39636 (2018).
63. Almanzar, N. *et al.* A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020).
64. Min, S. Y. *et al.* Diverse repertoire of human adipocyte subtypes develops from transcriptionally distinct mesenchymal progenitor cells. *Proc Natl Acad Sci U S A* **116**, 17970–17979 (2019).
65. Teo, J.-L. & Kahn, M. The Wnt signaling pathway in cellular proliferation and differentiation: A tale of two coactivators. *Adv Drug Deliv Rev* **62**, 1149–1155 (2010).
66. Chen, N. & Wang, J. Wnt/β-Catenin Signaling and Obesity. *Front Physiol* **9**, 792 (2018).
67. Binnerts, M. E. *et al.* R-Spondin1 regulates Wnt signaling by inhibiting internalization of LRP6. *Proc Natl Acad Sci U S A* **104**, 14700–14705 (2007).
68. Gore, A. V. *et al.* Rspo1/Wnt signaling promotes angiogenesis via Vegfc/Vegfr3. *Development* **138**, 4875–4886 (2011).
69. Chen, D., Liu, Q., Cao, G. & Zhang, W. TYRO3 facilitates cell growth and metastasis via activation of the Wnt/β-catenin signaling pathway in human gastric cancer cells. *Aging (Albany NY)* **12**, 2261–2274 (2020).
70. Zhu, Y.-Z., Wang, W., Xian, N. & Wu, B. Inhibition of TYRO3/Akt signaling participates in hypoxic injury in hippocampal neurons. *Neural Regen Res* **11**, 752–757 (2016).
71. Chen, H. *et al.* REGγ accelerates melanoma formation by regulating Wnt/β-catenin signalling pathway. *Exp Dermatol* **26**, 1118–1124 (2017).
72. Li, L. *et al.* REGγ is critical for skin carcinogenesis by modulating the Wnt/β-catenin pathway. *Nat Commun* **6**, 6875 (2015).
73. Ganner, A. *et al.* Regulation of ciliary polarity by the APC/C. *Proc Natl Acad Sci U S A* **106**, 17799–17804 (2009).
74. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* **28**, 166–174 (2019).
75. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from dbGaP Accession phs000424.v8.p2on 10/01/2020.
76. Fischer-Posovszky, P., Newell, F. S., Wabitsch, M. & Tornqvist, H. E. Human SGBS Cells – a Unique Tool for Studies of Human Fat Cell Biology. *Obes Facts* **1**, 184–189 (2008).
77. Ackers, I. & Malgor, R. Interrelationship of canonical and non-canonical Wnt signalling pathways in chronic metabolic diseases. *Diab Vasc Dis Res* **15**, 3–13 (2018).
78. Akoumianakis, I., Polkinghorne, M. & Antoniades, C. Non-canonical WNT signalling in cardiovascular disease: mechanisms and therapeutic implications. *Nat Rev Cardiol* **19**, 783–797 (2022).
79. Feng, H. *et al.* CTRP3 promotes energy production by inducing mitochondrial ROS and up-expression of PGC-1α in vascular smooth muscle cells. *Exp Cell Res* **341**, 177–186 (2016).
80. Zhang, C.-L. *et al.* Globular CTRP3 promotes mitochondrial biogenesis in cardiomyocytes through AMPK/PGC-1α pathway. *Biochim Biophys Acta Gen Subj* **1861**, 3085–3094 (2017).
81. Gao, J., Qian, T. & Wang, W. CTRP3 Activates the AMPK/SIRT1-PGC-1α Pathway to Protect Mitochondrial Biogenesis and Functions in Cerebral Ischemic Stroke. *Neurochem Res* **45**, 3045–3058 (2020).
82. Liu, X.-M. *et al.* Mitoguardin-1 and -2 promote maturation and the developmental potential of mouse oocytes by maintaining mitochondrial dynamics and functions. *Oncotarget* **7**, 1155–1167 (2016).

83. Zhang, Y. *et al.* Mitoguardin Regulates Mitochondrial Fusion through MitoPLD and Is Required for Neuronal Homeostasis. *Molecular Cell* **61**, 111–124 (2016).
84. Xie, Y. *et al.* The proteasome activator REGγ accelerates cardiac hypertrophy by declining PP2Acα-SOD2 pathway. *Cell Death Differ* **27**, 2952–2972 (2020).
85. Moncsek, A. *et al.* Evidence for anti-apoptotic roles of proteasome activator 28γ via inhibiting caspase activity. *Apoptosis* **20**, 1211–1228 (2015).
86. Tran, A. The N-end rule pathway and Ubr1 enforce protein compartmentalization via P2-encoded cellular location signals. *J Cell Sci* **132**, jcs231662 (2019).
87. Metzger, M. B., Scales, J. L., Dunklebarger, M. F., Loncarek, J. & Weissman, A. M. A protein quality control pathway at the mitochondrial outer membrane. *Elife* **9**, e51065 (2020).
88. Yamano, K. & Youle, R. J. PINK1 is degraded through the N-end rule pathway. *Autophagy* **9**, 1758–1769 (2013).
89. Ding, W., Yousefi, K. & Shehadeh, L. A. Isolation, Characterization, And High Throughput Extracellular Flux Analysis of Mouse Primary Renal Tubular Epithelial Cells. *J Vis Exp* 57718 (2018) doi:10.3791/57718.
90. Rose, S. *et al.* Oxidative stress induces mitochondrial dysfunction in a subset of autism lymphoblastoid cell lines in a well-matched case control cohort. *PLoS One* **9**, e85436 (2014).
91. Son, J. M. *et al.* Mitofusin 1 and optic atrophy 1 shift metabolism to mitochondrial respiration during aging. *Aging Cell* **16**, 1136–1145 (2017).
92. Bertholet, A. M. *et al.* Mitochondrial uncouplers induce proton leak by activating AAC and UCP1. *Nature* **606**, 180–187 (2022).
93. Franzén, O. *et al.* Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**, 827–830 (2016).
94. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
95. Raulerson, C. K. *et al.* Adipose Tissue Gene Expression Associations Reveal Hundreds of Candidate Genes for Cardiometabolic Traits. *Am J Hum Genet* **105**, 773–787 (2019).
96. Fuster, J. J. *et al.* Noncanonical Wnt signaling promotes obesity-induced adipose tissue inflammation and metabolic dysfunction independent of adipose tissue expansion. *Diabetes* **64**, 1235–1248 (2015).
97. Kang, Y. E. *et al.* Serum R-Spondin 1 Is a New Surrogate Marker for Obesity and Insulin Resistance. *Diabetes Metab J* **43**, 368–376 (2019).
98. Sun, Y. *et al.* Human RSPO1 Mutation Represses Beige Adipocyte Thermogenesis and Contributes to Diet-Induced Adiposity. *Adv Sci (Weinh)* e2207152 (2023) doi:10.1002/advs.202207152.
99. Yu, H. *et al.* Identification of a cullin homology region in a subunit of the anaphase-promoting complex. *Science* **279**, 1219–1222 (1998).
100. Zhao, M. L. *et al.* Molecular Competition in G1 Controls When Cells Simultaneously Commit to Terminally Differentiate and Exit the Cell Cycle. *Cell Rep* **31**, 107769 (2020).
101. Du, Y., Liu, Y., Zhou, Y. & Zhang, P. Knockdown of CDC20 promotes adipogenesis of bone marrow-derived stem cells by modulating β-catenin. *Stem Cell Res Ther* **13**, 443 (2022).
102. Lee, J., Lee, C. G., Lee, K.-W. & Lee, C.-W. Cross-talk between BubR1 expression and the commitment to differentiate in adipose-derived mesenchymal stem cells. *Exp Mol Med* **41**, 873–879 (2009).
103. Wong, A. C. H. The Role Of Intron Retention In Acute Myeloid Leukaemia. (2021).
104. Funderburk, K. E., Kang, J. & Li, H. J. Regulation of Life & Death by REGγ. *Cells* **11**, 2281 (2022).
105. Willemsen, N., Arigoni, I., Studencka-Turski, M., Krüger, E. & Bartelt, A. Proteasome dysfunction disrupts adipogenesis and induces inflammation via ATF3. *Mol Metab* **62**, 101518 (2022).
106. Kitamura, A. *et al.* A mutation in the immunoproteasome subunit PSMB8 causes autoinflammation and lipodystrophy in humans. *J Clin Invest* **121**, 4150–4160 (2011).
107. Lee, H. J. *et al.* Cold temperature extends longevity and prevents disease-related protein aggregation through PA28γ-induced proteasomes. *Nat Aging* **3**, 546–566 (2023).
108. Dong, S. *et al.* The REGγ proteasome regulates hepatic lipid metabolism through inhibition of autophagy. *Cell Metab* **18**, 380–391 (2013).

109. Freyre, C. A. C., Rauher, P. C., Ejsing, C. S. & Klemm, R. W. MIGA2 Links Mitochondria, the ER, and Lipid Droplets and Promotes De Novo Lipogenesis in Adipocytes. *Mol Cell* **76**, 811-825.e14 (2019).
110. Boutant, M. *et al.* Mfn2 is critical for brown adipose tissue thermogenic function. *EMBO J* **36**, 1543–1558 (2017).
111. Mancini, G. *et al.* Mitofusin 2 in Mature Adipocytes Controls Adiposity and Body Weight. *Cell Rep* **26**, 2849-2858.e4 (2019).
112. Tomizuka, K. *et al.* R-spondin1 plays an essential role in ovarian development through positively regulating Wnt-4 signaling. *Hum Mol Genet* **17**, 1278–1291 (2008).
113. Huang, L., Haratake, K., Miyahara, H. & Chiba, T. Proteasome activators, PA28γ and PA200, play indispensable roles in male fertility. *Sci Rep* **6**, 23171 (2016).
114. McGuinness, B. E. *et al.* Regulation of APC/C activity in oocytes by a Bub1-dependent spindle assembly checkpoint. *Curr Biol* **19**, 369–380 (2009).
115. Geng, A. *et al.* A novel function of R-spondin1 in regulating estrogen receptor expression independent of Wnt/β-catenin signaling. *Elife* **9**, e56434 (2020).
116. Blake, J. A. *et al.* Mouse Genome Database (MGD): Knowledgebase for mouse-human comparative biology. *Nucleic Acids Res* **49**, D981–D987 (2021).
117. Groza, T. *et al.* The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Research* **51**, D1038–D1045 (2023).
118. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682 (2012).

3.6.1 Supplemental Bibliographies

Supplemental Bibliography 3.1- Literature Search 45 Adipocyte Genes:

1. Ibrahim S, Temtem T. Medium-Chain Acyl-COA Dehydrogenase Deficiency. 2021 Jul 26. In: StatPearls
2. Lim SC, et al. Loss of the Mitochondrial Fatty Acid β-Oxidation Protein Medium-Chain Acyl-Coenzyme A Dehydrogenase Disrupts Oxidative Phosphorylation Protein Complex Stability and Function. Sci Rep. 2018 Jan 9;8(1):153.
3. Houten SM, Wanders RJ. A general introduction to the biochemistry of mitochondrial fatty acid β-oxidation. J Inherit Metab Dis. 2010 Oct;33(5):469-77.
4. Vega RB, Kelly DP. A role for estrogen-related receptor alpha in the control of mitochondrial fatty acid beta-oxidation during brown adipocyte differentiation. J Biol Chem. 1997 Dec 12;272(50):31693-9.
5. Huang LH, et al. Myeloid-specific Acat1 ablation attenuates inflammatory responses in macrophages, improves insulin sensitivity, and suppresses diet-induced obesity. Am J Physiol Endocrinol Metab. 2018 Sep 1;315(3):E340-E356.
6. Xu Y, et al. Enhanced acyl-CoA:cholesterol acyltransferase activity increases cholesterol levels on the lipid droplet surface and impairs adipocyte function. J Biol Chem. 2019 Dec 13;294(50):19306-19321.
7. Zhu Y, et al. In vitro exploration of ACAT contributions to lipid droplet formation during adipogenesis. J Lipid Res. 2018 May;59(5):820-829.
8. Moreno M, et al. Cytosolic aconitase activity sustains adipogenic capacity of adipose tissue connecting iron metabolism and adipogenesis. FASEB J. 2015 Apr;29(4):1529-39.
9. Yogosawa S, et al. Activin receptor-like kinase 7 suppresses lipolysis to accumulate fat in obesity through downregulation of peroxisome proliferator-activated receptor γ and C/EBPα. Diabetes. 2013 Jan;62(1):115-23.
10. Suchý T, et al. The repertoire of Adhesion G protein-coupled receptors in adipocytes and their functional relevance. Int J Obes (Lond). 2020 Oct;44(10):2124-2136.
11. Paik J, Haenisch M, Muller CH, Goldstein AS, Arnold S, Isoherranen N, Brabb T, Treuting PM, Amory JK. Inhibition of retinoic acid biosynthesis by the bisdichloroacetyldiamine WIN 18,446 markedly suppresses spermatogenesis and alters retinoid metabolism in mice. J Biol Chem. 2014 May 23;289(21):15104-17.
12. Kang S. Adipose Tissue Malfunction Drives Metabolic Dysfunction in Alström Syndrome. Diabetes. 2021 Feb;70(2):323-325.
13. Geberhiwot T, et al. Relative Adipose Tissue Failure in Alström Syndrome Drives Obesity-Induced Insulin Resistance. Diabetes. 2021 Feb;70(2):364-376.
14. Hearn T. ALMS1 and Alström syndrome: a recessive form of metabolic, neurosensory and cardiac deficits. J Mol Med (Berl). 2019 Jan;97(1):1-17.
15. Huang-Doran I, Semple RK. Knockdown of the Alström syndrome-associated gene Alms1 in 3T3-L1 preadipocytes impairs adipogenesis but has no effect on cell-autonomous insulin action. Int J Obes (Lond). 2010 Oct;34(10):1554-8.
16. Favaretto F, et al. GLUT4 defects in adipose tissue are early signs of metabolic alterations in Alms1GT/GT, a mouse model for obesity and insulin resistance. PLoS One. 2014 Oct 9;9(10):e109540.
17. Claussnitzer M, et al. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N Engl J Med. 2015 Sep 3;373(10):895-907.
18. Lee KY, et al. The differential role of Hif1β/Arnt and the hypoxic response in adipose function, fibrosis, and inflammation. Cell Metab. 2011 Oct 5;14(4):491-503.
19. Kälin S, et al. A Stat6/Pten Axis Links Regulatory T Cells with Adipose Tissue Function. Cell Metab. 2017 Sep 5;26(3):475-492.e7.
20. Valerio CM, et al. Dipeptidyl peptidase-4 levels are increased and partially related to body fat distribution in patients with familial partial lipodystrophy type 2. Diabetol Metab Syndr. 2017 Apr 24;9:26.
21. Marques AP, et al. Dipeptidyl peptidase IV (DPP-IV) inhibition prevents fibrosis in adipose tissue of obese mice. Biochim Biophys Acta Gen Subj. 2018 Mar;1862(3):403-413.
22. Zilleßen P, et al. Metabolic role of dipeptidyl peptidase 4 (DPP4) in primary human (pre)adipocytes. Sci Rep. 2016 Mar 17;6:23074.

23. Han J, et al. ER stress signalling through eIF2α and CHOP, but not IRE1α, attenuates adipogenesis in mice. Diabetologia. 2013 Apr;56(4):911-24.
24. Kobori M, et al. Dietary Intake of Curcumin Improves eIF2 Signaling and Reduces Lipid Levels in the White Adipose Tissue of Obese Mice. Sci Rep. 2018 Jun 13;8(1):9081.
25. Dagon Y, Avraham Y, Berry EM. AMPK activation regulates apoptosis, adipogenesis, and lipolysis by eIF2alpha in adipocytes. Biochem Biophys Res Commun. 2006 Feb 3;340(1):43-7.
26. Inoue M, et al. Compartmentalization of the exocyst complex in lipid rafts controls Glut4 vesicle tethering. Mol Biol Cell. 2006 May;17(5):2303-11.
27. Lizunov VA, et al. Insulin regulates fusion of GLUT4 vesicles independent of Exo70-mediated tethering. J Biol Chem. 2009 Mar 20;284(12):7914-9.
28. Wang S, et al Inducible *Exoc7/Exo70* knockout reveals a critical role of the exocyst in insulin-regulated GLUT4 exocytosis. J Biol Chem. 2019 Dec 27;294(52):19988-19996.
29. Miao Y, et al. Novel adipokine asprosin modulates browning and adipogenesis in white adipose tissue. J Endocrinol. 2021 May;249(2):83-93.
30. Muthu ML, Reinhardt DP. Fibrillin-1 and fibrillin-1-derived asprosin in adipose tissue function and metabolic disorders. J Cell Commun Signal. 2020 Jun;14(2):159-173.
31. Passarge E, Robinson PN, Graul-Neumann LM. Marfanoid-progeroid-lipodystrophy syndrome: a newly recognized fibrillinopathy. Eur J Hum Genet. 2016 Aug;24(9):1244-7.
32. Takenouchi T, et al. Severe congenital lipodystrophy and a progeroid appearance: Mutation in the penultimate exon of FBN1 causing a recognizable phenotype. Am J Med Genet A. 2013 Dec;161A(12):3057-62.
33. Zhao L, et al. Fibroblast growth factor 1 ameliorates adipose tissue inflammation and systemic insulin resistance via enhancing adipocyte mTORC2/Rictor signal. J Cell Mol Med. 2020 Nov;24(21):12813-12825.
34. Sun K, Scherer PE. The PPARγ-FGF1 axis: an unexpected mediator of adipose tissue homeostasis. Cell Res. 2012 Oct;22(10):1416-8.
35. Wang S, et al. Adipocyte Piezo1 mediates obesogenic adipogenesis through the FGF1/FGFR1 signaling pathway in mice. Nat Commun. 2020 May 8;11(1):2303.
36. Jonker JW,et al. A PPARγ-FGF1 axis is required for adaptive adipose remodelling and metabolic homeostasis. Nature. 2012 May 17;485(7398):391-4.
37. Nies VJ, et al. Fibroblast Growth Factor Signaling in Metabolic Regulation. Front Endocrinol (Lausanne). 2016 Jan 19;6:193.
38. Lindegaard B, et al. Expression of fibroblast growth factor-21 in muscle is associated with lipodystrophy, insulin resistance and lipid disturbances in patients with HIV. PLoS One. 2013;8(3):e55632.
39. Wu L, et al. *GNPDA2* Gene Affects Adipogenesis and Alters the Transcriptome Profile of Human Adipose-Derived Mesenchymal Stem Cells. Int J Endocrinol. 2019 Aug 1;2019:9145452.
40. Chan CY, et al. Transcription factor HMG box-containing protein 1 (HBP1) modulates mitotic clonal expansion (MCE) during adipocyte differentiation. J Cell Physiol. 2018 May;233(5):4205-4215.
41. Zhang X, et al. HSPA12A is required for adipocyte differentiation and diet-induced obesity through a positive feedback regulation with PPARγ. Cell Death Differ. 2019 Nov;26(11):2253-2267.
42. Vietor I, et al. The negative adipogenesis regulator DLK1 is transcriptionally regulated by TIS7 (IFRD1) and translationally by its orthologue SKMc15 (IFRD2). BioRXiv Preprint, 2020 Oct 15.
43. Vietor I, et al. TIS7 and SKMc15 Regulate Adipocyte Differentiation and Intestinal Lipid Absorption. BioRXiv Preprint, 2019 Jul 30.
44. Moure R, et al. Levels of β-klotho determine the thermogenic responsiveness of adipose tissues: involvement of the autocrine action of FGF21. Am J Physiol Endocrinol Metab. 2021 Apr 1;320(4):E822-E834.
45. Ogawa Y, et al. BetaKlotho is required for metabolic activity of fibroblast growth factor 21. Proc Natl Acad Sci U S A. 2007 May 1;104(18):7432-7.
46. Simon MF, et al. Lysophosphatidic acid inhibits adipocyte differentiation via lysophosphatidic acid 1 receptor-dependent down-regulation of peroxisome proliferator-activated receptor gamma2. J Biol Chem. 2005 Apr 15;280(15):14656-62.
47. Chabowski DS, et al. Lysophosphatidic acid acts on LPA_1 receptor to increase $H_2 O_2$ during flow-induced dilation in human adipose arterioles. Br J Pharmacol. 2018 Nov;175(22):4266-4280.

48. Wang J, et al. miR-30e reciprocally regulates the differentiation of adipocytes and osteoblasts by directly targeting low-density lipoprotein receptor-related protein 6. Cell Death Dis. 2013 Oct 10;4(10):e845.
49. Liu W, et al. Low density lipoprotein (LDL) receptor-related protein 6 (LRP6) regulates body fat and glucose homeostasis by modulating nutrient sensing pathways and mitochondrial energy expenditure. J Biol Chem. 2012 Mar 2;287(10):7213-23.
50. Zhao C, et al. MAT2B promotes adipogenesis by modulating SAMe levels and activating AKT/ERK pathway during porcine intramuscular preadipocyte differentiation. Exp Cell Res. 2016 May 15;344(1):11-21.
51. Li C, et al. Adipose-derived mesenchymal stem cells attenuate ischemic brain injuries in rats by modulating miR-21-3p/MAT2B signaling transduction. Croat Med J. 2019 Oct 31;60(5):439-448.
52. Kim JY, et al. ER Stress Drives Lipogenesis and Steatohepatitis via Caspase-2 Activation of S1P. Cell. 2018 Sep 20;175(1):133-145.e15.
53. Takahashi Y, et al. Perilipin-mediated lipid droplet formation in adipocytes promotes sterol regulatory element-binding protein-1 processing and triacylglyceride accumulation. PLoS One. 2013 May 29;8(5):e64605.
54. Ostrakhovitch EA, et al. 3-Mercaptopyruvate sulfurtransferase disruption in dermal fibroblasts facilitates adipogenic trans-differentiation. Exp Cell Res. 2019 Dec 15;385(2):111683.
55. Ying W, et al. MiR-690, an exosomal-derived miRNA from M2-polarized macrophages, improves insulin sensitivity in obese mice. Cell Metab. 2021 Apr 6;33(4):781-790.e5.
56. Navas LE, Carnero A. NAD$^+$ metabolism, stemness, the immune response, and cancer. Signal Transduct Target Ther. 2021 Jan 1;6(1):2.
57. Katwan OJ, et al. AMP-activated protein kinase complexes containing the β2 regulatory subunit are up-regulated during and contribute to adipogenesis. Biochem J. 2019 Jun 26;476(12):1725-1740.
58. Ding Q, Wang Z, Chen Y. Endocytosis of adiponectin receptor 1 through a clathrin- and Rab5-dependent pathway. Cell Res. 2009 Mar;19(3):317-27.
59. Tessneer KL, et al. Rab5 activity regulates GLUT4 sorting into insulin-responsive and non-insulin-responsive endosomal compartments: a potential mechanism for development of insulin resistance. Endocrinology. 2014 Sep;155(9):3315-28.
60. Karvela A, et al. Adiponectin Signaling and Impaired GTPase Rab5 Expression in Adipocytes of Adolescents with Obesity. Horm Res Paediatr. 2020;93(5):287-296.
61. Xie L, O'Reilly CP, Chapes SK, Mora S. Adiponectin and leptin are secreted through distinct trafficking pathways in adipocytes. Biochim Biophys Acta. 2008 Feb;1782(2):99-108.
62. Chun KH, et al. Regulation of glucose transport by ROCK1 differs from that of ROCK2 and is controlled by actin polymerization. Endocrinology. 2012 Apr;153(4):1649-62.
63. Lee DH, et al. Targeted disruption of ROCK1 causes insulin resistance in vivo. J Biol Chem. 2009 May 1;284(18):11776-80.
64. Dankel SN, et al. The Rho GTPase RND3 regulates adipocyte lipolysis. Metabolism. 2019 Dec;101:153999.
65. Imai T, Jiang M, Chambon P, Metzger D. Impaired adipogenesis and lipolysis in the mouse upon selective ablation of the retinoid X receptor alpha mediated by a tamoxifen-inducible chimeric Cre recombinase (Cre-ERT2) in adipocytes. Proc Natl Acad Sci U S A. 2001 Jan 2;98(1):224-8.
66. Shoucri BM, et al. Retinoid X Receptor Activation During Adipogenesis of Female Mesenchymal Stem Cells Programs a Dysfunctional Adipocyte. Endocrinology. 2018 Aug 1;159(8):2863-2883.
67. Lefebvre B, et al. Proteasomal degradation of retinoid X receptor alpha reprograms transcriptional activity of PPARgamma in obese mice and humans. J Clin Invest. 2010 May;120(5):1454-68.
68. Garg A, et al. A gene for congenital generalized lipodystrophy maps to human chromosome 9q34. J Clin Endocrinol Metab. 1999 Sep;84(9):3390-4.
69. Mizuarai S, et al. Identification of dicarboxylate carrier Slc25a10 as malate transporter in de novo fatty acid synthesis. J Biol Chem. 2005 Sep 16;280(37):32434-41.
70. Fukunaka A, et al. Zinc transporter ZIP13 suppresses beige adipocyte biogenesis and energy expenditure by regulating C/EBP-β expression. PLoS Genet. 2017 Aug 30;13(8):e1006950.
71. Liu C, et al. Fat-Specific Knockout of Mecp2 Upregulates Slpi to Reduce Obesity by Enhancing Browning. Diabetes. 2020 Jan;69(1):35-47.
72. Adapala VJ, Buhman KK, Ajuwon KM. Novel anti-inflammatory role of SLPI in adipose tissue and its regulation by high fat diet. J Inflamm (Lond). 2011 Feb 28;8:5.

