THE UNIVERSITY OF CHICAGO

CHARACTERIZING ALTERNATIVE SPLICING IN ADIPOSE TISSUE FUNCTION AND METABOLIC DISEASE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF HUMAN GENETICS

BY

KATHRYN MARIE FARRIS

CHICAGO, ILLINOIS

MARCH 2024

Table of Contents

Some supplementary tables are available as files online. In those cases, the List of Tables gives the page number for each table's caption.

List of Figures

List of Tables

Acknowledgments

Thank you to my advisor, Dr. Marcelo Nobrega, without whom none of this work would be possible. Thank you for your guidance and support throughout my time in graduate school, and for sticking with me through multiple iterations of these projects. You have always fostered a collaborative, honest approach towards science in the lab, and I am grateful to have had the opportunity to train in such a positive lab environment. I also want to thank our amazing collaborators in the GECKO consortium, especially Dr. Angela Crean, Dr. Alistair Senior, Dr. Stephen Simspon, and Dr. Romain Barrès, all of whom were instrumental to the formation and completion of this work.

Thank you to my committee members, Dr. Xin He, Dr. Carole Ober, and Dr. Francois Spitz, for their ongoing feedback and advice throughout my PhD.

Thank you to all current and past members of the Nobrega lab. From my first day to the last, everyone in the lab has supported my research and my growth as a scientist. I want to especially thank Dr. Débora Sobreira and Dr. Amelia Joslin, who taught me everything I know about molecular biology. I'd also like to thank Dr. Isabella Salamone, Alexis Thornburg, and Zach Weber for many enriching post-lab meeting discussions.

To Sue Levison, thank you for all of your unwavering support and help through every stage of my graduate education.

I'd also like to thank my fellow graduate students and friends, who enriched my life outside of the lab and made graduate school so much more fun than it would have been without them. Thank you to Dr. Katie Aracena, Dr. Selene Clay, Dr. Sammy Keyport Kik, and Dr. Grace Hansen for listening to me complain and making me laugh. Thank you to the Bachelors of Science, Katie, Selene, Shreya, JoJo, Jennifer, Maggie, and Renée, for being a breath of fresh air

every week and for nurturing my love of reality TV. Thank you to Juju, Jacqui, Yuge, and Dr. Lily for ten (or nearly ten) years of friendship, love, and laughter.

Thank you to my parents, for their love and support throughout my PhD and my life. Mom, Dad, Ashley, Karen – you have all always supported my education and have helped shape me into the person I am today. Thank you for that, and for all the support I know you will offer me in this next chapter of my life as well.

Finally, thank you to my partner and spouse, Allen. You have been endlessly supportive of me throughout this process, and I can't imagine doing it without you. Coming home to you is always my favorite part of every day. Thank you for everything. I am eternally grateful for our little family, and would be remiss if I didn't also thank Boba and Bowser for their unwavering support and commitment to my success.

Abstract

Obesity is a complex disease, with both environmental and genetic causes, and it confers a significant global health burden while remaining difficult to treat and prevent. A better understanding of the risk factors that lead to disease and the underlying regulatory responses of specific disease states can provide insight into possible new treatments and interventions to improve health outcomes worldwide. Here, I provide insight into both environmental and genetic causes for obesity through two parallel studies. First, I investigate one of the main environmental factors leading to obesity: diet. In particular, I dissect the impact of differences in dietary macronutrient composition on metabolic measures and gene regulation in adipose tissue, measuring both gene expression and splicing changes. I identify thousands of genes and exons that are responsive to dietary macronutrient composition in adipose tissue, and link them to specific macronutrient patterns and cellular functions. One particularly strong gene regulatory response is the differential expression of genes associated with Bardet-Biedl syndrome in response to dietary fat content. In my second study, I expand our understanding of the contribution of genetics to obesity through assaying alternative splicing across the differentiation of preadipocytes isolated from lean, obese, and obese with type 2 diabetes (T2D) individuals. I find that splicing is highly dynamic across adipocyte differentiation and is impacted by metabolic phenotype. I also find that there is significant enrichment for an overlap between regions that are differentially spliced across adipocyte differentiation and variants that are associated with T2D. In both studies, I find that there is very little overlap between genes that are differentially expressed in response to the perturbation of interest and those that are differentially spliced. These results suggest that alternative splicing and expression may represent largely separate modes of gene regulation, and that studies that seek to describe gene regulatory

responses to stimuli should strive to measure alternative splicing in addition to gene expression to capture a more complete picture of the gene regulatory change. Overall, these studies provide insight into adipose tissue function and both environmental and genetic risk for obesity, and can serve as a resource to guide future variant-to-function studies.

Chapter I: Introduction

The etiology of obesity

Obesity is a complex disease that presents a major public health challenge. In the past four decades, the prevalence of obesity has risen substantially across the world, with more than a third of adults now being classified as overweight or obese.[1] The rise in the prevalence of obesity confers a significant health burden, as obesity is associated with increased mortality and a wide range of co-morbidities, including cardiovascular disease, diabetes, musculoskeletal dysfunction, and some cancers.[2–5] The health risk of obesity is further confounded by the fact that it remains difficult to prevent and treat.[1,6,7] There are both environmental and genetic contributions to obesity, and a better understanding of the precise causes and consequences of obesity can provide insight into the etiology of the disease and provide avenues for possible treatment and prevention.

The main environmental factors that have been associated with obesity risk are energy expenditure and energy intake, with an imbalance between the two leading to weight gain.[8–10] However, there has also been considerable evidence to suggest that it is not just how much we eat that matters, but also what and when we eat. For example, consumption of high amounts of ultra-processed food has been associated with increased risk for obesity, possibly due to a decrease in diet quality.[11–13] Other research has suggested that intermittent fasting can improve cardiometabolic outcomes and increase weight loss when carried out on its own or in combination with other dietary treatments.[14–16] The impact of environmental factors, including diet, on obesity is therefore multifactorial and complex, with many different aspects of an individual's lifestyle contributing to obesity risk and outcomes.

In addition to environmental risk factors, there are also significant genetic contributions to obesity. Heritability estimates of obesity from family studies cluster around 40-50%,[17,18] suggesting that a large proportion of variation in body weight is due to genetic factors. In recent years, genome-wide association studies (GWAS) for body mass index (BMI) and waist-to-hip ratio (WHR) have identified thousands of single nucleotide polymorphisms (SNPs) across hundreds of loci that are associated with obesity risk.[19,20] These GWAS provide a rich body of evidence for the role of genetics in obesity, but stop short of providing clear mechanistic links between specific genes and disease risk or development.

Genetic contributions to obesity: from GWAS to function

In the past two decades, an ever-growing repertoire of GWAS has contributed to our understanding of the genetic underpinnings of human disease. The result of these studies is tens of thousands of genetic variants that have been associated with thousands of traits in millions of individuals,[21,22] including obesity-related traits such as BMI, WHR, and T2D.[19,20,23] As larger and better-powered GWAS have identified more and more variants associated with disease, the question becomes how to link these numerous variants to specific genes and functions to provide insight into disease etiology and possible therapeutic targets.

There are a number of key intermediate questions that must be answered to definitively link a GWAS result to a specific gene or mechanism. Important considerations include which variant in the locus is the causal variant (or variants), what gene (or genes) that variant is acting on, what the impact of the variant on the gene is, and what tissue (or tissues) and developmental time point or cell state the gene-variant interaction is relevant in.[24–26] In the post-GWAS era, researchers in the field have developed a number of tools and assays to answer these questions. For example, fine-mapping algorithms can help identify the likely causal variants in a locus,[27,28]

often aided by annotation data.[29,30] Assays such as promoter-capture HiC can provide insight into what genes a variant may be regulation,[31–33] and a growing collection of datasets across tissues and developmental time points can help identify the tissue and cell state of interest.

By combining many of these tools and datasets with functional validation, it is possible to precisely link a GWAS signal to a specific gene or genes driving disease risk. One example of this is the dissection of the obesity-associated *FTO* locus. The *FTO* locus is so named because the association signal falls within an intron of the gene *FTO*, and initially it was thought that *FTO* itself was the gene driving the association.[34,35] However, subsequent work dissecting the locus has revealed a more complex story. Annotation data and function characterization identified the locus as an enhancer, and promoter capture HiC identified long range connections between the locus and two other genes, *IRX3* and *IRX5*.[36–38] The signal was further narrowed down to four putatively causal SNPs using a massively parallel reporter assay (MPRA).[38] Confirmation of the role of *IRX3* and *IRX5* in body weight and the effect of the causal SNPs was obtained through functional characterization of both mouse and cell line models. This story illustrates the power of existing genetic datasets and tools to dissect and causally link variants to function. However, there remain many more GWAS variants that have not been linked to specific genes or function despite our best effort.

Although existing tools can help narrow down and direct the search for genetic mechanisms, linking variants to genes and functions remains a difficult process and to do so systematically requires accurate annotation data across may cell types and conditions. Therefore, one avenue to aid in our ability to link variants to function is to develop a deeper and broader catalogue of genetic and molecular annotations. One modality of annotation data that has proven particularly fruitful in recent years is the catalogue of quantitative trait loci (QTLs).[39–41] QTL

analysis allows us to link a particular genetic variant to a specific molecular trait of interest, often gene expression.[42,43] QTLs can therefore serve as a key intermediary link between a GWAS locus or a fine-mapped variant and a potential causal gene. Many studies to date have focused on expression QTLs (eQTLs) and have used this data to identify genetic variants that confer disease risk through perturbing gene expression.[40,44,45]

However, there are other molecular phenotypes, such as splicing, that may be equally impactful in disease risk but that have not been studied as extensively. Alternative splicing is a key source of functional complexity in human tissues and is known to play an important role in tissue identity and development.[46,47] Alternative splicing is also a highly regulated process, and perturbations in specified splicing patterns can lead to disease.[48–51] Although there has been an increase in recent years in studies that investigate sQTLs and their role in disease,[52–54] there are challenges in interpreting splicing events and sQTL results that must be overcome before we can access their full potential for elucidating mechanisms of disease risk. In particular, it remains difficult to determine the impact of a specific splicing event on gene function or disease state, as many splicing events still lack annotation or experimental data linking them to specific cellular or regulatory functions.

One mechanism that could underlie some functionally uncharacterized variants associated with obesity or other metabolic disorders is that the variant perturbs splicing in a tissue specific or developmental specific manner, such as during adipocyte differentiation, leading to impaired tissue development and function. However, to investigate this mechanism and definitively link variants to splicing events that are integral for tissue identity and development we need a deeper understanding of what splicing events are occurring in relevant tissue types and time points, and which are important for tissue function or associated with

disease. Therefore, an atlas of splicing events across adipocyte differentiation, their role in adipocyte function, and their perturbation in metabolic disease could provide a valuable resource to link GWAS variants to specific molecular functions.

Environmental contributions to obesity: insights from Nutritional Geometry

In addition to genetic contributions to obesity, there are also significant environmental contributions to disease risk. One important environmental risk factor to consider is diet. However, research into the impact of diet on obesity risk is complicated by the fact that diet is an incredibly complex environmental variable, made up of not just how much we eat, but also what and when we eat.[55,56] To simplify this complex space, many dietary studies focus on a single nutrient and compare across a small number of diets, for example by investigating the impact of a low fat diet or a low protein diet compared to a control "Western" diet or through the use of a high fat vs. chow diet in mice.[56–60] Although these studies have generated important insights into the impact of diet on obesity, they are limited by their ability to assess only a small amount of the nutritional space that diets encompass and often conflate separate dietary components such as dietary composition and energy density. To gain a more complete understanding of the impact of diet on metabolic health, we must move beyond a two-diet-at-a-time paradigm and capture data across a more complete dietary space.

One approach to address this issue that has proven successful in recent years is the Nutritional Geometry framework. The Nutritional Geometry framework conceptualizes diets as falling in a nutritional space defined by the amount of fat, carbohydrates, and protein in a diet and the energy density of a diet.[61,62] Using this framework, studies can be designed to systematically explore the dietary space and capture a more complete picture of the impact of specific dietary parameters and their interactions on health outcomes.[63,64] By considering a larger

number of diets and ensuring that differences in energy density are not confounded with differences in macronutrient composition, this framework can help tease out complex dietary effects on metabolic health.

Previous work in the field of Nutritional Geometry has demonstrated that both macronutrient composition and energy density have important and sometimes contrasting effects on metabolic health. For example, a large mouse study that considered 25 diets that differed systematically in their macronutrient composition and energy density found that lifespan and late-life health were maximized on low-protein, low-carbohydrate diets, and were not significantly impacted by caloric restriction.[65] Interestingly, subsequent work showed that many measures of reproductive function were optimized on high-protein, low-carbohydrate diets,[66] indicating a potential tradeoff between environments that encourage high reproductive rates and those that extend lifespan. These studies demonstrate the power of the Nutritional Geometry framework to dissect complex relationships between diet and health, and to provide important insights that can guide future research and dietary recommendations.

These studies also provide ample evidence that changes in macronutrient composition, even in the absence of changes in caloric intake, can have profound and sometimes contrasting impacts on metabolic health and overall lifespan. These results therefore have important implications for public health guidelines and for our understanding of what constitutes a healthy diet. However, the mechanism underlying the impact of macronutrient composition on metabolic health remains largely unexplored and unclear. A deeper understanding of *how* macronutrient composition leads to the observed changes in metabolic function can provide insights into the genes and pathways involved in response to diet and potentially elucidate new mechanisms of metabolic disease.

To begin to answer this question, it will be important to capture the gene regulatory response of key metabolic tissues to changes in dietary macronutrient composition. By identifying genes and pathways that are responsive to different macronutrients, macronutrient interactions, or individual dietary states we can begin to establish links between the observed phenotypic changes and specific dietary parameters. A better understanding of the mechanism underlying metabolic response to diet can help us more clearly define what makes a diet healthy, guide future dietary studies in humans, and possibly provide therapeutic targets to help improve metabolic health.

The role of adipose tissue in obesity and metabolic function

Obesity and metabolic function are complex phenotypes regulated jointly by many tissues, including the brain, liver, and adipose tissue.[67,68] Adipose tissue was once considered a fairly static tissue that played a passive role in fat storage, but we now know that adipose tissue is a dynamic regulatory tissue involved in endocrine signaling.[69,70] The dysregulation of adipose tissue is a hallmark of obesity, and is a complex phenotype involving tissue remodeling (hypertrophy), changes in cell type composition (immune cell invasion), and changes in insulin sensitivity.[71,72] These changes in adipose tissue function are also involved in many co-morbidities associated with obesity, such as insulin resistance leading to T2D and an increase in free fatty acids leading to fatty liver disease.[71–73] The gene regulatory landscape of adipose tissue is also known to be responsive to dietary change, exhibiting dynamic gene expression and alternative splicing changes after exposure to high fat diet.[74] Adipose tissue is therefore an important site of both physiological and gene regulatory changes in response to obesity and in response to upstream risk factors such as diet and genetic risk.

Although much is known about the physiological response of adipose tissue to various environmental and genetic perturbations, the gene regulatory response, especially with regards to alternative splicing, is less well studied. Previous work has established the importance of some individual splicing events in adipocyte function and identified specific splicing regulators involved in adipogenesis.[74–76] This work has established the importance of alternative splicing in adipose tissue function and development, but has stopped short of fully characterizing the role of alternative splicing in adipocyte function in relevant phenotypic states, such as in response to complex dietary perturbation and across differentiation time.

The role of alternative splicing in adipocyte differentiation is of particular interest because adipose tissue is a heterogeneous tissue, comprised of many cell types including mature adipocytes and differentiating preadipocytes.[69,77] Many of the key characteristics of adipose tissue dysfunction and eventual disease, such as hypertrophy, insulin resistance, and decreased differentiation capacity of adipose tissue, may be due to gene regulatory changes that occur in preadipocytes or during the process of differentiation to mature adipocytes.[78–80] We can therefore gain insight into adipose dysregulation and its role in obesity through investigating gene regulatory changes across the differentiation process from preadipocytes to adipocytes. A comprehensive catalog of gene regulation changes, including splicing changes, across adipocyte differentiation could provide insight into genes and pathways involved in the development of healthy mature adipocytes as well as into ways in which the gene regulatory landscape is perturbed in metabolic disease.

A more comprehensive characterization of changes in alternative splicing in adipocytes across a broad range of states would provide important insights into adipose function and

dysfunction, and possibly provide therapeutic targets to help treat many aspects of metabolic dysregulation.

Dissertation overview

In this thesis, I investigate the etiology of obesity through the lens of both environmental and genetic contributions. In Chapter 2, I use the Nutritional Geometry framework to systematically characterize the impact of macronutrient composition on gene expression and splicing in adipose tissue in mice. I identify gene expression and alternative splicing changes across ten diets in sixty mice, and identify clusters of gene regulation changes that respond to different macronutrients and macronutrient interactions. I provide evidence for the role of both gene expression and splicing changes in the response of adipose tissue to macronutrient composition and expand our understanding of the mechanisms behind metabolic response to diet. In Chapter 3, I analyze alternative splicing across adipocyte differentiation in preadipocytes isolated from lean, obese, and obese with T2D individuals. I show that alternative splicing is incredibly dynamic across adipocyte differentiation, is impacted by metabolic phenotype, and may be involved in genetic risk for T2D. These results can serve as a resource to help dissect previously uncharacterized GWAS variants associated with metabolic disorders and expand our understanding of the role of splicing in adipocyte biology. Together, these studies provide insight into the gene regulatory landscape of adipose tissue in multiple conditions (across a large dietary space and across differentiation time), and improve our understanding of the role of both environment and genetics in metabolic disease risk.

Chapter II: Dietary macronutrient composition impacts gene regulation in adipose tissue

Note: The following chapter is reproduced from the manuscript titled “Dietary macronutrient composition impacts gene regulation in adipose tissue”. An earlier version of this manuscript was published on Research Square in 2023.[81]

Abstract

Diet is a key lifestyle component that influences metabolic health through several factors, including total energy intake and macronutrient composition. While the impact of caloric intake on gene expression and physiological phenomenon in various tissues is well described, the influence of dietary macronutrient composition on these parameters is less well studied. Here, we use the Nutritional Geometry framework to investigate the role of macronutrient composition on metabolic function and gene regulation in adipose tissue. Using ten isocaloric diets that vary systematically in their proportion of energy from fat, protein, and carbohydrates, we find that gene expression and splicing are highly responsive to macronutrient composition, with distinct sets of genes regulated by different macronutrient interactions. Specifically, the expression of many genes associated with Bardet-Biedl syndrome is responsive to dietary fat content. Splicing and expression changes occur in largely separate gene sets, highlighting distinct mechanisms by which dietary composition influences the transcriptome and emphasizing the importance of considering splicing changes to more fully capture the gene regulation response to environmental changes such as diet. Our study provides insight into the gene regulation plasticity of adipose tissue in response to macronutrient composition, beyond the already well-characterized response to caloric intake.

Introduction

Diet and nutrition are key determinants of metabolic health, with implications for both personal and public health. However, what defines a metabolically healthy diet remains elusive. Some studies have focused on total energy and caloric restriction as the most impactful components of a healthy diet,[82,83] while others have argued for the importance of particular nutrients such as fat, carbohydrates, or protein.[84,85] One contested component of a healthy diet is therefore macronutrient composition, namely the ratio of fat, carbohydrates, and protein in a diet. Therefore, understanding the impact of these macronutrients on metabolic health is important for defining a healthy lifestyle and may lead to better-informed nutritional guidelines.

Studies that investigate the impact of diet on metabolic health often rely on a high-fat diet paradigm, where a high-fat, energy-dense diet is compared to a control diet. This study design focuses on a single macronutrient and conflates changes in macronutrient composition with changes in the energy density of the diet. The Nutritional Geometry framework moves beyond this single-macronutrient-at-a-time paradigm by considering a wide range of diets that vary systematically in their ratios of fat, carbohydrates, and protein.[61,62,86,87] By considering a large number of isocaloric diets, this framework allows us to determine the metabolic impact of each individual macronutrient and interactions between macronutrients, while controlling for caloric density through titrating indigestible cellulose. Previous work using the Nutritional Geometry framework has shown that both total energy intake and dietary macronutrient composition impact metabolic health, lifespan, and fertility,[65,66,88,89] but the mechanisms underlying these effects are not fully known.