73. Kim JH, J. C-terminus of HSC70-Interacting Protein (CHIP) Inhibits Adipocyte Differentiation via Ubiquitin- and Proteasome-Mediated Degradation of PPARγ. Sci Rep. 2017 Jan 6;7:40023.
74. Lim CY, et al. Tropomodulin3 is a novel Akt2 effector regulating insulin-stimulated GLUT4 exocytosis through cortical actin remodeling. Nat Commun. 2015 Jan 9;6:5951.
75. Zhang Y, Gu M, Ma Y, Peng Y. LncRNA TUG1 reduces inflammation and enhances insulin sensitivity in white adipose tissue by regulating miR-204/SIRT1 axis in obesity mice. Mol Cell Biochem. 2020 Dec;475(1-2):171-183.
76. Long J, et al. Role for carbohydrate response element-binding protein (ChREBP) in high glucose-mediated repression of long noncoding RNA Tug1. J Biol Chem. 2020 Nov 20;295(47):15840-15852.
77. Zhang Y, Ma Y, Gu M, Peng Y. lncRNA TUG1 promotes the brown remodeling of white adipose tissue by regulating miR-204-targeted SIRT1 in diabetic mice. Int J Mol Med. 2020 Dec;46(6):2225-2234.
78. Peterson JM, et al. CTRP2 overexpression improves insulin and lipid tolerance in diet-induced obese mice. PLoS One. 2014 Feb 20;9(2):e88535.
79. Lei X, Wong GW. C1q/TNF-related protein 2 (CTRP2) deletion promotes adipose tissue lipolysis and hepatic triglyceride secretion. J Biol Chem. 2019 Oct 25;294(43):15638-15649.
80. Ou CY, et al. Coregulator cell cycle and apoptosis regulator 1 (CCAR1) positively regulates adipocyte differentiation through the glucocorticoid signaling pathway. J Biol Chem. 2014 Jun 13;289(24):17078-86.
81. Moreno-Navarrete JM, et al. Deleted in breast cancer 1 plays a functional role in adipocyte differentiation. Am J Physiol Endocrinol Metab. 2015 Apr 1;308(7):E554-61.
82. Escande C, et al. Deleted in breast cancer 1 limits adipose tissue fat accumulation and plays a key role in the development of metabolic syndrome phenotype. Diabetes. 2015 Jan;64(1):12-22.
83. Moreno-Navarrete JM, et al. DBC1 is involved in adipocyte inflammation and is a possible marker of human adipose tissue senescence. Obesity (Silver Spring). 2015 Mar;23(3):519-22.
84. Able AA, Richard AJ, Stephens JM. Loss of DBC1 (CCAR2) affects TNFα-induced lipolysis and *Glut4* gene expression in murine adipocytes. J Mol Endocrinol. 2018 Oct 15;61(4):195-205.
85. Botero V, et al. Neurofibromin regulates metabolic rate via neuronal mechanisms in Drosophila. Nat Commun. 2021 Jul 13;12(1):4285.
86. Souza MLR, et al. Increased resting metabolism in neurofibromatosis type 1. Clin Nutr ESPEN. 2019 Aug;32:44-49.
87. Summers MA, et al. Dietary intervention rescues myopathy associated with neurofibromatosis type 1. Hum Mol Genet. 2018 Feb 15;27(4):577-588.
88. Wei X, et al. Cell autonomous requirement of neurofibromin (Nf1) for postnatal muscle hypertrophic growth and metabolic homeostasis. J Cachexia Sarcopenia Muscle. 2020 Dec;11(6):1758-1778.
89. Greenberg CC, Danos AM, Brady MJ. Central role for protein targeting to glycogen in the maintenance of cellular glycogen stores in 3T3-L1 adipocytes. Mol Cell Biol. 2006 Jan;26(1):334-42.
90. Printen JA, Brady MJ, Saltiel AR. PTG, a protein phosphatase 1-binding protein with a role in glycogen metabolism. Science. 1997 Mar 7;275(5305):1475-8.
91. Matsui Y, et al. Overexpression of TNF-α converting enzyme promotes adipose tissue inflammation and fibrosis induced by high fat diet. Exp Mol Pathol. 2014 Dec;97(3):354-8.
92. Yong SB, Song Y, Kim YH. Visceral adipose tissue macrophage-targeted TACE silencing to treat obesity-induced type 2 diabetes. Biomaterials. 2017 Dec;148:81-89.
93. Menghini R, et al. The role of ADAM17 in metabolic inflammation. Atherosclerosis. 2013 May;228(1):12-7.
94. Wang Y, Sul HS. Ectodomain shedding of preadipocyte factor 1 (Pref-1) by tumor necrosis factor alpha converting enzyme (TACE) and inhibition of adipocyte differentiation. Mol Cell Biol. 2006 Jul;26(14):5421-35.
95. Gelling RW, et al. Deficiency of TNFalpha converting enzyme (TACE/ADAM17) causes a lean, hypermetabolic phenotype in mice. Endocrinology. 2008 Dec;149(12):6053-64.
96. Lownik JC, et al. Adipocyte ADAM17 plays a limited role in metabolic inflammation. Adipocyte. 2020 Dec;9(1):509-522.
97. Takahashi A, et al. The CCR4-NOT Deadenylase Complex Maintains Adipocyte Identity. Int J Mol Sci. 2019 Oct 24;20(21):5274.
98. Deconinck AE, et al. Utrophin-dystrophin-deficient mice as a model for Duchenne muscular dystrophy. Cell. 1997 Aug 22;90(4):717-27.

Supplemental Bibliography 3.2: Evidence of Key Driver Involvement in Wnt Signaling

1. Ganner A, et al. Regulation of ciliary polarity by the APC/C. Proc Natl Acad Sci U S A. 2009 Oct 20;106(42):17799-804.
2. Chen D, et al. ANTXR1, a stem cell-enriched functional biomarker, connects collagen signaling to cancer stem-like cells and metastasis in breast cancer. Cancer Res. 2013 Sep 15;73(18):5821-33.
3. Cheng B, et al. The role of anthrax toxin protein receptor 1 as a new mechanosensor molecule and its mechanotransduction in BMSCs under hydrostatic pressure. Sci Rep. 2019 Sep 2;9(1):12642.
4. Verma K, Gu J, Werner E. Tumor endothelial marker 8 amplifies canonical Wnt signaling in blood vessels. PLoS One. 2011;6(8):e22334.
5. Ding C, et al. Tumor Endothelial Marker 8 Promotes Proliferation and Metastasis *via* the Wnt/β-Catenin Signaling Pathway in Lung Adenocarcinoma. Front Oncol. 2021 Oct 14;11:712371.
6. Sheen VL, et al. Mutations in ARFGEF2 implicate vesicle trafficking in neural progenitor proliferation and migration in the human cerebral cortex. Nat Genet. 2004 Jan;36(1):69-76.
7. Li CC, et al. Enhancement of β-catenin activity by BIG1 plus BIG2 via Arf activation and cAMP signals. Proc Natl Acad Sci U S A. 2016 May 24;113(21):5946-51.
8. Lalli MA, et al. Haploinsufficiency of BAZ1B contributes to Williams syndrome through transcriptional dysregulation of neurodevelopmental pathways. Hum Mol Genet. 2016 Apr 1;25(7):1294-306.
9. Schepeler T, et al. Attenuation of the beta-catenin/TCF4 complex in colorectal cancer cells induces several growth-suppressive microRNAs that target cancer promoting genes. Oncogene. 2012 May 31;31(22):2750-60.
10. Jiang S, et al. KIAA1522 Promotes the Progression of Hepatocellular Carcinoma via the Activation of the Wnt/β-Catenin Signaling Pathway. Onco Targets Ther. 2020 Jun 16;13:5657-5668.
11. Silhankova M, et al. Wnt signalling requires MTM-6 and MTM-9 myotubularin lipid-phosphatase function in Wnt-producing cells. EMBO J. 2010 Dec 15;29(24):4094-105.
12. Chen H, et al. REGγ accelerates melanoma formation by regulating Wnt/β-catenin signalling pathway. Exp Dermatol. 2017 Nov;26(11):1118-1124.
13. Li L, et al. REGγ is critical for skin carcinogenesis by modulating the Wnt/β-catenin pathway. Nat Commun. 2015 Apr 24;6:6875.
14. Essien BE, et al. Transcription Factor ZBP-89 Drives a Feedforward Loop of β-Catenin Expression in Colorectal Cancer. Cancer Res. 2016 Dec 1;76(23):6877-6887.
15. Ocadiz-Ruiz R, et al. ZBP-89 function in colonic stem cells and during butyrate-induced senescence. Oncotarget. 2017 Oct 9;8(55):94330-94344.
16. Mirra S, et al. Function of Armcx3 and Armc10/SVH Genes in the Regulation of Progenitor Proliferation and Neural Differentiation in the Chicken Spinal Cord. Front Cell Neurosci. 2016 Mar 3;10:47.
17. Pan M, et al. BNIP-2 retards breast cancer cell migration by coupling microtubule-mediated GEF-H1 and RhoA activation. Sci Adv. 2020 Jul 31;6(31):eaaz1534.
18. Soh UJ, Low BC. BNIP2 extra long inhibits RhoA and cellular transformation by Lbc RhoGEF via its BCH domain. J Cell Sci. 2008 May 15;121(Pt 10):1739-49.
19. Wen J, et al. IL-8 promotes cell migration through regulating EMT by activating the Wnt/β-catenin pathway in ovarian cancer. J Cell Mol Med. 2020 Jan;24(2):1588-1598.
20. Ruffner H, et al. R-Spondin potentiates Wnt/β-catenin signaling through orphan receptors LGR4 and LGR5. PLoS One. 2012;7(7):e40976.
21. Lee S, et al. WNT Signaling Driven by R-spondin 1 and LGR6 in High-grade Serous Ovarian Cancer. Anticancer Res. 2020 Nov;40(11):6017-6028.
22. Gore AV, et al. Rspo1/Wnt signaling promotes angiogenesis via Vegfc/Vegfr3. Development. 2011 Nov;138(22):4875-86.
23. Geng A, et al. A novel function of R-spondin1 in regulating estrogen receptor expression independent of Wnt/β-catenin signaling. Elife. 2020 Aug 4;9:e56434.
24. Binnerts ME, et al. R-Spondin1 regulates Wnt signaling by inhibiting internalization of LRP6. Proc Natl Acad Sci U S A. 2007 Sep 11;104(37):14700-5.
25. Chen D, et al. TYRO3 facilitates cell growth and metastasis via activation of the Wnt/β-catenin signaling pathway in human gastric cancer cells. Aging (Albany NY). 2020 Feb 4;12(3):2261-2274.

26. Al Kafri N, Hafizi S. Galectin-3 Stimulates Tyro3 Receptor Tyrosine Kinase and Erk Signalling, Cell Survival and Migration in Human Cancer Cells. Biomolecules. 2020 Jul 11;10(7):1035.
27. Zhu YZ, et al. Inhibition of TYRO3/Akt signaling participates in hypoxic injury in hippocampal neurons. Neural Regen Res. 2016 May;11(5):752-7.

Supplemental Bibliography 3.3: Evidence of Key Driver Involvement in Mitochondrial Activity

1. Feng H, et al. CTRP3 promotes energy production by inducing mitochondrial ROS and up-expression of PGC-1α in vascular smooth muscle cells. Exp Cell Res. 2016 Feb 15;341(2):177-86.
2. Gao J, Qian T, Wang W. CTRP3 Activates the AMPK/SIRT1-PGC-1α Pathway to Protect Mitochondrial Biogenesis and Functions in Cerebral Ischemic Stroke. Neurochem Res. 2020 Dec;45(12):3045-3058.
3. Zhang CL, et al. Globular CTRP3 promotes mitochondrial biogenesis in cardiomyocytes through AMPK/PGC-1α pathway. Biochim Biophys Acta Gen Subj. 2017 Jan;1861(1 Pt A):3085-3094.
4. Xie Y, et al. The proteasome activator REGγ accelerates cardiac hypertrophy by declining PP2Acα-SOD2 pathway. Cell Death Differ. 2020 Oct;27(10):2952-2972.
5. Moncsek A, et al. Evidence for anti-apoptotic roles of proteasome activator 28γ via inhibiting caspase activity. Apoptosis. 2015 Sep;20(9):1211-28.
6. Gustafsson ÅB, Dorn GW 2nd. Evolving and Expanding the Roles of Mitophagy as a Homeostatic and Pathogenic Process. Physiol Rev. 2019 Jan 1;99(1):853-892.
7. Metzger MB, et al. A protein quality control pathway at the mitochondrial outer membrane. Elife. 2020 Mar 2;9:e51065.
8. Emfinger CH, et al. β Cell-specific deletion of Zfp148 improves nutrient-stimulated β cell Ca2+ responses. JCI Insight. 2022 May 23;7(10):e154198.
9. Hartmann B, et al. Homozygous YME1L1 mutation causes mitochondriopathy with optic atrophy and mitochondrial network fragmentation. Elife. 2016 Aug 6;5:e16078.
10. Ruan Y, et al. Loss of Yme1L perturbates mitochondrial dynamics. Cell Death Dis. 2013 Oct 31;4(10):e896.
11. MacVicar T, et al. Lipid signalling drives proteolytic rewiring of mitochondria by YME1L. Nature. 2019 Nov;575(7782):361-365.
12. Gowans GJ, et al. INO80 Chromatin Remodeling Coordinates Metabolic Homeostasis with Cell Division. Cell Rep. 2018 Jan 16;22(3):611-623.
13. Diquigiovanni C, et al. A novel mutation in *SPART* gene causes a severe neurodevelopmental delay due to mitochondrial dysfunction with complex I impairments and altered pyruvate metabolism. FASEB J. 2019 Oct;33(10):11284-11302.
14. Seo BA, et al. TRIP12 ubiquitination of glucocerebrosidase contributes to neurodegeneration in Parkinson's disease. Neuron. 2021 Dec 1;109(23):3758-3774.e11.
15. Maruyama H, et al. Medullary thick ascending limb impairment in the Gla^{tm}Tg(CAG-A4GALT) Fabry model mice. FASEB J. 2018 Aug;32(8):4544-4559.
16. Serrat R, et al. The non-canonical Wnt/PKC pathway regulates mitochondrial dynamics through degradation of the arm-like domain-containing protein Alex3. PLoS One. 2013 Jul 2;8(7):e67773.
17. Mou Z, Tapper AR, Gardner PD. The armadillo repeat-containing protein, ARMCX3, physically and functionally interacts with the developmental regulatory factor Sox10. J Biol Chem. 2009 May 15;284(20):13629-13640.

18. Howells CC, et al. The Bcl-2-associated death promoter (BAD) lowers the threshold at which the Bcl-2-interacting domain death agonist (BID) triggers mitochondria disintegration. J Theor Biol. 2011 Feb 21;271(1):114-23.
19. Liu XM, et al. Mitochondrial Function Regulated by Mitoguardin-1/2 Is Crucial for Ovarian Endocrine Functions and Ovulation. Endocrinology. 2017 Nov 1;158(11):3988-3999.
20. Liu XM, et al. Mitoguardin-1 and -2 promote maturation and the developmental potential of mouse oocytes by maintaining mitochondrial dynamics and functions. Oncotarget. 2016 Jan 12;7(2):1155-67.
21. Xu L, et al. Miga-mediated endoplasmic reticulum-mitochondria contact sites regulate neuronal homeostasis. Elife. 2020 Jul 10;9:e56584.
22. Deng L, et al. NMT1 inhibition modulates breast cancer progression through stress-triggered JNK pathway. Cell Death Dis. 2018 Nov 16;9(12):1143.

Chapter 4: Combined effects of genetic background and diet on mouse metabolism and gene expression

Adapted from a manuscript in preparation by:

Jordan N. Reed[1,2], Faten Hasan[3], Abhishek Karkar[4], Dhanush Banka[1,2], Jameson Hinkle[2], Preeti Shastri[4], Navya Srivastava[5], Sarah E. Newkirk[4], **Heather A. Ferris**[4,6]**, Bijoy K. Kundu**[5,6]**, Sibylle Kranz**[3,6]**, Mete Civelek**[1,2,6,*]**, Susanna R. Keller**[4,6,7,*]

[1]Department of Biomedical Engineering, University of Virginia School of Medicine, Charlottesville VA 22903, USA

[2]Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville VA 22903, USA

[3]Department of Kinesiology, University of Virginia School of Education and Human Development, Charlottesville VA 22903, USA

[4]Department of Medicine-Division of Endocrinology and Metabolism, University of Virginia School of Medicine, Charlottesville VA 22903, USA

[5]Department of Radiology and Medical Imaging, University of Virginia School of Medicine, Charlottesville VA 22903, USA

[6]Senior Authors

[7]Lead contact

*Correspondence: srk4b@virginia.edu; mete@virginia.edu

4.1 Summary

In human subjects, different diet patterns affect weight and metabolism differentially. Wide variations in outcomes between individuals on the same diet suggest diet effects are modified by genetic background. To start uncovering the genetic basis for differential effects of diets on weight gain/loss and other metabolic parameters, we subjected different mouse strains (C57BL/6J, A/J, DBA/2J and SJL) to humanized diets (American, Mediterranean, Vegetarian, and Vegan), measured various metabolic parameters, and performed RNA-seq on muscle and fat tissues. We observed pronounced diet- and strain-dependent effects on body weight (despite similar caloric intake) and on triglyceride and insulin levels. Glucose uptake into skeletal muscle and brown adipose tissue and total body fat showed predominantly strain-dependent differences. RNA-seq analysis revealed strain-dependent differences in gene expression in most tissues. Conversely, ~400 genes in visceral adipose tissue differentially responded to diet and strain. These genes are in metabolite transport and lipid metabolism pathways and thus affect metabolic parameters. The results suggest differential impact of genetic backgrounds on metabolism with nutrient sources in diet modifying effects.

4.2 Introduction

4.2.1 Diet modification used to prevent and treat cardio-metabolic disease

In the United States, 73% of the adult population is obese (42%) or overweight (31%)[1]. Diet modification is considered key to weight management and prevention of the complications of excess weight, prominently type 2 diabetes and cardiovascular disease[2,3].

The most common dietary patterns investigated for effects on weight loss and cardio-metabolic disease are Mediterranean style[4,5], vegetarian and vegan[6–9], low and very low fat[10], low and very low carbohydrates[10–14], Dietary Approaches to Stop Hypertension (DASH)[15], and Paleo diets[16]. Mediterranean diet is typically high in polyphenols and omega-3s, and plant-based diets are high in fiber and anti-oxidant nutrients; both are low in saturated fat and red meats. Because of the different nutrient composition, dietary patterns of Mediterranean, DASH, and plant-based diets improve weight loss, insulin sensitivity, and circulating metabolite levels in the general population[4,17]. These diets were associated with favorable cardio-metabolic disease outcomes[18–22] and are prescribed by clinicians to prevent/treat diabetes and cardiovascular disease[2,3].

However, diets caused variable degrees of improvement between individuals; for low fat and low carbohydrate-containing diets, study subjects within diet groups with similar caloric intake exhibited weight changes ranging from 25 kg weight loss to 5 kg weight gain[10]. Iso-caloric Mediterranean and Chinese plant-based diets in pre-diabetic patients also caused high variability in weight changes between individuals[23].

4.2.2 Diet effects on cardio-metabolic disease traits are modified by genetic variation

These suggested that genetic background may modify an individuals' responses to diets; indeed, others find that diet significantly modifies the genetic risk imparted by disease-associated single nucleotide variants (SNPs)[24]. The effects of Mediterranean diet were modified SNPs in transcription factor 7-like 2, *TCF7L2,* on body composition[25–28], in apo-lipoprotein E, *APOE,* on triglycerides[29], and in cholesteryl ester transfer protein, *CETP,* on cholesterol[30,31]. A combination of Mediterranean-DASH modified the effects of a SNP in caveolin 1, *CAV1,* on lipids[32,33]. A SNP in carnitine palmitoyltransferase 1A, *CPT1A,* modified the effect of Inuit diet on fatty acids[34], and a SNP in glucokinase regulatory protein, *GCKR,* modified the effect of high-fish diets on circulating triglycerides[35,36].

Genotype can also modify the relationship between the macronutrient content and metabolic outcomes. SNPs in fat mass and obesity associated, *FTO*[37], neuropeptide Y, *NPY*[38], Perilipin, *PLIN*[39–41], fibroblast growth factor 21, *FGF21*[42], adiponectin, *ADIPOQ*[43], Melatonin Receptor 1B, *MTNR1B*[44], fatty acid desaturase, *FADS2*[45–47], hepatocyte nuclear factor 1α, *HNF1A*[48], 3-hydroxy-3-methylglutaryl-coenzyme A reductase gene, *HMGCR*[49], and apo-lipoprotein A2, *APOA2*[50], modified the relationship between dietary macronutrients and anthropometric traits or circulating metabolites. Finally, SNPs in insulin receptor substrate 1, *IRS1*[51], uncoupling protein 2 *UCP2*[52], and beta-2 adrenergic receptor *ADRB2*[53], modified the effect of overall energy intake or caloric content on various metabolic parameters.

Analysis of the UK Biobank identified common variation near 4 genes (Solute carrier family 12 member 3, *SLC12A3,* ATP Binding Cassette Subfamily A Member 6, *ABCA6,* MLX Interacting Protein Like, *MLXIPL,* and the gap junction *GJB6-GJB2-GJA3* cluster) that modified the effects of dietary fish oil on cholesterol or triglycerides[54]. Many SNPs also interact with diet to influence cardiovascular disease outcomes[55,56].

4.2.3 Genetically diverse mouse models are used to study diverse responses to diet

While many gene-diet interactions have been identified, studies in humans are limited because it can be hard to include enough human subjects, get accurate self-reporting of or adherence to diet, and control the environment to tease out subtle genotype-diet interaction effects[57]. The smaller studies, referenced above, have primarily identified SNPs that edit the coding region of the genes and likely have large effect size; few have enough power to detect interactions between diet and common regulatory variants[54].

Mouse populations with genetic diversity equivalent to human genetic diversity provide a renewable resource for controlled studies, the results of which can be tested in humans at limited scale[58]. Mouse models allow deeper tissue and cellular phenotyping, and the controlled environment reduces noise and ensures adherence to diet. Studies using genetically diverse mice show that, when challenged with high fat or high carbohydrate diet, individual genetic background causes diverse responses in body weight gain[58–60], atherosclerosis development[58,61], insulin resistance[58,62,63], circulating metabolites[60,63], glucose uptake[63], adipocyte size[63], and protein abundances[63]. In populations of ~100 mouse strains, genetic loci that favor body fat gain on a high-fat high-sucrose diet and loci associated with atherosclerosis overlapped with the loci identified in humans[58,61,64]. Further, many lipid metabolism genes and pathways in adipose tissue are conserved between human and mouse[65]. In all, the effect of macronutrient content in diets on metabolic parameters is significantly modified by genetic variation in mice, and the genetic and cellular mechanisms are likely conserved in humans. However, most studies were performed under high fat diet conditions, and do not test the more subtle effects of nutrient source.

Recently, Barrington et al. used humanized mouse diets when studying effects of diets on metabolic parameters and DNA methylation, thereby further expanding the possibilities for mouse studies relevant to human diet patterns[66]. They used mouse diets representative of human American, Mediterranean, Ketogenic, and Japanese diets in four genetically diverse strains of mice. They showed that body fat, cholesterol, glucose tolerance, and liver triglycerides were all affected by significant interactions of between diet and the genetic background of the mice. Though iso-caloric, the diets were formulated to represent human diet composition, and were therefore not similar in macronutrient composition, which could drive differences.

The goal of this study was to investigate metabolic effects of different diets and evaluate how genetic background modifies their effects. We used common human diets (American, Mediterranean, vegetarian, and vegan) and studied their effects on weight and other metabolic parameters as well as gene transcription in key metabolic tissues in four mouse strains (A/J, C57BL/6J, DBA/2J and SJL). We observed clear strain-dependent differences in metabolic parameters and distinct strain-dependent effects of different diets suggesting an interplay between diets and genetic background.

4.3 Results

4.3.1 Design of mouse diets that reflect clinically relevant human nutrient sources

To evaluate how the metabolic responses to clinically relevant diets are mediated by genetic background, we selected four genetically diverse strains of mice, A/J, C57BL/6J, DBA/2J, and SJL[67]. These four strains showed widely varying weight gains and insulin resistance in response to high fat diets[58,62]. SJL is resistant to weight gain, C57BL/6J and A/J mice have intermediate weight gains, while DBA/2J has the most extreme weight gain. C57BL/6J and A/J are insulin sensitive, while DBA/2J mice are insulin resistant. We hypothesized that the metabolic responses of these mice could be modified by the nutrient sources in the diet they consume.

The chosen diets (Mediterranean, Vegetarian, and Vegan) are the most prevalent human diet patterns consumed with beneficial effects, compared to an American diet pattern most associated with negative health effects. To avoid variability in responses due to variations in macronutrient contents in diets, we equalized the amount of fat, carbohydrate and protein to the average consumption of these nutrients in the American population,[68] 35% fat, 50% carbohydrate, and 15% protein (Supplemental Table 4.1). Importantly, the macronutrients were derived from different sources. Human versions of each diet were generated, then translated into mouse chow by sourcing the nutrients from the same foods as the human diet (Figure 4.1). The American diet contained the highest amounts of nutrients sourced from animal products (beef, milk, etc.) and had with high amounts of saturated fatty acids, animal proteins, and refined carbohydrates. The Mediterranean diet was made from plant and animal products (fish, vegetables, and olive oil, etc), was high in fiber, unsaturated fatty acids, and cholesterol. Vegetarian diet was made from plants and animal products (wheat, milk, eggs, etc) and was high in fiber, unsaturated fatty acids, and low in cholesterol. Vegan diet was made from only plants products (grain, vegetables, etc) and was high in fiber, unsaturated fatty acids, and lowest in cholesterol.

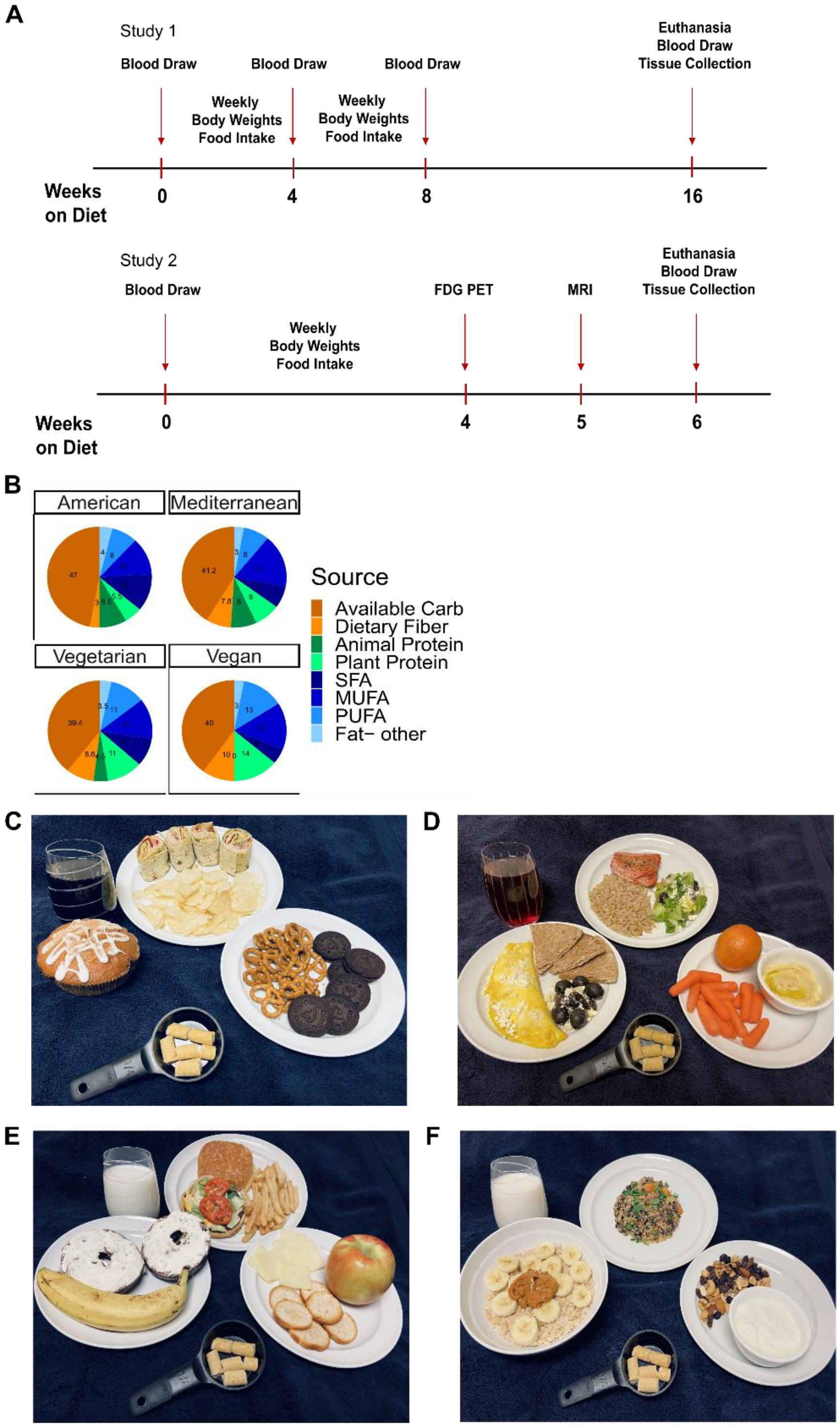

Figure 4.1: Overview of experimental procedures and design of diets

- *A. Experimental design of study 1 and 2*
- *B. Major macronutrient sources for each of the 4 constructed diets*
- *C. Representative photo of the American diet*

D. Representative photo of the Mediterranean diet
E. Representative photo of the Vegetarian diet
F. Representative photo of the Vegan diet
MUFA = Mono-unsaturated fatty acid, PUFA = Poly-unsaturated fatty acid, SFA = Saturated fatty acid

4.3.2 Genetic background and diet interaction effects body weight gain and metabolic parameters

An initial study was carried out using male mice, who fed ad libitum on one of the humanized diets or mouse chow diet, starting at 14-15 weeks of age. Body weights and food intake were determined weekly, while blood samples for measuring glucose, insulin, triglycerides, and non-esterified fatty acids were taken before starting the diets and at 4, 8, and 16 weeks on the diet (Figure 4.1).

Over 16 weeks, C57BL/6J and DBA/2J mice showed the largest weight gains (Figure 4.2 A). This effect was modified by the diet, where mice fed American or Mediterranean diet continued to gain weight over 16 weeks, while mice fed Vegetarian or Vegan diets reached a maximum weight around week 4 (Figure 4.3 A). A/J and SJL mice did not gain much weight over the 16-week period, and this effect was unchanged by the diet consumed.

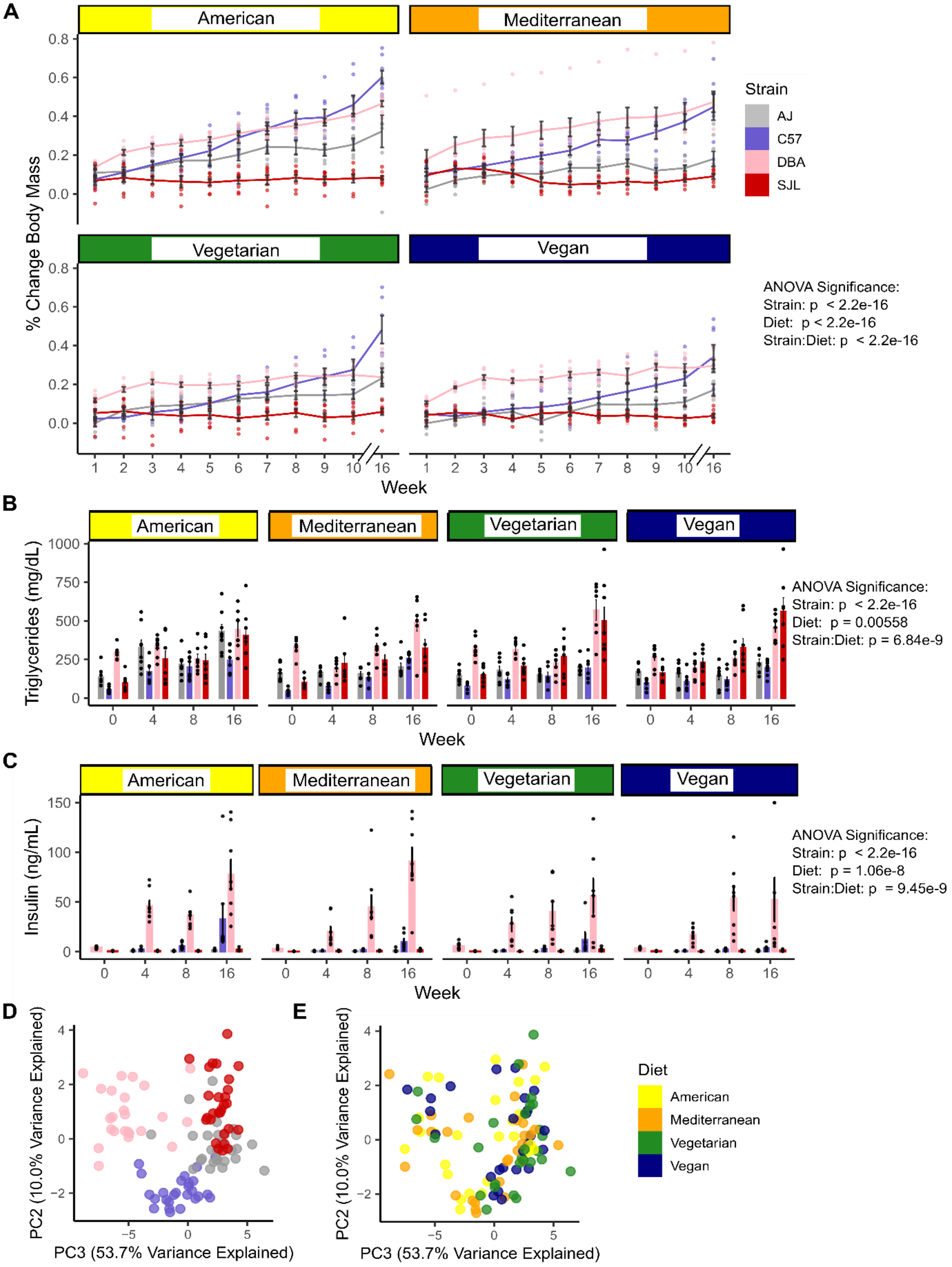

Figure 4.2: Diet and genetic background impact body weight and metabolic parameters in male mice

A. *Body weight in grams after 16 weeks on diet. (n = 8, 14 weeks old at week 0, male mice)*
B. *Circulating levels of triglycerides in blood after 16 weeks on diet. (n = 8, 14 weeks old at week 0, male mice)*
C. *Circulating levels of insulin in blood after 16 weeks on diet. (n = 8, 14 weeks old at week 0, male mice)*
D. *Principal components analysis of all traits measured in study 1 using all samples, PC 1 and 2 shown with samples colored by genetic background*
E. *Principal components analysis of all traits measured in study 1 using all samples, PC 1 and 2 shown with samples colored by diet*

Error bars represent mean ± SEM, 3-way ANOVAs performed considering the effects of Strain, Diet, Time, and all interaction effects.

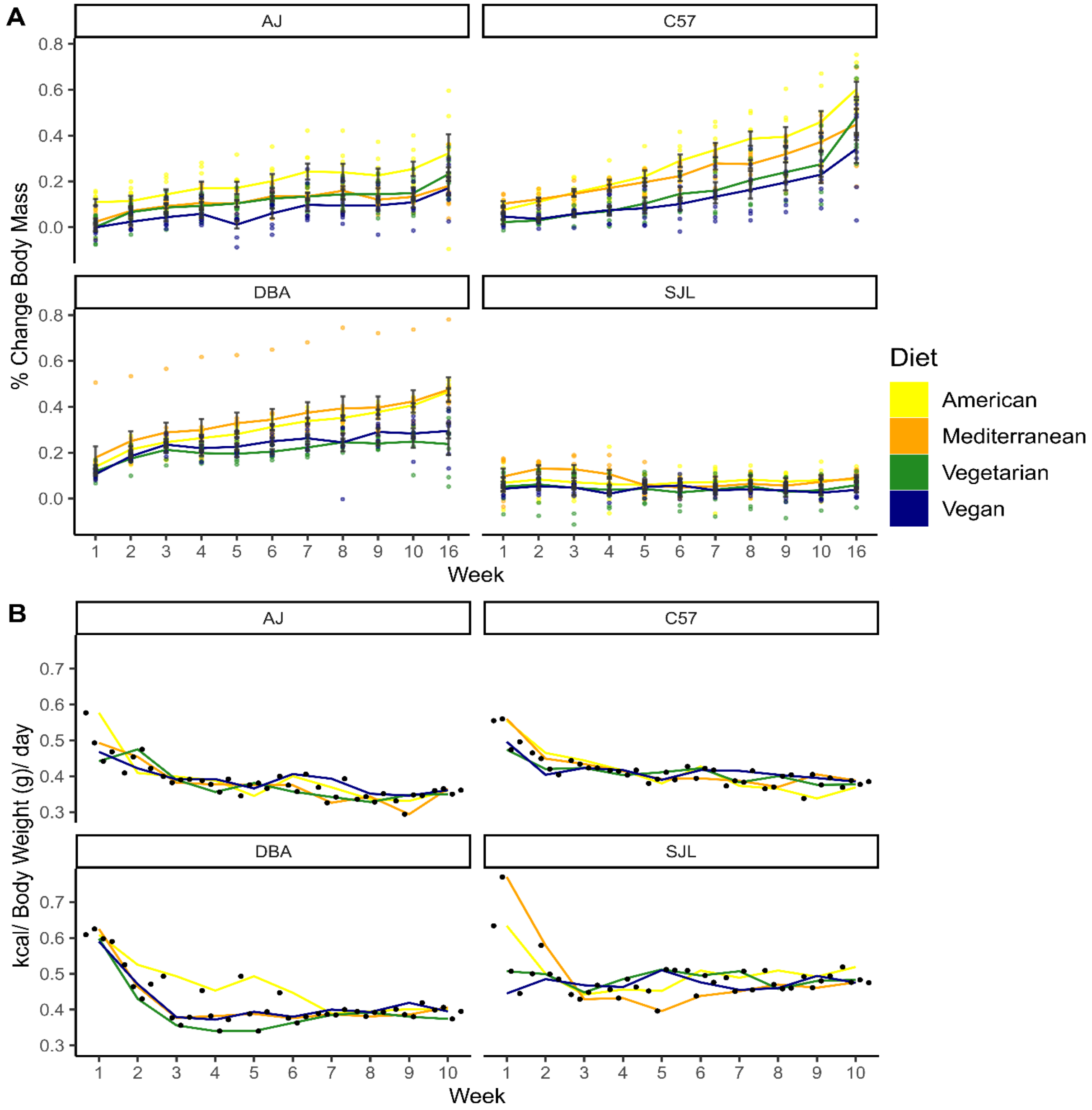

Figure 4.3: Diet and genetic background impact body weight in male mice

A. *Body weight (g) percent change after 16 weeks on diet. (n = 8, 14 weeks old at week 0, male mice)*

B. Food intake in kcal per gram body weight per day over 10 weeks on diet. (n = 8, 14 weeks old at week 0, male mice, average per cage)
Error bars represent mean ± SEM

Since we were interested in studying the unique effects a diet can cause on different genetic backgrounds, we used the Analysis of Variance (ANOVA) tests to determine the independent effects of strain, diet, and the interaction effects of strain and diet (Strain:Diet). Interactions occur when the effects of one variable depends on state of the other. The Strain:Diet interaction term is significant if there the effect of one or more diets significantly differ between one or more strains, or vice versa. Using a 3-way ANOVA account for the effect of time, we determined that there is a significant effect of strain, diet, and Strain:Diet interactions on body weight (Table 4.1). DBA/2J and C57BL/6J mice gain the most weight and American and Mediterranean diets promote weight gain most. The effect of diet was blunted in A/J mice and was not observed in SJL mice. Post-hoc tests showed many significant differences between groups, which were not plotted due to space.

Table 4.1: P-values derived from three-way ANOVA interaction test in cohort 1 phenotypic data

Term	**Strain**	**Diet**	**Week**	**Strain:Diet**	**Strain:Week**	**Diet:Week**	**Strain:Diet:Week**
Body Weight	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	2.68E-16
NEFA	6.76E-15	4.36E-05	1.12E-14	1.35E-05	2.56E-13	4.95E-09	3.90E-08
Glucose	< 2.2e-16	7.53E-07	9.50E-13	0.001803	3.19E-14	0.114361	0.001352
Insulin	< 2.2e-16	1.06E-08	< 2.2e-16	9.45E-09	< 2.2e-16	1.44E-05	0.0001823
Triglycerides	< 2.2e-16	0.005577	< 2.2e-16	6.84E-09	3.54E-15	0.009593	0.003218

We observed no large changes in caloric intake between our diet groups (Figure 4.3). We observed small differences between strains, with SJL mice eating more and DBA/2J mice eating less than average. Since these effects are opposite in direction to the observed body weight change, we conclude that the differences in food intake are unlikely the cause of differences in weight gain between strains.

To understand if altered metabolism accompanied the altered body weights observed in these mice, we measured blood insulin, triglycerides, glucose, and non-esterified fatty acids (NEFA) after challenge with the “humanized” diets. DBA/2J mice had elevated triglycerides (Figure 4.2 B) at baseline and on the diets, while SJL mice have elevated triglycerides only after challenge with the new diet. The Vegan diet causes the largest increase in triglycerides in SJL mice, while the American diet causes the largest increase in triglycerides in C57BL/6J mice (Figure 4.4). We also observed that DBA/2J mice have elevated insulin (Figure 4.2 C) at baseline and on the “humanized” diets, and A/J mice have consistently low insulin levels. C57BL/6J and SJL insulin levels are responsive to diet, with Vegan being the most protective (Figure 4.4). Using the 3-way ANOVA, we determined that there were significant effects due to Strain:Diet interactions for both triglycerides and insulin (Table 4.1). Blood glucose and NEFA measurements also varied significantly by strain, diet, and the interaction effects of Strain:Diet (Figure 4.5, Table 4.1).

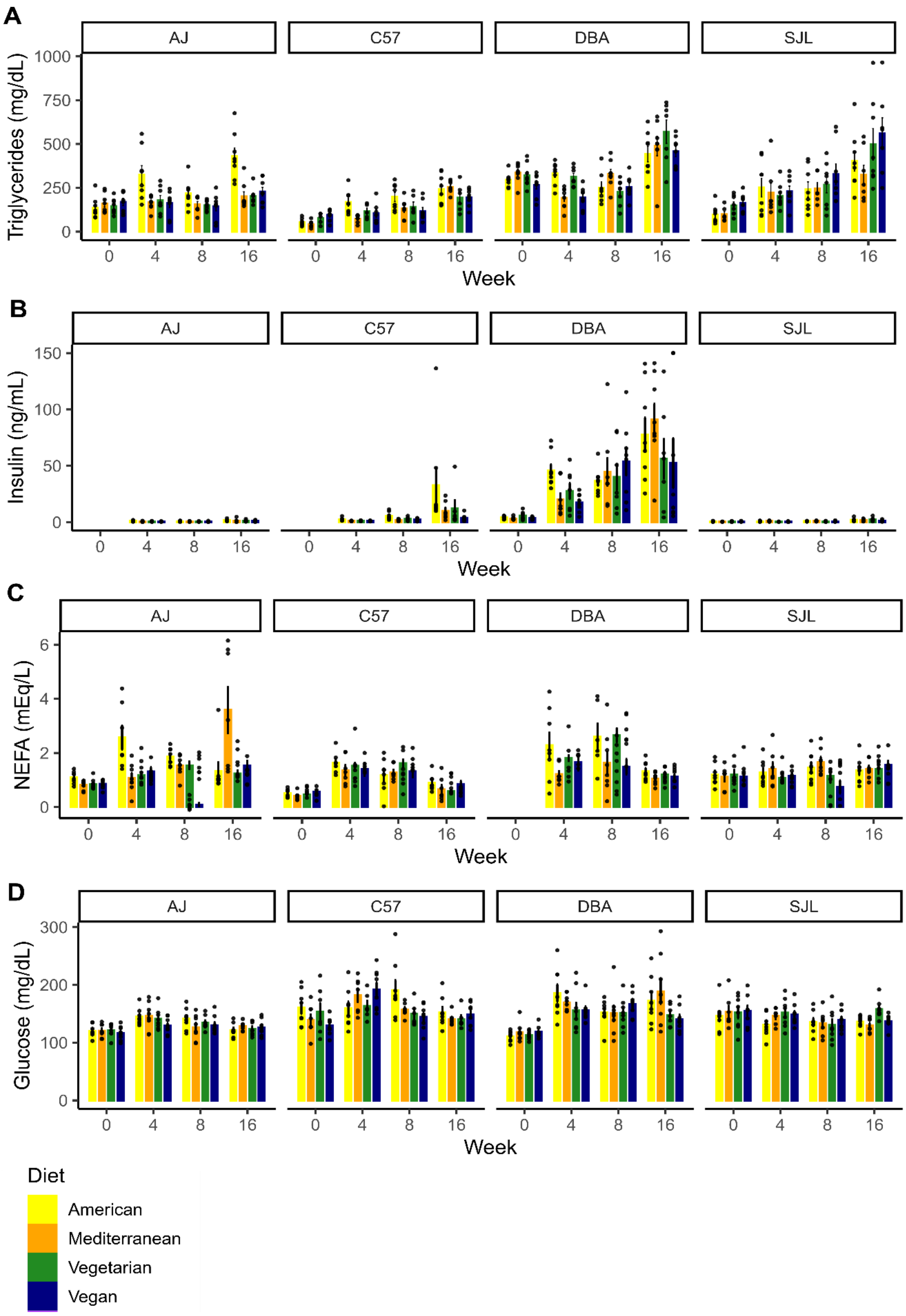

Figure 4.4: Diet and genetic background impact metabolic parameters in male mice

A. Triglycerides after 16 weeks on diet. (n = 8, 14 weeks old at week 0, male mice)

B. Insulin over 16 weeks on diet. (n = 8, 14 weeks old at week 0, male mice)
C. Blood glucose after 16 weeks on diet. (n = 8, 14 weeks old at week 0, male mice)
D. Non-esterified fatty acids (NEFA) over 16 weeks on diet. (n = 8, 14 weeks old at week 0, male mice)
Error bars represent mean ± SEM

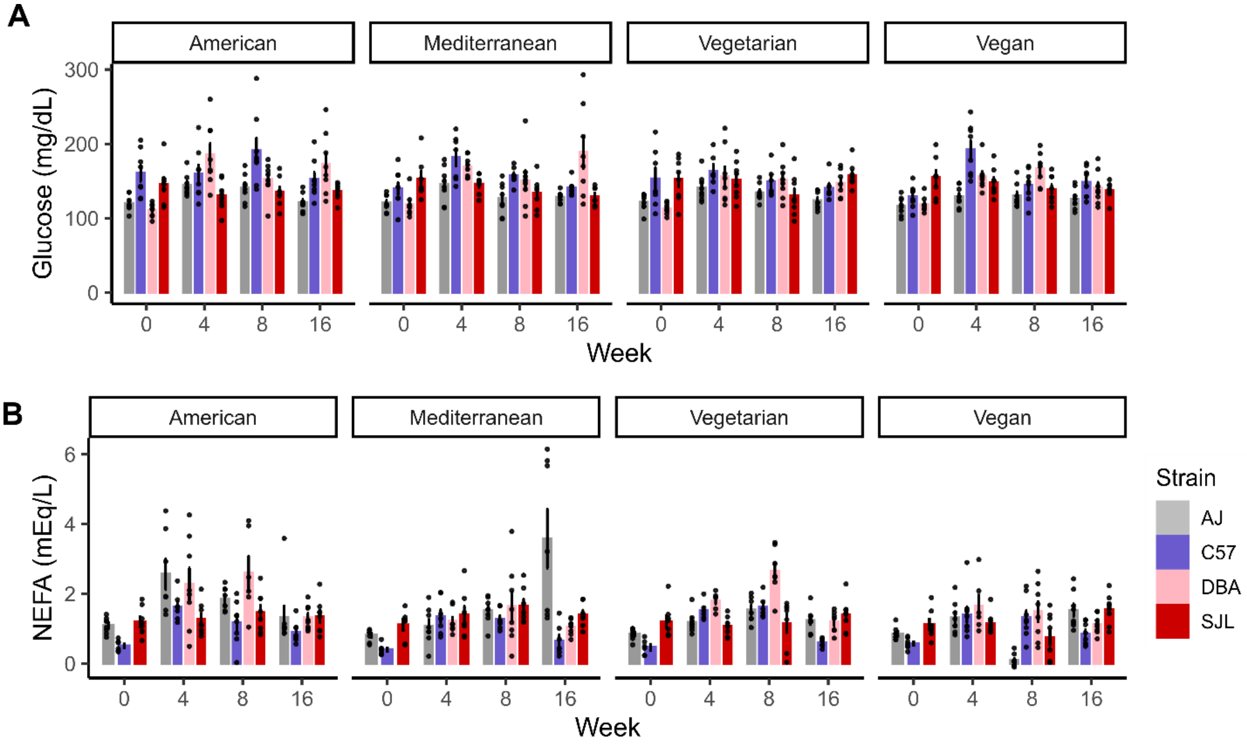

Figure 4.5: Genetic background impacts glucose and fatty acids in male mice
A. Blood glucose after 16 weeks on diet. (n = 8, 12 weeks old at week 0, male mice)
B. Non-esterified fatty acids (NEFA) over 16 weeks on diet. (n = 8, 12 weeks old at week 0, male mice)
Error bars represent mean ± SEM

We used principal components analysis (PCA) to explore our multi-dimensional metabolic phenotype data in two dimensions on calculated axes of highest variability. The axis of highest variability through the data, principal component one, separates DBA/2J samples and C57BL/6J samples into clear groups, while A/J and SJL are less discriminated on components one and two (Figure 4.2 D, E). Neither axis groups the samples by diet, meaning the strain can explain more of the variability in the metabolic phenotypes than the diet.

4.3.3 Deeper phenotyping of body fat and glucose uptake reveal variation due to genetic background

Using the information obtained from the first study, we placed a second cohort of mice, 16 each of A/J, C57BL/6J, DBA/2J, and SJL, on the four diets at 12 weeks of age. We took weekly body weight and food intake measurements and blood samples at the start of the diet and at euthanasia after 6 weeks on the diet. Glucose uptake was measured at 4 weeks of age and total body fat was imaged at 5 weeks of age. At euthanasia, blood was collected for analysis of metabolic parameters and tissues for RNA-seq analysis (Figure 4.1).

Similar to the first study, we saw that the DBA/2J and C57BL/6J mice gained weight rapidly, while the A/J and SJL mice were more protected (Figure 4.6). We also observed again that DBA/2J and C57BL/6J were most responsive to diet, with American causing more weight gain and Vegan causing less in these mice. These effects were less pronounced than the first study, but using 3-way ANOVA, we show that the effects of strain, diet, and the interaction effects of Strain:Diet were all significant (Table 4.2). We also show that the blood insulin and triglycerides, at baseline and after 6 weeks on the diets, are similar to the first study (Figure 4.6). Blood insulin, triglycerides, and glucose are all significantly affected by strain and the Strain:Diet interactions, though diet only had a significant main effect on body weight (Table 4.2).

After 5 weeks on the “humanized” diet, we used magnetic resonance imaging (MRI) to measure the total body fat (Figure 4.7 A). We observed that C57BL/6J and DBA/2J had the highest body fat, and in C57BL/6J, Vegan diet was more protective than the other three diets. We used a 2-way ANOVA test to determine the effects of strain, diet and Strain:Diet interaction effects. We found that the effects of Strain:Diet were suggestive but not significant; only strain-dependent effects were significant (Table 4.2).

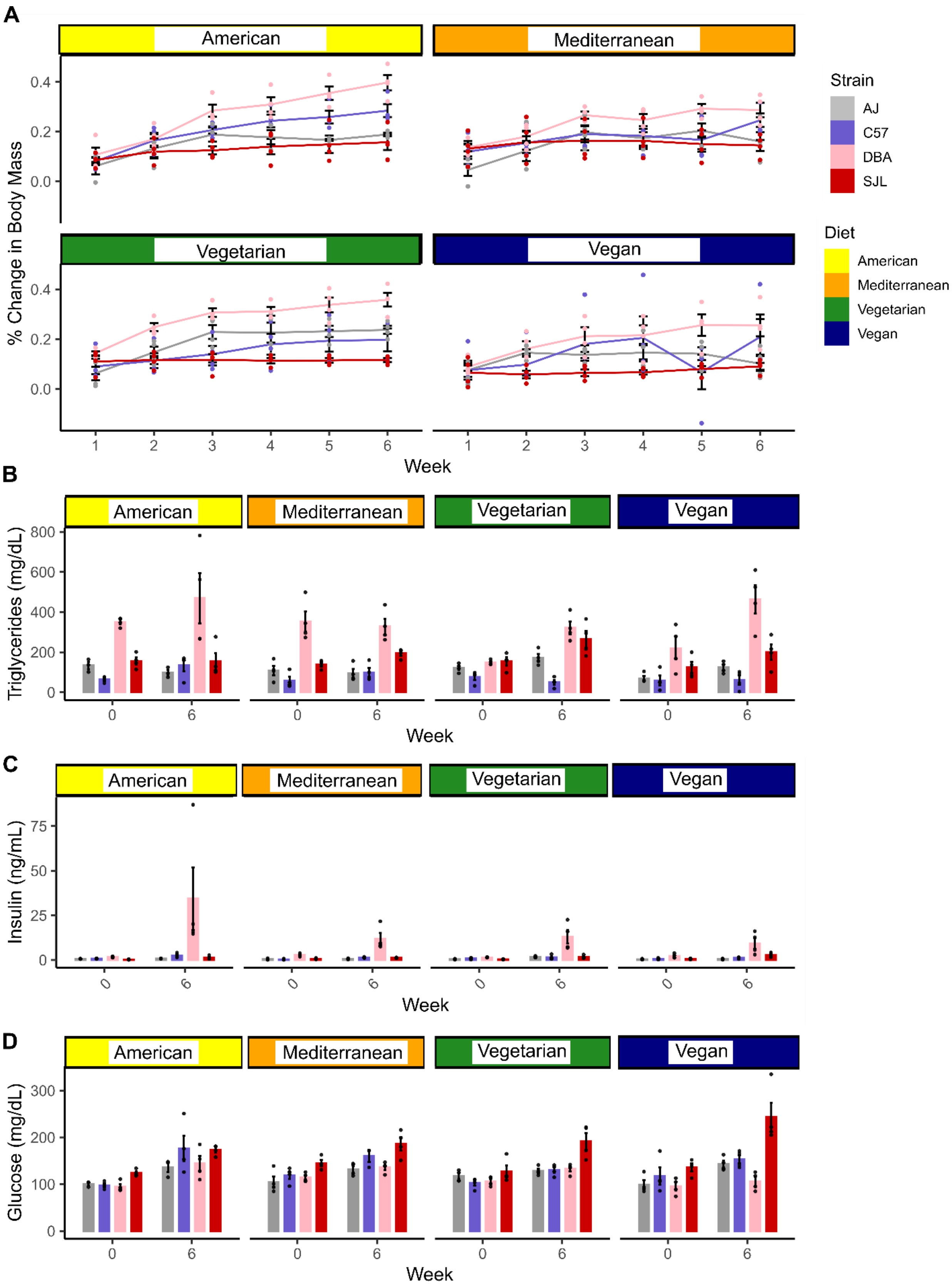
A
American
Mediterranean
Vegetarian
Vegan
% Change in Body Mass
0.4
0.2
0.0
1
2
3
4
5
6
Week
Strain
AJ
C57
DBA
SJL
Diet
American
Mediterranean
Vegetarian
Vegan
B
Triglycerides (mg/dL)
800
600
400
200
0
0
6
Week
C
Insulin (ng/mL)
75
50
25
0
0
6
Week
D
Glucose (mg/dL)
300
200
100
0
0
6
Week

Figure 4.6: Diet and genetic background impact metabolic parameters in male mice

A. *Body weight in grams after 6 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)*
B. *Triglycerides after 6 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)*
C. *Insulin over 6 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)*
D. *Blood glucose after 6 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)*
Error bars represent mean ± SEM

Table 4.2: P-values derived from two- and three-way ANOVA interaction tests of cohort 2 phenotypic data

Term	Strain	Diet	Strain:Diet	Week	Strain:Week	Diet:Week	Strain:Diet:Week
Body Weight	< 2.2e-16	1.26E-07	4.94E-06	< 2.2e-16	0.0010	0.9399	1.0000
Glucose	1.06E-13	0.4692	0.0173	< 2.2e-16	0.0039	0.1525	0.0996
Insulin	< 2.2e-16	0.4710	0.0014	< 2.2e-16	0.0004	0.1402	0.0639
Triglycerides	< 2.2e-16	0.2650	0.0042	2.50E-05	0.0104	0.1225	0.0902
MRI-Total Fat	1.42E-09	0.9533	0.0611				
FDG-PET-VAT	1.07E-13	0.1583	0.9668				
FDG-PET-BAT	0.0070	0.5012	0.1759				
FDG-PET-Quad	0.0145	0.4809	0.1481				
VAT Weight / Body Weight	2.90E-11	0.1161	0.3730				
SAT Weight / Body Weight	7.90E-08	0.3660	0.6787				
BAT Weight / Body Weight	1.82E-14	0.1177	0.8495				
Quadriceps Weight / Body Weight	9.39E-05	0.4665	0.4320				
Gastrocnemius Weight / Body Weight	4.39E-08	0.6367	0.2555				
Pancreas Weight / Body Weight	2.05E-07	0.6602	0.6538				
Brain Weight / Body Weight	0.0385	0.8781	0.5050				
Liver Weight / Body Weight	0.0303	0.8587	0.4358				

At 4 weeks on the “humanized” diet, we used positron emission tomography (PET) imaging to track radiolabeled glucose uptake (FGD-PET) in visceral adipose tissue (VAT), brown adipose tissue (BAT), and quadriceps muscle. We found that SJL mice uptake the most glucose in VAT, while SJL Vegan mice take up the most in skeletal muscle (Figure 4.7 B,C). DBA/2J and A/J uptake the most glucose in BAT (Figure 4.8). Using the 2-way ANOVA test, while skeletal muscle was suggestive, none had a significant Strain:Diet effect; only strain-dependent effects were significant (Table 4.2).