A deeper understanding of the molecular mechanisms underlying changes in metabolic function in response to dietary macronutrient composition may provide insights into what

constitutes a healthy diet and possible interventions to maintain a healthy metabolic profile. One mechanism that may underlie the observed changes in metabolic function is changes in gene regulation in metabolic tissues, which may lead to changes in tissue function and overall health. Adipose tissue is a key metabolic tissue that is highly functionally dynamic in response to metabolic change[90,91] and is known to have dynamic gene regulation after exposure to high fat diet.[74] By investigating gene regulation changes in adipose tissue in response to differences in macronutrient composition, we can gain further insights into the impact of diet on adipose tissue function and possibly undercover mechanisms underlying previously reported effects of dietary macronutrients on metabolic health.

Many studies that consider gene regulation change focus on gene expression alone, but it is important to also consider the role of other forms of gene regulation, such as alternative splicing, in the metabolic response to diet. Alternative splicing is a fundamental source of functional complexity in tissues and contributes to tissue identity and development.[46,47] Alternative splicing is also a highly regulated process, and its misregulation can lead to developmental defects and disease.[92,93] However, in the context of gene regulation under environmental effects such as diet, splicing remains relatively understudied compared to other mechanisms of gene regulation, such as transcriptional regulation.

Here, we used the Nutritional Geometry framework to investigate the effects of dietary macronutrient composition on metabolic function and gene regulation in the fat pads of male mice. This framework provides insight into a more complete dietary space than previously considered, allowing us to determine the impact of each macronutrient singly and in combination. Using RNA-seq data collected from the fat pads of mice fed one of ten isocaloric diets ad libitum, we identified extensive differences in both gene expression and splicing in

response to dietary composition and determined the primary macronutrients driving the observed differences. The majority of alternative splicing events we identified are in genes whose expression is not significantly different in response to dietary composition, highlighting a pervasive and complementary mechanism by which cells regulate their transcriptome beyond regulation of gene expression. Using this comprehensive dietary paradigm we are able to cluster the gene regulation changes on the basis of their functional response to macronutrients and identify several common patterns of gene regulation associated with dietary macronutrient composition, providing insight into the effect of different macronutrients and macronutrient interactions on adipose tissue function.

Results

Body composition and metabolic health

To measure the impact of dietary macronutrient composition on metabolic health, we fed 60 male mice one of 10 isocaloric diets that differed systematically in their ratios of protein, carbohydrates, and fat (Figure 2-1a, Table 2-1). For each mouse, we collected data on body composition, including body weight, fat mass, and lean mass, as well as other measures of metabolic health such as glucose tolerance (Supplementary Figure 2-1). To analyze these data we used a mixture-model framework, where models were fitted for each metabolic response over the dietary space, exploring linear, non-linear, and interactive effects of the macronutrients. Predictions from fitted models were then plotted as a right-angled mixture triangle with the percent dietary protein on the x-axis, percent dietary carbohydrate on the y-axis, and percent dietary fat as the distance from the hypotenuse to the origin.[94]

Table 2-1. **Dietary macronutrient composition.** The macronutrient composition of each experimental diet as a percent of total energy.

Diet	Protein (%)	Carbohydrate (%)	Fat (%)
1	7	33	60
2	7	78	15
3	14	56	30
4	14	26	60
5	21	64	15
6	21	34	45
7	30	40	30
8	35	20	45
9	42	43	15
10	50	20	30

Using this framework, we found that dietary macronutrient composition had a significant impact on body composition and metabolic health (Figure 2-1b). Body weight, fat mass, lean mass, and glucose tolerance all differed significantly across the diets. Body weight and fat mass were both maximized on a moderate fat, moderate carb, and moderate protein diet (diet 7, Supplementary Table 2-1) and minimized on a low protein, moderate carb, high fat diet (diet 1, Supplementary Table 2-1). When considering associations with single macronutrients, body weight was positively correlated with protein content (r = 0.26, P = 0.043) and negatively correlated with fat content (r = -0.40, P = 0.0017) whereas fat mass was positively correlated with carbohydrate content (r = 0.40, P = 0.0015) and negatively correlated with fat content (r = -0.39, P = 0.0019) (Supplementary Figure 2-2).

In contrast, lean mass was positively correlated with protein content (r = 0.52, P = 2.0e-05) and not correlated with fat or carbohydrate content (Supplementary Figure 2-2). Lean mass

was maximized on a moderate protein, high carb, and low fat diet (diet 5) and minimized on low protein diets (diets 1 and 2) (Supplementary Table 2-1). Glucose tolerance was also impacted by differences in dietary macronutrient composition (Figure 2-1b). The incremental area under the curve (iAUC) in an oral glucose tolerance test was positively correlated with protein content ($r = 0.40$, $P = 0.0017$) and negatively correlated with carbohydrate content ($r = -0.28$, $P = 0.033$), but not correlated with fat content (Supplementary Figure 2-2).

Dietary macronutrient composition therefore had diverse impacts on body composition and metabolic health. Using the Nutritional Geometry framework, we are able to determine the impact of differences in macronutrient composition alone, in the absence of differences in caloric density. This allows us to ask more precise questions about the impact of individual macronutrients on metabolic health without confounding with the energy density of the diet, as is common in a high fat diet context. In this context, we found evidence for both linear and non-linear effects of dietary macronutrients on various metabolic measures. Body weight and fat mass were both negatively correlated with fat content, with some interactions with protein and carbohydrate content as well. On the other hand, lean mass and glucose tolerance were correlated with protein content but not fat content. Overall, in an isocaloric context the ratio of dietary macronutrients significantly altered the metabolic profiles of these mice.

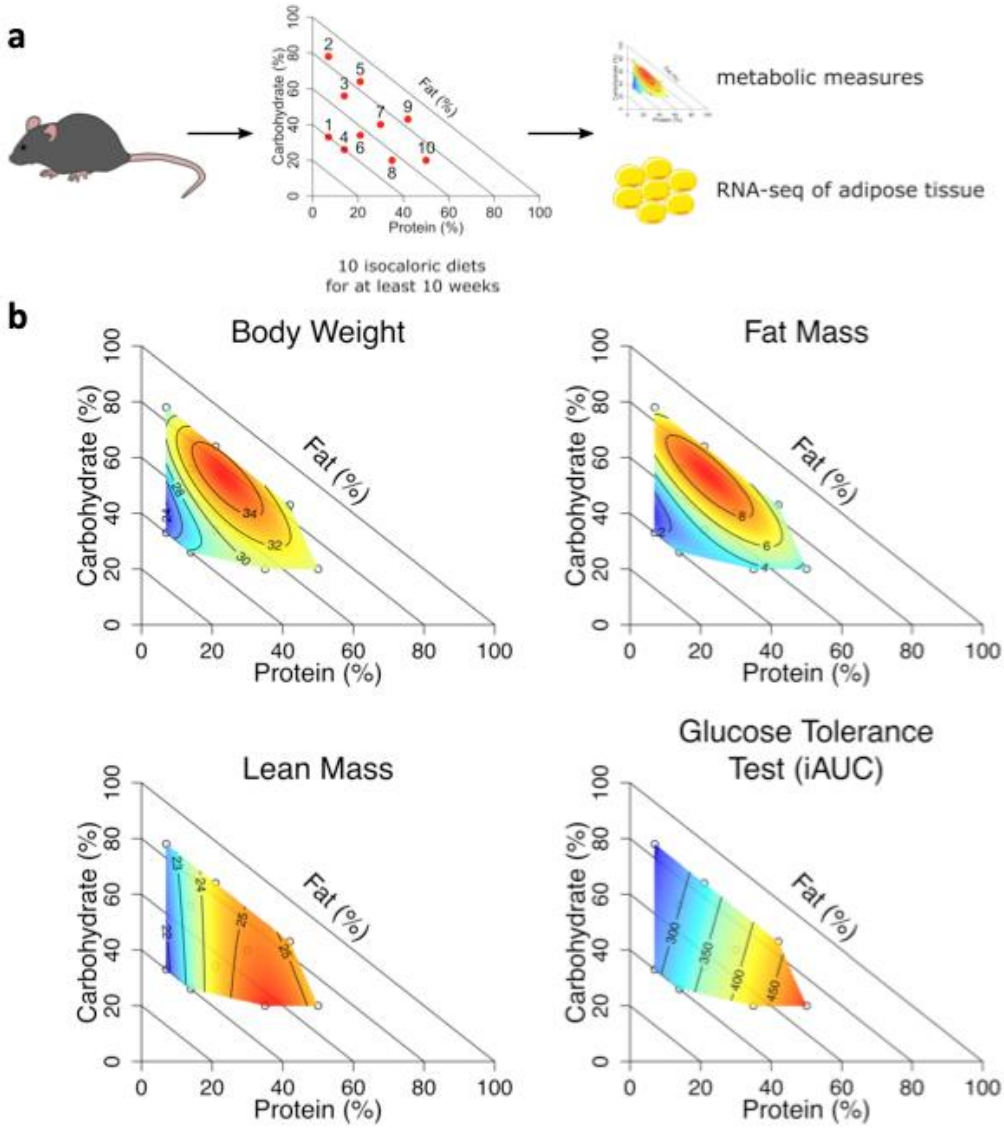


Figure 2-1. Metabolic response to dietary macronutrient composition. a. A diagram of the experimental setup and data collection. b. Surfaces of metabolic measures across the 10 diets plotted as a right-angled mixture triangle, with color indicating the level of the measured variable (red = higher, blue = lower) and isolines showing the model predicted response. The diagonal lines are included to help visualize fat content and are isolines of dietary fat content. At the origin, dietary fat content is 100% and it decreases to 0% as you move away from the origin along the y = x line. n = 6 mice per diet.

Changes in gene regulation in response to diet

To better understand how differences in dietary macronutrient composition led to the observed effects on metabolic health *in vivo*, we investigated alterations in gene regulation

programs associated with the observed changes in body composition and metabolic parameters. We performed RNA-seq in the inguinal fat pads of each of the 60 mice to measures changes in gene expression and splicing across the 10 diets. Following quality assessment, 57 samples were retained for all genomics analyses (n = 5 or 6 per diet). We tested the response of each gene or exon across the macronutrient space, and found that there were 4,308 differentially spliced exons in 2,615 unique genes (Figure 2-2a, Supplementary Figure 2-3a, Supplementary Table 2-2) and 5,644 differentially expressed genes (Figure 2-2c, Supplementary Figure 2-3b, Supplementary Table 2-2). Only 967 genes were both differentially expressed and differentially spliced, with the majority of genes that underwent gene regulatory changes being acted on by only one of the two measured mechanisms (Figure 2-2b).

We were also interested in whether any of the splicing changes we identified may be driven by expression changes, in particular by the differential expression of splicing factors known to be involved in adipogenesis or adipocyte function. We therefore asked whether splicing factors that have previously been identified as having a role in adipocyte function were differentially expressed across the diets, possibly leading to associated changes in alternative splicing. We considered six splicing factors with known roles in adipogenesis and adipocyte function, and found that none were significantly differentially expressed in response to dietary macronutrient composition (Supplementary Table 2-3). This indicates that the splicing changes we see may be due to the perturbation of other, unknown splicing factors or may be responding to diet without the intermediary effect of a differentially expressed splicing factor.

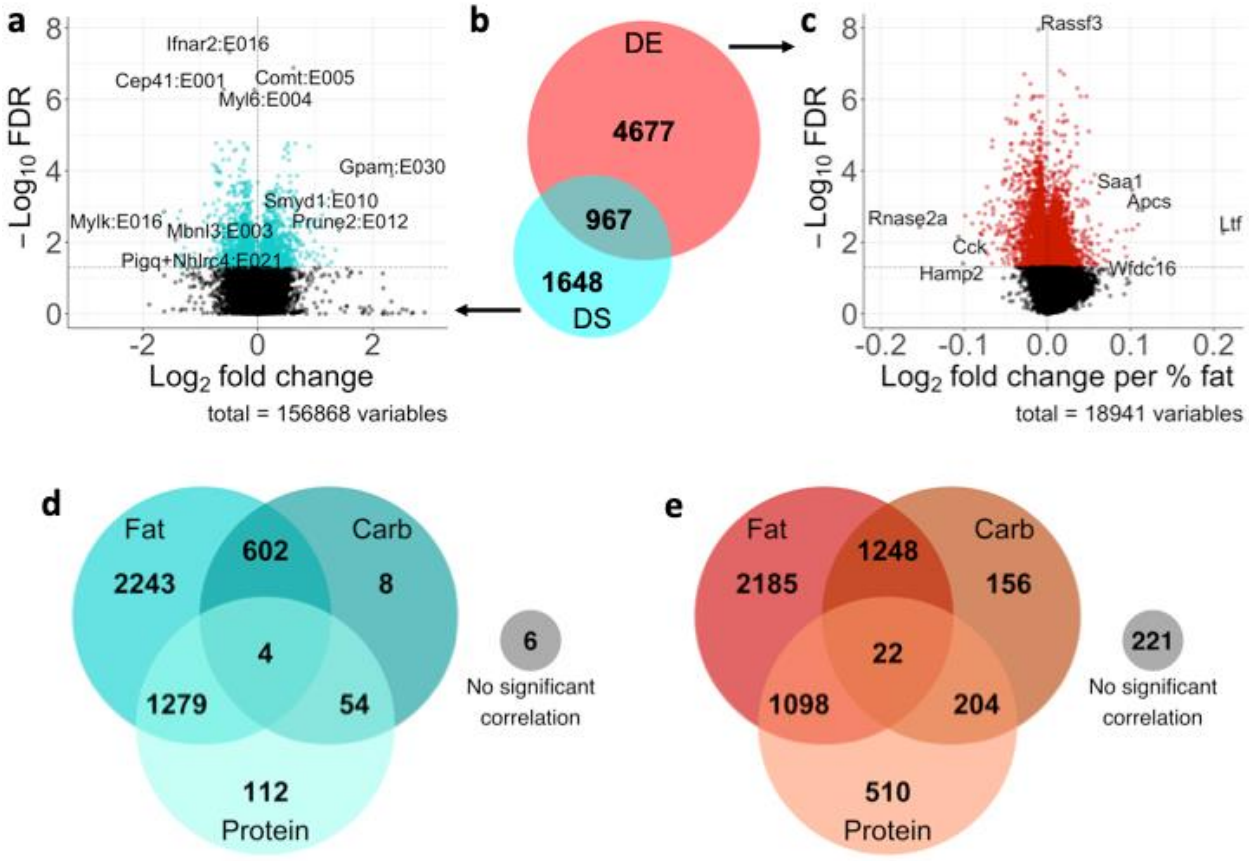


Figure 2-2. **Significant regulatory response to dietary macronutrient composition.** a. Volcano plot of differential splicing changes, plotting the log fold change between 15% dietary fat and 60% dietary fat. Blue dots are significant, black are non-significant. Extreme exons in terms of log fold change or p-value are labeled. b. Venn diagram of differentially expressed and/or differentially spliced genes. c. Volcano plot of differential expression changes, plotting the log fold change per percent dietary fat. Red dots are significant, black are non-significant. d. Venn diagram of the correlation of each differentially spliced exon with the three macronutrients. E. Venn diagram of the correlation of each differentially expressed gene with the three macronutrients. n = 57 mice.

We therefore observed abundant changes in gene regulation in response to differences in dietary macronutrient composition, with expression and splicing changes largely occurring in distinct genes, underscoring how distinct gene regulatory strategies may impact the transcriptome in cells quantitatively (through transcription regulation) and qualitatively (through differential usage of exons encoding specific protein domains). This analysis identifies genes and

exons that were significantly impacted by diet, but doesn't provide insight into what macronutrient or macronutrients these genes and exons might be responding to. We found that metabolic measures can have disparate responses to macronutrient composition (Figure 2-1b), and therefore sought to better understand what macronutrient interactions might be driving the observed changes in gene expression and splicing.

Correlation between gene regulation changes and individual macronutrients

To quantify the impact of individual macronutrients on gene expression and splicing, we calculated the correlation of each differentially expressed gene or differentially spliced exon with fat, protein, and carbohydrate content. We found that dietary fat content is the predominant driver of the observed gene expression and splicing changes (Figures 2-2d, 2-2e). This was particularly true for the differential splicing changes, where 4,128 differentially spliced exons in 2,510 genes (96% of all differentially spliced exons) were correlated with dietary fat content (Figure 2-2d). Of note, the diets in this study contained varying amounts of non-digestible cellulose to maintain their caloric density. Cellulose content of the diet is positively correlated with fat content, so it is possible that some of the gene regulation changes that are correlated with fat content are actually responding to fiber content of the diet. However, due to the indigestible nature of the cellulose we believe that fat content is the main driver of the observed gene regulation changes.

Although fat content was the strongest driver of gene regulation changes, protein and carbohydrate content were also correlated with many of the changes in gene expression and splicing, often in conjunction with fat content. Gene regulation changes correlated with more than one macronutrient represent 46% of all differentially expressed genes, and 45% of all differentially spliced exons. While correlations with individual macronutrients can capture some

of the dynamics of this dietary space, they do not necessarily capture the full response of genes and exons across all 10 diets, especially for gene regulation changes that respond to multiple macronutrients or interactions between macronutrients. To better capture these complex dietary responses, we sought to categorize the gene regulation changes that we observed in terms of their holistic response across the nutrient space encompassed by the 10 diets, as opposed to focusing on each macronutrient separately.

Clustering analysis of differentially spliced exons

We identified complex gene regulation responses to differences in dietary macronutrient composition, including many genes and exons that responded to multiple macronutrients or possibly interactions between macronutrients. To better partition these complex responses, we can quantify the response of each differentially spliced exon to macronutrient composition using the regression coefficients from a mixture model. Nutritional Geometry then allows these responses to be visualized as response surfaces (topologies) mapped onto dietary macronutrient space. We therefore clustered all the differentially spliced exons based on the regression coefficients for all three nutrients using fuzzy c-means clustering and visualized each cluster using a response surface generated from the mean exon usage for all exons assigned to that cluster. Using this method, we can observe a more representative range of dietary response landscapes than simple linear correlations with individual macronutrients. The results of this analysis with five clusters are shown (Figure 2-3a).

Consistent with the results from the correlation analysis, we see that the three largest clusters (clusters 3 – 5) show a predominant response to fat content, either positive (clusters 3 and 4, which were closely similar in topology) or negative (cluster 5). The remaining two clusters capture interaction effects, namely a positive carbohydrate by negative protein gradient

(cluster 1) and a positive protein by negative fat gradient (cluster 2). These interaction effects were not identified as strong signals from the single-nutrient correlation analyses (Figure 2-2d) and would most likely have been missed had we not considered the full response of each exon across nutrient space and instead considered one macronutrient at a time, as is conventional.

We next investigated whether these groups of exons that were clustered based on their response to diet also fell into shared functions or pathways. Using functional enrichment analysis,[95] we found that the clusters were significantly enriched for distinct functional terms (Figure 2-3b). Cluster 1, which demonstrated a primarily carb by protein gradient, was enriched for terms related to cell adhesion, such as cell-cell adhesion and focal adhesion. The protein by fat cluster (cluster 2) was also enriched for focal adhesion, but showed stronger enrichment for regulation of cell morphogenesis involved in differentiation and regulation of protein catabolic processes. In contrast, the exons responding more to fat content, such as cluster 5, show enrichment for terms related to intracellular transport and organization.

Overall, we observe distinct functional enrichment in groups of exons that respond differently to dietary macronutrient composition, demonstrating the importance of capturing the full dietary response to understand gene regulation changes in response to diet. Further, most of these genes would not have been identified as undergoing gene regulation change if we had considered expression differences alone, emphasizing the need to consider splicing as well as expression changes when analyzing gene regulation responses.