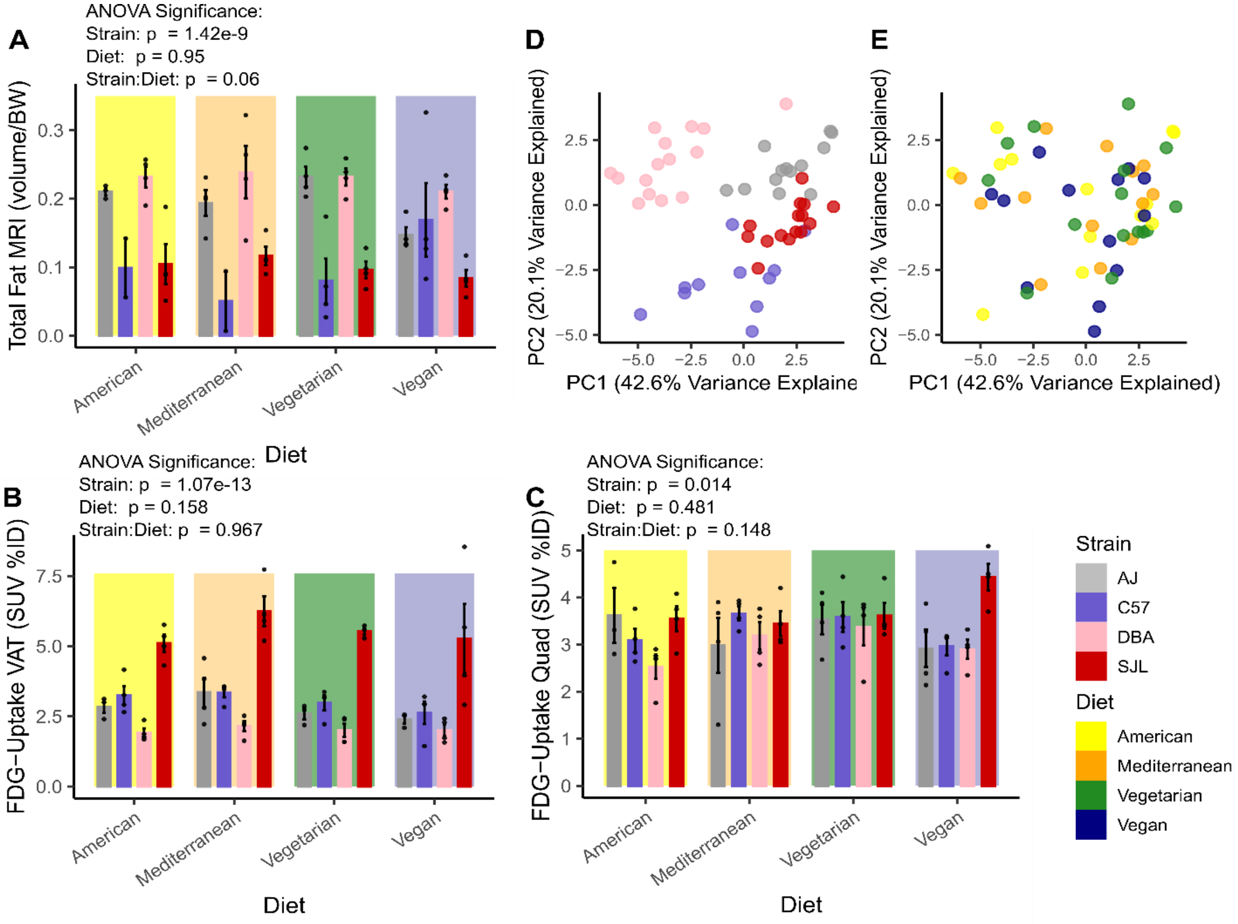

Figure 4.7: Genetic background impacts glucose-related metabolic parameters in male mice

- *A. Total body fat measured via MRI after 5 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)*
- *B. Glucose uptake in visceral adipose tissue measured via FDG-PET imaging at 4 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)*
- *C. Glucose uptake in quadriceps skeletal muscle measured via FDG-PET imaging at 4 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)*
- *D. Principal components analysis of all traits measured in study 2 using all samples, PC 1 and 2 shown with samples colored by genetic background*
- *E. Principal components analysis of all traits measured in study 2 using all samples, PC 1 and 2 shown with samples colored by diet*
  Error bars represent mean ± SEM

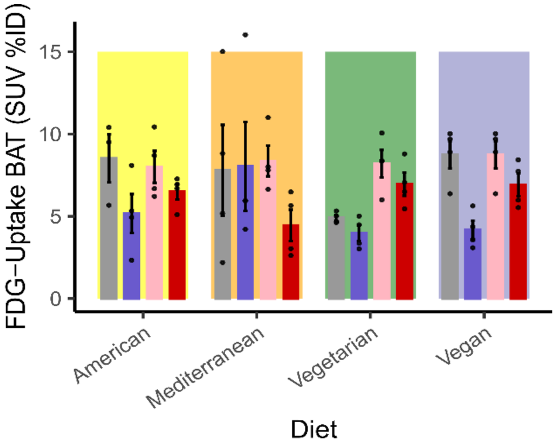

Figure 4.8: Glucose uptake in brown adipose tissue, as measured via FDG-PET imaging after 4 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)

After 6 weeks the mice were sacrificed, and selected metabolic organs were harvested and weighed. DBA/2J mice had the heaviest fat pads (Figure 4.9); subcutaneous adipose tissue (SAT), VAT, and BAT were consistently largest in DBA/2J and smallest in SJL. C57BL/6J mice had the largest quadriceps and gastrocnemius muscles, and the largest brain weights. Liver weights were consistent among diets and strains, while pancreas weights were highest in DBA/2J and C57BL/6J. Using the 2-way ANOVA test, no organ weights significantly differed due to Strain:Diet interactions; only strain-dependent effects were significant (Table 4.2).

Using the basic and extended phenotype data from the second cohort, we performed PCA and plotted the data on principal components 1 and 2, the axes of highest variability. We show again that this axis is able to separate the samples based on strain, but not diet (Figure 4.7 D,E), showing that strain is able to explain more of the variability seen in the measured traits.

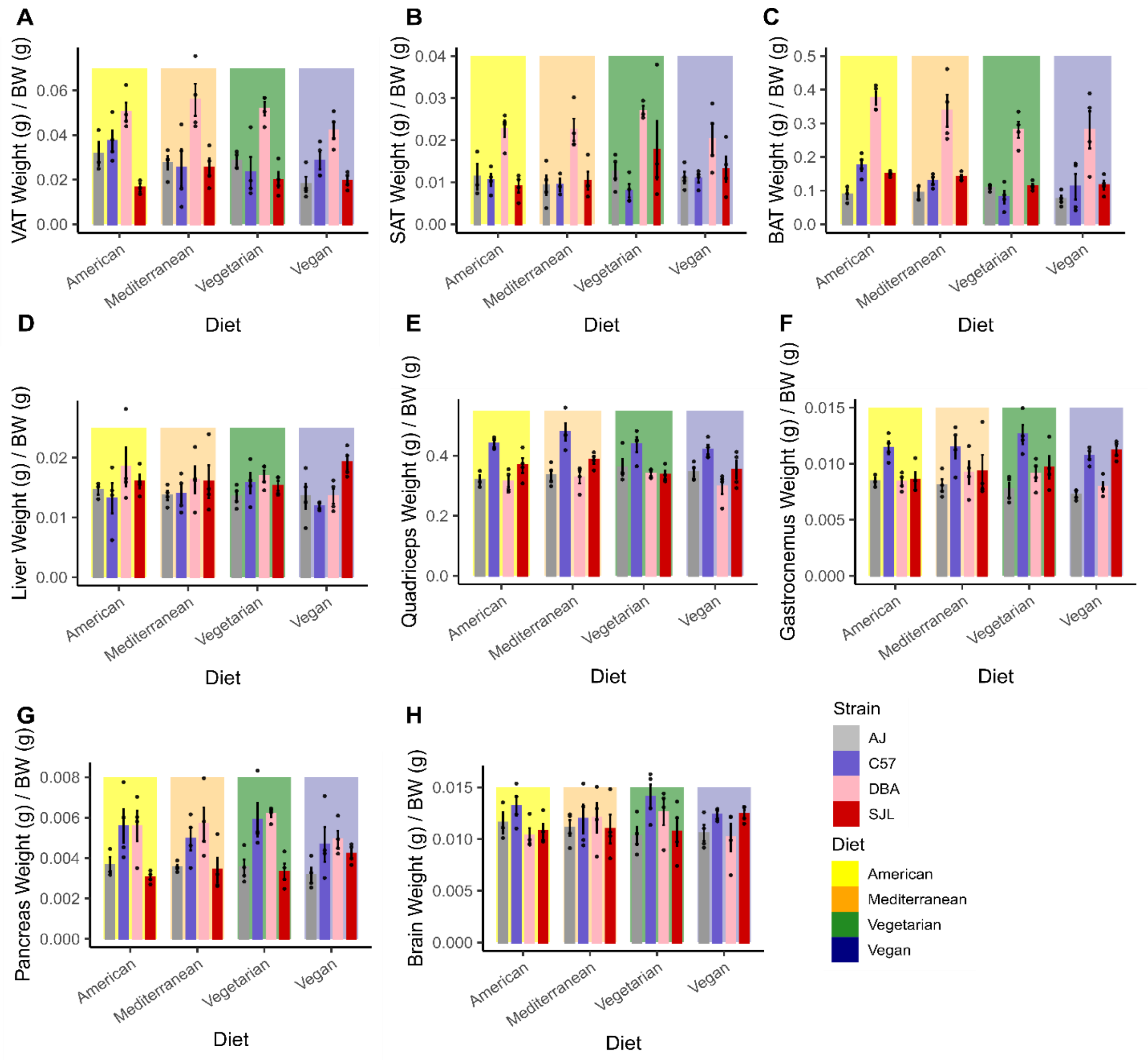

Figure 4.9: Diet and genetic background impact orgain weights in male mice

A. *Visceral adipose tissue weight per gram body weight after 6 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)*
B. *Subcutaneous adipose tissue weight per gram body weight after 6 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)*
C. *Brown adipose tissue weight per gram body weight after 6 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)*
D. *Liver weight per gram body weight after 6 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)*
E. *Quadriceps muscle weight per gram body weight after 6 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)*
F. *Gastrocnemius muscle weight per gram body weight after 6 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)*
G. *Brain weight per gram body weight after 6 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)*
H. *Pancreas weight per gram body weight after 6 weeks on diet. (n = 4, 12 weeks old at week 0, male mice)*
Error bars represent mean ± SEM

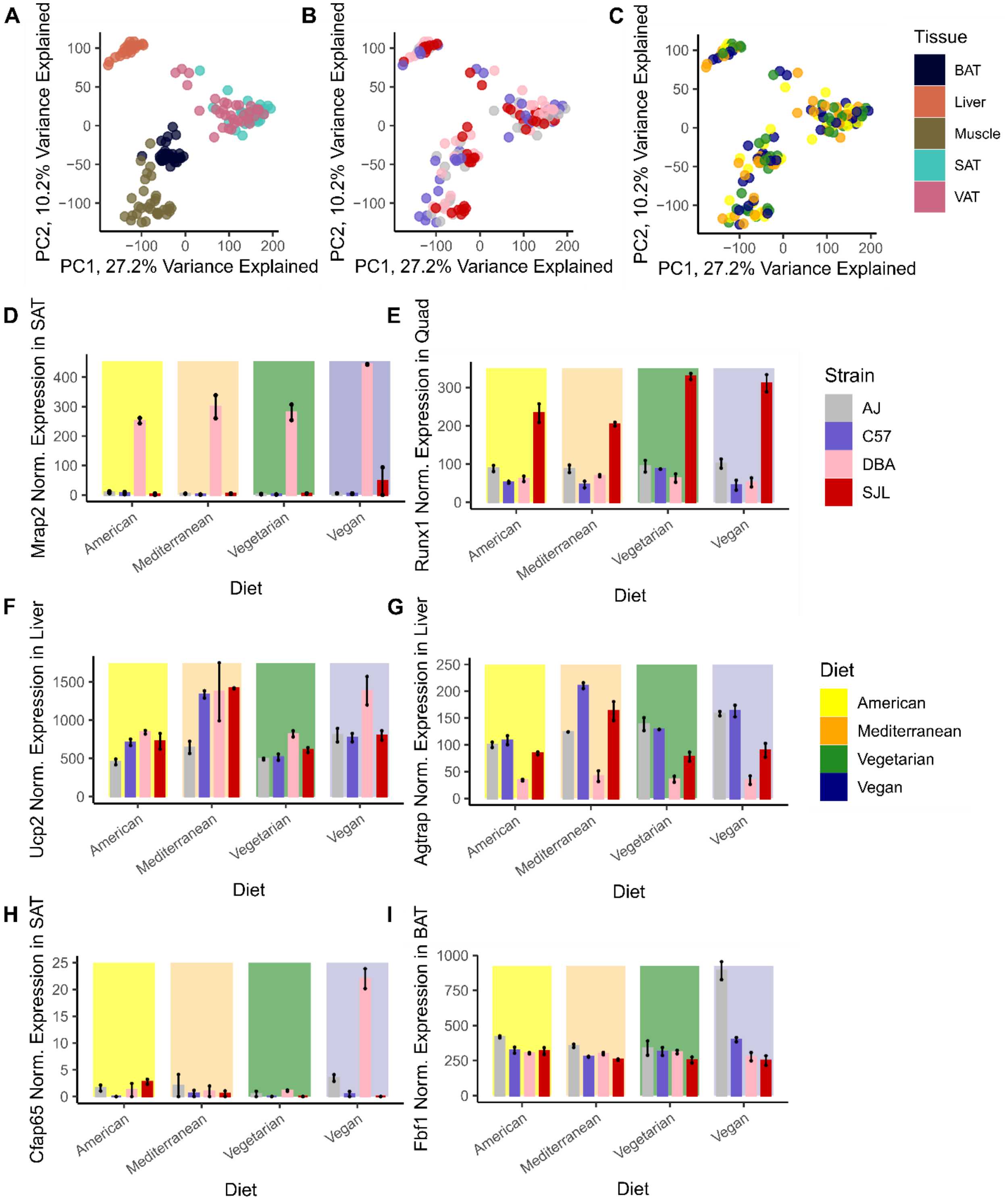

Figure 4.10: Genetic background impacts gene expression in liver, muscle, SAT, and BAT

A. Principal components analysis of expressed RNA species using all samples, PC 1 and 2 shown with samples colored by tissue

B. Principal components analysis of expressed RNA species using all samples, PC 1 and 2 shown with samples colored by genetic background

C. Principal components analysis of expressed RNA species using all samples, PC 1 and 2 shown with samples colored by diet
D. Expression of gene Mrap2 in subcutaneous adipose tissue, with significant strain-dependent effects. (n = 2, 18 weeks old, male mice)
E. Expression of gene Mrap2 in quadriceps skeletal muscle, with significant strain-dependent effects. (n = 2, 18 weeks old, male mice)
F. Expression of gene Ucp2 in liver, with significant diet-dependent effects. (n = 2, 18 weeks old, male mice)
G. Expression of gene Agtrap in liver, with significant diet-dependent effects. (n = 2, 18 weeks old, male mice)
H. Expression of gene Cfap65 in subcutaneous adipose tissue, with significant strain:diet-dependent effects. (n = 2, 18 weeks old, male mice)
I. Expression of gene Fbf1 in brown adipose tissue, with significant strain:diet-dependent effects. (n = 2, 18 weeks old, male mice)
Error bars represent mean ± SEM

4.3.4 Gene expression differences are predominantly dependent on genetic background

To determine which genes are expressed in response to diet, in combination with genetic background, we isolated and sequenced mature RNA species from the VAT, SAT, BAT, liver, and quadriceps skeletal muscle of two mice from each strain and diet group (Methods). We found that between ~24,000 (Liver) and ~31,000 (SAT) genes were expressed in each tissue.

We used PCA to reduce the dimensionality of the gene expression data. When viewing all samples on principal components 1 and 2, we observe clear separation of the samples by tissue, with strong overlap between the two white adipose tissues (SAT and VAT) (Figure 4.10 A). We considered principal components 3 and 4, which explain a smaller portion of the variation in the data (Figure 4.11). These components also separate the samples by tissue, including SAT and VAT. As expected, differences between tissue types account for the most variation in the expressed genes. We found that principal components 5, 6, and 7 separate the samples by strain (Figure 4.11), but we did not find any axes that were able to separate samples by diet.

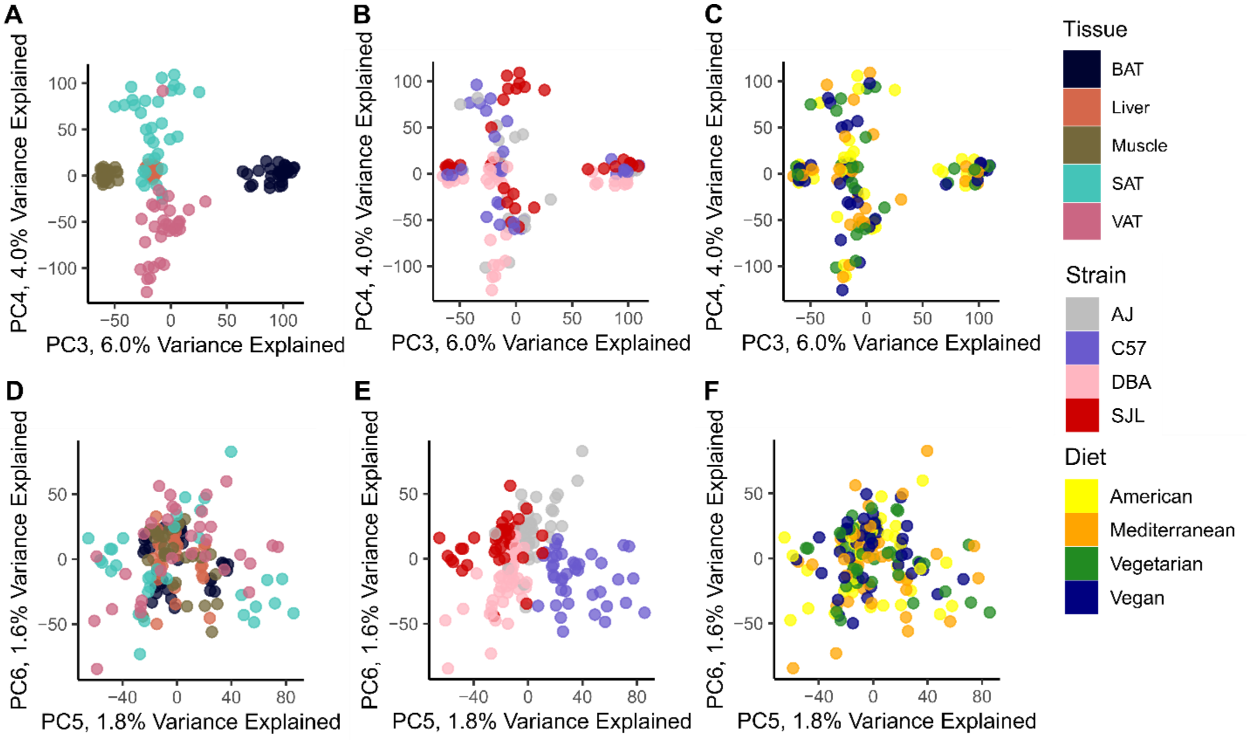

Figure 4.11: RNA expression separates samples by tissue, genetic background

- *A. Principal components 3 and 4, with samples colored by tissue*
- *B. Principal components 3 and 4, with samples colored by genetic background*
- *C. Principal components 3 and 4, with samples colored by diet*
- *D. Principal components 5 and 6, with samples colored by tissue*
- *E. Principal components 5 and 6, with samples colored by genetic background*
- *F. Principal components 5 and 6, with samples colored by diet*

Because of the dominating differences in gene expression due to tissue type, we separated the data by tissue and performed the PCA per tissue. In all tissues, principal components 1 and 2 separated samples by strain, but no components could separate the samples by diet (Figure 4.12). Strain is also more explanatory of the variation in gene expression than diet.

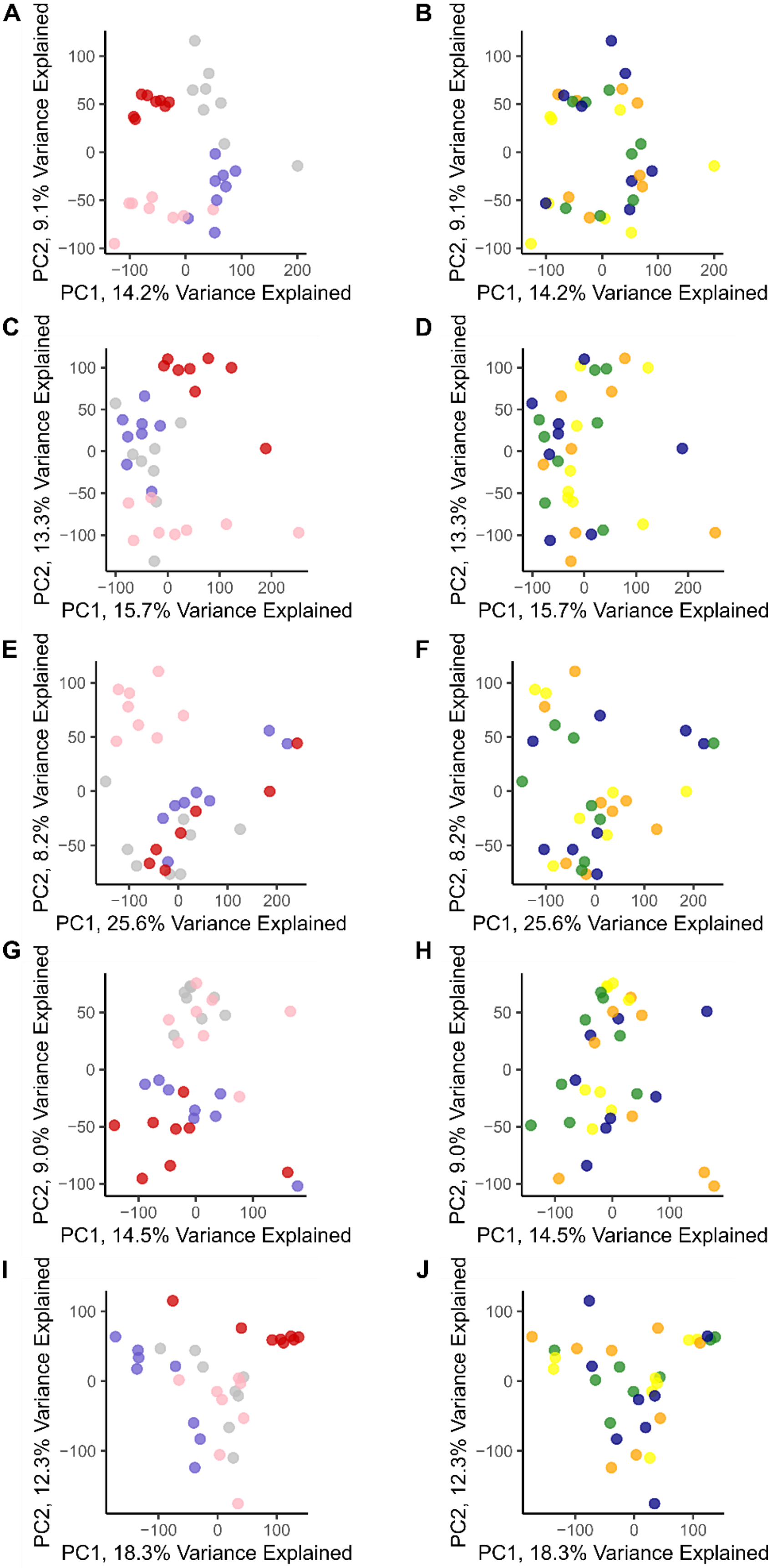
A
B
C
D
E
F
G
H
I
J
PC1, 14.2% Variance Explained
PC2, 9.1% Variance Explained
PC1, 15.7% Variance Explained
PC2, 13.3% Variance Explained
PC1, 25.6% Variance Explained
PC2, 8.2% Variance Explained
PC1, 14.5% Variance Explained
PC2, 9.0% Variance Explained
PC1, 18.3% Variance Explained
PC2, 12.3% Variance Explained
−100
0
100
200
−50
50

Figure 4.12: RNA expression separates samples by genetic background when considering only the samples of one tissue

- A. *Brown adipose tissue principal components 1 and 2, with samples colored by genetic background*
- B. *Brown adipose tissue principal components 1and 2, with samples colored by diet*
- C. *Subcutaneous adipose tissue principal components 1 and 2, with samples colored by genetic background*
- D. *Subcutaneous adipose tissue principal components 1and 2, with samples colored by diet*
- E. *Visceral adipose tissue principal components 1 and 2, with samples colored by genetic background*
- F. *Visceral adipose tissue principal components 1and 2, with samples colored by diet*
- G. *Liver principal components 1 and 2, with samples colored by genetic background*
- H. *Liver principal components 1and 2, with samples colored by diet*
- I. *Quadriceps muscle principal components 1 and 2, with samples colored by genetic background*
- J. *Quadriceps muscle principal components 1and 2, with samples colored by diet*

For each tissue, we used the 2-way ANOVA test to identify genes whose expression varies across samples due to the effects of strain, diet, and the Strain:Diet interactions. We identified between ~8,300 (Quad) and ~13,600 (SAT) genes with a significant strain effect (Supplemental Tables 4.2-4.6). Melanocortin 2 Receptor Accessory Protein 2, *Mrap2,* expression in SAT (Figure 4.10 D) and RUNX Family Transcription Factor 1, *Runx1,* expression in skeletal muscle are significantly affected by strain (Figure 4.10 E). Far fewer genes were impacted by diet; we identified between 0 (BAT) and 50 (VAT) genes with a significant diet effect (Supplemental Tables 4.2-4.6). Angiotensin II Receptor Associated Protein, *Agtrap,* and Uncoupling Protein 2, *Ucp2,* expression in liver were significantly affected by diet (Figure 4.10 F,G). After VAT, liver showed the largest response to diet (Supplemental Table 4.3).

In most tissues, we identified few genes with significant Strain:Diet interaction effects (Table 4.3, Supplemental Table 4.7). Fas Binding Factor 1, *Fbf1,* is the only gene that shows interaction effects in BAT (Figure 4.10 I), while SAT had six genes including Cilia and Flagella Associated Protein 65, *Cfap65,* (Figure 4.10 H); liver had five genes, and muscle had two genes (Table 4.3). In VAT, however, 421 genes had significant interaction effects between diet and genetic background on gene expression (Table 4.3, Supplemental Table 4.7).

Table 4.3: 2-way ANOVA reveals genes whose expression depends on strain and diet interactions

Tissue	Ensembl ID	Gene Symbol	strain:diet adjusted *P* value
BAT	ENSMUSG00000020776	Fbf1	1.33E-02
Liver	ENSMUSG00000028610	Dmrtb1	7.50E-05
Liver	ENSMUSG00000112384	Gm34921	7.50E-05
Liver	ENSMUSG00000107964	4930447C11Rik	5.65E-04
Liver	ENSMUSG00000047222	Rnase2a	3.27E-02
Liver	ENSMUSG00000031860	Pbx4	5.04E-02
Quadriceps	ENSMUSG00000097149	G630030J09Rik	1.44E-05
Quadriceps	ENSMUSG00000109643	Gm31545	3.71E-03
SAT	ENSMUSG00000096824	Ighv2-7	2.86E-05
SAT	ENSMUSG00000110123	Gm18562	2.86E-05
SAT	ENSMUSG00000031881	Cdh16	3.49E-04
SAT	ENSMUSG00000047021	Cfap65	5.68E-04
SAT	ENSMUSG00000108947	Or4d29-ps1	2.26E-02
SAT	ENSMUSG00000103939	Ighv3-4	3.82E-02
VAT	ENSMUSG00000024987	Cyp26a1	4.80E-05
VAT	ENSMUSG00000042129	Rassf4	4.80E-05
VAT	ENSMUSG00000094732	1500015L24Rik	4.80E-05
VAT	ENSMUSG00000110537	Gm4316	4.80E-05
VAT	ENSMUSG00000041468	Gpr12	1.69E-04
VAT	ENSMUSG00000089873	Mup13	2.45E-04
VAT	ENSMUSG00000095385	D630033O11Rik	2.45E-04
VAT	ENSMUSG00000103706	6820402A03Rik	2.45E-04
VAT	ENSMUSG00000104200	Gm37399	2.45E-04
VAT	ENSMUSG00000085818	Gm13267	4.23E-04
VAT	ENSMUSG00000020926	Adam11	4.83E-04
VAT	ENSMUSG00000103258	Gm37518	4.83E-04
VAT	ENSMUSG00000115919	Gm31583	4.83E-04
VAT	ENSMUSG00000024130	Abca3	5.36E-04
VAT	ENSMUSG00000069170	Adgrv1	7.18E-04
VAT	ENSMUSG00000120238	Gm57333	7.18E-04
VAT	ENSMUSG00000061356	Nuggc	7.79E-04
VAT	ENSMUSG00000118345	Gm7105	7.79E-04
VAT	ENSMUSG00000043671	Dpy19l3	8.24E-04
VAT	ENSMUSG00000084132	Gm15982	8.87E-04
VAT	ENSMUSG00000091157	Serpina3l-ps	9.71E-04
VAT	ENSMUSG00000035864	Syt1	1.18E-03
VAT	ENSMUSG00000085123	Rubie	1.18E-03
VAT	ENSMUSG00000109480	Gm45042	1.18E-03
VAT	ENSMUSG00000022003	Slc25a30	1.31E-03

**Top 20 most significant interactions per tissue show, full list in Supplemental Table 4.7*

4.3.5 Lipid metabolism and transport genes in are regulated by strain and diet interactions in VAT

To understand which pathways were effected by Strain:Diet interactions, we identified gene ontology (GO) biological process pathways that were enriched within our gene set. We found that fatty acid and sterol transport and metabolism pathways and P450 cytochrome pathways were enriched in our Strain:Diet responsive gene set (Figure 4.13).

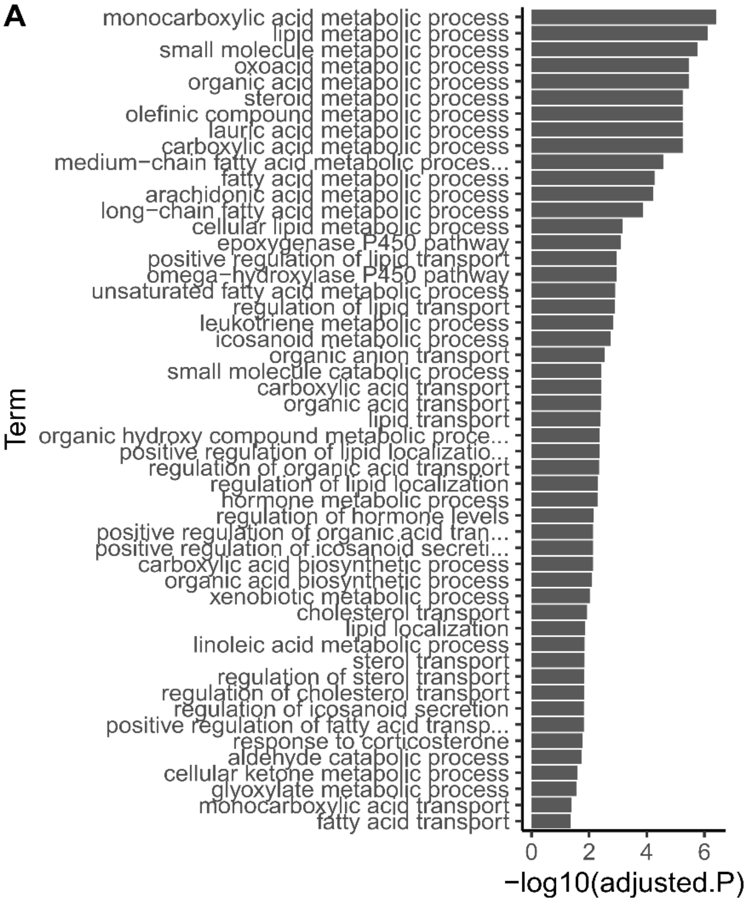

Figure 4.13: Pathway analysis of 421 visceral adipose tissue genes that exhibit with significant strain:diet-dependent effects

A. *Significantly enriched pathways from the Gene Ontology- biological processes database. P-values adjusted using FDR correction for multiple tests.*

We observe various patterns of expression in the 421 genes (Figure 4.14 A). To reconcile the patterns with our observed phenotypes, we investigated a small subset of the 421 genes whose expression varies in a similar way to the certain traits. Some genes, such as High-mobility group box 2, *Hmgb2,* (Figure 4.14 B) and Neurocan, *Ncan,* (Figure 4.14 C), were higher in DBA/2J mice, and diets affected expression in each strain in a similar pattern to body weight or insulin phenotypes, respectively. Fat pad weights somewhat corresponded with both genes.

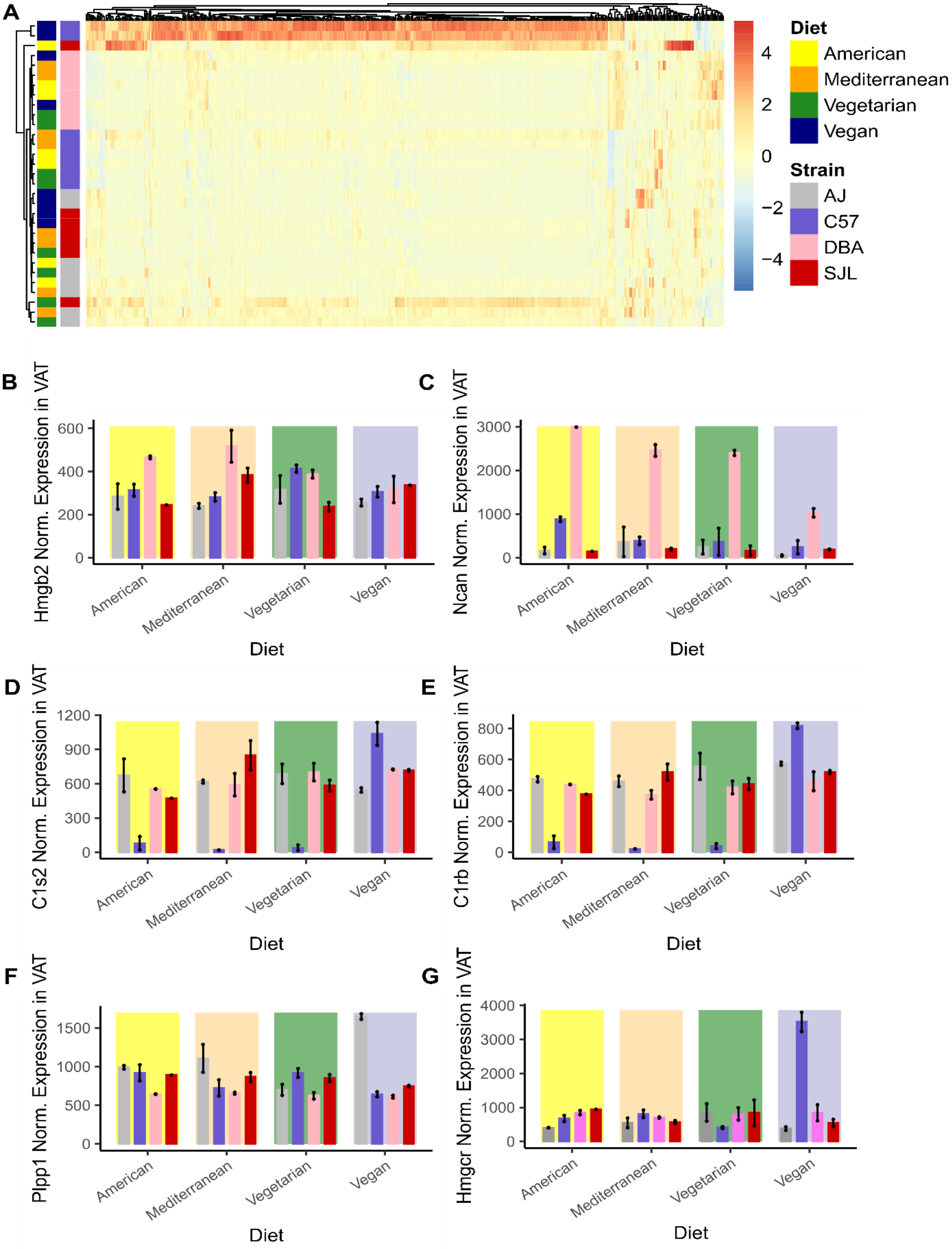

Figure 4.14: 421 genes in visceral adipose tissue are driven by strain:diet interaction effects

A. *Heatmap of 421 visceral adipose tissue genes that exhibit with significant strain:diet-dependent effects*

B. Expression of gene Hmgb2 in visceral adipose tissue, with significant strain:diet-dependent effects. (n = 2, 18 weeks old, male mice)
C. Expression of gene Ncan in visceral adipose tissue, with significant strain:diet-dependent effects. (n = 2, 18 weeks old, male mice)
D. Expression of gene C1s2 in visceral adipose tissue, with significant strain:diet-dependent effects. (n = 2, 18 weeks old, male mice)
E. Expression of gene C1rb in visceral adipose tissue, with significant strain:diet-dependent effects. (n = 2, 18 weeks old, male mice)
F. Expression of gene Plpp1 in visceral adipose tissue, with significant strain:diet-dependent effects. (n = 2, 18 weeks old, male mice)
G. Expression of gene Slc2a2 in visceral adipose tissue, with significant strain:diet-dependent effects. (n = 2, 18 weeks old, male mice)
Error bars represent mean ± SEM

Other genes also follow the pattern of highest expression in DBA/2J mice on American diet, and often, for the same gene, the SJL samples have the highest expression on Vegan diet (Figure 4.15). These genes respond to diet on some genetic backgrounds, and may directly contribute to the observed weight gain and blood metabolite levels. Expression of complement Factor 1 genes *C1s2* and *C1rb* (Figure 4.14 D,E) partially reflected MRI total body fat patterns. Vegan diet increased expression of these genes on SJL, and to a larger degree, on C57BL/6J backgrounds.

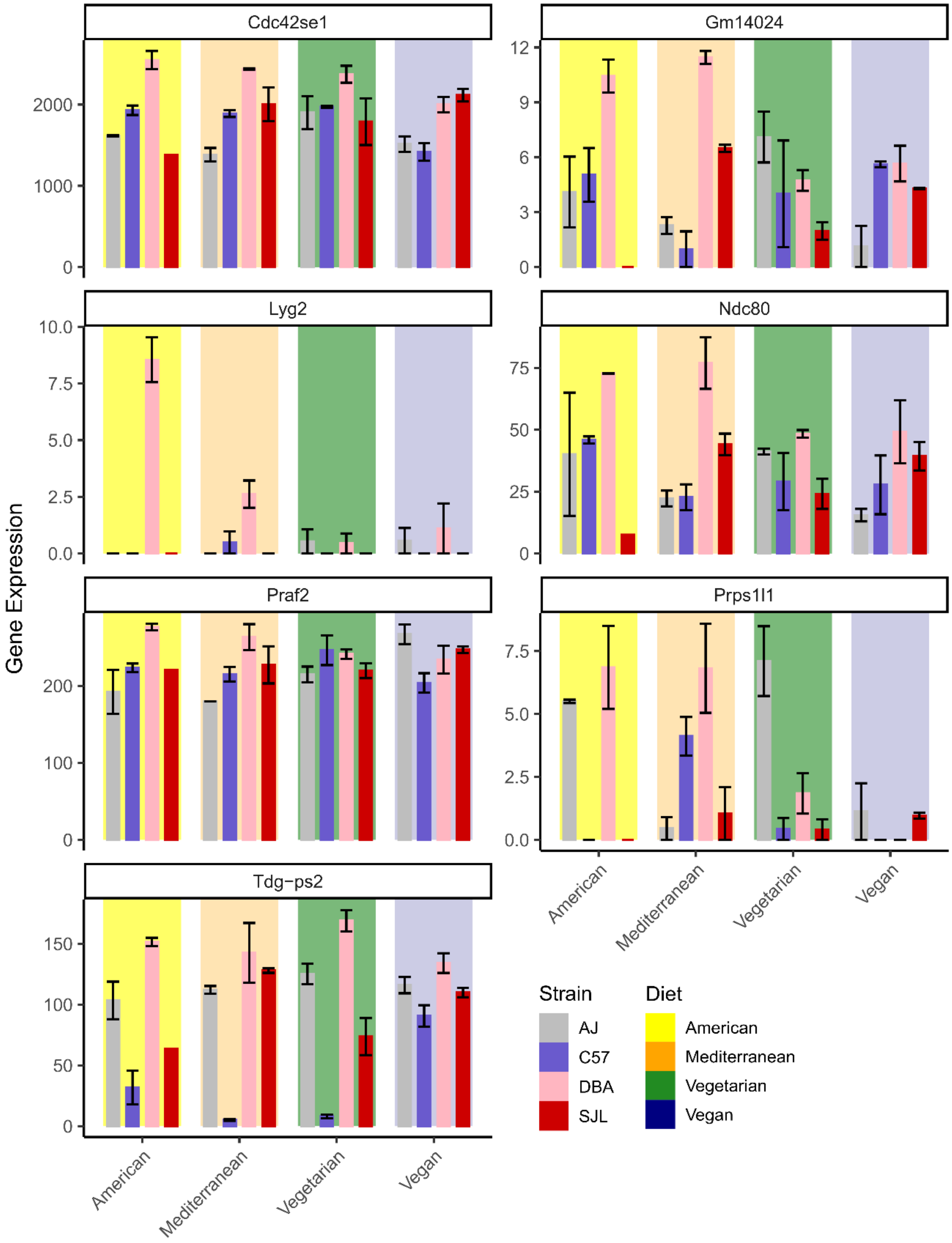

Figure 4.15: Expression of 7 select genes that exhibit DBA-strain effects of the 421 visceral adipose tissue, with significant strain:diet-dependent effects

Other genes were expressed specifically in some strains or diets in a way that did not reflect any observed phenotypes, such as Phospholipid phosphatase 1, *Plpp1* (Figure 4.14 F), was specific to A/J Vegan samples. A large portion of Strain:Diet interaction genes was dominated by C57BL/6J Vegan-specific effects, which did not reflect any observed phenotypes. *Hmgcr* followed this pattern of expression (Figure 4.14 G), and other C57BL/6J Vegan-specific genes fell into broad categories, such as solute carriers and transporters (Figure 4.16). These encoded transporters of metabolically relevant molecules [*Slc2a2-* glucose, *Slc17a8-* glutamate, *Slc13a2-* citrate, *Slc22a7-* prostaglandins*, Slc22a1-* Na^+/vitamin C, *Slc30a1-* manganese*, Slc35a2-* nucleotide sugars*, Slc36a1-* apolar amino acids (aa), *Slc39a2-* zinc, *Slc40a1-* iron*, Slc66a1-* cationic aas, *Kcnk10-* Potassium]. Other categories of genes included ATP-binding cassettes (Figure 4.17), cytochrome P450s (Figure 4.18), hydroxysteroid dehydrogenases (Figure 4.19), and serine protease inhibitors (Figure 4.20). Other notable genes include sterol and fatty acid metabolism genes, including fatty acid binding proteins *Fabp1* and *Fabp2,* lipases *Lipc* and *Lipg,* and insulin like growth factor binding protein 2 *Igfbp2* (Figure 4.21). This pattern of expression was unique to C57BL/6J Vegan diet-fed mice, which did not correspond specifically with any observed phenotypes.

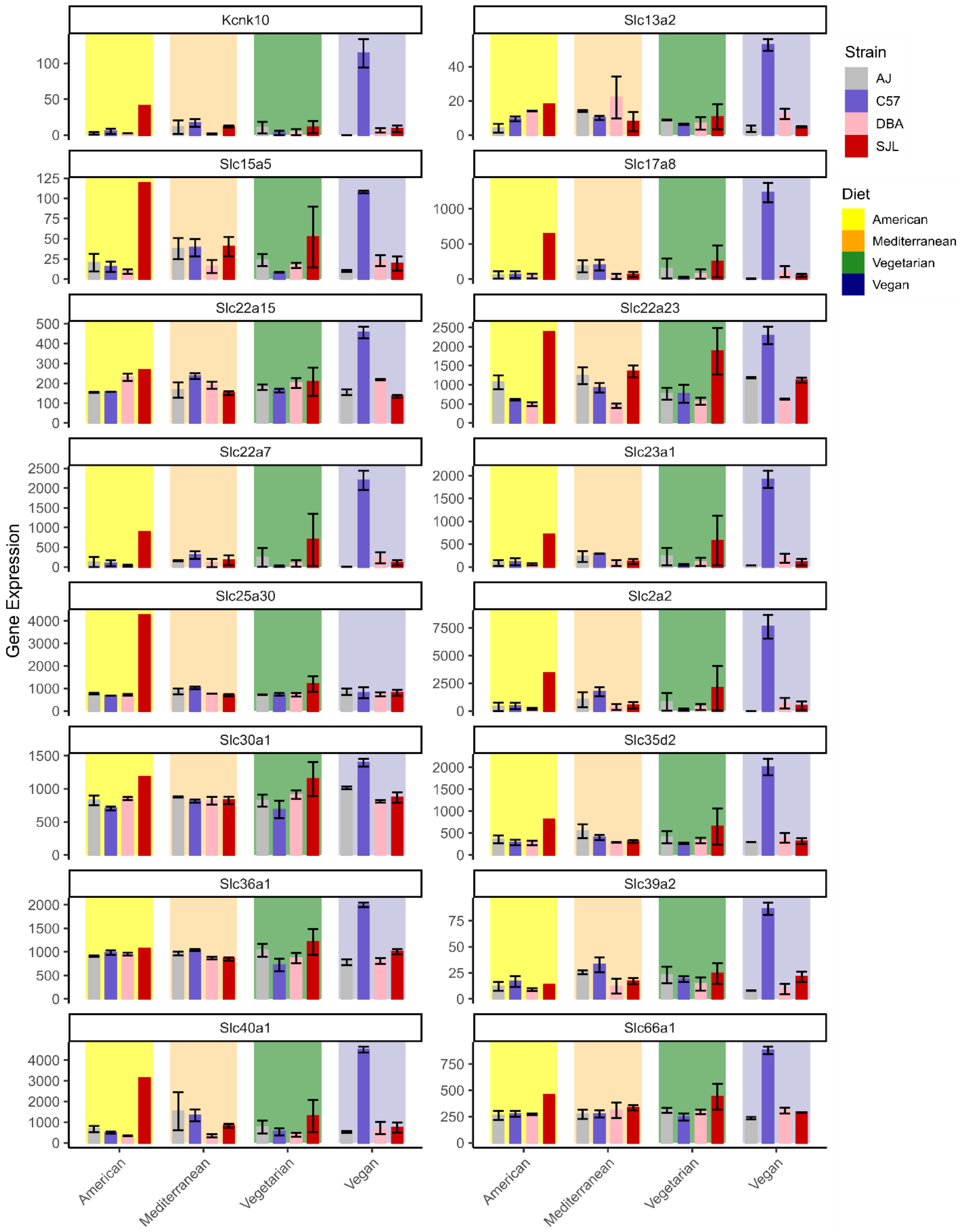

Figure 4.16: Expression of transporter genes that exhibit C57BL/6J-Vegan effects of the 421 visceral adipose tissue genes with significant strain:diet-dependent effects

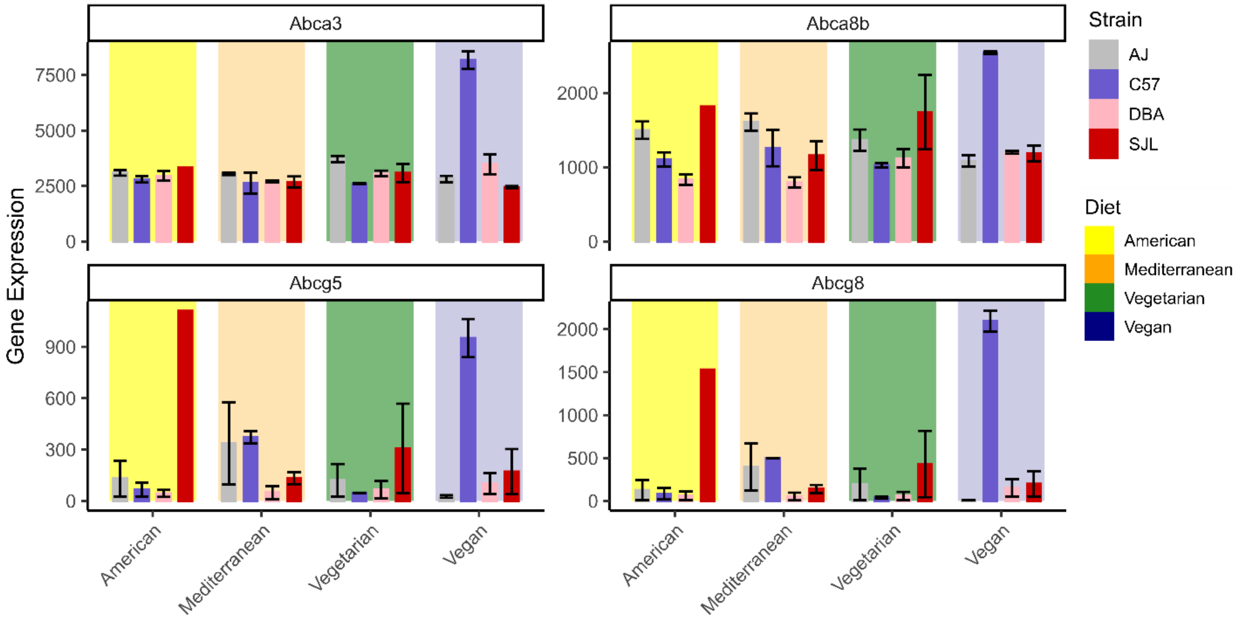

Figure 4.17: Expression of ATP-binding cassette genes that exhibit C57BL/6J-Vegan strain:diet-dependent effects

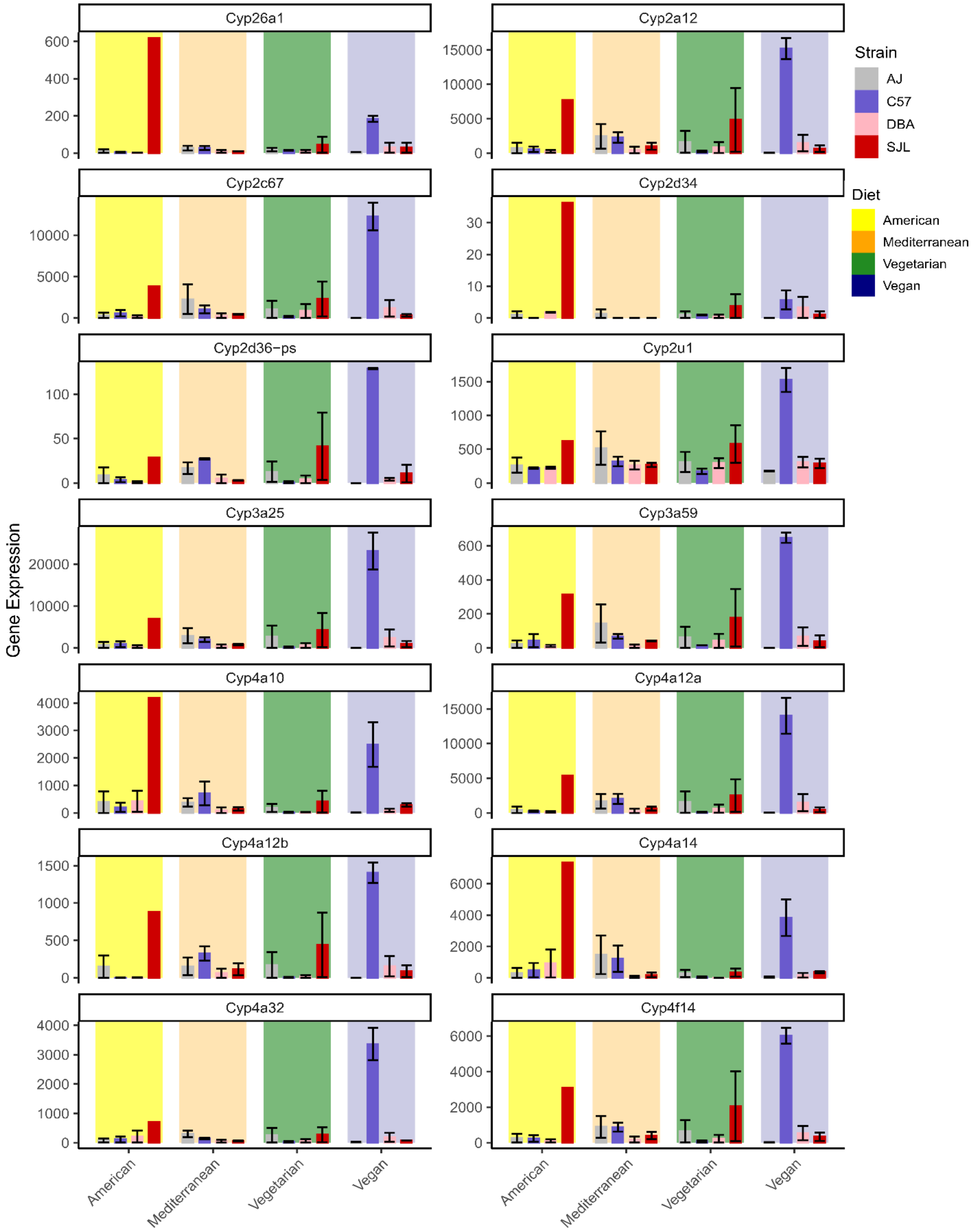

Figure 4.18: Expression of cytochrome P450 genes that exhibit C57BL/6J-Vegan strain:diet-dependent effects