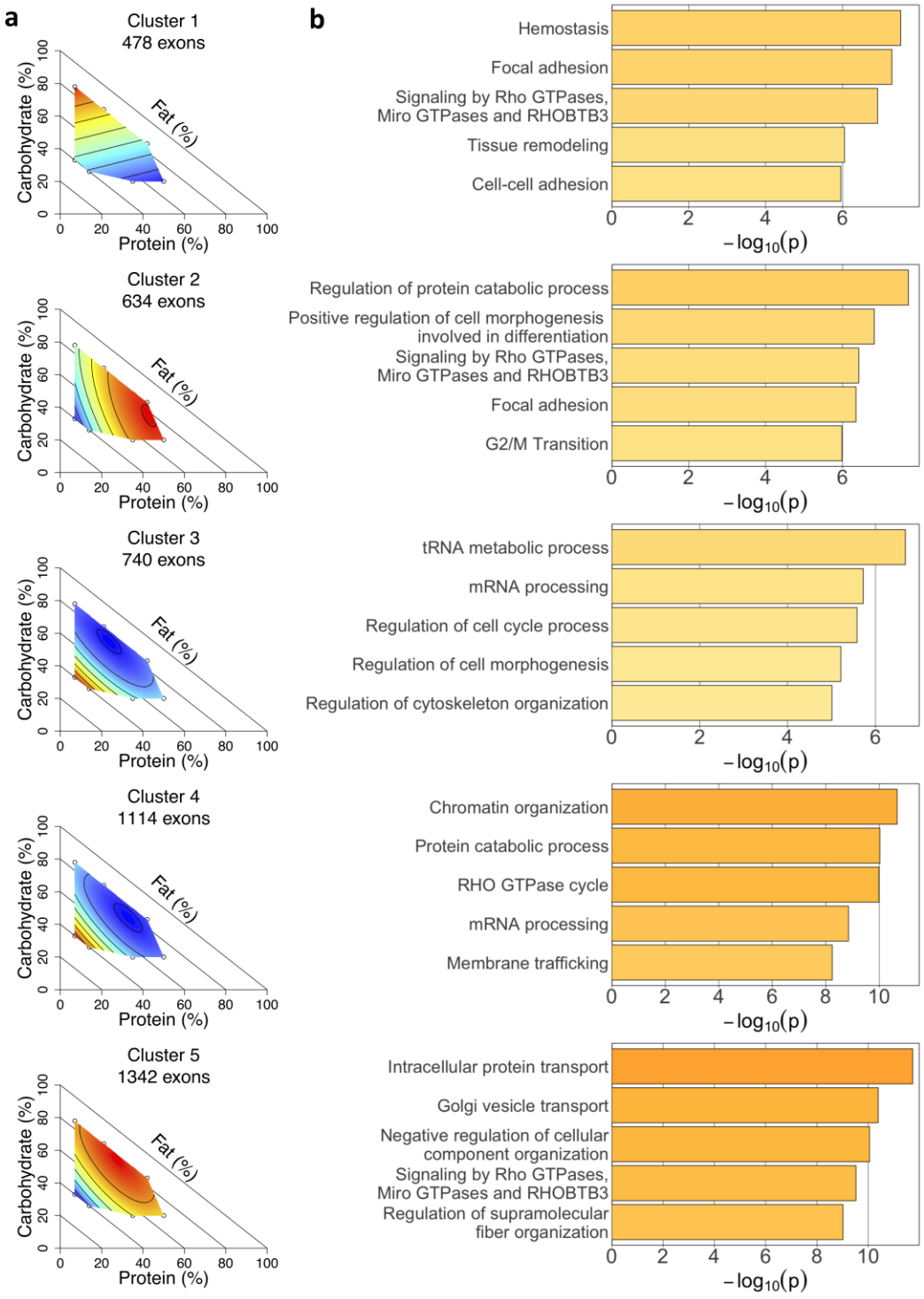


Figure 2-3. **Differential splicing changes clustered into five distinct groups.** a. Surfaces generated from the mean centered and scaled exon usage of each exon assigned to the cluster, with color indicating the level of exon usage (red = higher, blue = lower). b. The five most significantly enriched functional terms for each cluster. n = 6 for diets 1, 5, 6, 7, 8, 9, and 10 and n = 5 for diets 2, 3, and 4

Differential splicing of key adipocyte genes

In addition to considering the differential splicing changes at the level of clusters and functional groups, we also identified individual splicing events predicted to have significant impact on adipocyte function. These include differential splicing events in *Vegfa* and *Igf1*.

Vegfa regulates angiogenesis, and has been implicated in adipose tissue response to diet-induced obesity.[96,97] Specifically, overexpression of *Vegfa* in the fat pads of mice leads to increased vascularization and a healthier phenotype in response to high fat diet.[97] In our study, we found that exon 6 of *Vegfa* was differentially spliced in response to dietary composition, and demonstrated a predominant response to carb and protein content of the diet (Figure 2-4a). Exons 6 and 7 contain heparin-binding domains, and are known to be differentially spliced to produce isoforms that contain one, both, or neither domain[98] (Figure 2-4b). Transcripts lacking exon 6 and 7 produce a variant of *Vegfa* that does not bind heparin and is fully soluble, whereas heparin-binding variants of *Vegfa* bind to the cell surface and extracellular matrix, with different isoforms leading to different sites of angiogenesis.[99] Therefore, differential splicing of exon 6 in response to macronutrient composition suggests that changes in dietary carb and/or protein content may lead to changes in the angiogenic potential and patterns of fat tissue.

Another key adipocyte gene that we identified as differentially spliced is *Igf1*. *Igf1* regulates adipocyte differentiation[100,101] and controls the response of adipose tissue to metabolic stress.[102] Alternative splicing produces isoforms of IGF1 that differ in their N-terminus (known as the signal peptide) and C-terminus (known as the E peptide), which are removed post-transcriptionally to produce the same mature peptide.[103] In mice, there are two main variants of the E peptide based on whether exon 5 is spliced in or out[103] (Figure 2-4d). The E peptide that includes exon 5 has been functionally implicated in IGF1 bioavailability, *via* more strongly

facilitating IGF1 binding to the extracellular matrix than the shorter E peptide,[104] and increases the adipogenic potential of bone marrow mesenchymal stem cells.[105] Here, we found that exon 5 of *Igf1* was differentially spliced in response to dietary composition, and identified a specific response to the protein by fat ratio of the diet (Figure 2-4c).

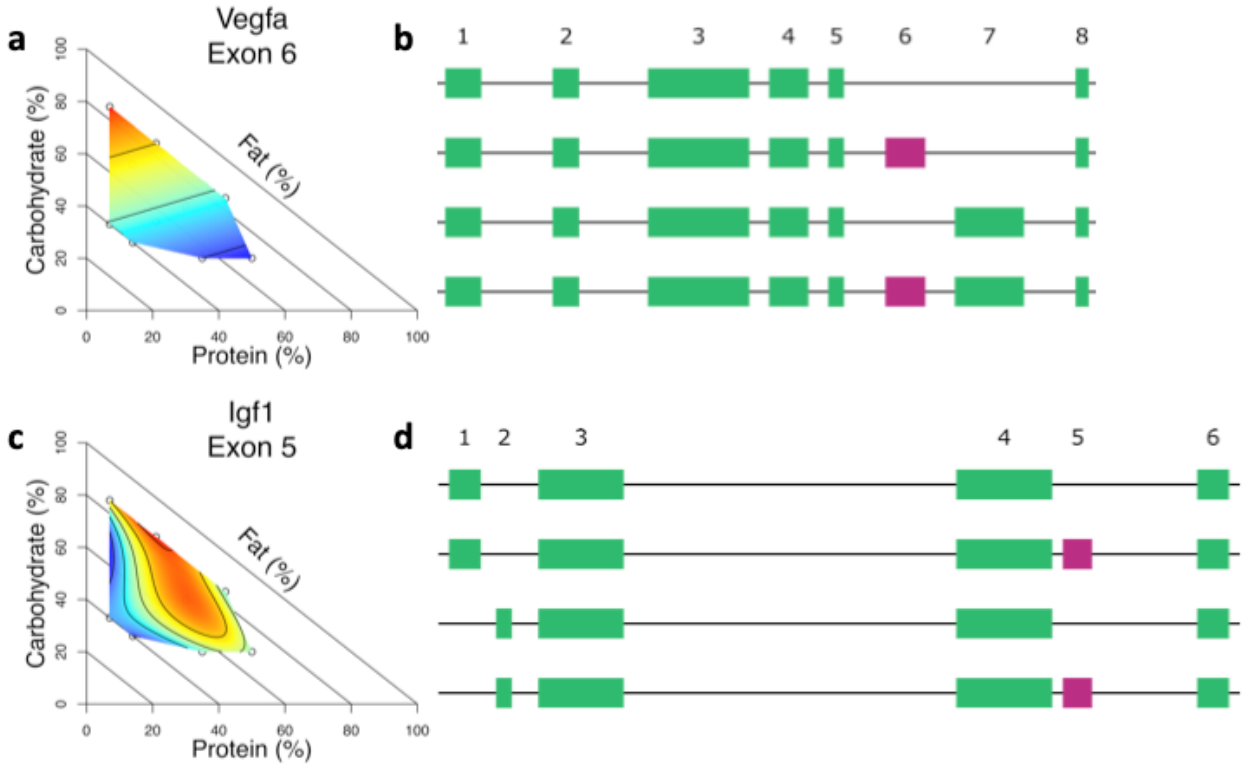


Figure 2-4. **Differential splicing of *Vegfa* and *Igf1*.** a. Surface of the centered and scaled exon usage of *Vegfa* exon 6. b. Diagram of selected *Vegfa* isoforms. Exon 6 is highlighted in pink. c. Surface of the centered and scaled exon usage of *Igf1* exon 5. d. Diagram of selected *Igf1* isoforms. Exon 5 is highlighted in pink. n = 6 for diets 1, 5, 6, 7, 8, 9, and 10 and n = 5 for diets 2, 3, and 4

Although adipose tissue does express *Igf1*, it is not the main source of circulating IGF1. Rather, the main contributor to circulating IGF1 levels is hepatocytes.[106] To further support our finding that dietary macronutrient composition significantly impacts *Igf-1* splicing and adipose tissue function, we therefore asked whether *Igf-1* splicing in the liver was also responsive to diet. Using liver samples collected from the same 60 mice in which we assayed adipose gene

regulation changes, we used RNA-seq to measure the impact of dietary macronutrient composition on *Igf-1* splicing in the liver. We found that *Igf-1* exon 5 is indeed differentially spliced in response to diet in the liver, and responds primarily to protein content of the diet (Supplementary Figure 2-4). The *Igf-1* exon 5 splicing surfaces are distinct between adipose and liver tissue, with the liver splicing displaying a more marked response to dietary protein content. Notably, in both tissues exon 5 of *Igf-1* is responsive to diet and minimized on low protein diets. This result suggests that there may be differences in IGF1 bioavailability in response to different dietary macronutrient compositions in both liver and adipose tissue, potentially leading to changes in adipose tissue function.

Since neither *Vegfa* nor *Igf1* were differentially expressed in response to dietary macronutrient composition, the impact of macronutrient composition on these biological processes would not have been detected by measuring differential expression. By analyzing differential splicing changes, we were able to detect gene regulation changes that may alter adipocyte function and that would have been missed when considering gene expression alone.

Clustering analysis of differentially expressed genes

Next, we performed fuzzy c-means clustering on the responses of the 5,644 differentially expressed genes to capture the differential expression dynamics across all ten diets. The results of this analysis with five clusters are shown (Figure 2-5a). From these five clusters, the largest two clusters show a strong positive (cluster 4) or negative (cluster 5) fat gradient. This is consistent with the single-nutrient correlation results that identified fat content as the strongest driver of the observed expression changes (Figure 2-2e), as well as the splicing clustering results in which the three largest clusters (clusters 3 – 5) also show a predominant response to fat (Figure 2-3a). The remaining three clusters capture interaction effects between the

macronutrients, specifically fat by protein gradients (cluster 2 showing a negative fat by positive protein interaction and cluster 3 showing a positive fat by negative protein interaction) and a positive carbohydrate by negative protein gradient (cluster 1). Again, these dietary interactions were not apparent in single nutrient analyses.

We next investigated whether genes falling into these distinct clusters based on their response to diet also had distinct biological functions. Using functional enrichment analysis, we found that the clusters differed not just in their response to diet but also in their functional enrichment (Figure 2-5b). One particularly strong enrichment signal was for immune function in genes assigned to cluster 1, possibly representing an inflammatory phenotype associated with carb and protein content in the diet. We found another strong enrichment signal for cilium function in cluster 4, where the enriched categories included cilium organization, cilium assembly, and intraciliary transport. Differentiating preadipocytes are ciliated, and cilia function is known to be involved in adipocyte differentiation.[107,108] The observed gene expression differences in cilia function could therefore indicate changes in adipogenic potential in the fat pad in response to dietary fat content. Overall, these data demonstrate that there are distinct sets of genes that respond differently to dietary macronutrient composition and carry out distinct functions in adipose tissue.

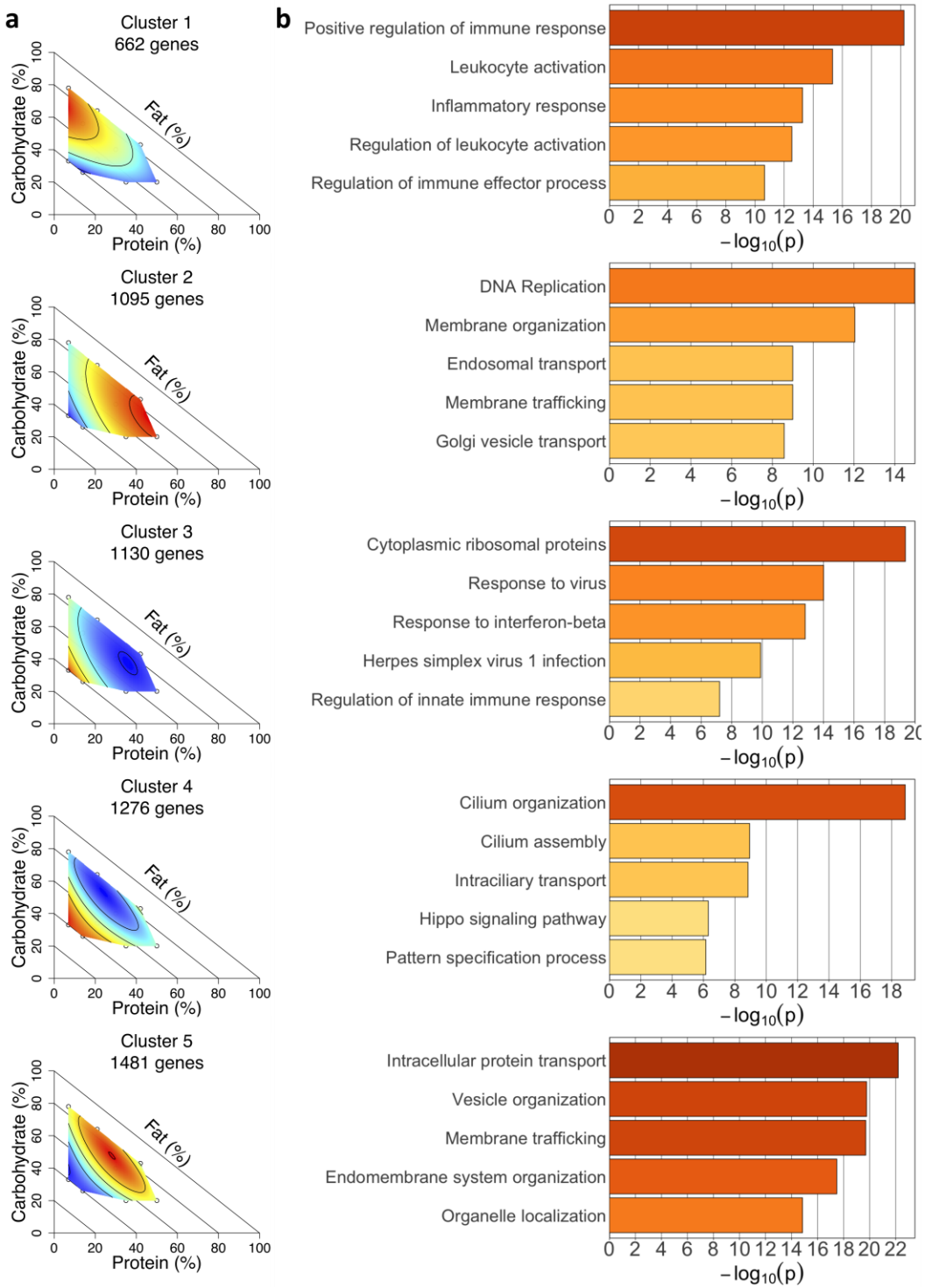


Figure 2-5. **Differential expression changes clustered into five distinct groups.** a. Surfaces generated from the mean centered and scaled expression of each gene assigned to the cluster, with color indicating the level of expression (red = higher, blue = lower). b. The five most significantly enriched functional terms for each cluster. n = 6 for diets 1, 5, 6, 7, 8, 9, and 10 and n = 5 for diets 2, 3, and 4

Changes in the expression of cilium-associated genes in response to dietary fat

One of the most striking enrichment signals that arose from the clustering analysis of differentially expressed genes was the enrichment in cluster 4 for genes involved in ciliary function (Figure 2-5b). When we investigated this signal more closely, we found that the signal was driven in part by a set of genes associated with Bardet-Biedl syndrome (BBS). BBS is an autosomal recessive ciliopathy with symptoms that include obesity.[109] At least 19 genes have been shown to cause BBS, many of which are associated with a structure called the BBSome, which is a protein complex that is involved in protein trafficking to the cilium.[110]

In our differential expression analysis, we found that nine BBS-associated genes were differentially expressed in response to dietary macronutrient composition. These nine genes have a variety of ciliary-related functions, including some that are components of the BBSome itself (Figure 2-6a). Eight of the nine differentially expressed BBS-associated genes were assigned to cluster 4 and one (*Ift27*) was assigned to cluster 3 (Figure 2-5a). As expected from the clustering analysis, the surface plots for each individual gene assigned to cluster 4 demonstrated a strong expression response to dietary fat content (Figure 2-6b). This may represent a novel association of BBS genes with diet-induced metabolic changes, in particular in response to differences in dietary fat content.

Many BBS genes are known to be upregulated during adipocyte differentiation.[111] It is therefore possible that the increase of BBS gene expression in response to dietary composition represents a change in cell type composition of the fat pads, with more differentiating preadipocytes present in diets with high fat content. We therefore sought to determine if there were changes in cell type composition associated with changes in BBS gene expression. All analyses in this study were done in bulk tissue samples, and we do not have direct access to

measurements of cell type composition of the tissue. Using the dampened weighted lease squares (DWLS) method,[112] we performed cellular deconvolution to computationally estimate the cell type composition of each bulk tissue sample based on bulk gene expression and a reference single-cell RNA-seq dataset of mouse adipose tissue.[77] We found that there were significant differences in predicted cell type composition in response to dietary composition, including a change in the estimated proportion of adipocyte progenitor cells that was predominately associated with protein content in the diet (Supplementary Figure 2-5).

We next asked whether BBS gene expression was correlated with cell type composition. We limited this analysis to the five cell types that were identified at greater than 1% frequency in at least one sample. We found that there were correlations between BBS gene expression and predicted cell type composition (Figure 2-6c). In particular, seven of the nine differentially expressed BBS genes were positively correlated with mAd3 proportion and negatively correlated with mAd5 proportion. Although this correlation may suggest an association between BBS gene expression and these specific adipose subclusters, it may also simply be due to the fact that both BBS genes and the proportion of adipocyte subclusters are independently correlated with fat content in the diet. These mouse adipocyte subclusters have previously been identified as responding to high fat diet, with mAd3 proportion reduced after high fat diet and mAd5 increased.[77] Here, we saw a negative association between mAd5 proportion and fat content, and a positive association between mAd3 proportion and fat content (Supplementary Figure 2-5). Given that our diets were isocaloric, our results suggest that these adipocyte subclusters may be responding to caloric density rather than fat content in the high fat diet context. Notably, none of the differentially expressed BBS genes were associated with the predicted proportion of

adipocyte progenitors, suggesting that the observed gene expression changes are not due to a change in cell type composition.