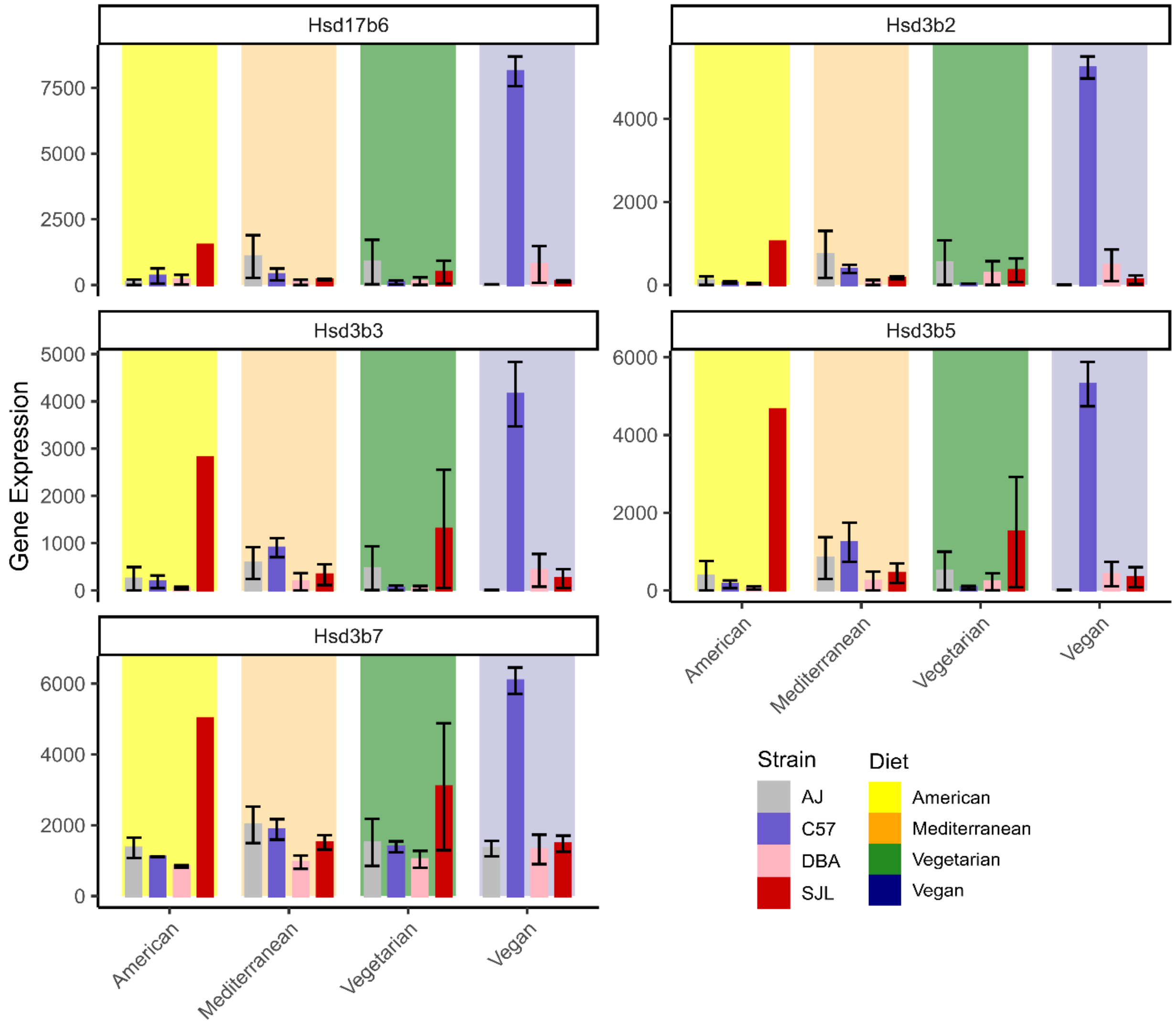

Figure 4.19: Expression of hydroxyl-steroid dehydrogenase genes that exhibit C57BL/6J-Vegan strain:diet-dependent effects

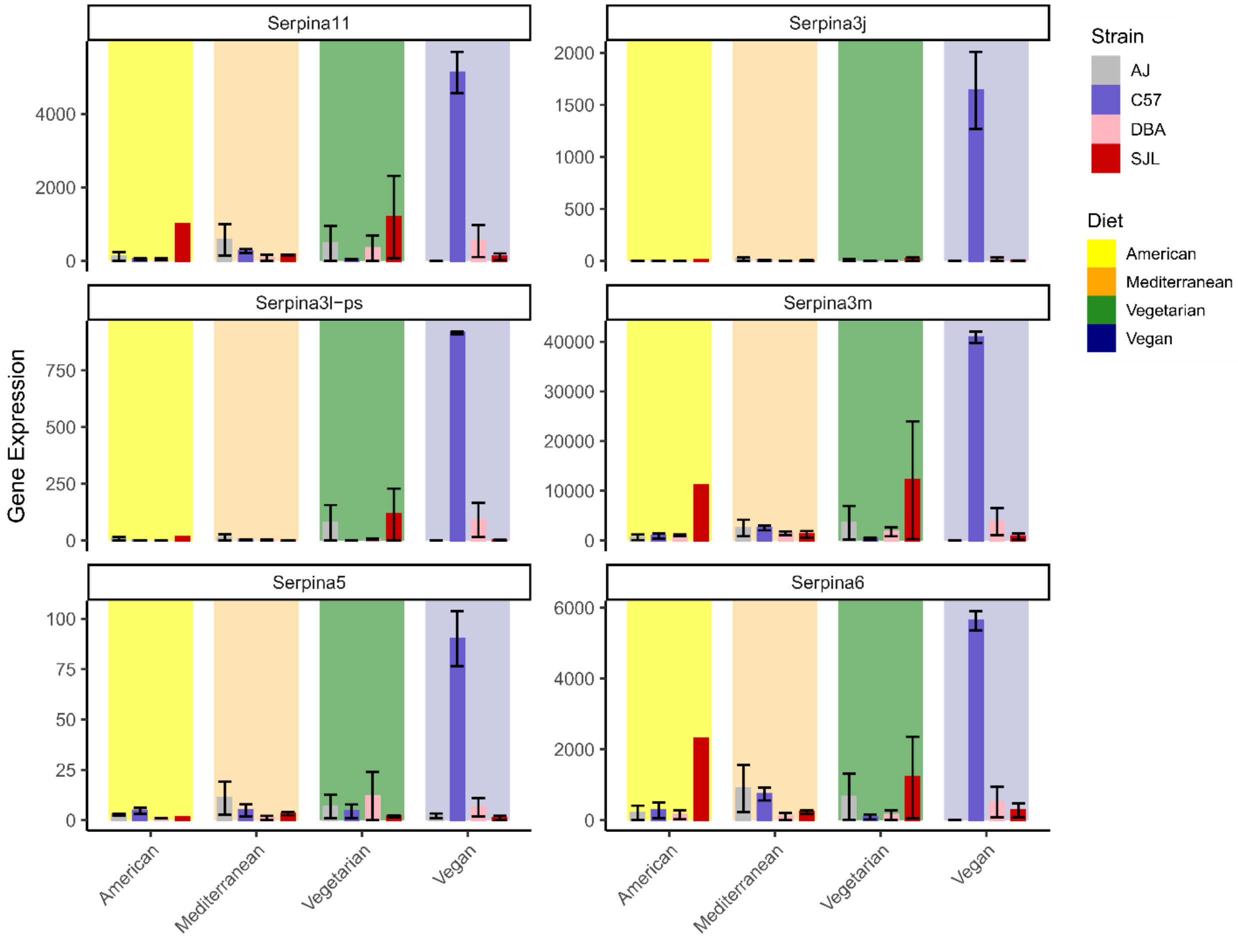

Figure 4.20: Expression of serine protease genes that exhibit C57BL/6J-Vegan strain:diet-dependent effects

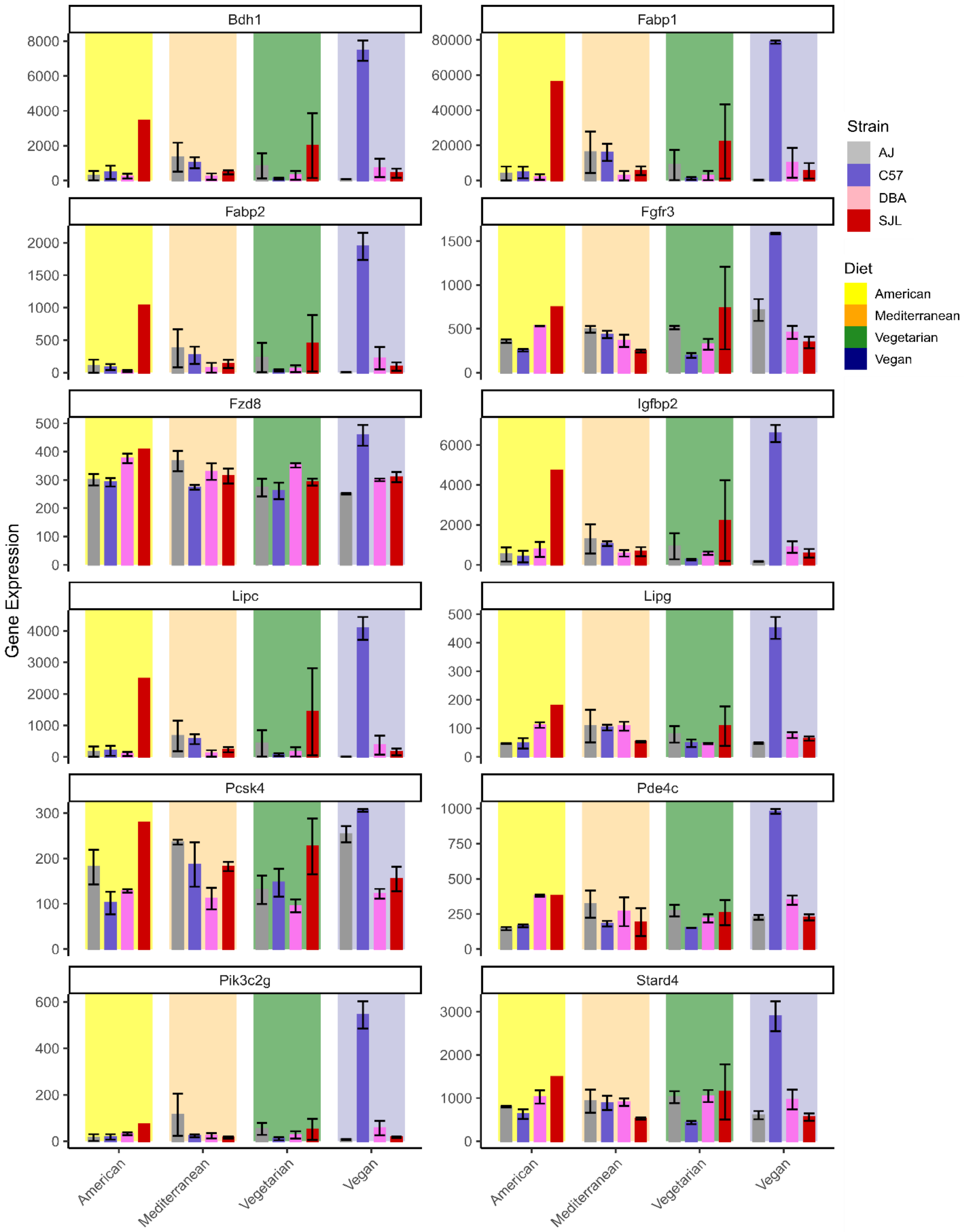

Figure 4.21: Expression of other genes that exhibit C57BL/6J-Vegan strain:diet-dependent effects

4.3.6 Patterns of visceral adipose tissue gene expression may control strain- and diet dependent metabolic responses

The patterns of gene expression were only partially explanatory of the strain and diet responses observed in metabolic phenotyping. Therefore we employed a computational tool called partial least-squares regression (PLSR) to investigate how the gene expression and metabolic traits are related using the underlying correlation structure within both datasets. Like PCA, this approach reduces the high dimensional data to one or two component axes. In PLSR, we reduce two high dimensional datasets, and the axes represent the direction in the predictor dataset that explains the most variability in the outcome dataset. Predictors, outcomes, and samples that are highly related are grouped together in the component space.

We used the expression of the 421 genes with significant Strain:Diet responses to predict the observed metabolic phenotypes (Figure 4.22 A). Like PCA, PLSR components 1 and 2 are able to separate the samples by strain. The phenotypes cluster into logical groups; body weight phenotypes are grouped on the bottom left, muscle weights at the bottom, glucose on the right, and fat pad weight, insulin, and triglyceride phenotypes are grouped at the top left. The y-axis appears to separate lean mass from fat mass, while the x-axis seems to control triglyceride storage versus glucose utilization. DBA/2J mice overlap with fat pad phenotypes in the component space, while C57BL/6J were spread between body weight and glucose phenotypes. Some SJL overlapped with glucose phenotypes, and A/J mice did not overlap with any measured phenotypes.

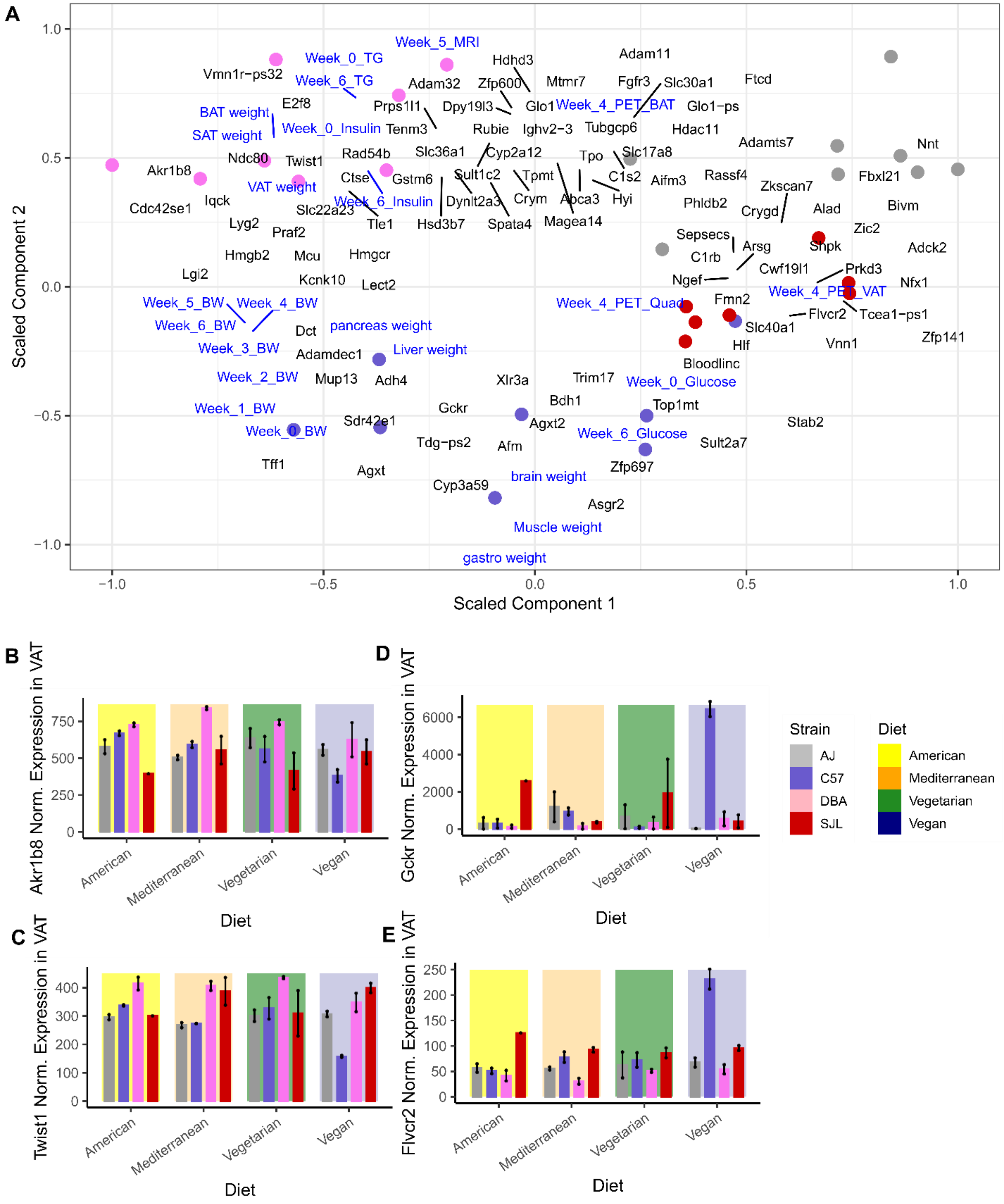

Figure 4.22: Dimensionality reduction predicts correlated genes and traits

- *A. PLSR components 1 and 2, scaled to [-1,1], with samples colored by genetic background. Loadings of predictor (black) and outcome (blue) variables shown.*
- *B. Expression of gene Akr1b8 in visceral adipose tissue, with significant strain:diet-dependent effects. (n = 2, 18 weeks old, male mice)*

C. Expression of gene Twist1 in visceral adipose tissue, with significant strain:diet-dependent effects. (n = 2, 18 weeks old, male mice)
D. Expression of gene Gckr in visceral adipose tissue, with significant strain:diet-dependent effects. (n = 2, 18 weeks old, male mice)
E. Expression of gene Flvcr2 in visceral adipose tissue, with significant strain:diet-dependent effects. (n = 2, 18 weeks old, male mice)
Error bars represent mean ± SEM

The 421 genes can be found near the phenotypes they are most predictive of. *Hmgcr* is found on the left, near both weight and adipose tissue phenotypes, as are other genes with known links to adipocyte expansion *Hmgb2*[69,70] and *Kcnk10*[71,72]*.* Aldo-keto reductase family 1 B8, *Akr1b8* is a gene involved in lipid processing in liver[73,74]; its expression in VAT shows strong strain and diet interactions (Figure 4.22 B). Other genes closer to the fat pad weight area include transcription factors *Twist1*[75] (Figure 4.22 C) and *E2f8*[76,77] that show similar patterns of expression to *Akr1b8*.

Gckr is found between body weight, liver, and muscle phenotypes, and its expression shows strong diet effects in C57BL/6J mice (Figure 4.22 D). Other genes in this area include metabolic genes alanine-glyoxylate aminotransferases *Agxt* and *Agxt2*[78]*,* and ketogenic gene 3-hydroxybutyrate dehydrogenase, *Bdh1*[79]. In the glucose uptake area on the right, we observe the iron exporter ferroportin, *Slc40a1*[80]*,* and heme/ choline transporter feline leukemia virus subgroup C cellular receptor 2 *Flvcr2*[81,82]*.* These show a similar pattern of expression to other C57BL/6J-Vegan-specific genes (Figure 4.22 E).

4.4 Discussion

We show that body weight gain and metabolic phenotypes in mice are a result of diet source, genetic background, and interaction effects between the two. We consistently find that the effect of genetic background is stronger than the effect of diet. We find that DBA/2J mice store the most weight and have the highest insulin and glucose, and this effect is modified by diet. We identify over 400 genes whose expression is responsive to the interaction effects of strain and diet in visceral adipose tissue. We find some genes with expression patterns that match the patterns of body weight gain, like *Hmgb2*, *Akr1b8,* and *Twist1* in VAT and, to some degree, *Ucp2* in liver. *Ncan* expression in VAT matches insulin measurements and fat pad weights*, Runx1* expression in muscle aligns with glucose uptake measurements, and *C1s2* and *C1rb* expression patterns in VAT partially match total fat. Using dimensionality reduction tools, we predict that *Hmgb2*, *Akr1b8,* and *Twist1* expression influence fat storage phenotypes, most strongly in DBA/2J mice.

Previous mouse studies show large differences in weight gain and insulin resistance between the A/J, C57BL/6J, DBA/2J and SJL strain at baseline and in response to high fat diet feeding[58]; we also capture a large effect of strain on body weight gain and insulin levels. Other studies that used humanized diets showed stronger effects of diet[66]; however, differences in macronutrient content between those diets could account for the observed differences. Keeping fat, carbohydrate, and protein levels the same between our diets perhaps blunted the benefits. Each diet contained around twice the fat content as mouse chow; abnormal macronutrient content could also mute the effects as well. Further, while we observed similar caloric intake between groups, we measured food consumption per cage (4-5 animals averaged). Individual-level phenotyping using metabolic cages would allow us to accurately measure food consumption and energy expenditure, as well as observe any confounding behavior such as coprophagy.

We also saw larger, more significant differences in metabolic parameters in study 1 than study 2, though the directionality and pattern between strains and diets was largely conserved. The study

2 mice we younger and on the diet for less time; age-related weight gain could magnify the differences. We also performed more intrusive tests in study 2 that could illicit stress responses and introduce noise. We only used four mice per group in study 2, and we are underpowered to fully detect differences. Similarly, with n=2 mice per group, RNA-seq was underpowered to identify the full spectrum of Strain:Diet responsive genes; only large magnitude changes in gene expression were identified as significant. Further, we only considered male mice in this study, which limit our conclusions to this population.

Our ANOVA analysis identified 400 genes in VAT whose expression is altered by Strain:Diet interactions, many of which were identified previously. Genetic variation in *HMGCR*[49] and *GCKR*[35,36] modify the effect of diet on metabolic parameters, here we show that the expression of these genes in visceral adipose is responsive to genetic background and diet interactions. We identify *Ucp2* as strain and diet responsive but without a significant interaction effect; previous studies show interactions between dietary energy intake and SNPs in *UCP2* on metabolic parameters[52]. *Ncan* is a gene previously identified in fatty liver disease[83].

Previous studies identified a mechanistic role in for *Hmgb2*[69] and *KcnK10*[71] in regulating early stages of adipogenesis, and for *Twist1* in regulating adipose tissue angiogenesis[75]. We find their expression is responsive to genetic background and diet. These genes may coordinate adipose expansion under diet challenge. We also identified genes that had been previously studied in the liver (*Agxt, Agxt2, Bdh1, Slc40a1, Hmgcr,* and *Gckr*)[79,80,84–87] that regulate fatty acid and sterol metabolism. These genes had highest expression in C57BL/6J-Vegan fed mice specifically. Interestingly, in cohort 1, the C57BL/6J mice had the lowest circulating NEFA at week 16, though this was not specific to Vegan diet. In the PLSR plot, we find the adipogenic/angiogenic genes among fat pad and body weight phenotypes and, at the opposite side of the plot, we find the liver-related genes among glucose phenotypes, though NEFAs were not measured in cohort 2, and thus could not be predicted.

We used multiple dimensionality reduction techniques to view our samples based on the underlying correlation structure within the data. PLSR was chosen for its ability to dissect relationships among multiple co-correlated variables. The model recapitulates relationships between samples and between observed phenotypes, and predicts known and novel regulators of fat storage. Mechanistic studies are required to determine if and how these genes respond to components of the diet.

These studies present a novel resource for the study precision nutrition, which attempts to apply diets in the populations in which they would be most beneficial[88]. By considering the genetic risk, among other factors, clinicians can prescribe diets to patients and populations who would see the most disease protection. We identify known and novel candidate genes that respond to nutrient source and genetic background, and lay the groundwork for identifying future mechanistic studies.

4.5 Methods

4.5.1 Experimental Details and Study Participant Details

Animals: Male mice were obtained from Jackson Laboratories at 9 weeks of age except where noted, C57BL/6J (Stock #000664), A/J (Stock #000646), DBA/2J (Stock #000671), and SJL/J (Stock #000686). For the first study SJL/JCrNTac (Model # SJL-M) were obtained from Taconic Biosciences. For the second study SJL/J mice were obtained from Jackson Laboratories, but at 5 weeks of age (as they were not available for purchase at a later age). Mice were housed under temperature-(22°C) and humidity-controlled conditions and a constant light-dark cycle with free access to water and chow. They were fed a standard chow diet (cat. no. 7912 Teklad LM-485; Harlan Laboratories) until the start of the humanized special diets at 14-15 weeks of age (first study) and at 12-13 weeks of age (second study). After this time mice were fed one of the 4 humanized diets (described in detail below under “Humanized Diets”), American, Mediterranean, Vegetarian, Vegan, until euthanasia (16 weeks on diets for study 1 and 6 weeks for study 2, see Figure 4.1). Before starting the diets, mice were weighed and mice from each strain were assigned to the different diets such that the average body weight of mice on each diet was similar. Mice were housed 4 per cage for all strains except for SJL/J that required single housing due to aggressive behavior against each other. Body weights and food intake were determined weekly in both studies until euthanasia for study 2 and up to 10 weeks of age and then at euthanasia for study 1. Blood draws for measurements of metabolic parameters (blood glucose, insulin, triglycerides, and free fatty acids) were obtained before starting the diets and then at 4, 8, and 16 weeks on diets for study 1 and for study 2 before starting the diets and at euthanasia. Glucose uptake in different tissues was measured in live mice at 4 weeks on diets in study 2 using FDG-PET. MRI was performed at 5 weeks of age in study 2 mice. The number of mice to begin diets was n=7-8 for study 1 and n=4 for study 2. There were a few deaths (for unknown reasons) during study 1 (C57BL/6J 1 on Mediterranean diet, 1 on Vegetarian diet, DBA/2J 1 on Mediterranean and 1 on Vegetarian diet and 1 death in study 2 A/J on American diet). Mice were euthanized and blood collected by cardiac puncture and tissues dissected. Tissues were frozen in liquid nitrogen and stored at -80^{o}C until further processing. All animal procedures were approved by the University of Virginia Institutional Animal Care and Use Committee.

4.5.2 Design of clinically relevant mouse diets

We designed four mouse versions of common human diets. To avoid issues with differences in macronutrient composition, we kept macronutrient composition the same for all diets. Average/mean macronutrient intake among adults in the US[68] served as a guide; 49-51% kcal from carbohydrates, 15-16% kcal from protein; 33-34% kcal from fat settling on a constant content of macronutrients in each diet of 35% fat, 15% protein, 50% carbohydrates. Meal plans for humans containing the desired macronutrient composition of 35% fat, 15% protein, 50% carbohydrate were designed for a “typical” American, Mediterranean, Vegetarian, and Vegan diets (Figure 4.1). The sources for macronutrients used in these human diets were translated into mouse diets (Supplemental Table 1). A “humanized” mouse version of each diet was milled. Each diet is iso-caloric and contains equivalent macronutrient content (Supplemental Table 1).

4.5.3 Food Intake Measurements

Food weights were measured before addition to top of cages. One week later remaining food was weighed and subtracted from food weight added a week later to obtain food intake/cage during the week.

4.5.4 Blood Collection

Blood was collected between 8-9 AM from nicked tail veins of mice fed ad libitum. The first drop was used for blood glucose measurements (as described below) and additional blood was then collected into heparinized and non-heparinized capillary tubes. Blood was transferred to 0.65 ml microfuge tubes and kept on ice before being centrifuged at 1,600 g for 30 min at 4^{o}C. Plasma or serum was collected and stored at -20^{o}C until further analysis (see below). Blood at euthanasia was collected by cardiac puncture (right after CO_2 asphyxia) with a 1 ml syringe connected to a 23-gauge needle, transferred to a 1.5 ml microcentrifuge tube, allowed to clot for 30 min on ice, and was then centrifuged at 1,600 g for 30 min at 4^{o}C. Serum was stored at -20 (short-term storage) and -80^{o}C (long-term storage).

4.5.5 Measurements of Metabolic Parameters

Blood glucose was measured using a glucometer in a drop of blood collected from the nicked tail vein (as described above). At euthanasia blood glucose measurements were performed before animals were euthanized. Insulin and triglycerides were measured in plasma or serum (terminal bleed only) using an Insulin Rodent Chemiluminescence ELISA (cat. no. 80-INSMR from ALPCO) and the L-Type Triglyceride M Assay (Triglyceride M Color A 994-02891, Triglyceride M Color B cat. no. 990-02991 and Multi-Lipid Calibrator cat. no. 464-01601 from FUJIFILM Wako Diagnostics USA Corporation). Non-esterified fatty acid levels (NEFA) were determined in serum using the HR Series NEFA-HR(2) assay (Color Reagent A cat. no. 999-34691, Solvent A cat. no. 995-34791, Color Reagent B cat. no. 991-34891, Solvent B cat. no. 993-35191, NEFA Standard Solution cat. no. 276-76491 from FUJIFILM Wako Diagnostics USA Corporation).