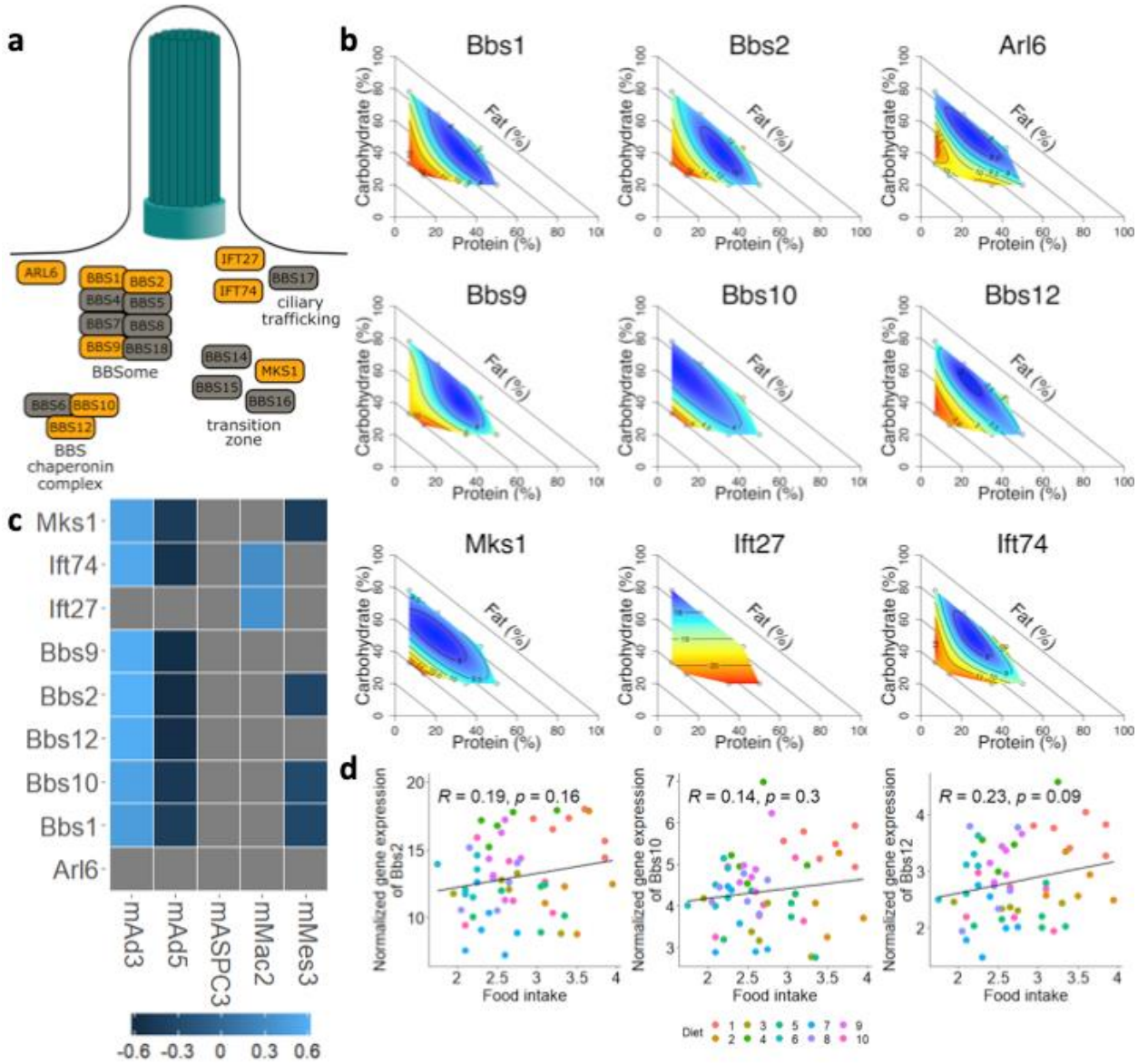


Figure 2-6. **Response of BBS genes to dietary fat.** a. Diagram of the primary cilia, with BBS genes organized by their role in cilia function. Differentially expressed BBS genes are highlighted in orange. BBS-associated gene functions adapted from previous work.[109,113,114] b. Surfaces generated from the expression of each differentially expressed BBS gene. c. Heatmap of the correlation of each differentially expressed BBS gene's expression with the cell type proportions estimated by deconvolution. d. Plots of the correlation between food intake (grams per day) and the expression of *Bbs2*, *Bbs10*, and *Bbs12*. n = 6 for diets 1, 5, 6, 7, 8, 9, and 10 and n = 5 for diets 2, 3, and 4

To complement and extend these results based on computational deconvolution of cell type composition, we also performed single-nucleus RNA-seq in adipose tissue from one mouse from a diet with high fat content (diet 4, 60% fat) and one mouse from a diet with low fat content (diet 7, 30% fat). After anchoring the resulting single-nucleus datasets to the same single-cell atlas that was used for deconvolution,[77] we identified clusters associated with all of the major cell types we expected to find in adipose tissue, such as adipocytes, adipose stem and progenitor cells (ASPCs), immune cells, and mesothelial cells (Supplementary Figure 2-6a). We do observe some differences in cell type composition between these two samples, most notably a decrease in the proportion of total cells identified as adipocytes in the diet 7 sample as compared to diet 4. To assess the expression of the differentially expressed BBS genes in each sample and cluster, we calculated a BBS expression score based on the sum of the normalized, centered, and scaled expression across the nine BBS-associated genes of interest. As expected from the bulk expression results, we see that the BBS genes are overall more lowly expressed in the diet 7 sample than in diet 4 (Supplementary Figure 2-6b), with a particular depletion of BBS gene expression in the adipocyte and mesothelial cell clusters in diet 7. Overall, these results suggest that BBS gene expression is lowered in multiple cell types, including adipocytes and ASPCs, in response to differences in dietary fat content, possibly leading to altered adipogenic potential in response to diet.

Finally, some BBS genes, such as *Bbs2*,[115] *Bbs10*,[116] and *Bbs12*,[117] are associated with changes in food intake. We therefore tested the association between the expression of these BBS genes and the food intake of each individual mouse. We found that there were no significant correlations (Figure 2-6d), suggesting that in this context BBS gene expression is not a significant driver of food intake.

Altogether, our results indicated that the observed changes in BBS gene expression may regulate ciliary function in response to dietary macronutrient composition, which does not appear to be caused by changes in cell type composition in adipose tissue. This suggests a possibly novel role for the BBSome and other BBS-associated genes in response to diet.

Discussion

In this study, we aimed to use the Nutritional Geometry framework to dissect the effects of dietary composition and interactions between macronutrients on gene regulation in adipocytes. By using ten isocaloric diets that cover a large range of the macronutrient space, we could precisely assign the observed gene regulation changes to specific macronutrient gradients and control for effects of caloric density and energy intake. Using this framework, we have generated a comprehensive analysis of gene regulation changes in fat tissue in response to differences in macronutrient composition, and identified key changes that may be important for adipocyte biology.

Our results illustrate the power of the Nutritional Geometry framework to identify patterns of regulation beyond linear relationships with single macronutrients. Using RNA-seq data from a broad range of diets, we are able to quantify the holistic response of each gene regulation change to macronutrient composition and cluster the gene regulation responses across nutrient space. This clustering analysis identified a positive carbohydrate by negative protein gradient (cluster 1) and a positive protein by negative fat gradient (cluster 2) as key patterns in both the differential splicing and differential expression analyses (Figures 2-3a, 2-5a). Interaction effects such as these would be difficult or impossible to identify from single nutrient analyses or a standard high fat diet paradigm. Clustering gene regulation responses across the nutrient space

encompassed by these 10 diets therefore allows for more precise interpretation of response to diet than was previously possible.

We also showed that both expression and splicing in fat tissue are dynamic in response to environmental change such as dietary composition, and that they act on largely separate gene sets (Figure 2-2). Many studies of gene regulatory responses focus solely on gene expression changes, and in doing so may miss a great deal of impactful regulatory changes that are due to alternative splicing. Here, we see large changes in the transcriptome in response to macronutrient composition that are regulated at the level of alternative splicing and would not have been detected by looking at gene expression alone. Of note, it is possible that the lack of overlap between gene expression and splicing changes is due to lower power to detect changes at the splicing level than at the expression level. However, if that were the case we would expect the splicing results to largely be a subset of the expression results, whereas we identify more genes that are acted on solely by splicing than genes that are impacted by both splicing and expression. We therefore conclude that these findings suggest that thousands of differentially spliced exons represent a concerted cellular response to dietary composition that impacts the transcriptome in a mechanism independent of gene expression, highlighting the importance of giving further consideration to the role of alternative splicing in adipocyte biology and in the response of other tissues to dietary composition.

Our results also have implications for the interpretation of the effects of genetic variants on metabolic traits and diseases. Metabolic disorders such as obesity and diabetes are complex diseases, with both genetic and environmental components.[118,119] When considering the contribution of genetics to these diseases, previous studies have demonstrated that splicing quantitative trait loci (sQTLs) play an important role in disease risk and etiology.[53,54,120] While

some genetic variants have been associated with gene expression and metabolic traits,[26,36,38,121] the functional impact of most GWAS variants associated with metabolic disease remains uncharacterized.[24,122,123] We found that splicing is highly dynamic in response to dietary composition, including alternative splicing changes that may have significant impacts on adipocyte biology. One possible mechanism underlying these uncharacterized GWAS variants is therefore that they modulate splicing, but not gene expression, in response to environmental inputs such as diet. These results highlight the role of splicing in the response of adipose tissue to dietary composition, and provide a foundation from which to consider the impact that genetic variation may have on these processes.

Further research will be required to determine the similarity between the changes seen here and the gene regulatory impact of differences in macronutrient composition in humans. Of particular interest is determining whether there may be any genotype by environment interactions governing these responses in humans, with implications for public dietary guidelines and personalized medicine approaches. Of note, this study was conducted in only young male mice of a single strain, and future studies are needed to determine whether these results replicate fully in both female mice and humans.

Finally, although the diets used in this study were isocaloric per gram of food (i.e. all of the diets had the same energy density), the mice were *ad libitum* fed and there was variation in food intake (and therefore energy intake) across the diets, as show in Figure 1d in Crean et al.[124] In particular, mice on the lowest protein diets (7% of energy from protein) showed increased food intake, as previously reported.[65,125] Although we cannot rule out the possibility that some of the gene regulation changes we see are due to differences in energy intake between the diets, food intake was not significantly correlated with fat content (Supplementary Figure 2-7), which

we identified as the strongest driver of gene regulation changes in this study. As this study focused on isocaloric diets, further research is necessary to consider the impact of energy level on these processes.

Overall, this study utilizes the Nutritional Geometry framework to expand our understanding of the impact of macronutrient composition on metabolic function and gene regulation in adipose tissue. We find that both expression and splicing are highly dynamic across the dietary space, and that considering multiple modes of gene regulation change provides novel insights into the processes underlying the metabolic response to macronutrient composition.

Methods

Animal husbandry

C57BL/6J male mice (n = 60), housed in the Charles Perkins Centre (Sydney, Australia) animal facility (24-26◦C, 44-46% humidity, 12h day/light cycle), were used in this study. Four-week-old mice were purchased from the Animal Resources Centre (Murdoch, Australia) and allowed to acclimate for 3 days before being randomly assigned to dietary treatments. Food and water were supplied ad libitum, mice were weighed weekly, and health checks performed at least twice weekly. Mice were anaesthetized with sodium pentobarbital (100mg/kg) and culled at 21 weeks of age for tissue collection. Gonadal white adipose tissue deposits and liver tissue deposits were weighed, snap frozen in liquid nitrogen, and stored at -80◦C until further use. All procedures were reviewed and approved by the University of Sydney animal ethics committee (project number 2019/1610). We have complied with all relevant ethical regulations for animal use.

Diets

Ten treatment diets, manufactured by Specialty Feeds (Glen Forrest, Australia), were designed to include ingredients of AIN-93G in varying proportions to cover the full range of physiologically viable macronutrient intake space (Figure 2-1). Non-digestible cellulose was included at varying amounts to maintain the net metabolizable energy of diets at 14.7MJ/kg (3.5 kcal/g). Micronutrient content was equal across diets. Protein content was exome-matched to the *Mus musculus* genome,[126] achieved by mixing casein and whey protein isolates supplemented with leucine, threonine, methionine, tyrosine, phenylalanine, tryptophan, alanine, aspartic acid, arginine, glycine, histidine and serine. Omega 3 to omega 6 fatty acid ratio was maintained at 1:3.7 using a combination of soybean oil, linseed oil and lard, with saturated fats making up 23.2% of dietary fats. Carbohydrate sources included wheat starch, dextrinised starch and sucrose at a ratio of 4: 1.3: 1. Individual food intake was measured at 16 and 20 weeks of age by weighing food before and after a 24- hour feeding period. Bedding was changed at the start of intake measures and sifted for food crumbs at the end of the feeding period to obtain as accurate measures of food consumed as possible.

Body composition and metabolic phenotyping

Metabolic phenotyping was completed at 18 weeks of age. Body composition was measured using an EchoMRI-900-A130 (EchoMRI, Houston, USA). Oral glucose tolerance tests were performed after 4h of fasting. Blood samples were obtained by tail tipping and blood glucose measured using a clinical glucometer (Accu-Chek Performa, Roche Diagnostics Australia Pty Ltd). Glucose (2g/kg lean mass) was administered via oral gavage and blood glucose was measured at baseline, 15, 30, 45, 60 and 90 min. Blood from tail tipping was also

used to measure blood insulin at baseline, 15 and 30 min, using an enzyme-linked immunosorbent assay (ELISA) following manufacturer's instructions (Crystal Chem IL).

Bulk RNA extraction and sequencing

Fat tissue was lysed with a 20G needle in Trizol (Life Technologies, #15596018) and total RNA was extracted using the Zymo Direct-zol RNA Miniprep kit (Zymo, #R2052). RNA quantity and quality were measured using the Agilent 2100 Bioanalyzer (Agilent). RNA-seq libraries were generated from 1 μg of total RNA using the NEBNext Ultra II Directional RNA library prep kit (NEBNext, #E7765) and NEBNext Poly(A) mRNA magnetic isolation module (NEBNext, #E7490) with a size selection step to generate 300 bp inserts. The libraries were sequenced using an Illumina NovaSeq 6000 machine (Illumina) with 100-bp paired end reads. Samples were sequenced to an average depth of 84,769,220 reads per sample (54,667,206 – 209,155,949).

10 - 20mg of flash-frozen liver tissue was disrupted using a Dounce homogenizer in ice-cold homogenization buffer (250 mM sucrose, 25 mM KCl, 5mM MgCl2, 20 mM Tricine pH 7.8). The homogenate was mixed 1:3 with Tri-Reagent (Zymo) and passed through a 20g-syringe 10X. The Zymo Direct-zol RNA Microprep kit (Zymo) was used to isolate total RNA according to the manufacturer's recommendations. RNA integrity and concentration was assessed with the Agilent Bioanalyzer Nano kit (Agilent) and 1 ug of mRNA was reverse transcribed and amplified using the NEB Ultra II Directional RNA library prep kit (NEBNext). Sequencing was performed on an Illumina NovaSeq 6000 (Illumina) with 50-bp paired end reads. Samples were sequenced to an average depth of 25,421,980 reads per sample (11,447,079 – 54,008,386).

Response surfaces

Response surfaces were created by fitting a series of mixture models (a.k.a, Scheffe's polynomials) to each response variable in R using the mixexp package. We started by fitting a null model (i.e., intercept only; y ~ 1), before also fitting a linear and a non-linear mixture model, equivalent to equations 1 and 2 in Lawson and Willden.[127] We then selected among models using Akaike Information Criterion, where the simplest model (i.e., fewest terms) within 2 points of the minimal AIC score was selected. In the event that a non-null model was favored we infer an effect of the diet composition on the outcome of interest. To visualize the effects to diet composition we created response surfaces by taking the predicted values from AIC-favored models, and projecting them in to the right-angle mixture triangle (RMT) compositional space.[94] AIC values and a summary of the selected model are provided for all variables in Supplementary Table 2-4.

Differential expression and splicing analysis

RNA-seq reads were aligned to the GRCm39 genome using STAR two-pass mapping[128] and read counts per gene were quantified. For the adipose tissue samples, on average 93.67% of reads per sample uniquely mapped to the genome, with a range of 89.39% – 94.53%, resulting in an average of 79,365,634 mapped reads per sample (51,015,437 – 196,899,410). For the liver tissue samples, on average 83.82% of reads per sample uniquely mapped to the genome, with a range of 78.91% – 90.20%, resulting in an average of 21,322,851 mapped reads per sample (9,577,376 – 43,882,813). Sample quality was assessed using principal components analysis (PCA) and three samples were removed from all genomics analyses in adipose tissue. To test for differential expression, we used edgeR[129] and treated the percent fat and percent carbohydrates in the diets as continuous variables, testing the model $\sim percent\ fat\ +\ percent\ carbohydrates$.

As the percent fat, carbohydrates, and proteins in each diet always sum to 100 the third macronutrient is redundant. Genes with an FDR < 0.05 in this analysis were considered significantly differentially expressed. To test for differential splicing, we used DEXSeq[130] and exons with an FDR < 0.05 in this analysis were considered significantly differentially spliced. In the liver analysis, RNA collection date was included as a covariate. When considering the overlap between differential expression and splicing, genes with at least one significantly differentially spliced exon were considered to be differentially spliced.

Macronutrient correlation

To identify genes or exons with a significant correlation with individual macronutrients, we calculated the Pearson's correlation between the expression of each differentially expressed gene or the exon usage of each differentially spliced exon and the percentage of each macronutrient in the diets. Multiple test correction was performed and correlations with FDR < 0.05 were considered significant.

Fuzzy c-means clustering

Two separate clustering analyses were performed, one for differentially expressed genes and one for differentially spliced exons. All genes that were significantly differentially expressed or exons that were significantly differentially spliced were included. Genes and exons were clustered based on their model coefficients from fitting the model $\sim 0 + percent\ fat + percent\ carbohydrates + percent\ protein$ in edgeR. Model coefficients were centered and scaled to account for differences in expression across genes. Using the e1071 package in R, fuzzy c-means clustering was performed on the centered and scaled model coefficients for every differentially expressed gene or every differentially spliced exon. Genes or exons were assigned

to the cluster for which they had the highest membership. Functional enrichment analysis of the gene sets associated with each cluster was performed using Metascape (http://metascape.org).[95]

Cell type deconvolution of bulk tissue samples

To estimate the cell type proportions of each bulk tissue sample, we performed cellular deconvolution based on gene expression signatures using the DWLS method[112] and a single-cell atlas of mouse white adipose tissue.[77] Any cell type that was estimated at > 1% proportion in at least one sample was considered in further analyses. To determine the relationship between BBS gene expression and cell type proportion, Pearson's correlation with two-sided hypothesis testing was calculated between each differentially expressed BBS gene and each estimated cell type. Multiple test correction was performed via FDR estimation.

Single nucleus extraction and sequencing

Nuclei were isolated from 200-300mg of flash-frozen mouse white adipose tissue according to Van Hauwaert, E. L. et al.[131] Briefly, adipose tissue was finely minced on a petri dish in nuclei isolation buffer (NIB, 250 mM sucrose, 10 mM HEPES, 1.5 mM MgCl2, 10 mM KCl, 0.001% triton-x100, 0.2 mM DTT, 0.5U ul RNase inhibitor). Minced tissue was disrupted with a Dounce homogenizer in NIB, filtered through a 70 um cell strainer and nuclei isolated from contaminating lipids and cellular debris by differential centrifugation. Nuclei were suspended in nuclei resuspension buffer (NRB, 1x PBS with 1% BSA, 2 mM MgCl2, and 0.04 U/ul RNase inhibitor) for 10X genomics. All steps were performed rapidly on ice and all reagents were ice-cold. The resulting suspensions were processed using the Chromium Next GEM Single Cell 3' Kit v3.1 according to the manufacturer's instructions (10X Genomics). Barcoding was performed using the Chromium Controller (10X Genomics) and sequencing was performed on an Illumina NovaSeq X (Illumina) with 100-bp paired end reads.

FASTQ files derived from sequencing were aligned to GRCm38/mm10 genome using 10X Cell Ranger v7.1.0. (10X Genomics). Filtered counts matrices were inspected and then re-filtered for minimum total read count of 1500 Unique Molecular Identifiers (UMIs) per nucleus. After filtering, 18,101 cells from the sample from diet 4 and 11,727 cells from the sample from diet 7 were retained for analysis. Count matrices for both sample data sets and for reference nuclei from Emont et al.[77] were normalized and regressed for cell cycle (s.score, g2m.score) and mitochondrial read percentage using the SCTransform algorithm included as a part of the Seurat R Package.[132] Reference nuclei were subset to include only murine nuclei derived from perigonadal adipose tissue of male mice fed chow diets. Cell-type labels for the two diet samples were learned using the reference dataset using Seurat's FindTransferAnchors and MapQuery functions. UMAP projections were generated using the first 30 principal components of gene expression computed on the normalized reference dataset.

For each sample data set, BBSome gene expression scores were calculated as the sum of normalized, centered, and scaled expression across nine BBS-associated genes of interest (*Bbs1*, *Bbs2*, *Arl6*, *Bbs9*, *Bbs10*, *Bbs12*, *Mks1*, *Ift27*, and *Ift74*) and plotted by the learned cell type groups for each diet.

Statistics and reproducibility

Biological replicates were employed in this study, using n = 6 per diet for all metabolic measures and the liver genomic analyses and n = 5 or 6 per diet for the bulk adipose genomic analyses (n = 6 for diets 1, 5, 6, 7, 8, 9, and 10 and n = 5 for diets 2, 3, and 4). Statistical analyses were performed using R (v. 4.1.1), specifically the DEXSeq (v 1.40.0), edgeR (v 3.36.0), e1071 (v 1.7-9), mixexp (v 1.2.7), and seurat (v 5.0.0) packages. In the differential expression analysis, lowly expressed genes were filtered out (less than 10 reads in 53 or more samples) and in the

differential splicing analysis, lowly expressed exons were filtered out (less than 10 reads across all samples). All data exclusions were performed before performing any analyses. Correlations were measured using Pearson's correlation. Results with an FDR < 0.05 were considered significant.