4.5.6 2-[^{18}F] Fluoro-2-Deoxy-D-Glucose Positron Emission Tomography (FDG-PET) Imaging in vivo

Mice fed ad libitum were injected intraperitoneally with 2-[^{18}F] fluoro-2-deoxy-D-glucose (100-200 µCi) between 8-9 AM. After 30 min, during which mice were allowed to move freely in their cages, mice were anesthetized with isoflurane (3% in oxygen/air (50/50, 1 L/min).) and imaged in a Bruker Albira Si PET-CT system (Bruker, Billerica, MA) for 10 min to acquire a single static PET image followed by a CT scan for 5 min to ensure attenuation correction and anatomical co-registration. Anesthesia was maintained during scanning using 1.5-2% isoflurane. The images were automatically reconstructed and registered using the Albira reconstruction tool box as both the PET and CT images are in the same image space. Visualization and quantification of PET/CT data were performed using Pixel wise MODeling (PMOD)'s image View tool (version 3.9, PMOD Technologies Ltd., Zurich, Switzerland). Briefly, the whole-body PET and CT images were read using the View tool and volumes of interest (VOI) drawn on the CT images on regions corresponding to the brown adipose tissue (BAT), quadriceps skeletal muscle (SM) and epididymal white adipose tissue (WAT). The same VOI were then automatically drawn on the PET images. The Contour tool with appropriate thresholding was then applied for accurate delineation of the boundaries on BAT, SM and WAT. Standardized uptake values (SUV) were computed for the VOI normalizing by the injected dose per gram body weight. The SUV computations were performed using PMOD.

4.5.7 Magnetic Resonance Imaging (MRI)

MRI was performed using a 9.4T horizontal bore Biospec AVANCE neo preclinical imaging system equipped with a 116 mm bore gradient insert (Bruker BioSpin GmbH, Germany, maximum gradient strength 660 mT/m). Between 8-10 AM mice fed ad libitum were anaesthetized with 3% isoflurane and secured in an MRI compatible cradle. Anesthesia was maintained during scanning using 1.5-2% isoflurane in oxygen/air (50/50, 1 L/min). The respiration rate and rectal temperature were monitored throughout, body temperature maintained at 37°C (SA Instruments Inc., Stony Brook, NY, USA). A 40-mm inner diameter quadrature radiofrequency coil was used for signal transmission and reception. Scout images were taken to confirm correct positioning and orientation of subsequent scans. A 2-dimensional 28-slice T2-weighted (RARE) 3-point DIXON scan was performed for water-fat separation (1 mm slice thickness, image matrix = 256 x 128, field of view = 50 x 35 mm, repetition time = 1404 ms, echo time = 18 ms, signal averages = 2, RARE-factor = 4). The fat fraction (FF) map was calculated from the separated fat images. Quantification of fat volumes (from the FF maps) were performed using PMOD's VOI and contour (with appropriate thresholding) tools as described above. The volumes were then normalized to the body weight for each mouse.

4.5.8 Isolation and Sequencing of Nature RNA Species

4.5.8.1 Homogenization of VAT, SAT and Liver- Because some of the fat pads, quad muscle and livers were very large and dense, we were concerned about the ability of Trizol reagent to permeate the tissue quickly. We homogenized the tissues at liquid nitrogen (LN2) temperatures using the Cell Crusher Tissue Pulverizer (from CELLCRUSHER, Portland Oregon, USA). Briefly, tissues were kept on dry ice until use, and all devices and tools were cooled in LN2. The whole tissue was inserted into the cell crusher device and crushed, then the powdered tissue was collected in 1.5ml tubes in LN2. Tubes were used in tissue isolation or frozen at -70C until use.

4.5.8.2 RNA Isolation- All tissues were weighed before powdering or digestion, and all tissues were kept on dry ice until digestion. For BAT and small SAT tissues, the whole tissue was added to 1 mL Trizol. For powdered tissues (VAT, liver, quad muscle, large SAT), added ~.3 g (not weighed to keep samples cold) of powdered tissue to 1 mL Trizol. Samples were spun at 4C, 15,000xg, for 15 minutes to remove lipid and cell debris. We then extracted RNA using the RNeasy Micro Kit (Qiagen, Velno, Netherlands), following manufacturer's protocol (Cat# 74004). We digested the DNA species on the Qiagen spin column using the RNAse-free DNAse kit (Qiagen, Velno, Netherlands) following manufacturer's protocol (Cat# 79254). We quantified the isolated RNA using the Qubit with RNA Broad Range assay kit (ThermoFisher Scientific, Waltham, Massachusetts, USA), following manufacturer's protocol (Cat# Q10210). We assayed the RNA quality using the TapeStation RNA Broad Range assay (Aligent, Santa Clara California, USA). RNA species with RNA Integrity Number (RIN) > 7.5 were used in analyses.

4.5.8.3 Sequencing and Quantification- RNA species were sequenced by Psomagen. Raw fastq files will be available on GEO upon the publication of the manuscript. Using fastQC, we assessed the quality of the sequenced reads, and we removed low quality reads (Phred < 28) with trimGalore. We aligned our sequenced reads to the mouse genome GRCm39 using hisat2 and quantified the aligned reads using htseq.

4.5.9 Quantification and Statistical Analyses:

All data are presented as mean ± SEM. All gene expression plots show normalized counts.

4.5.9.1 Analysis of Variance- Gene expression was normalized using R package *DESeq2* with a design matrix Strain*Diet. Genes with zero values in more than 75% of the samples were removed. Statistical analyses were performed using R, ANOVA tests were performed using aov(). [2-way: trait or gene expression ~ Strain*Diet, 3-way: trait ~ Strain*Diet*Time]. P-values were adjusted using FDR correction for the number traits or genes tested.

4.5.9.2 Principal components analysis- Principal components analysis was performed using base R pca() on transcripts per million normalized gene expression data or phenotypic traits. We removed genes with less than 0.1 TPMs in 80% of the samples. For phenotypic data, we removed traits with many zeros, then removed samples with any zeros.

4.5.9.3 Partial Least Squares Regression- Partial least squares regression was performed using R package mixOmics. We used transcripts per million normalized gene expression data for 421 genes in visceral adipose tissue as the X predictor variable. We used phenotypic traits for cohort 2 as the Y response variables. Pls() performs a linear regression to identify the combination of predictor variables, or X-latent variable 1, that explain the axis of most variance within the response variables, the Y-latent variable 1. We performed validation to include the optimal number of latent variables in the model. These axes in X- and Y- latent variables are averaged to create a shared component space. Pls regression creates loadings, or mappings of the predictor and response variables to the latent space, and variates, which map each sample to the latent space. We use a biplot to visualize the predictor and response loadings and the variates in the shared component space, with each variable scaled [-1, 1].

4.5.10 Data and Code Availability

Counts, TPMs and fastq files from the RNA-seq generated in this study will be publicly available as of the date of publication. Code to perform the analyses in this study will be available on https://github.com/jnr3hh/.

Acknowledgements

We thank Stephen C. Scherping, Alice C. Innis, Jie Li, Soumen Paul, Jack Roy. Molecular Imaging Core grant: Imaging data was acquired through the University of Virginia Molecular Imaging Core Laboratory, with NIH S10OD025024 funding for the Bruker 9.4T MRI scanner and through the University of Virginia Molecular Imaging Core Laboratory with National Institutes of Health S10OD021672 funding for the Albira Si trimodal scanner.

4.6 Bibliography

1. Hales, C. M., Carroll, M. D., Fryar, C. D. & Ogden, C. L. Prevalence of Obesity and Severe Obesity Among Adults: United States, 2017-2018. *NCHS Data Brief* 1–8 (2020).
2. Gardner, C. D. *et al.* Popular Dietary Patterns: Alignment With American Heart Association 2021 Dietary Guidance: A Scientific Statement From the American Heart Association. *Circulation* **147**, 1715–1730 (2023).
3. Evert, A. B. *et al.* Nutrition Therapy for Adults With Diabetes or Prediabetes: A Consensus Report. *Diabetes Care* **42**, 731–754 (2019).
4. Guasch-Ferré, M. & Willett, W. C. The Mediterranean diet and health: a comprehensive overview. *J Intern Med* **290**, 549–566 (2021).
5. Estruch, R. & Ros, E. The role of the Mediterranean diet on weight loss and obesity-related diseases. *Rev Endocr Metab Disord* **21**, 315–327 (2020).
6. Kahleova, H. *et al.* Effect of a Low-Fat Vegan Diet on Body Weight, Insulin Sensitivity, Postprandial Metabolism, and Intramyocellular and Hepatocellular Lipid Levels in Overweight Adults: A Randomized Clinical Trial. *JAMA Netw Open* **3**, e2025454 (2020).
7. Termannsen, A.-D. *et al.* Effects of vegan diets on cardiometabolic health: A systematic review and meta-analysis of randomized controlled trials. *Obes Rev* **23**, e13462 (2022).
8. Yokoyama, Y., Levin, S. M. & Barnard, N. D. Association between plant-based diets and plasma lipids: a systematic review and meta-analysis. *Nutr Rev* **75**, 683–698 (2017).
9. Pilis, W., Stec, K., Zych, M. & Pilis, A. Health benefits and risk associated with adopting a vegetarian diet. *Rocz Panstw Zakl Hig* **65**, 9–14 (2014).
10. Gardner, C. D. *et al.* Effect of Low-Fat vs Low-Carbohydrate Diet on 12-Month Weight Loss in Overweight Adults and the Association With Genotype Pattern or Insulin Secretion: The DIETFITS Randomized Clinical Trial. *JAMA* **319**, 667–679 (2018).
11. Paoli, A., Rubini, A., Volek, J. S. & Grimaldi, K. A. Beyond weight loss: a review of the therapeutic uses of very-low-carbohydrate (ketogenic) diets. *Eur J Clin Nutr* **67**, 789–796 (2013).
12. Kosinski, C. & Jornayvaz, F. R. Effects of Ketogenic Diets on Cardiovascular Risk Factors: Evidence from Animal and Human Studies. *Nutrients* **9**, 517 (2017).
13. Li, M. & Yuan, J. Effects of very low-carbohydrate ketogenic diet on lipid metabolism in patients with type II diabetes mellitus: a meta-analysis. *Nutr Hosp* **39**, 916–923 (2022).
14. Rosenbaum, M. *et al.* Glucose and Lipid Homeostasis and Inflammation in Humans Following an Isocaloric Ketogenic Diet. *Obesity (Silver Spring)* **27**, 971–981 (2019).
15. Chiu, S. *et al.* Comparison of the DASH (Dietary Approaches to Stop Hypertension) diet and a higher-fat DASH diet on blood pressure and lipids and lipoproteins: a randomized controlled trial. *Am J Clin Nutr* **103**, 341–347 (2016).
16. Ghaedi, E. *et al.* Effects of a Paleolithic Diet on Cardiovascular Disease Risk Factors: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Adv Nutr* **10**, 634–646 (2019).
17. Adeva-Andany, M. M. *et al.* Effect of diet composition on insulin sensitivity in humans. *Clin Nutr ESPEN* **33**, 29–38 (2019).
18. Martín-Peláez, S., Fito, M. & Castaner, O. Mediterranean Diet Effects on Type 2 Diabetes Prevention, Disease Progression, and Related Mechanisms. A Review. *Nutrients* **12**, 2236 (2020).
19. Sofi, F. *et al.* Low-Calorie Vegetarian Versus Mediterranean Diets for Reducing Body Weight and Improving Cardiovascular Risk Profile: CARDIVEG Study (Cardiovascular Prevention With Vegetarian Diet). *Circulation* **137**, 1103–1113 (2018).
20. Golzarand, M., Mirmiran, P. & Azizi, F. Adherence to the MIND diet and the risk of cardiovascular disease in adults: a cohort study. *Food Funct* **13**, 1651–1658 (2022).
21. Glenn, A. J. *et al.* Relationship Between a Plant-Based Dietary Portfolio and Risk of Cardiovascular Disease: Findings From the Women's Health Initiative Prospective Cohort Study. *J Am Heart Assoc* **10**, e021515 (2021).
22. Kahleova, H. *et al.* Dietary Patterns and Cardiometabolic Outcomes in Diabetes: A Summary of Systematic Reviews and Meta-Analyses. *Nutrients* **11**, 2209 (2019).
23. Luo, Y. *et al.* Isocaloric-restricted Mediterranean Diet and Chinese Diets High or Low in Plants in Adults With Prediabetes. *J Clin Endocrinol Metab* **107**, 2216–2227 (2022).

24. Han, H. Y., Paquet, C., Dubé, L. & Nielsen, D. E. Diet Quality and Food Prices Modify Associations between Genetic Susceptibility to Obesity and Adiposity Outcomes. *Nutrients* **12**, 3349 (2020).
25. Sotos-Prieto, M. *et al.* Mediterranean Diet Adherence Modulates Anthropometric Measures by TCF7L2 Genotypes among Puerto Rican Adults. *J Nutr* **150**, 167–175 (2020).
26. Mattei, J., Qi, Q., Hu, F. B., Sacks, F. M. & Qi, L. TCF7L2 genetic variants modulate the effect of dietary fat intake on changes in body composition during a weight-loss intervention. *Am J Clin Nutr* **96**, 1129–1136 (2012).
27. Bodhini, D. *et al.* Interaction between TCF7L2 polymorphism and dietary fat intake on high density lipoprotein cholesterol. *PLoS One* **12**, e0188382 (2017).
28. Grau, K. *et al.* TCF7L2 rs7903146-macronutrient interaction in obese individuals' responses to a 10-wk randomized hypoenergetic diet. *Am J Clin Nutr* **91**, 472–479 (2010).
29. Gomez-Delgado, F. *et al.* Apolipoprotein E genetic variants interact with Mediterranean diet to modulate postprandial hypertriglyceridemia in coronary heart disease patients: CORDIOPREV study. *Eur J Clin Invest* **49**, e13146 (2019).
30. Campos-Perez, W. *et al.* Physical inactivity and excessive sucrose consumption are associated with higher serum lipids in subjects with Taq1B CETP polymorphism. *J Hum Nutr Diet* **33**, 299–307 (2020).
31. Wuni, R., Kuhnle, G. G. C., Wynn-Jones, A. A. & Vimaleswaran, K. S. A Nutrigenetic Update on CETP Gene-Diet Interactions on Lipid-Related Outcomes. *Curr Atheroscler Rep* **24**, 119–132 (2022).
32. Khatibi, N. *et al.* Interactions between caveolin 1 polymorphism and the Mediterranean and Mediterranean-DASH Intervention for Neurodegenerative Delay diet (MIND) diet on metabolic dyslipidemia in overweight and obese adult women: a cross-sectional study. *BMC Res Notes* **14**, 364 (2021).
33. Abaj, F. *et al.* Interactions between Caveolin-1 (rs3807992) polymorphism and major dietary patterns on cardio-metabolic risk factors among obese and overweight women. *BMC Endocr Disord* **21**, 138 (2021).
34. Senftleber, N. *et al.* Genetic study of the Arctic CPT1A variant suggests that its effect on fatty acid levels is modulated by traditional Inuit diet. *Eur J Hum Genet* **28**, 1592–1601 (2020).
35. Tam, C. H. T. *et al.* Early gene-diet interaction between glucokinase regulatory protein (GCKR) polymorphism, vegetable and fish intakes in modulating triglyceride levels in healthy adolescents. *Nutr Metab Cardiovasc Dis* **25**, 951–958 (2015).
36. Fernandes Silva, L., Vangipurapu, J., Kuulasmaa, T. & Laakso, M. An intronic variant in the GCKR gene is associated with multiple lipids. *Sci Rep* **9**, 10240 (2019).
37. Zhang, X. *et al.* FTO genotype and 2-year change in body composition and fat distribution in response to weight-loss diets: the POUNDS LOST Trial. *Diabetes* **61**, 3005–3011 (2012).
38. Lin, X. *et al.* Neuropeptide Y genotype, central obesity, and abdominal fat distribution: the POUNDS LOST trial. *Am J Clin Nutr* **102**, 514–519 (2015).
39. Corella, D. *et al.* Obese subjects carrying the 11482G>A polymorphism at the perilipin locus are resistant to weight loss after dietary energy restriction. *J Clin Endocrinol Metab* **90**, 5121–5126 (2005).
40. Smith, C. E. *et al.* Perilipin polymorphism interacts with dietary carbohydrates to modulate anthropometric traits in hispanics of Caribbean origin. *J Nutr* **138**, 1852–1858 (2008).
41. Corella, D. *et al.* Perilipin gene variation determines higher susceptibility to insulin resistance in Asian women when consuming a high-saturated fat, low-carbohydrate diet. *Diabetes Care* **29**, 1313–1319 (2006).
42. Heianza, Y. *et al.* Macronutrient Intake-Associated FGF21 Genotype Modifies Effects of Weight-Loss Diets on 2-Year Changes of Central Adiposity and Body Composition: The POUNDS Lost Trial. *Diabetes Care* **39**, 1909–1914 (2016).
43. de Luis, D. A., Primo, D., Izaola, O. & Aller, R. Adiponectin Gene Variant rs266729 Interacts with Different Macronutrient Distribution of Two Different Hypocaloric Diets. *Lifestyle Genom* **13**, 20–27 (2020).
44. Goni, L. *et al.* A circadian rhythm-related MTNR1B genetic variant modulates the effect of weight-loss diets on changes in adiposity and body composition: the POUNDS Lost trial. *Eur J Nutr* **58**, 1381–1389 (2019).

45. Mahmoudinezhad, M. *et al.* The integrative panel of fatty acid desaturase-2 (FADS2) rs174583 gene polymorphism and dietary indices (DQI-I and HEI) affects cardiovascular risk factors among obese individuals. *BMC Endocr Disord* **23**, 41 (2023).
46. Khodarahmi, M., Javidzade, P., Farhangi, M. A., Hashemzehi, A. & Kahroba, H. Interplay between fatty acid desaturase2 (FADS2) rs174583 genetic variant and dietary antioxidant capacity: cardio-metabolic risk factors in obese individuals. *BMC Endocr Disord* **22**, 167 (2022).
47. Huang, T. *et al.* Fish and marine fatty acids intakes, the FADS genotypes and long-term weight gain: a prospective cohort study. *BMJ Open* **9**, e022877 (2019).
48. Huang, T. *et al.* HNF1A variant, energy-reduced diets and insulin resistance improvement during weight loss: The POUNDS Lost trial and DIRECT. *Diabetes Obes Metab* **20**, 1445–1452 (2018).
49. Freitas, R. N. *et al.* A single nucleotide polymorphism in the 3-hydroxy-3-methylglutaryl-coenzyme A reductase gene (HMGCR) influences the serum triacylglycerol relationship with dietary fat and fibre in the European Prospective Investigation into Cancer and Nutrition in Norfolk (EPIC-Norfolk) study. *British Journal of Nutrition* **104**, 765–772 (2010).
50. Corella, D. *et al.* APOA2, dietary fat, and body mass index: replication of a gene-diet interaction in 3 independent populations. *Arch Intern Med* **169**, 1897–1906 (2009).
51. Qi, Q. *et al.* Insulin receptor substrate 1 gene variation modifies insulin resistance response to weight-loss diets in a 2-year randomized trial: the Preventing Overweight Using Novel Dietary Strategies (POUNDS LOST) trial. *Circulation* **124**, 563–571 (2011).
52. Muhammad, H. F. L., Sulistyoningrum, D. C., Huriyati, E., Lee, Y. Y. & Muda, W. A. M. W. The interaction between energy intake, physical activity and UCP2 -866G/A gene variation on weight gain and changes in adiposity: an Indonesian Nutrigenetic Cohort (INDOGENIC). *Br J Nutr* **125**, 611–617 (2021).
53. de Luis, D. A., Izaola, O., Primo, D. & Gómez, J. J. L. Role of beta-2 adrenergic receptor polymorphism (rs1042714) on body weight and glucose metabolism response to a meal-replacement hypocaloric diet. *Nutrition* **116**, 112170 (2023).
54. Francis, M. *et al.* Genome-wide association study of fish oil supplementation on lipid traits in 81,246 individuals reveals new gene-diet interaction loci. *PLoS Genet* **17**, e1009431 (2021).
55. Roa-Díaz, Z. M. *et al.* Gene-diet interactions and cardiovascular diseases: a systematic review of observational and clinical trials. *BMC Cardiovasc Disord* **22**, 377 (2022).
56. Giovanella, J. *et al.* Diet-gene interaction: effects of polymorphisms in the ACE, AGT and BDKRB2 genes and the consumption of sodium, potassium, calcium, and magnesium on blood pressure of normotensive adult individuals. *Mol Cell Biochem* **476**, 1211–1219 (2021).
57. Vitolins, M. Z. & Case, T. L. What Makes Nutrition Research So Difficult to Conduct and Interpret? *Diabetes Spectr* **33**, 113–117 (2020).
58. Lusis, A. J. *et al.* The Hybrid Mouse Diversity Panel: a resource for systems genetics analyses of metabolic and cardiovascular traits. *J Lipid Res* **57**, 925–942 (2016).
59. Griffin, L. E. *et al.* Diet-induced obesity in genetically diverse collaborative cross mouse founder strains reveals diverse phenotype response and amelioration by quercetin treatment in 129S1/SvImJ, PWK/EiJ, CAST/PhJ, and WSB/EiJ mice. *J Nutr Biochem* **87**, 108521 (2021).
60. Yam, P. *et al.* Genetic Background Shapes Phenotypic Response to Diet for Adiposity in the Collaborative Cross. *Front Genet* **11**, 615012 (2020).
61. Bennett, B. J. *et al.* Genetic Architecture of Atherosclerosis in Mice: A Systems Genetics Analysis of Common Inbred Strains. *PLoS Genet* **11**, e1005711 (2015).
62. Parks, B. W. *et al.* Genetic architecture of insulin resistance in the mouse. *Cell Metab* **21**, 334–347 (2015).
63. Nelson, M. E. *et al.* Systems-level analysis of insulin action in mouse strains provides insight into tissue- and pathway-specific interactions that drive insulin resistance. *Cell Metab* **34**, 227-239.e6 (2022).
64. Parks, B. W. *et al.* Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metab* **17**, 141–152 (2013).
65. Börgeson, E., Boucher, J. & Hagberg, C. E. Of mice and men: Pinpointing species differences in adipose tissue biology. *Front Cell Dev Biol* **10**, 1003118 (2022).
66. Barrington, W. T. *et al.* Improving Metabolic Health Through Precision Dietetics in Mice. *Genetics* **208**, 399–417 (2018).

67. Kirby, A. *et al.* Fine mapping in 94 inbred mouse strains using a high-density haplotype resource. *Genetics* **185**, 1081–1095 (2010).
68. Table 056. Hyattsville, MD. 2017. Available from: https://www.cdc.gov/nchs/hus/data-finder.htm.
69. Chen, K. *et al.* HMGB2 orchestrates mitotic clonal expansion by binding to the promoter of C/EBPβ to facilitate adipogenesis. *Cell Death Dis* **12**, 666 (2021).
70. Morinaga, H. *et al.* High-mobility group box 2 protein is essential for the early phase of adipogenesis. *Biochem Biophys Res Commun* **557**, 97–103 (2021).
71. Nishizuka, M., Hayashi, T., Asano, M., Osada, S. & Imagawa, M. KCNK10, a Tandem Pore Domain Potassium Channel, Is a Regulator of Mitotic Clonal Expansion during the Early Stage of Adipocyte Differentiation. *Int J Mol Sci* **15**, 22743–22756 (2014).
72. Sato, S. *et al.* RNA interference-mediated knockdown of the mouse gene encoding potassium channel subfamily K member 10 inhibits hormone-induced differentiation of 3T3-L1 preadipocytes. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **157**, 46–53 (2010).
73. Endo, S., Matsunaga, T. & Nishinaka, T. The Role of AKR1B10 in Physiology and Pathophysiology. *Metabolites* **11**, 332 (2021).
74. Pastel, E., Pointud, J.-C., Volat, F., Martinez, A. & Lefrançois-Martinez, A.-M. Aldo-Keto Reductases 1B in Endocrinology and Metabolism. *Front Pharmacol* **3**, 148 (2012).
75. Hunyenyiwa, T. *et al.* Obesity Inhibits Angiogenesis Through TWIST1-SLIT2 Signaling. *Frontiers in Cell and Developmental Biology* **9**, (2021).
76. Lv, Y., Xiao, J., Liu, J. & Xing, F. E2F8 is a Potential Therapeutic Target for Hepatocellular Carcinoma. *J Cancer* **8**, 1205–1213 (2017).
77. Yang, C.-H. *et al.* E2f8 and Dlg2 genes have independent effects on impaired insulin secretion associated with hyperglycaemia. *Diabetologia* **63**, 1333–1348 (2020).
78. Hu, X.-L., Li, M.-P., Song, P.-Y., Tang, J. & Chen, X.-P. AGXT2: An unnegligible aminotransferase in cardiovascular and urinary systems. *Journal of Molecular and Cellular Cardiology* **113**, 33–38 (2017).
79. Xu, B. *et al.* Bdh1 overexpression ameliorates hepatic injury by activation of Nrf2 in a MAFLD mouse model. *Cell Death Discov.* **8**, 1–10 (2022).
80. Fernández-Real, J. M., McClain, D. & Manco, M. Mechanisms Linking Glucose Homeostasis and Iron Metabolism Toward the Onset and Progression of Type 2 Diabetes. *Diabetes Care* **38**, 2169–2176 (2015).
81. Duffy, S. P. *et al.* The Fowler Syndrome-Associated Protein FLVCR2 Is an Importer of Heme. *Mol Cell Biol* **30**, 5318–5324 (2010).
82. Kenny, T. C. *et al.* Integrative genetic analysis identifies FLVCR1 as a plasma-membrane choline transporter in mammals. *Cell Metab* **35**, 1057-1071.e12 (2023).
83. Gorden, A. *et al.* Genetic Variation at NCAN Locus is Associated with Inflammation and Fibrosis in Non-alcoholic Fatty Liver Disease in Morbid Obesity. *Hum Hered* **75**, 10.1159/000346195 (2013).
84. Gianmoena, K. *et al.* Epigenomic and transcriptional profiling identifies impaired glyoxylate detoxification in NAFLD as a risk factor for hyperoxaluria. *Cell Rep* **36**, 109526 (2021).
85. Barbara, M., Scott, A. & Alkhouri, N. New insights into genetic predisposition and novel therapeutic targets for nonalcoholic fatty liver disease. *Hepatobiliary Surg Nutr* **7**, 372–381 (2018).
86. Mooli, R. G. R. & Ramakrishnan, S. K. Emerging Role of Hepatic Ketogenesis in Fatty Liver Disease. *Frontiers in Physiology* **13**, (2022).
87. Nagashima, S. *et al.* Liver-specific deletion of 3-hydroxy-3-methylglutaryl coenzyme A reductase causes hepatic steatosis and death. *Arterioscler Thromb Vasc Biol* **32**, 1824–1831 (2012).
88. Voruganti, V. S. Precision Nutrition: Recent Advances in Obesity. *Physiology (Bethesda)* **38**, 0 (2023).

Chapter 5: Discussion-

5.1 Systems Biology Approaches are characterize fat storage and distribution

Biological processes are carefully regulated by the complex interactions between many individual parts[1,2]. We identify conserved, organized systems at all scales of biology[3,4] (Figure 5.1 A). In diseases of overnutrition, the brain and adipose tissue work together to control energy storage and expenditure[5,6]. Adipose tissue stores fat and releases signaling molecules[7–9] accomplished by adipocytes and other cell types[10]. Adipocyte differentiation and fatty acid uptake contribute to lipid accumulation, while lipolysis and thermogenesis reduce fat storage[9]. Key molecular regulators, like PPARG[11] orchestrate molecular signaling that drives these cellular phenotypes[12]. While we are limited by models[13–15] considering these processes in isolation is reductive. The full body phenotype fat distribution result from linear and non-linear effects between expressed genes, pathways, and cell phenotypes that result in asymmetrical lipid storage between two adipose tissue depots. To fully characterize their effects in context, we can employ a systems biology approach.