Data availability

All RNA-seq data generated by this study is available through SRA (bulk and snRNA-seq adipose samples: PRJNA987348, liver samples: PRJNA1043119). All other source data have been provided in the Supplementary Information and Source Data.

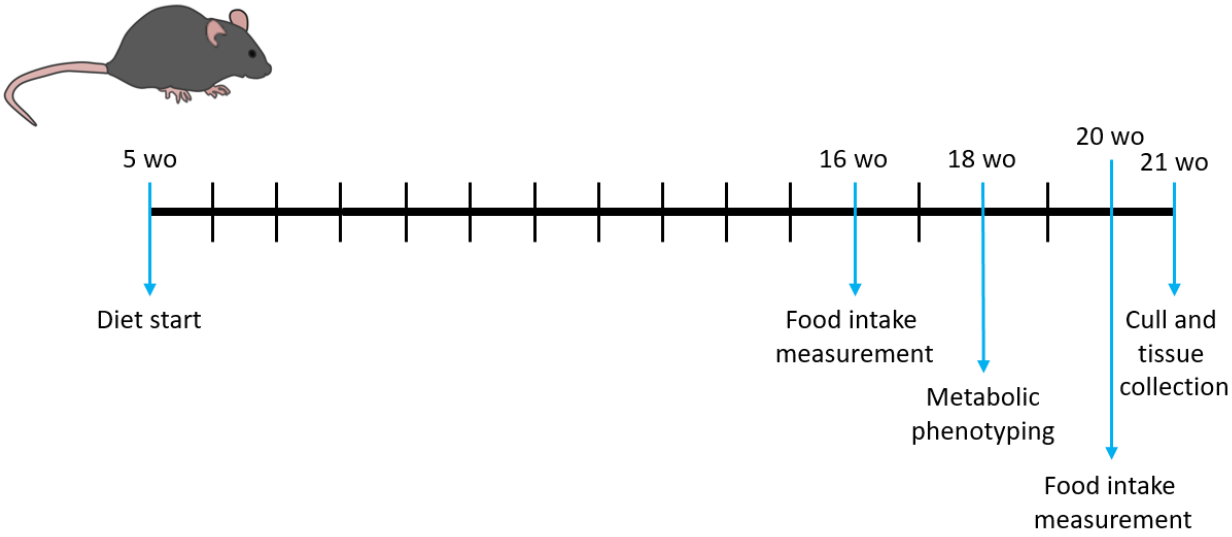
Acknowledgements

This research was funded by a Challenge Programme Grant from the Novo Nordisk Foundation under grant agreement NNF18OC0033754 and the National Institutes of Health (NIH) grants HL128075, U19AI162310 and P30DK020595. The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent research center at the University of Copenhagen, partially funded by an unrestricted donation from the Novo Nordisk Foundation (NNF18CC0034900). K. M. F. was supported by an NIH training grant (T32GM007197).
We thank all members of the GECKO consortium for their comments and valuable suggestions. We thank the University of Chicago Genomics Facility for their assistance with next-generation sequencing. Tamara Pulpitel and Victoria Pye are acknowledged for their assistance with animal work. In Figure 2-1A and Supplementary Figure 2-1, the mouse was sourced from SciDraw (https://scidraw.io/drawing/549) and the adipose cells in Figure 2-1A were adapted from Bioicons (adipocyte-5 icon by Servier). Adipocyte-5 icon by Servier (https://smart.servier.com/) is licensed under CC-BY 3.0 Unported (https://creativecommons.org/licenses/by/3.0/).

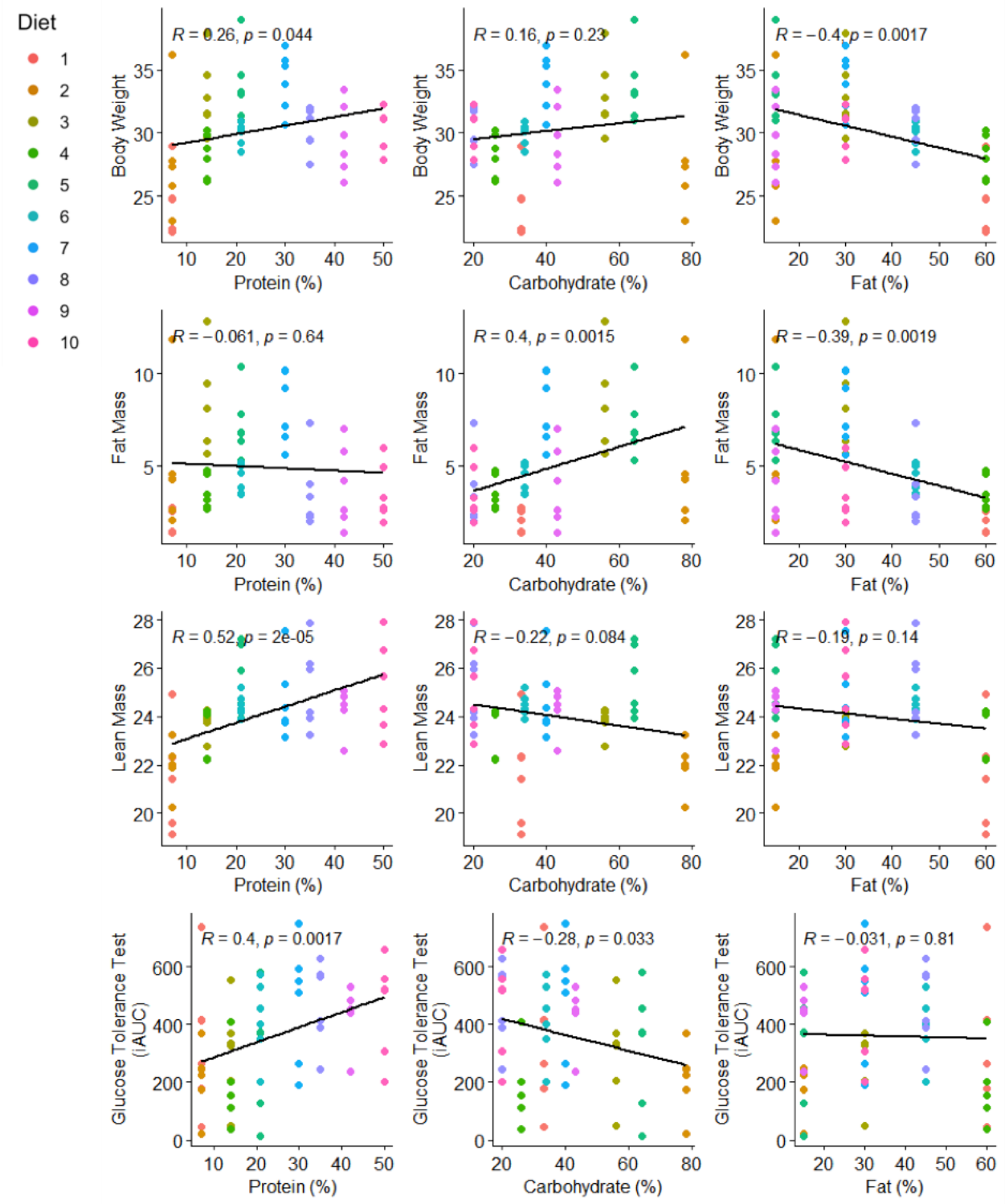
Author contributions

K.M.F. contributed to experimental design, data collection and analysis, and manuscript drafting. A.M.S. contributed to experiment design, data analysis, and manuscript drafting. D.R.S. contributed to data collection. R.M.M. contributed to data collection and manuscript drafting. Z.T.W. contributed to data analysis and manuscript drafting. L.R.I. contributed to data analysis. R.B., S.J.S., and M.A.N. contributed to experimental design, interpretation of results, and manuscript drafting, and provided resources. A.J.C. contributed to experimental design, data collection and analysis, interpretation of results, and manuscript drafting

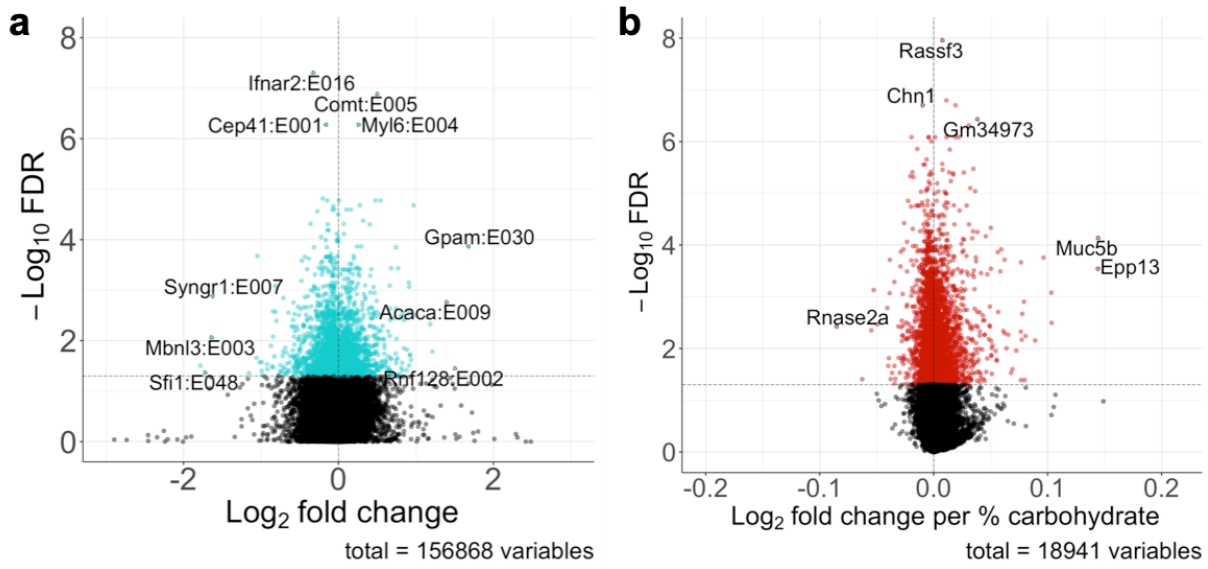
Supplementary Figures for Chapter II



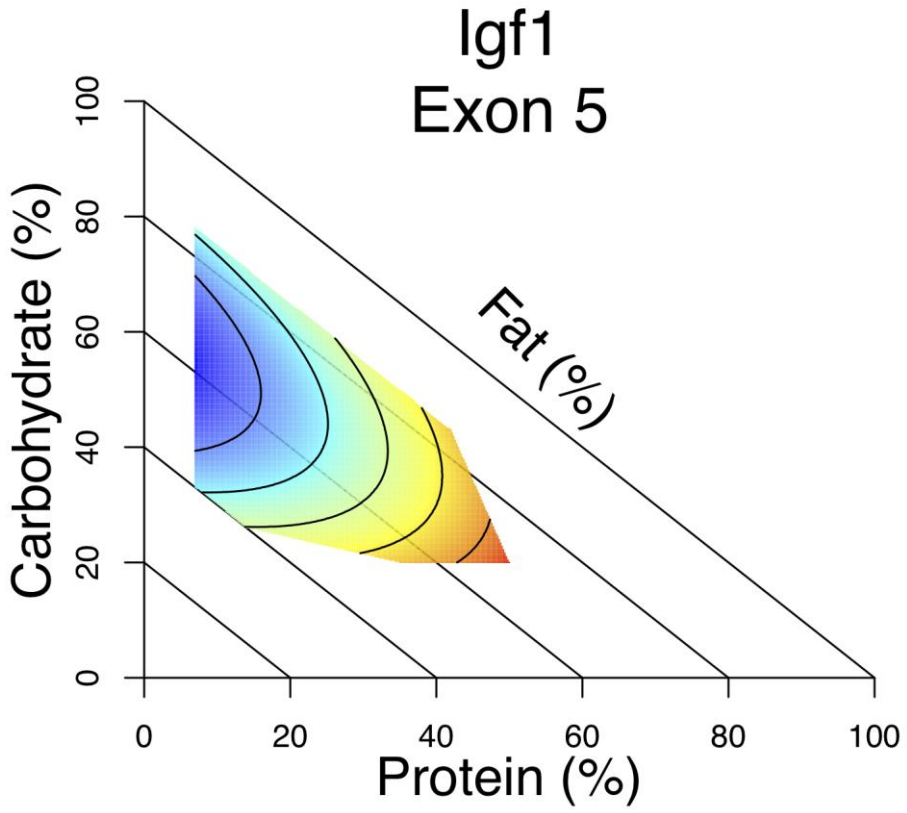
Supplementary Figure 2-1. **Experimental timeline.** Timeline showing the age of the mice (in weeks) when each measurement or intervention was carried out (wo = weeks old).



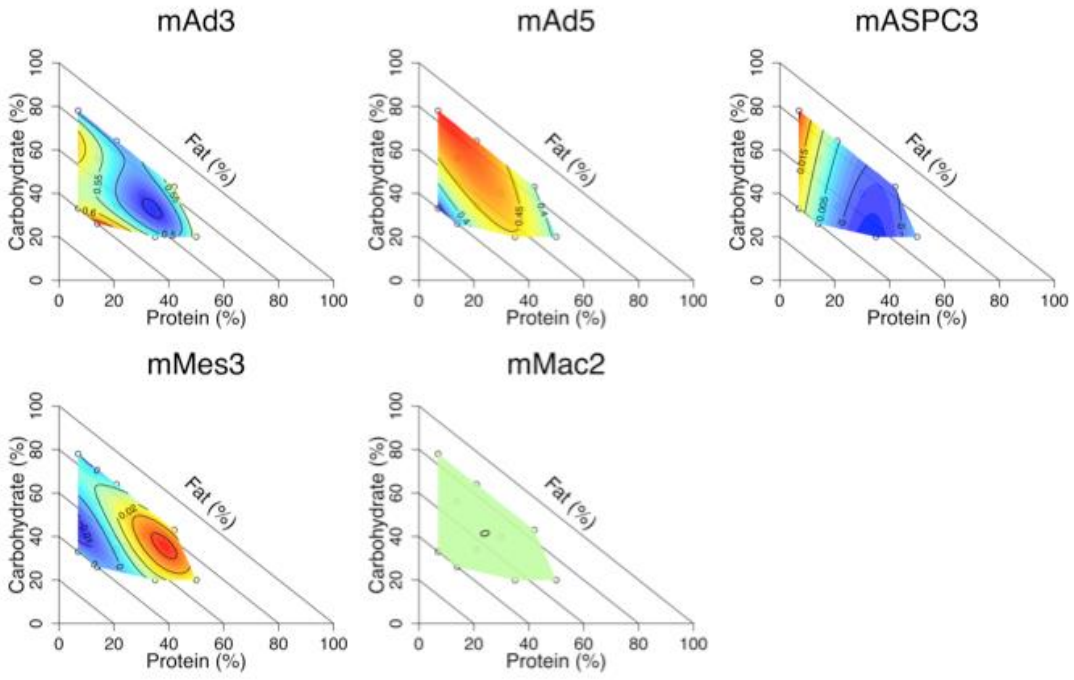
Supplementary Figure 2-2. **Correlations of metabolic traits with each dietary macronutrient.** Plots of the correlation of each dietary macronutrient with each metabolic trait in Figure 2-1. Correlation was assessed using Pearson's correlation with two-sided hypothesis testing. Each point is a mouse, colored by diet. n = 6 mice per diet.



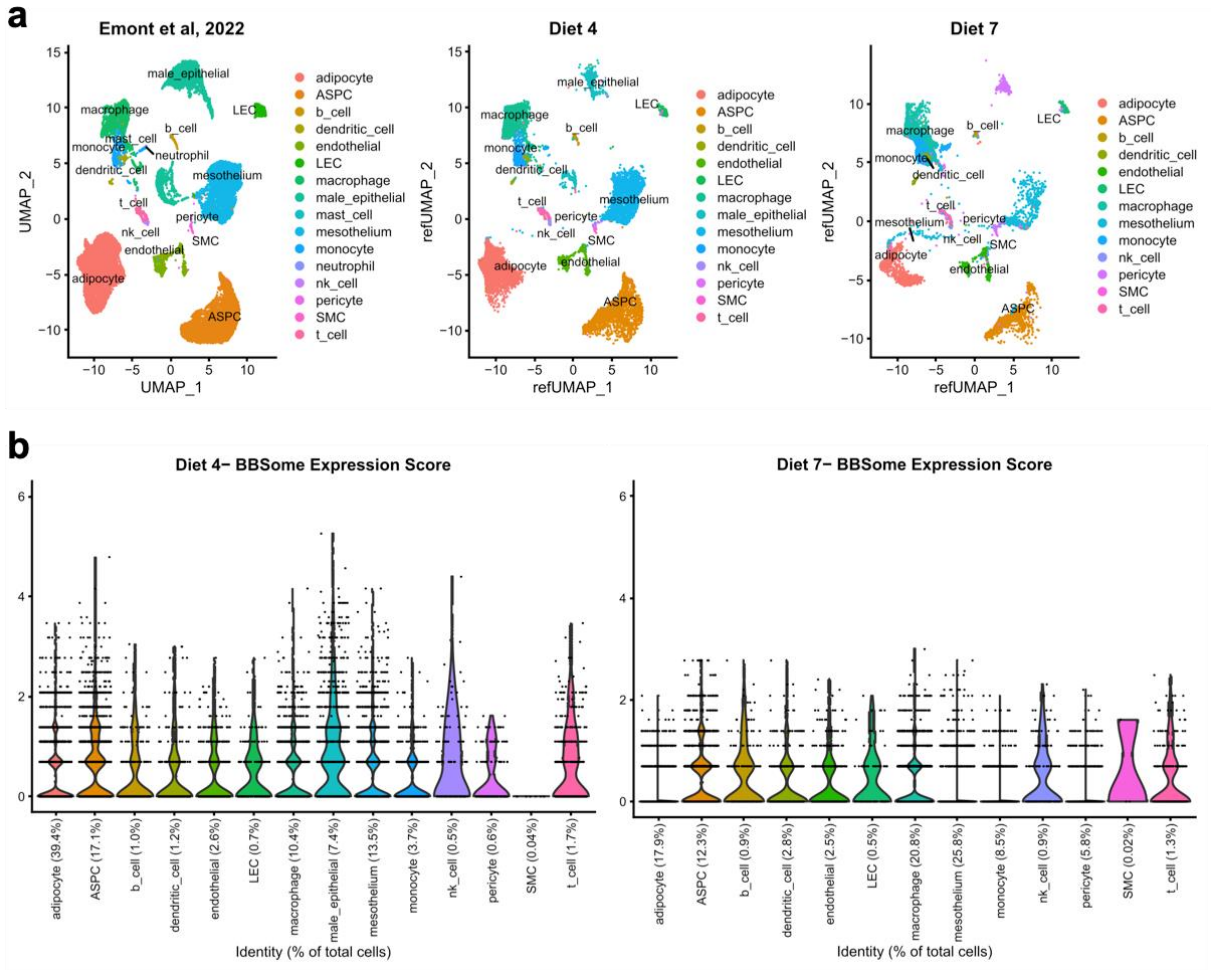
Supplementary Figure 2-3. **Gene regulatory changes relative to dietary carbohydrates.** a. Volcano plot of differential splicing changes, plotting the log fold change between 20% dietary carbohydrates and 78% dietary carbohydrates. Blue dots are significant, black are non-significant. Extreme exons in terms of log fold change or p-value are labeled. b. Volcano plot of differential expression changes, plotting the log fold change per percent dietary carbohydrates. Red dots are significant, black are non-significant. n = 57 mice.



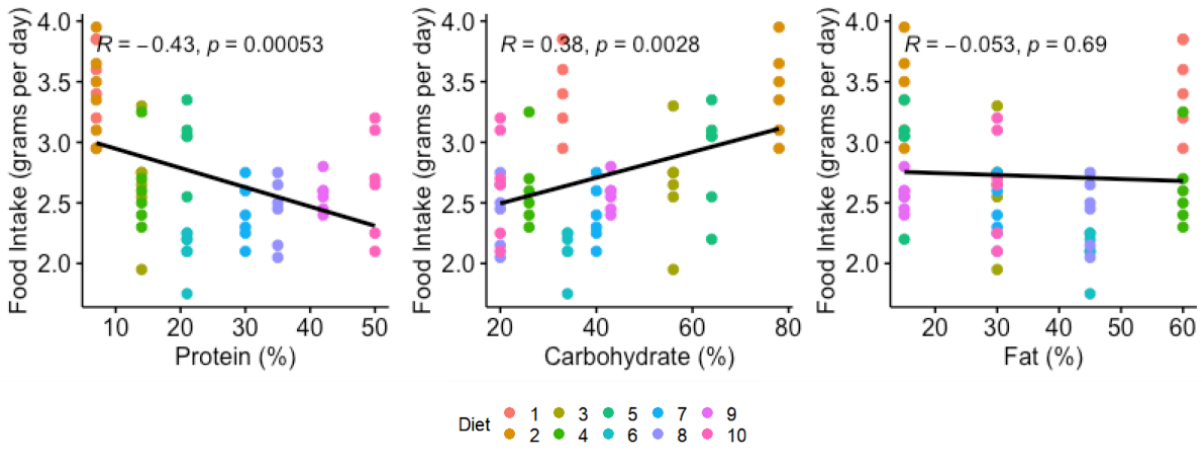
Supplementary Figure 2-4. **Differential splicing of *Igf1* in the liver.** Surface of the centered and scaled exon usage of *Igf1* exon 5 in the liver. n = 6 mice per diet.



Supplementary Figure 2-5. **Computationally estimated cell type proportions.** Surfaces generated from the computationally estimated cell type proportions of each fat tissue sample. For the proportion of macrophages (mMac2), the data supported a null model with no impact of dietary macronutrients. n = 6 for diets 1, 5, 6, 7, 8, 9, and 10 and n = 5 for diets 2, 3, and 4



Supplementary Figure 2-6. **Single nucleus RNA-seq results from diets 4 and 7.** a. UMAP projections of the reference dataset (perigonadal adipose tissue from chow-fed male mice from Emont et al.[77]), the diet 4 cells, and the diet 7 cells. Each point is a cell, colored by cell type. b. Violin plots of the BBSome expression score for each cell type in diet 4 and diet 7. The expression score is based on the nine BBS genes that were identified as differentially expressed in the bulk analysis.



Supplementary Figure 2-7. **Correlations of food intake with each dietary macronutrient.** Plots of the correlation of each dietary macronutrient with food intake. Correlation was assessed using Pearson's correlation with two-sided hypothesis testing. Each point is a mouse, colored by diet. n = 6 mice per diet.

Supplementary Tables for Chapter II

Supplementary Table 2-1. **Metabolic trait values.** Mean and standard deviation for the measurements of each metabolic trait in Figure 2-1 in each diet. n = 6 mice per diet.

Diet	Body Weight	Fat Mass	Lean Mass	Glucose iAUC
1	24.7 ± 2.5	2.04 ± 0.55	21.6 ± 2.1	343 ± 240
2	28.0 ± 4.4	4.95 ± 3.52	21.9 ± 1.0	214 ± 115
3	33.0 ± 2.9	8.03 ± 2.80	23.8 ± 0.5	307 ± 170
4	28.2 ± 1.7	3.58 ± 0.89	23.5 ± 1.0	186 ± 125
5	33.7 ± 2.9	7.23 ± 1.74	25.5 ± 1.4	320 ± 211
6	29.9 ± 0.9	4.28 ± 0.74	24.5 ± 0.4	420 ± 133
7	34.1 ± 2.3	8.15 ± 1.95	24.7 ± 1.6	476 ± 211
8	30.2 ± 1.7	3.57 ± 2.01	25.2 ± 1.8	469 ± 145
9	29.5 ± 2.8	3.88 ± 2.20	24.3 ± 0.9	397 ± 127
10	30.4 ± 1.6	3.59 ± 1.53	25.2 ± 1.9	462 ± 171

Supplementary Table 2-2. **Differential splicing and expression results for all genes in adipose tissue.** Available as an excel file online.

Supplementary Table 2-3**. Differential expression results for known adipose splicing factors.**

Splicing Factor	Differential expression corrected p-value	References
Nova1	0.125	Vernia et al[74]
Nova2	0.283	Vernia et al[74]
Sam68/Khdrbs1	0.175	Huot et al,[133] Song and Richard[134]
Srsf1	0.158	Aprile et al[135]
Srsf10	0.851	Li et al[136]
Tra2b/Sfrs10	0.498	Patel et al,[137] Mikoluk et al[138]

Supplementary Table 2-4**. AIC values and mixture model summaries for all variables.** Available as an excel file online.

Chapter III: Splicing across adipocyte differentiation is highly dynamic and impacted by metabolic phenotype

Note: The following chapter is reproduced from the manuscript titled "Splicing across adipocyte differentiation is highly dynamic and impacted by metabolic phenotype". An earlier version of this manuscript was published on Research Square in 2023.[139]

Abstract

Adipose tissue dysfunction underlies many of the metabolic complications associated with obesity. A better understanding of the gene regulation differences present in metabolically unhealthy adipose tissue can provide insights into the mechanisms underlying adipose tissue dysfunction. Here, we used RNA-seq data from a differentiation time course of lean, obese, and obese with T2D individuals to characterize alterative splicing in adipocyte differentiation and function. Splicing was highly dynamic across adipocyte differentiation in all three cohorts, and the dynamics of splicing were significantly impacted by metabolic phenotype. There was very little overlap between genes that were differentially spliced in adipocyte differentiation and those that were differentially expressed, positioning alternative splicing as a largely independent gene regulatory mechanism whose impact would be missed when looking at gene expression changes alone. To assess the impact of alternative splicing across adipocyte differentiation on genetic risk for metabolic diseases, we integrated our differential splicing results with GWAS results for BMI and T2D, and found that variants associated with T2D were enriched in regions that were differentially spliced in early differentiation. These findings provide insight into the role of alternative splicing in adipocyte differentiation and can serve as a resource to guide future variant-to-function studies.

Introduction

Metabolic diseases represent a significant global health burden, with obesity and its often-associated cardiometabolic complications representing the largest contributor of this health burden.[140] A key characteristic of the metabolic complications of obesity is adipose tissue dysfunction, with adipose tissue from obese individuals exhibiting abnormal phenotypes such as hypertrophy, increased inflammation, decreased differentiation capacity, and impaired insulin sensitivity.[71,72,78] These differences in obese adipose tissue function may already be present in adipocyte precursors such as preadipocytes, or may develop over the course of differentiation or in mature adipocytes.[78–80] Therefore, by considering the transcriptomic and epigenomic differences between lean and obese individuals across the differentiation of preadipocytes to mature adipocytes we can gain insights into the pathophysiological mechanisms of obesity and its comorbidities.

In a previous study, our collaborators demonstrated that preadipocytes from obese subjects with or without type 2 diabetes (T2D) have distinct transcriptomic and epigenomic changes across differentiation when compared to lean individuals.[79] In all three cohorts, the transcriptome was significantly remodeled across differentiation, with thousands of differentially expressed genes between each time point comparison. Further, preadipocytes isolated from obese individuals exhibited reduced adipogenic potential that may be linked to differences in DNA methylation, suggesting that preadipocytes may be epigenetically reprogrammed *in vivo* in response to obesity and T2D. This work therefore showed that gene expression and DNA methylation are significantly altered in preadipocytes isolated from obese and obese with T2D individuals when compared to lean individuals, and that there are significant transcriptomic differences between these cohorts across adipocyte differentiation that impact genes involved in

important pathways such as PPAR signaling, oxidative phosphorylation, fatty acid metabolism, and insulin signaling. These results provided insights into possible mechanisms underlying adipose tissue dysfunction in obesity and T2D.

It has therefore been shown that gene expression and DNA methylation changes across adipocyte differentiation are important for adipocyte biology and function. However, other forms of gene regulation may also play important roles in adipocyte function, such as alternative splicing. Specifically, alternative splicing can lead to variations in exon usage in mature mRNAs, resulting in distinct protein domains being present in the cell. This represents a regulatory strategy for modifying the cellular proteome which is independent of detectable gene expression differences measured by RNA-seq. Previous work has shown that alternative splicing in adipocytes is dynamic in response to diet-induced obesity in mice, and can be critical for key adipocyte functions such as thermogenesis.[74,141] However, relative to other modes of gene regulation, such as transcriptional regulation, the role of alternative splicing in adipocyte differentiation and function remains understudied. Splicing may play an essential role in the function of preadipocytes and adipocytes, and may underlie some differences between adipose tissue function in lean and obese individuals. In this study, we aim to systematically characterize the dynamic scope of alternative splicing during adipocyte differentiation in health and disease.

In addition to providing insights into adipocyte function and dysfunction, a better understanding of the role of splicing in adipocyte differentiation may also provide insights into the genetic underpinnings of metabolic disorders. Genome-wide association studies (GWAS) have identified hundreds of loci associated with metabolic disorders such as obesity and T2D, but the mechanisms of action for most of these genetic associations remain unclear.[20,23] By expanding our understanding of the scope of splicing in adipocyte differentiation and key

adipocyte functions, we may be able to identify variants associated with metabolic disorders that confer risk through disrupting splicing during adipocyte differentiation, thus leading to impaired or altered adipocyte function and eventually disease. A better understanding of the splicing events that occur across adipocyte differentiation and how they are perturbed in obesity may therefore provide insights into the genetic underpinnings of metabolic disorders.

Here, we extend our collaborators' previous study to consider changes in splicing across adipocyte differentiation between lean, obese, and obese with T2D humans and investigate whether any identified splicing differences may play a role in disease risk. Focusing on a subset of individuals that were deeply sequenced to identify a large number of splicing events, we identified thousands of differential splicing events across adipocyte differentiation in each cohort. We found that these splicing events largely occurred in genes that were not differentially expressed and would thus not be detectable by gene expression analyses, emphasizing the importance of considering splicing in addition to expression when measuring gene regulation changes. Finally, we considered the role these splicing events may play in the genetic risk for obesity and T2D, providing a resource to potentially help disentangle previously uncharacterized genetic associations with metabolic disease.

Results

Identifying splicing changes across adipocyte differentiation

To investigate the role of splicing in adipocyte differentiation, we used RNA-seq data collected from a differentiation time course of preadipocytes isolated from visceral adipose tissue of lean, obese, and obese with T2D subjects[79] (Supplementary Table 3-1). Isolated preadipocytes were plated and proliferated to confluence, and once they reached confluence they were induced to differentiate into adipocytes. Aliquots for RNA-seq were collected at confluence, day 3, and

day 15 of differentiation. Using data collected from a subset of the original cohort (three individuals per cohort per time point, see Supplementary Table 3-2), we tested for differential splicing and differential expression across adipocyte differentiation in each of the three cohorts. For each cohort, we used DEXSeq[130] to test for differential splicing between confluence and day 3 and between day 3 and day 15 of differentiation, and for each time point comparison in each cohort exons with an FDR < 0.05 were considered significantly differentially spliced.

We found that alternative splicing is dynamic across adipocyte differentiation, with the number of differential splicing events identified at each time point comparison ranging from 1999 (in the obese with T2D cohort at confluence vs. day 3) to 9229 (in the obese with T2D cohort at day 3 vs. day 15) (Figure 3-1A, Supplementary Table 3-3). However, the dynamics of splicing were different between the groups, with an especially pronounced difference in the obese with T2D cohort, where we saw a large number of differential splicing events during late differentiation. This pattern was the reverse of what we saw in the other two cohorts, in which there were more differential splicing events during early differentiation and fewer during late differentiation. This pattern was not apparent in the differential expression data, where all three cohorts had largely similar numbers of differentially spliced genes in later differentiation (Figure 3-1B). These results demonstrate that differential splicing is pervasive and dynamic across adipocyte differentiation, that those dynamics are impacted by diseases status, and that they often are not mimicked or observable in expression data.

We also investigated whether we identified any splicing events with previous evidence of a role in adipocyte biology and adipogenesis. Previous work has identified naturally occurring alternative splicing variants in *PPARG*, one of the master regulators of adipocyte differentiation. In particular, exon skipping of exon 5 produces an isoform of *PPARG* that lacks the ligand-

binding domain and leads to impaired differentiation. Further, expression of this isoform in subcutaneous adipose tissue has been shown to positively correlated with BMI.[135] In our data, consistent with previous reports, we saw that exon 5 of *PPARG* was significantly differentially spliced between day 3 and day 15 in our obese cohort, with exon 5 showing lower exon usage at day 15 than day 3 (Supplementary Table 3-3). In addition to *PPARG*, we also identified a known alternative splicing event in the gene *LPIN1*. *LPIN1* is involved in adipogenesis and mutations in the gene in mice lead to lipodystrophy.[142] Two isoforms of *LPIN1* have been identified, lipin-α and lipin-β, that differ in exon 7, with lipin-β containing exon 7 and lipin-α lacking it. These two isoforms have unique temporal expression patterns across adipocyte differentiation, with lipin-α expression peaking in early differentiation and thought to be involved in adipocyte differentiation and lipin-β expression peaking in late differentiation and thought to be involved in lipogenesis.[143] In each of our cohorts, we see increased usage of exon 7 between day 3 and day 15 of differentiation, representing this previously identified isoform switch (Supplementary Table 3-3) Some previously identified splicing events, such as alternative splicing of *DLK1/PREF1*[144] and *INSR*,[145] were not identified in our data.

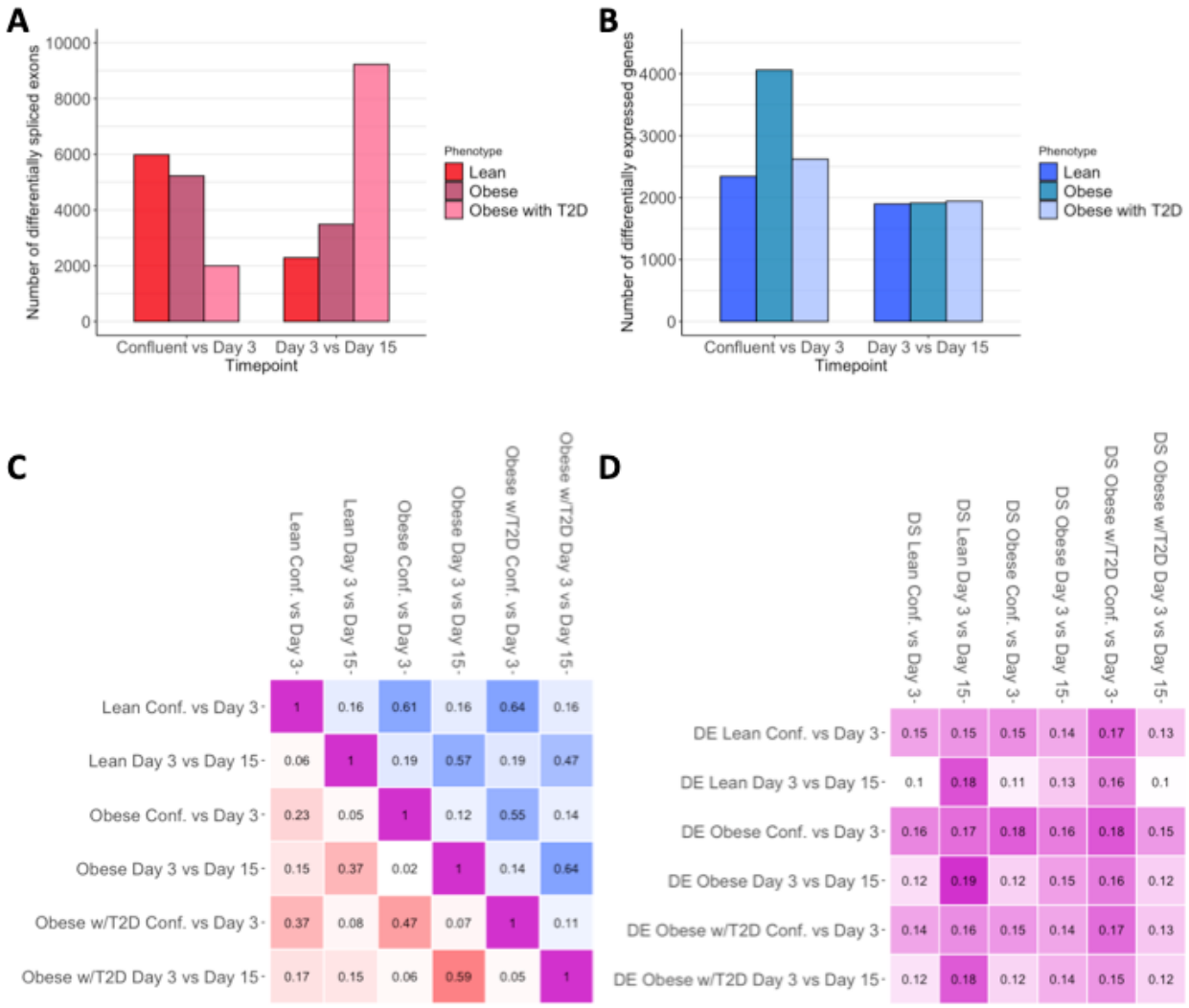


Figure 3-1. **Characterizing differential splicing across adipocyte differentiation in lean, obese, and obese with T2D cohorts.** A. Bar plot of the number of significantly differentially spliced exons in each cohort and each time point comparison. B. Bar plot of the number of significantly differentially expressed genes in each cohort and each time point comparison. C. Heat map of the pairwise sharing between the differential splicing sets (red) or between the differential expression sets (blue). D. Heat map of the pairwise sharing between differential expression sets and differential splicing sets. Pairwise sharing was calculated using MASH.

Investigating patterns of gene regulation sharing between groups

After observing that there was abundant differential splicing across adipocyte differentiation in each of the three cohorts, we next investigated whether the differential splicing changes we observed were shared across the groups, and how the dynamics of splicing sharing

compared to expression. Using MASH,[146] we quantified the degree of sharing between each cohort both within and across time points (Figure 3-1C). We found that, overall, most of the differential splicing events identified were unique to each cohort. For example, at confluence vs. day 3 the splicing changes in the lean cohort have an estimated pairwise sharing of only 23% and 37% with the obese and obese with T2D groups respectively. The obese and obese with T2D cohorts showed higher sharing overall, with an estimated pairwise sharing of 47% at confluence vs. day 3 and 59% at day 3 vs. day 15, indicating more similarity between the splicing patterns in the two obese cohorts than between either obese cohort and the lean cohort. These results indicate that there are distinct differential splicing profiles between all three cohorts, with the lean cohort in particular showing low degrees of sharing with either of the obese cohorts.

We also investigated the degree of sharing among the differentially expressed genes, and found that they showed overall higher degrees of sharing than differential splicing (Figure 3-1C). For differential expression, the pairwise sharing between any two cohorts at the same time point ranged from 47-64% as compared to 15-59% for differential splicing. For differential expression, the obese cohorts had higher degrees of sharing at later differentiation than early differentiation (64% vs 55%), in contrast to the lean cohort which had higher pairwise sharing with the obese cohorts at early differentiation as opposed to late (61% and 64% vs 57% and 47%). The amount and patterns of sharing between the cohorts therefore differ between differential splicing and differential expression, suggesting that these two forms of regulation are being affected differently by changes in phenotype.

In addition to investigating the degree of sharing within differential expression and differential splicing, we also estimated the degree of sharing between these two modes of gene regulation. We summarized the differential splicing results to the gene level by considering the

most significant differential splicing change for each gene and used MASH to determine the pairwise sharing between each set of differential splicing changes and each set of differential expression changes. Overall, we found that the sharing between expression and splicing changes in this context is very low, ranging from 10-19% (Figure 3-1D), indicating that analysis of gene expression alone would miss thousands of splicing events that lead to quantitative and qualitative changes in the proteome of adipocytes. These results underscore the importance of considering splicing in addition to expression when measuring the transcriptomic response of adipocytes, as splicing changes can represent a largely orthogonal form of genetic regulation that is not captured by looking at expression changes alone.

Characterizing splicing dynamics by clustering analysis

After detecting pervasive differences in alternative splicing across differentiation that were distinct between the cohorts and largely not overlapping with gene expression changes, we next asked whether there might be distinct groups of alternative splicing events involved in specific biological processes in each cohort. We therefore sought to classify the differential splicing events we identified in each cohort into distinct response patterns. This analysis allows us to ask not just whether individual splicing events are shared or not, but whether the dynamic patterns of splicing changes overall are different between the groups. To address this question, we used fuzzy c-means clustering to cluster all the differential splicing events from each cohort across the three differentiation time points considered (confluence, day 3, and day 15). We performed a separate clustering analysis for each cohort that assigned each differentially spliced exon to one of six distinct clusters based on its dynamics across the time course (Supplementary Table 3-4).