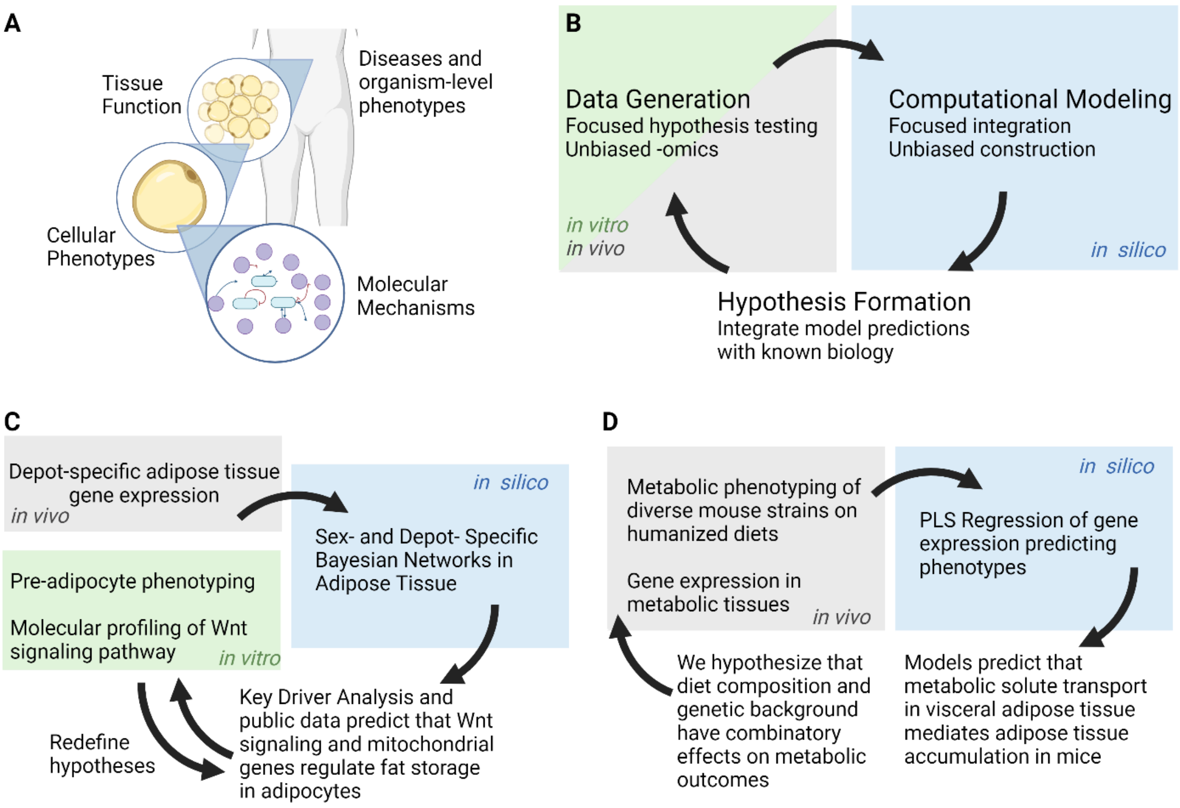

Figure 5.1: Applications of systems biology in adipose tissue health and disease

(D) Biological systems control health and disease at many scales, from the organism-level to the molecular level.

(E) Systems biology employs multiple methodologies to iteratively generate and test data-driven hypotheses.

(F) In Chapter 2, we use Bayesian networks to model adipose tissue gene regulation in silico. This enables us to predict key driver genes, and in Chapter 3, we integrate public data to prioritize key drivers in relevant pathways. We use gain and loss of function experiments in (pre)-adipocytes to show that key driver genes impact adipogenesis and mitochondrial function. We align these with human phenotype data to refine predictions.

(G) In Chapter 4, we investigate the effect of genetic background and diet composition on metabolic outcomes and on gene expression in metabolic tissues. We use the gene expression to predict the observed phenotypes using partial least squares regression (PLSR) models, which leads to the generation of new hypotheses.

Figure Created using BioRender.

Systems biology is a field in which we attempt to understand these systems of natural phenomena through experimental and computational techniques[16–18]. In these studies, we used computational techniques to model the interactions between the components of large biological systems, then used these models to make predictions about the underlying biology (Figure 5.1 B). We generated experimental data to validate these predictions. We used unbiased methods to generate models, then focused our hypotheses using public data and previous literature[19].

In Chapters 2 and 3, we use Bayesian networks to model the gene regulation in adipose tissue (Figure 5.1 C). We compare sex- and depot-specific networks to identify unique putative regulators of fat storage. Using publicly available data, we prioritize a subset of network regulators that likely act in adipocytes. Finally, we test the role of each gene in fat storage, mitochondrial function, and Wnt signaling. We use the results of these experiments to refine our hypotheses about predicted gene function and about the predictive power of the networks.

In Chapter 4, we start by generating phenotypic and gene expression data to understand the effects of diet composition and genetic background on metabolism at multiple scales (Figure 5.1 D). We use partial least squares (PLSR) regression to model the predicted effects of gene expression in metabolic tissues on fat storage and circulating metabolites. We use these models to make predictions about the genes and pathways that regulate these phenotypes.

These studies integrate large and small data at multiple biological scales. We observed *in vivo* whole body phenotypes in mouse, and used human *in vivo* expression data to build causal gene regulatory networks. We conducted *in vitro* cellular phenotyping and characterized molecular pathways. We built *in silico* models of gene-gene regulation in human data and modeled the gene-phenotype associations in mouse data. By using the framework of systems biology, we generate and made meaningful conclusions from big data.

5.2 Multi-scale mechanisms of fat distribution

By taking a holistic, systems approach, we recapitulated known mechanisms of disease biology and nominated novel regulators. Throughout our studies, we confirmed the dominating effect of male visceral adipose tissue that contributes to the harmful effects of excess adiposity[9]. In response to diet challenge, mouse visceral adipose tissue (VAT) showed a large strain and diet specific response in gene expression, which could explain part of the strain and diet specific variation seen in metabolic phenotypes. *in silico* Bayesian network analyses nominated 119 key drivers of male VAT gene expression. Interestingly, these contained over 30 genes primarily expressed in immune cells. We also identified that two complement 1 genes were responsive to strain and diet in mouse VAT, confirming that VAT is capable of robust immune response[7,20]. We showed, in subcutaneous cells, that four key drivers of visceral networks affected adipocyte function, identifying a novel role for these genes.

We also observed a dominating effect of genetic background on metabolic outcomes. Certainly, we know that obesity and fat distribution are highly heritable[21,22] and are caused by the combined small effects of many genetic variants[21,23,24]. We observed that genetic background explained many of the observed differences in metabolic phenotypes measured in mice, and had a strong effect on gene expression in metabolic disease. Conversely, the effects of non-linear interactions between genetic background and diet composition were apparent in whole body phenotypes, but not in organ level phenotypes or in gene expression, with the exception of VAT gene expression. These differences may be explained by assaying other tissues or time-points, as discussed in Chapter 4.

We also know that sex-differences in $\mathrm{WHR}_{\mathrm{adjBMI}}$ and in cardio-metabolic disease outcomes are thought to be driven by differences in visceral adipose tissue between sexes[25,26]. $\mathrm{WHR}_{\mathrm{adjBMI}}$ has different heritability between men and women[27] and different genetic architecture at almost a third of the GWAS loci[24,28]. Though our *in vitro* and *in* vivo studies were limited by using cells and mice of one sex, our network models predict sex-specific regulation of adipose tissue gene expression. Key driver analysis and prioritization identified putative key driver genes that were largely unique to one sex or depot. Interestingly, most validated key drivers were predicted to regulate female networks.

At the cell and molecular scale, we observed mechanisms of both hyperplastic and hypertrophic adipocyte expansion[9]. The humanized diets used in Chapter 4 contained more fat content than chow, and were obesogenic on some genetic backgrounds, likely inducing hyperplastic adipose expansion. Within the strain and diet responsive genes identified in mouse VAT, we find enrichment for fatty acid and sterol transport and metabolism, and the upregulation of cholesterol releasing ATP-binding cassette genes. This may indicate increased lipolysis in the tissue, which may explain lower fat storage seen in these mice.

Our studies in Chapter 3 identified genes that effected adipogenesis and hypertrophic adipocyte expansion, *ANAPC2, PSME3* and *RSPO1*. These genes were predicted to regulate female networks, and two were predicted in female visceral networks specifically. This could indicate that hypertrophic expansion is driven more by putative female genes; mechanistic studies of adipocyte hyperplastic processes might validate more putative male driver genes. Both visceral female network key driver genes, *PSME3* and *RSPO1*, inhibit non-canonical Wnt signaling, favoring canonical pathways. Because non-canonical Wnt signaling has disease implications[29,30], these genes may act in female visceral adipocyte to exert protective, anti-inflammatory effects.

Following the systems biology framework, we attempted to align our findings with other human data and refine our hypotheses. *RSPO1*'s effect on fat storage in adipocytes, and correlation with $\mathrm{WHR}_{\mathrm{adjBMI}}$ could explain its effect on $\mathrm{WHR}_{\mathrm{adjBMI}}$; we hypothesize that *RSPO1* expression inhibits storage in both depots and may have different magnitude of effect in each depot that would create fat distribution differences. We also find associations between genetic variants in an *ANAPC2* intron and $\mathrm{WHR}_{\mathrm{adjBMI}}$; the same variant causes excision of the intron from *ANAPC2* transcripts.

We hypothesize that the alternate transcript may have a role in WHR_{adjBMI} regulation or in regulating expression of the main transcript. In Chapter 4, we identify three genes that respond to diet and genetic background, and previous studies show that genetic variation in these genes in humans modify diet effects on metabolic traits. We make predictions that genes with similar patterns of expression may function in the same pathways.

5.2 Limitations

5.2.1 Cell and mouse model limitations

While studies that model human body fat distribution in cells and mice have large predictive and descriptive power[31], they are limited by the choice of models used[13]. Our cell studies of adipogenesis in subcutaneous cells are only partially predictive of WHR_{adjBMI}; future experimentation in primary cells of both depots would fully characterize these effects. Further, sex differences in WHR_{adjBMI} presentation and genetic regulation necessitate[26,28,32] modeling and testing these predictions in a sex-specific manner. Our focus on primarily male cells and mice neglects female disease and protection mechanisms.

Because of the challenges working with primary cells, summarized in Chapter 3, immortalizing the cells using hTERT and SV40 could help them retain adipogenic potential at higher passages[13,33,34]. Adipocytes are spherical cells that normally grow in 3-D space with other cell types. Models of 3-D culture and co-culture with other cell types can improve the application of cell model predictions to organism health[35,36].

5.2.1 Limitations of computational models

Computational models of biological processes can be incredible tools for describing and predicting natural phenomena[1,2]; however, model predictions are that- predictions- and must be biologically validated[16,31]. Models are by necessity approximations of a more complex process and make assumptions that sometimes don’t represent biology. We find that discretization, an approximation of gene expression, has a large effect on the Bayesian network structure. Even completely unbiased model construction from observed data is skewed by observed and underlying characteristics of the samples used to build the model, as well as noise introduced during sample isolation and analyte quantification. Large data in genomics research is dominated by European ancestry subjects, which hinders the application of their conclusions in other populations.

5.3 Validated, functional key driver genes are robust to sub-sampling

Reproducibility of data is essential in biology, and often, networks can be descriptive of the specific population used to generate the model[42]. We find that subsampled networks partially recapitulate the key driver predictions of the full networks and that STARNET and GTEx share about 10% of key driver predictions; therefore, a large subset of key drivers aren't robust or meaningful.

However, literature studies summarized in Supplemental Bibliography 3.1 show that 45 key driver genes already have a known role in adipocyte biology. While these point to the predictive power of the Bayesian network approach, literature searches are biased by well-studied genes and pathways[43], e.g. *CTNN1B* is one of the most cited genes on PubMed. To assess the power of networks in predicting key driver genes of unknown function, we used our unpublished negative results. Studies in Chapter 3 validated *ANAPC2, MIGA1, PSME3, RSPO1*, and *UBR1*, and found negative results for *TYRO3* and *C1QTNF3*. Using previous iterations of the prioritization pipeline, we identified key driver genes *MCC, MTARC1, ANTXR1, AKAP9, ARAP1, LDHD*, and *QTRT1* as genes of interest (Table 5.1). When tested in cells, we found that these genes did not affect the tested phenotype, either adipogenesis or mitochondrial function (not shown).

Table 5.1: Characteristics of validated and non-functional key driver genes

Key Driver Gene	Network		Public Data		Network Predictions		Validation
	Depot	Sex	Cell Type	Evidence	Replication in both datasets	Reproduction in 90% subsampled networks, both datasets	Functional
MIGA1	Visceral	Male	Adipocyte	Mouse	Replicated	Neither	Functional
ANAPC2	Subcutaneous	Both	Adipocyte	GWAS	Replicated	Reproduced in both male subq.	Functional
PSME3	Visceral	Female	Adipocyte	GWAS	Replicated	Reproduced	Functional
RSPO1	Visceral	Female	Multiple	Mouse	Replicated	Reproduced	Functional
UBR1	Visceral	Female	Multiple	Mouse	Replicated	Reproduced	Functional
TYRO3	Subcutaneous	Male	Unknown	GWAS	Replicated	GTEx only	X
QTRT1	Subcutaneous	Female	Adipocyte	GWAS	Replicated	GTEx only	X
ARAP1	Subcutaneous	Female	Adipocyte	Both	Replicated	STARNET only	X
AKAP9	Visceral	Male	Adipocyte	Mouse	Replicated	STARNET only	X
ANTXR1	Visceral	Male	Multiple	GWAS	Replicated	Neither	X
LDHD	Visceral	Male	Unknown	GWAS	Replicated	GTEx only	X
C1QTNF3	Visceral	Female	Adipocyte	GWAS	Replicated	GTEx only	X
MCC	Visceral	Both	Multiple	GWAS	STARNET only	STARNET only	X
MTARC1	Both	Both	Adipocyte	GWAS	GTEx only	Reproduced in both male visc.	X

We find that positive results in functional studies are frequently predicted by visceral and by female networks, while negative results are found mostly by visceral networks of both depots. We tested few key drivers predicted by subcutaneous networks. Many of the tested genes were prioritized for their location in significant GWAS loci, only validated key driver genes are found outside genetic loci as well. Genes that were strong key drivers in multiple networks of one dataset but were not replicated in the other (*MCC, MTARC1*) did not have a role in adipogenesis. This may indicate over-fitting of the model, and highlights the need to use multiple cohorts. Sub-sampling may be the strongest indication of functional validation in cells. When considering networks constructed using 90% of the original samples, we identify 4/5 validated genes as key drivers of the sub-sampled network. Using a predictive pipeline that considers whether a replicated key driver gene is robust to subsampling in both networks, we would make only one false negative prediction, *MIGA1*. While the 90% networks have significant predictive power, the

50% and 10% sub-sampled networks have only some (Table 5.2). *RSPO1* and *PSME3* were identified in 50% subsampled networks, while *MIGA1* was reproduced in the 10%. In GTEx female visceral networks (n = 3), predictions were largely conserved between iterations (Table 5.2).

Table 5.2: Tested key driver reproduction by percent sub-sampling

			10% Subsampled		**50% Subsampled**		**90% Subsampled**	
Key Driver	Depot	Sex	**STARNET**	**GTEx**	**STARNET**	**GTEx**	**STARNET**	**GTEx**
Female Subcutaneous								
MTARC1	*Both*	*Both*	0/1	1/1	0/1	1/1	0/1	1/1
ANAPC2	Subcutaneous	Both	1/1	0/1	0/1	0/1	1/1	0/1
ARAP1	*Subcutaneous*	*Female*	0/1	1/1	0/1	0/1	1/1	0/1
QTRT1	*Subcutaneous*	*Female*	0/1	0/1	0/1	0/1	0/1	1/1
Male Subcutaneous								
MTARC1	*Both*	*Both*	0/1	1/1	0/1	1/1	0/1	1/1
ANAPC2	Subcutaneous	Both	0/1	0/1	0/1	0/1	1/1	1/1
TYRO3	*Subcutaneous*	*Male*	0/1	0/1	0/1	1/1	0/1	1/1
Female Visceral								
MTARC1	*Both*	*Both*	1/1	2/3	0/1	3/3	0/1	3/3
MCC	*Visceral*	*Both*	0/1	0/3	0/1	0/3	1/1	0/3
C1QTNF3	*Visceral*	*Female*	0/1	0/3	0/1	2/3	0/1	3/3
PSME3	Visceral	Female	0/1	2/3	1/1	2/3	1/1	1/3
RSPO1	Visceral	Female	0/1	1/3	1/1	3/3	1/1	3/3
UBR1	Visceral	Female	0/1	0/3	0/1	1/3	1/1	2/3
Male Visceral								
MTARC1	*Both*	*Both*	0/1	1/1	0/1	1/1	1/1	1/1
MIGA1	Visceral	Male	1/1	1/1	0/1	0/1	0/1	0/1
AKAP9	*Visceral*	*Male*	1/1	0/1	1/1	0/1	1/1	0/1
ANTXR1	*Visceral*	*Male*	0/1	0/1	0/1	0/1	0/1	1/1
LDHD	*Visceral*	*Male*	0/1	1/1	0/1	1/1	0/1	1/1

Using the systems biology framework, we can update our models to include some metric of robustness to sub-sampling in the key driver identification or prioritization pipeline. Since we know that some key driver predictions and therefore some network connections are false, we may be able to identify features of the original network that would indicate non-robust predictions. Simultaneously, we should pinpoint sub-sampling parameters that optimize predictive power and reduce the associated computation time. Interestingly, the network construction method already uses 1,000 random-seed iterations and a score penalty-term to avoid over-fitting the data. We found that the networks were likely still overfit, and only made accurate predictions only when comparing two datasets and sub-sampling.

5.4 Future Directions

Based on the results in table 5.1, we conclude that both robustness to subsampling and replication across datasets is necessary for a key driver to be validated in cells. We will incorporate both of these requirements when performing experiments based on Bayesian network predictions. Because of the time required to construct networks, it might be useful to find the optimal subsampling threshold. We could randomly subsample the network to 50, 55, 60 ,65, 70, 75, 80, 85, 90, and 95% of original samples, repeating this random subsampling 3 times per condition. Then, as we did before, we would determine the key drivers replicated in both the STARNET and the GTEx networks. We would continue to use the subsampling percentage that predicts the most true positive key driver genes without falsely predicting the non-functional genes. So far, we have performed experimental validation on 14 key driver genes. While this is a very small sample size that was not chosen randomly, we may be able to make rudimentary precision-recall curves to mathematically determine the “best” subsampling percentage.

In chapter 3, we show that five of the predicted key driver genes affect the function of pre-adipocytes or adipocytes *in vitro.* To make conclusions about overall fat distribution, we must show that there is a difference in fat accumulation between adipocytes derived from subcutaneous and visceral adipose tissue. Since there is high variability between individuals, we would ideally obtain primary subcutaneous and visceral pre-adipocytes from 3-6 donors each. We would immortalize the cells and then characterize their baseline proliferation and adipogenic rates. Then, we could introduce extra copies of the gene of interest and assess the difference between controls and overexpressing cells. Much like in chapter 4, we would use ANOVA to test if there is an interaction effect between the gene of interest and the depot of origin. For example, if the gene of interest promotes adipogenesis more strongly in adipocytes from visceral fat, we could conclude that that gene promotes abdominal obesity. These studies would also allow us to determine if *MIGA1* and *UBR1* cause changes in lipid accumulation in either depot, in addition to their effects on cellular respiration. Ideally, we would obtain these primary cells from at least 3 males and 3 females. Using 3-way ANOVA, we could then determine the effects of the variables sex, depot, and gene of interest. Based on other mechanistic studies and GWAS signals, we predict that *RSPO1* and *PSME*3 may have female specific effects, and we would expect to see larger magnitude changes in adipogenesis in female cells than in male cells.

In chapter 4, we observed large changes in gene expression in mouse visceral adipose tissue due to the interaction effects of genetic background and diet composition. However, due to our small sample size, we were only able to identify large magnitude changes as significant. Further, all 4 mice in strain/diet group (n = 2 used for RNA-seq) were housed in the same cage besides SJL, who were singly-housed. In future studies, we would focus only on the visceral adipose tissue, ideally sequencing RNA from ≥6 mice. In this way, we would be powered to detect more subtle differences, and we could better account for the batch effects resulting from cage assignments. We might even consider replacing the SJL mice with a different strain that can be co-housed, to reduce noise and strain differences. We would ideally perform this experiment using a sex-balanced cohort, with ≥2 cages of mice in each sex/strain/diet group. However, since we know that male visceral adipose is more active than female visceral adipose, we might consider a smaller pilot experiment to first determine if there are detectable differences in gene expression in female visceral adipose between strain and diet groups.

5.5 Conclusions

Our studies represent significant contributions to our understanding of the genetic basis of abdominal obesity. Using systems biology, we investigated the causes of these diseases using multiple data types and methodologies. Networks and other models integrated the large amount of unknowns into coherent structures of gene regulation. We used these models to inform mechanistic studies and draw conclusions. Further advances in this field will require dedicated research in the computational analysis of big data and in the experimental validation of predicted disease drivers using cell and mouse models; more than ever, we need interdisciplinary researchers who can apply these tools in biologically meaningful ways.

5.6 Bibliography:

1. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**, 101–113 (2004).
2. Dančík, V., Basu, A. & Clemons, P. Properties of Biological Networks. in *Systems Biology: Integrative Biology and Simulation Tools* (eds. Prokop, A. & Csukás, B.) 129–178 (Springer Netherlands, 2013). doi:10.1007/978-94-007-6803-1_5.
3. Dada, J. O. & Mendes, P. Multi-scale modelling and simulation in systems biology. *Integrative Biology* **3**, 86–96 (2011).
4. Montagud, A., Ponce-de-Leon, M. & Valencia, A. Systems biology at the giga-scale: Large multiscale models of complex, heterogeneous multicellular systems. *Current Opinion in Systems Biology* **28**, 100385 (2021).
5. Massa, M. G. & Correa, S. M. Sexes on the brain: Sex as multiple biological variables in the neuronal control of feeding. *Biochim Biophys Acta Mol Basis Dis* **1866**, 165840 (2020).
6. Park, H.-K. & Ahima, R. S. Leptin signaling. *F1000Prime Rep* **6**, 73 (2014).
7. Kawai, T., Autieri, M. V. & Scalia, R. Adipose tissue inflammation and metabolic dysfunction in obesity. *Am J Physiol Cell Physiol* **320**, C375–C391 (2021).
8. Kwon, H. & Pessin, J. E. Adipokines mediate inflammation and insulin resistance. *Front Endocrinol (Lausanne)* **4**, 71 (2013).
9. Haczeyni, F., Bell-Anderson, K. S. & Farrell, G. C. Causes and mechanisms of adipocyte enlargement and adipose expansion. *Obes Rev* **19**, 406–420 (2018).
10. Emont, M. P. *et al.* A single-cell atlas of human and mouse white adipose tissue. *Nature* **603**, 926–933 (2022).
11. Lefterova, M. I., Haakonsson, A. K., Lazar, M. A. & Mandrup, S. PPARγ and the global map of adipogenesis and beyond. *Trends Endocrinol Metab* **25**, 293–302 (2014).
12. Wen, X. *et al.* Signaling pathways in obesity: mechanisms and therapeutic interventions. *Signal Transduct Target Ther* **7**, 298 (2022).
13. Ruiz-Ojeda, F. J., Rupérez, A. I., Gomez-Llorente, C., Gil, A. & Aguilera, C. M. Cell Models and Their Application for Studying Adipogenic Differentiation in Relation to Obesity: A Review. *Int J Mol Sci* **17**, 1040 (2016).
14. Doulberis, M. *et al.* Rodent models of obesity. *Minerva Endocrinol* **45**, 243–263 (2020).
15. Börgeson, E., Boucher, J. & Hagberg, C. E. Of mice and men: Pinpointing species differences in adipose tissue biology. *Front Cell Dev Biol* **10**, 1003118 (2022).
16. Civelek, M. & Lusis, A. J. Systems genetics approaches to understand complex traits. *Nat Rev Genet* **15**, 34–48 (2014).
17. Nielsen, J. Systems Biology of Metabolism. *Annu Rev Biochem* **86**, 245–275 (2017).
18. Ma'ayan, A. Complex systems biology. *J R Soc Interface* **14**, 20170391 (2017).
19. Papin, J. A., Reed, J. L. & Palsson, B. O. Hierarchical thinking in network biology: the unbiased modularization of biochemical networks. *Trends Biochem Sci* **29**, 641–647 (2004).
20. Alvehus, M., Burén, J., Sjöström, M., Goedecke, J. & Olsson, T. The human visceral fat depot has a unique inflammatory profile. *Obesity (Silver Spring)* **18**, 879–883 (2010).
21. Bouchard, C. Genetics of Obesity: What We Have Learned Over Decades of Research. *Obesity (Silver Spring)* **29**, 802–820 (2021).
22. Schleinitz, D., Böttcher, Y., Blüher, M. & Kovacs, P. The genetics of fat distribution. *Diabetologia* **57**, 1276–1286 (2014).
23. Lappalainen, T. & MacArthur, D. G. From variant to function in human disease genetics. *Science* **373**, 1464–1468 (2021).
24. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* **28**, 166–174 (2019).
25. Link, J. C. & Reue, K. Genetic Basis for Sex Differences in Obesity and Lipid Metabolism. *Annu Rev Nutr* **37**, 225–245 (2017).
26. Hansen, G. T. *et al.* Genetics of sexually dimorphic adipose distribution in humans. *Nat Genet* **55**, 461–470 (2023).
27. Zhu, C. *et al.* Amplification is the primary mode of gene-by-sex interaction in complex human traits. *Cell Genom* **3**, 100297 (2023).

28. Pulit, S. L., Karaderi, T. & Lindgren, C. M. Sexual dimorphisms in genetic loci linked to body fat distribution. *Biosci Rep* **37**, BSR20160184 (2017).
29. Akoumianakis, I., Polkinghorne, M. & Antoniades, C. Non-canonical WNT signalling in cardiovascular disease: mechanisms and therapeutic implications. *Nat Rev Cardiol* **19**, 783–797 (2022).
30. Ackers, I. & Malgor, R. Interrelationship of canonical and non-canonical Wnt signalling pathways in chronic metabolic diseases. *Diab Vasc Dis Res* **15**, 3–13 (2018).
31. Sieberts, S. K. & Schadt, E. E. Moving toward a system genetics view of disease. *Mamm Genome* **18**, 389–401 (2007).
32. Winkler, T. W. *et al.* The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet* **11**, e1005378 (2015).
33. Darimont, C. & Macé, K. Immortalization of human preadipocytes. *Biochimie* **85**, 1231–1233 (2003).
34. Wang, L. *et al.* hTERT gene immortalized human adipose-derived stem cells and its multiple differentiations: a preliminary investigation. *Appl Biochem Biotechnol* **169**, 1546–1556 (2013).
35. Bellas, E., Marra, K. G. & Kaplan, D. L. Sustainable three-dimensional tissue model of human adipose tissue. *Tissue Eng Part C Methods* **19**, 745–754 (2013).
36. Choi, J. H., Bellas, E., Gimble, J. M., Vunjak-Novakovic, G. & Kaplan, D. L. Lipolytic function of adipocyte/endothelial cocultures. *Tissue Eng Part A* **17**, 1437–1444 (2011).
37. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from dbGaP Accession phs000424.v8.p2on 10/01/2020.
38. Franzén, O. *et al.* Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**, 827–830 (2016).
39. Friedman, N., Linial, M., Nachman, I. & Pe'er, D. Using Bayesian networks to analyze expression data. *J Comput Biol* **7**, 601–620 (2000).
40. Malone, J. H. & Oliver, B. Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biology* **9**, 34 (2011).
41. Xu, X. *et al.* Parallel comparison of Illumina RNA-Seq and Affymetrix microarray platforms on transcriptomic profiles generated from 5-aza-deoxy-cytidine treated HT-29 colon cancer cells and simulated datasets. *BMC Bioinformatics* **14 Suppl 9**, S1 (2013).
42. Cohain, A. *et al.* Exploring the reproducibility of probabilistic causal molecular network models. in *Biocomputing 2017* 120–131 (WORLD SCIENTIFIC, 2016). doi:10.1142/9789813207813_0013.
43. Rocha, J. J. *et al.* Functional unknomics: Systematic screening of conserved genes of unknown function. *PLOS Biology* **21**, e3002222 (2023).