Using this approach, we find that the clusters are largely shared between cohorts, although with some notable differences (Supplementary Figures 3-1, 3-2, and 3-3). Specifically, cluster 6 in the lean cohort, which contained exons whose usage was largely unchanged from confluence to day 3 and increased from day 3 to day 15, was not identified as one of the six strongest clusters in the obese cohort, where we instead identified the opposite - a cluster of exons whose usage was unchanged from confluence to day 3 and then decreased from day 3 to day 15. The other five clusters that were identified in the lean cohort represent dynamic patterns that were the same or very similar to clusters that were identified in the other two cohorts.

However, although five of the six clusters were largely similar across the three cohorts, there were notable differences in the prevalence of each splicing pattern between the cohorts (Figure 3-2A). When we consider the percent of total differential splicing events assigned to each cluster for each group, we find that cluster 2, which is comprised of exons whose usage decreases between confluence and day 3 and increases between day 3 and day 15, is much more prevalent in the obese and obese with T2D cohorts as compared to the lean (23% and 22% of exons compared to 15%). There were also differences in cluster prevalence between the two obese cohorts, with the obese with T2D cohort having a marked decrease in exons assigned to cluster 5 (exons whose usage decreases across all three time points) compared to both the lean and obese cohort, and the obese cohort having a marked increase in exons assigned to cluster 1 (exons whose usage increases between confluence and day 3 and is largely unchanged from day 3 to day 15) compared to both the lean and obese with T2D cohort. The three metabolic cohorts therefore exhibited distinct dynamics of alternative splicing across adipocyte differentiation.

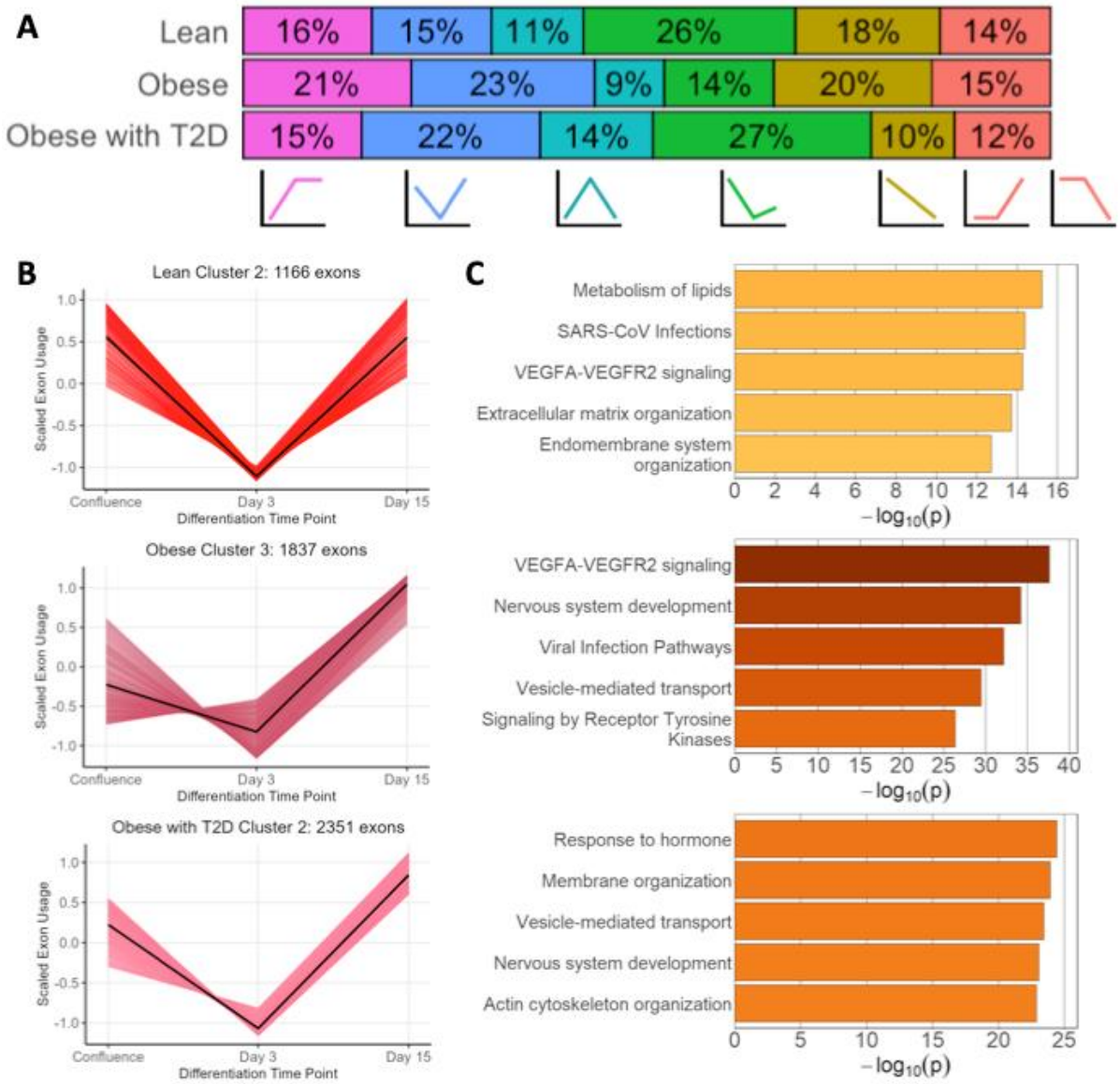


Figure 3-2. **Functional enrichment differences between lean, obese, and obese with T2D cohorts within a shared splicing cluster.** A. Plot showing the percent of all differentially spliced exons assigned to each of the six clusters for each cohort. Clusters are identified by a simplified representation of the cluster dynamics. B. Plots of the scaled and centered exon usage of each exon assign to lean cluster 2, obese cluster 3, and obese with T2D cluster 2. For each plot, the black line connects the cluster centroids of the cluster. C. The five most significantly enriched functional terms for each cluster, plotted against the $-\log_{10}$ of the enrichment p-value for each term.

Finally, in addition to changes in splicing dynamics across differentiation, we also observed differences in the functional enrichment of splicing events, even within a shared cluster. Lean cluster 2, which represents exons whose usage decreases between confluence and

day 3 and then increases between day 3 and day 15, is also found in the obese and obese with T2D cohorts (Figure 3-2B). However, functional enrichment analysis revealed key differences between the exons assigned to these three clusters (Figure 3-2C). Notably, the exons assigned to lean cluster 2 are enriched for functional terms such as metabolism of lipids, VEGFA-VEGFR2 signaling, and extracellular matrix organization, indicating that this may be an important splicing dynamic for key adipocyte functions. However, the obese and obese with T2D cohorts enriched for less relevant terms such as nervous system development and viral infection pathways (although the obese cohort is also enriched for VEGFA-VEGFR2 signaling). These results demonstrate that there are both qualitative and quantitative changes in the overall dynamics of splicing between the three cohorts, and that the genes and functions being acted on by these splicing patterns have shifted in ways that may have implications for adipocyte function and development.

Role of alternative splicing across adipocyte differentiation in GWAS for BMI and T2D

After observing that splicing is highly dynamic across adipocyte differentiation and involved in key adipocyte functions that may be perturbed in obesity, we next asked whether these splicing events might play a role in the genetic risk for metabolic diseases such as obesity and T2D. GWAS have identified hundreds of SNPs associated with metabolic disorders, and the mechanisms by which these SNPs confer risk remains largely unknown.[24,122] One possible mechanism of action for some SNPs associated with metabolic disorders is that they disrupt key splicing events during adipocyte differentiation, leading to impaired adipocyte differentiation or function. To assess whether any SNPs associated with BMI or T2D may be impacting splicing across adipocyte differentiation, we overlapped the flanking introns of differentially spliced exons in each cohort and time point with SNPs significantly associated with BMI or T2D in two

large meta-analyses[20,23] (Supplementary Table 3-5). For each differential splicing analysis, we compared the number of introns overlapping at least one SNP with a null distribution generated from the flanking introns of randomly selected exons. The exons in each control set were randomly sampled from all tested exons in that analysis, removing exons or genes that did not meet the minimum expression threshold for differential splicing testing.

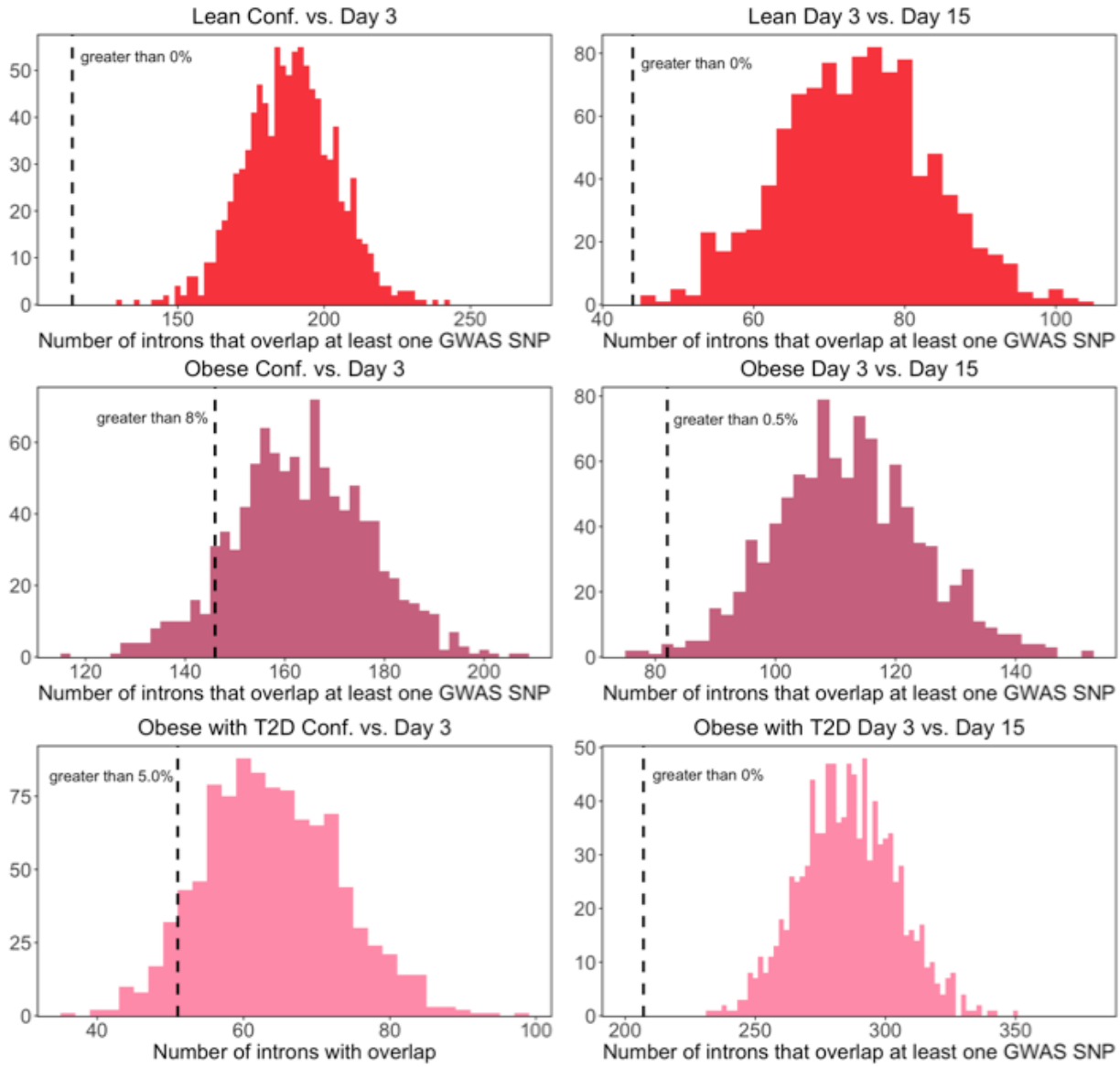


Figure 3-3. **Overlap between BMI GWAS SNPs and differentially spliced exons.** Histograms showing the distribution of the number of introns that overlap at least one BMI SNP across 1000 control sets of randomly selected exons. The dotted black line indicates the number of introns that overlap at least one BMI SNP in each set of differentially spliced exons.

When we performed this analysis using SNPs significantly associated with BMI, we found that there was no significant enrichment for SNPs associated with BMI in the introns of any of the differential splicing analyses tested (Figure 3-3). Instead, we found that in almost every comparison there was a significant depletion of SNPs associated with BMI in the introns of differential spliced exons. This result is consistent with previous work that has found that genetic associations with BMI are strongly enriched around genes that are specifically expressed in the central nervous system.[38,121,147] This suggests that gene sets that are enriched for genes with specific functions in adipocytes would likely be depleted in SNPs associated with BMI. These results emphasize the complex nature of metabolic disorders and the many possible modes of action that can underlie GWAS associations.

Next, we tested for significant enrichment of SNPs associated with T2D in each of our differential splicing analyses (Figure 3-4). We found that SNPs associated with T2D were significantly enriched in the flanking introns of exons that were differentially spliced between confluence and day 3 of differentiation in each of the three cohorts, but not significantly enriched in any cohort at day 3 vs. day 15 of differentiation. These results indicate that exons that are differentially spliced in the early stages of adipocyte differentiation are more likely to be functionally involved in disease risk for T2D than those that are differentially spliced in the later stages of differentiation. The differential splicing events identified here therefore provide a resource to investigate the role of differential splicing in T2D disease risk and potentially unravel the function of previously uncharacterized genetic associations with T2D. Overall, these results emphasize the importance of considering splicing as a possible disease mechanism in metabolic disorders, as well as the need to collect data across developmental time to capture possible transient associations or gene regulation changes.

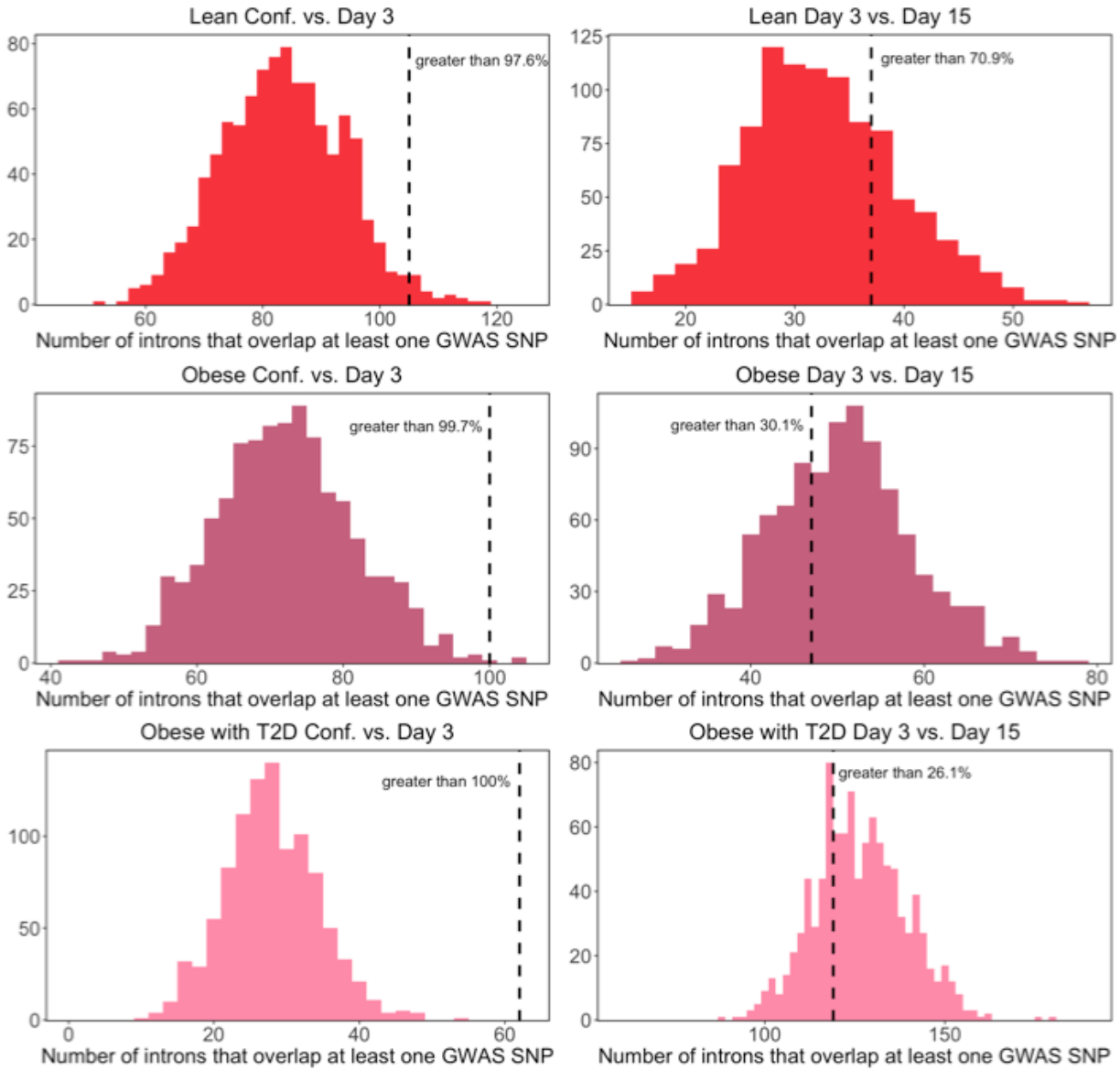


Figure 3-4. **Overlap between T2D GWAS SNPs and differentially spliced exons.** Histograms showing the distribution of the number of introns that overlap at least one T2D SNP across 1000 control sets of randomly selected exons. The dotted black line indicates the number of introns that overlap at least one T2D SNP in each set of differentially spliced exons.

Discussion

In this study, we aimed to expand our understanding of the impact of splicing in adipocyte biology and metabolic disease. Using isolated preadipocytes collected from lean, obese, and obese with T2D individuals, we were able to generate a comprehensive catalog of

alternative splicing changes across adipocyte differentiation in three different metabolic states. We identified shared and divergent splicing patterns between the three cohorts, as well as functional differences within a shared splicing pattern. Finally, we integrated our splicing results with GWAS results for BMI and T2D, generating a resource to help untangle the association between some genetic variants and disease.

Our results emphasize the importance of considering splicing in addition to expression when measuring gene regulation, as it can act as an independent regulatory mechanism from expression. We found very little overlap between differentially spliced and differentially expressed genes in every cohort and adipocyte differentiation time point assayed, with no comparison having a pairwise sharing estimate higher than 20%. This is consistent with the small overlaps found in comparisons between alternative splicing and expression in other contexts, such as genes identified as differentially spliced and expressed across the cell cycle,[148] in the aging hippocampus in mice,[149] and in a rodent model of sarcopenia.[150] This suggests that when considering the gene regulatory response to a perturbation or time course, it is essential to consider splicing as well as expression to capture the full spectrum of gene regulation events and changes that shape the cellular proteome, both quantitatively (measured by gene expression differences) and qualitatively (measured by splicing differences).

By expanding our understanding of the role of splicing in adipocyte differentiation and function, we can gain insight into the mechanisms of genetic risk for metabolic diseases. In the past two decades, GWAS for metabolic traits such as BMI, T2D, and WHR have provided evidence for the association of hundreds of noncoding variants with metabolic disease.[24,26,118,122] In an effort to elucidate the mechanism underlying these associations, previous work has used sQTL mapping to established the regulation of alternative splicing as an important linking

mechanism between variant and disease.[52,53,151] However, some molecular QTLs have been shown to be transient over differentiation time, and would likely not be detectable in mature tissues.[152] Here, we collected data from multiple time points across the differentiation of preadipocytes to adipocytes, allowing us to capture splicing events and possible variant-splice event pairs that may have been missed by looking at preadipocytes or adipocytes alone. By linking GWAS variants to splicing events across adipocyte differentiation, we can identify otherwise undetectable mechanisms possibly underlying the association of those variants with metabolic disorders. Future studies that expand the number of individuals for which we have RNA-seq time-course data could extend these results through formal sQTL mapping.

Of note, we found that there was a significant depletion of SNPs associated with BMI in the introns of differential spliced exons in many of the cohorts and time points. One possible explanation for this depletion is the previously characterized enrichment of genetic associations with BMI in genes highly expressed in the central nervous system.[121,147] If our sets of differentially spliced genes across adipocyte differentiation are enriched for genes highly expressed and functional in adipose tissue, they would therefore likely be depleted in BMI-associated SNPs. It is also important to note that we performed the GWAS-intron intersection analyses using the entire flanking introns of each exon in order to capture all possible intronic variants. However, previous work has shown that splicing variants tend to occur more frequently closer to the intron-exon border, and that the most proximal 200-400 bp of intronic sequence may be most important for splicing regulation.[153–155] Although deep intronic variants can be involved in splicing, it is possible that by including the entire intron we are diluting the signal of enrichment of GWAS variants around differentially spliced exons by counting deep intronic variants in the control sets and weighing them equally with more proximal variants. Of note,

although we do not detect a significant enrichment of BMI-associated variants in the differentially spliced regions, we still identify hundreds of exons whose flanking introns contain BMI-associated variants that may act by perturbing splicing in the region.

In conclusion, we have demonstrated that splicing is highly dynamic in adipocyte differentiation, is impacted by metabolic phenotype, and is acting on important adipocyte functions such as lipid metabolism and angiogenesis. We have also integrated our splicing results with genetic variants associated with BMI or T2D and generated a set of putative variant-splice event pairs that may be relevant in disease function. These results expand our understanding of the role of splicing in adipocyte differentiation and metabolic disease etiology, and can act as a resource to guide further research into the genetic underpinnings of metabolic disorders and the impact of individual splicing events on adipocyte biology.

Methods

Study participants

The study was approved by the Ethics Committee from the Capital Region of Denmark (reference H-1-2011-077) and informed consent was obtained from all participants. This study included a total of five lean controls, five obese subjects with T2D according to ICPC-2-DK, and four obese subjects with no history of diabetes. The participants were recruited from Surgical Gastrointestinal Department, Hvidovre Hospital, Denmark. The lean controls were subjects undergoing surgery for laparoscopic inguinal hernia repair. Individuals of both the Obese T2D and Obese groups were subjects to laparoscopic gastric bypass operation. Prior to surgery, all study participants were measured and weighted. Exclusion criteria for all three groups were: alcohol consumption of more than 14 units/week, smoking, daily intake of medicine and presence of chronic/acute diseases. Lean men with diagnosed hypercholesterolemia,

hypertension and/or diabetes were excluded. Participants were fasted for at least 12 hrs and blood was drawn before undergoing anesthetics. Blood was analyzed at the Clinical Biochemistry Department, Hvidovre Hospital. Visceral adipose tissue was collected from the omental fat pat with laparoscopic surgery instruments under full narcosis during surgery.

Isolation and culture of human preadipocytes

Isolation and culture of preadipocytes was performed as previously described.[156] The adipose tissue biopsy was immediately rinsed in Phosphate-buffered saline (PBS), minced and digested by collagenase for 2½ hours in 37°C water bath shaking. Digestion was stopped by adding Dulbecco's Modified Eagle Medium (DMEM) media supplemented with 10% Fetal Bovine Serum (FBS). The suspension was passed through a 200-µm sterile nylon filter (Spectrum Laboratories). The stromal vascular fraction (SVF) from the infranatant and the mature adipocytes from the upper fraction were washed 3 times with DMEM. The SVF was further processed through a 40-µm cell strainer and washed once in DMEM. Cells were plated at 75x106 cells/80 cm2 flask and cultured at 37°C (95% air/5% CO2) in DMEM/F12, 10% (v/v) FBS, 100 U/ml penicillin and 100 mg/ml streptomycin until 3 days prior to induction of differentiation, where FBS was removed from the media. At day 0, cells were differentiated in 5 µg/ml insulin, 10 µg/ml transferrin, 0.2 nM tri-iodothyronine (T3), 1 µM rosiglitazone, 50 µM 3-isobutyl-1-methylxanthine (IBMX) and 1 µM dexamethasone for the first 3 days. Thereafter, IBMX and dexamethasone were removed. Insulin was removed at day 12 and cells were processed at day 15.

Nucleic acid purification and RNA sequencing

RNA and DNA were isolated with AllPrep DNA/RNA/miRNA Universal Kit (QIAGEN) according to the manufacturer’s protocols. Quality and yield were assessed by NanoDrop and

Qubit dsDNA HS Assay Kit (Life Technologies). For RNA undergoing RNA-seq library preparation, RIN value was determined by Bioanalyzer instrument (Agilent Genomics), using the Agilent RNA 6000 Pico Kit. RNA-sequencing libraries were prepared using the Illumina TruSeq Stranded Total RNA with Ribo-Zero Gold protocol (Illumina) and performed as described[157]. Libraries were sequenced on a NextSeq500 instrument (Illumina) with 38-bp paired end.

Splicing and gene expression analysis

RNA-seq reads were aligned to NCBI GRCh38 using STAR.[128] Each sample was sequenced twice, and the two sequencing runs were treated as technical replicates, with sequencing run included as covariate in both the splicing and expression analyses. Differential splicing and differential expression were calculated for each cohort (lean, obese, and obese with T2D) and for two time point comparisons (confluence vs. day 3 of differentiation and day 3 vs. day 15 of differentiation). Differential splicing was assessed at the exon level using DEXSeq[130] and for each time point comparison in each cohort exons with an FDR < 0.05 were considered significantly differentially spliced. Differential expression was assessed using limma[158] and for each time point comparison in each cohort genes with an FDR < 0.05 and a fold-change of at least 1.25 were considered significantly differentially expressed.

Pairwise sharing analysis

To assess the degree of sharing between the differential splicing and expression datasets, we performed three separate analyses – one comparing all differential splicing analyses generated here, one comparing all differential expression analyses, and one comparing differential splicing and differential expression for each cohort and time point. We estimated the degree of sharing in each of these sets of analyses using mashr[146] and a matrix of *Z* scores (for the differential splicing analysis and the joint splicing-expression analysis) or matrixes of effect

sizes and standard errors (for the differential expression analysis). For the joint splicing-expression analysis, we aggregated the splicing results to the gene level by considering the most significant exon for each gene. Using mash, significant effects were considered shared if they had the same sign and were within a factor of 0.5 of each other.

Fuzzy c-means clustering

We performed three independent clustering analyses, one for each cohort. In each, all exons that were differentially spliced in either time point comparison for that cohort were included. Differentially spliced exons were then clustered based on exon usage coefficients calculated using DEXSeq. The goal of this analysis was to capture overall patterns of expression, so exon usage coefficients were centered and scaled to account for differences in magnitude. Fuzzy c-means clustering with k = 6 was then performed on each of the three resulting datasets (for the lean, obese, and obese with T2D cohorts) using the e1071 package in R and each exon was assigned to the cluster for which it had the highest membership. Metascape[95] (http://metascape.org) was used to perform functional enrichment analysis on the gene sets associated with each cluster in each cohort.

GWAS enrichment analysis

We first assessed the degree of overlap between each differential splicing comparison and genetic variants associated with metabolic disorder using all significant SNPs from recent GWAS of BMI[20] and T2D.[23] SNP coordinates were converted from hg19 to hg38 using the LiftOver tool in the UCSC Genome Browser[159] and intersected with the flanking introns of differentially spliced exons using bedtools.[160] To determine if the resulting overlap represented a significant enrichment, we generated control sets of exons to compare to. For each differential splicing analysis, we generated 1000 control sets that had the same number of exons as were

differentially spliced and were randomly sampled from all tested exons in that analysis. We then intersected each of those control sets with the same GWAS SNPs to generate a null distribution of the number of introns that overlap at least one GWAS SNP. Intron overlap values that were greater than 95% of the control sets were considered significantly enriched and values that were less than 5% of the control sets were considered significantly depleted.

Acknowledgements

We thank all members of the GECKO consortium for their comments and valuable suggestions.

Funding

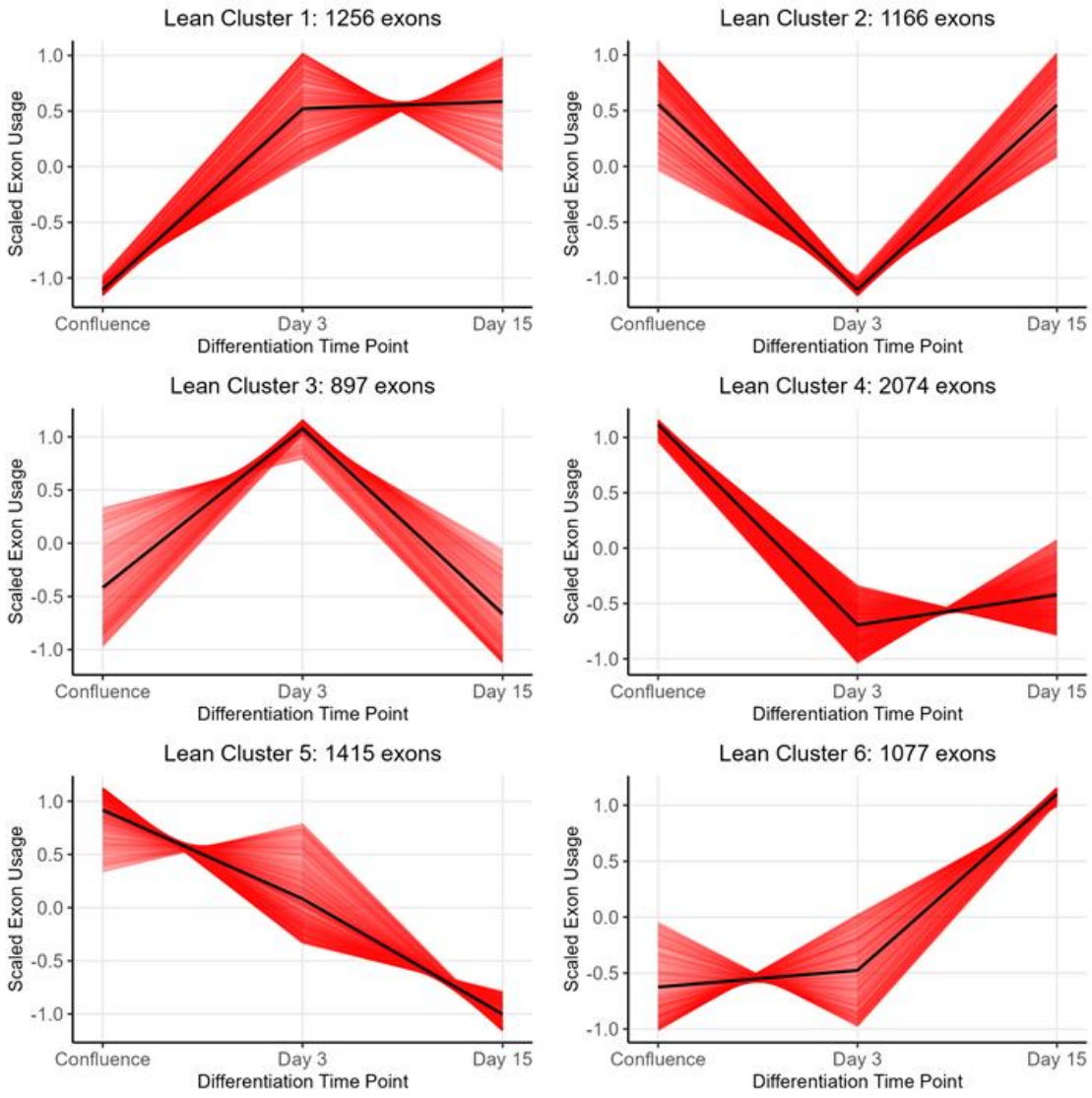
This study was supported by a Novo Nordisk Foundation Challenge Grant (NNF18OC0033754). K. M. F. was supported by an NIH training grant (T32GM007197).

Author contributions

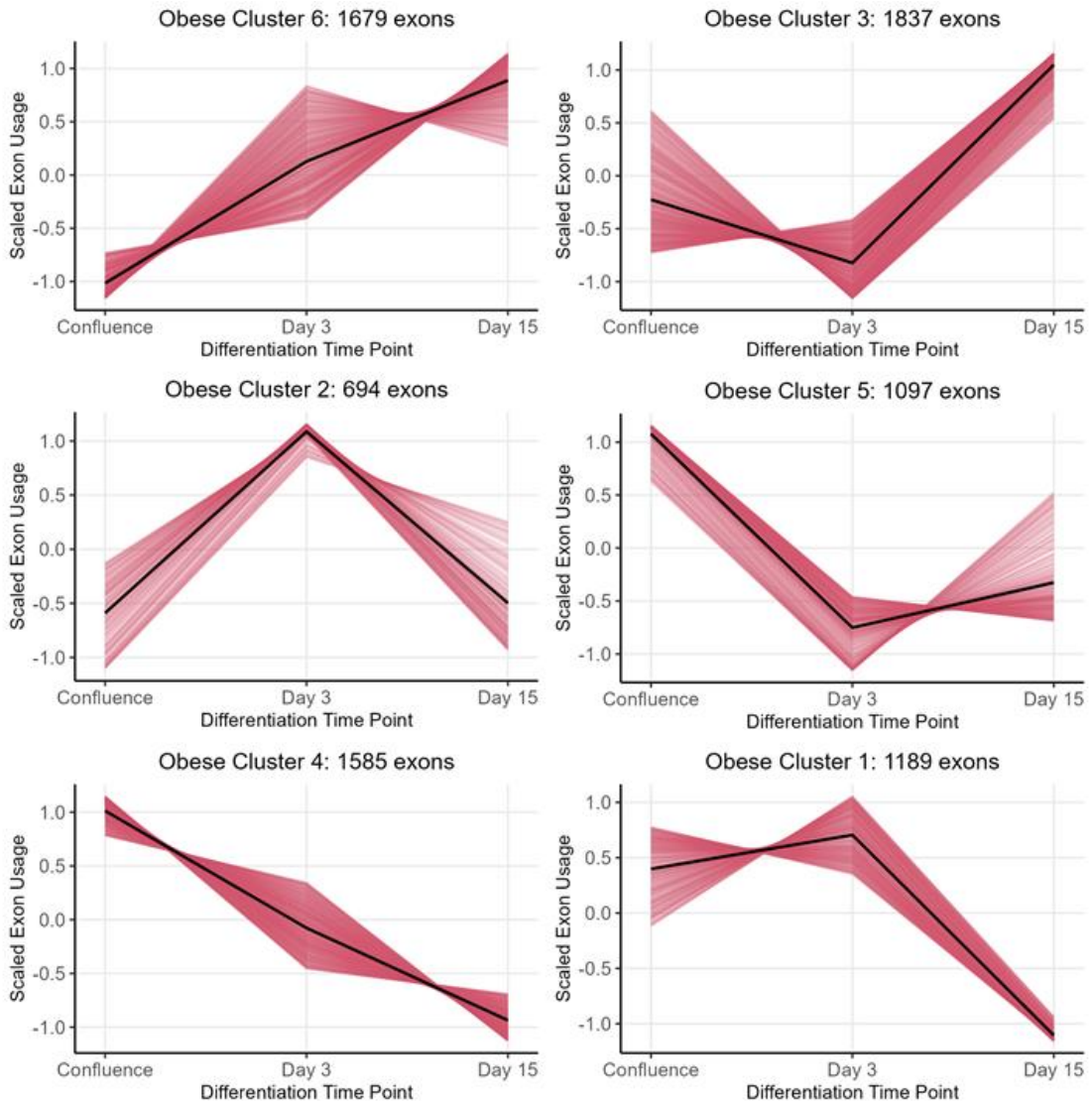
KMF contributed to study design, analyzed the data, and wrote the manuscript. EA contributed to study design, collected human samples, and edited the manuscript. ID, SV, and VBK collected human samples and contributed to study design. SJS provided expert advice and edited the manuscript. RB and MAN designed the study and edited the manuscript. All authors reviewed the manuscript.

Data availability

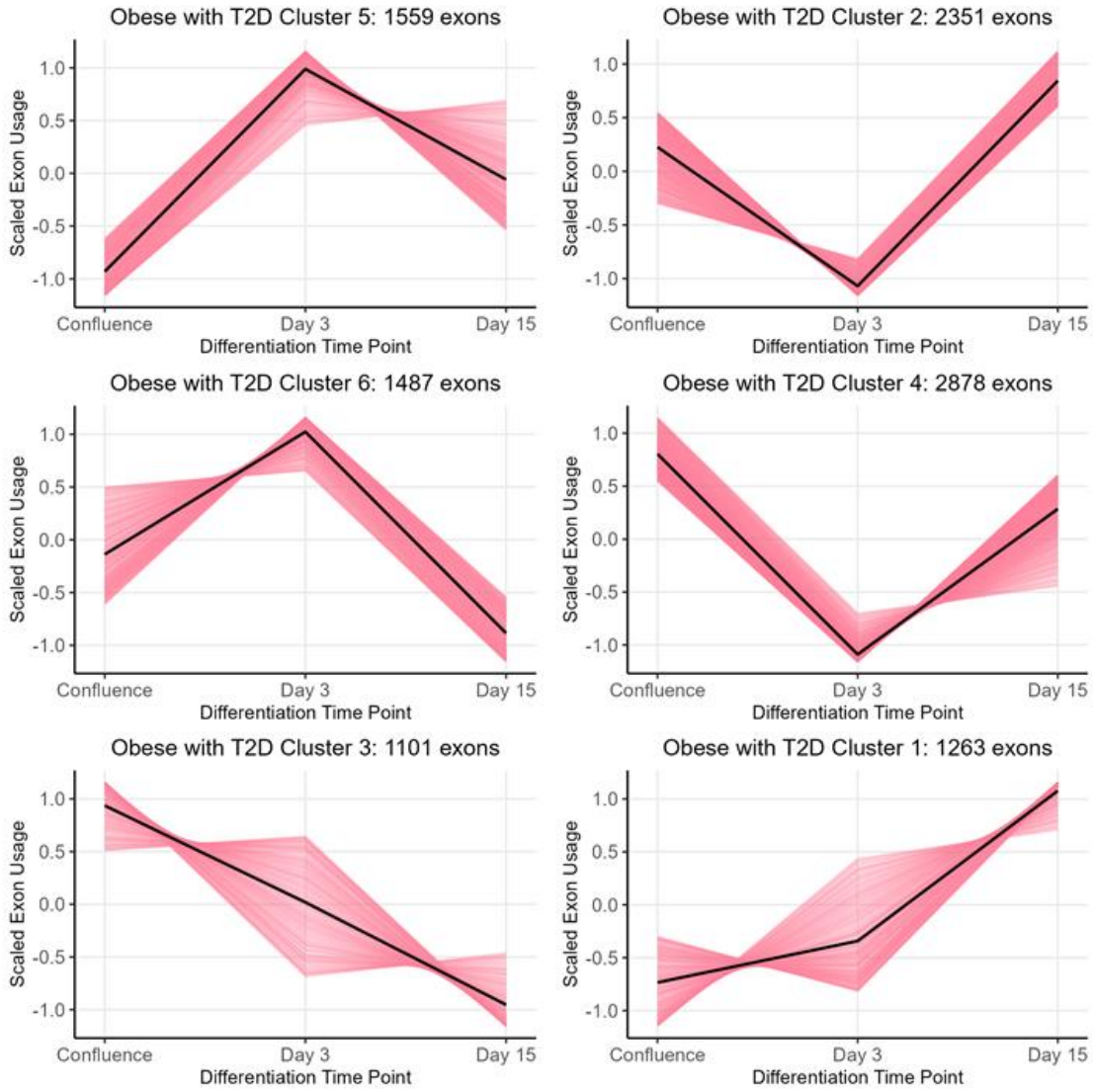
All differential expression and differential splicing results generated in this study are available as supplementary information. The RNA-seq datasets generated in this study are available from the corresponding author upon request.



Supplementary Figure 3-1. **Lean splicing clusters.** Plots showing the centered and scaled exon usage for each exon assigned to one of six clusters. For each plot, the black line connects the cluster centroids of the cluster. Clusters were generated by using fuzzy c-means clustering to cluster all differentially spliced exons from the lean cohort with k = 6.



Supplementary Figure 3-2. **Obese splicing clusters.** Plots showing the centered and scaled exon usage for each exon assigned to one of six clusters. For each plot, the black line connects the cluster centroids of the cluster. Clusters were generated by using fuzzy c-means clustering to cluster all differentially spliced exons from the obese cohort with k = 6.



Supplementary Figure 3-3. **Obese with T2D splicing clusters.** Plots showing the centered and scaled exon usage for each exon assigned to one of six clusters. For each plot, the black line connects the cluster centroids of the cluster. Clusters were generated by using fuzzy c-means clustering to cluster all differentially spliced exons from the obese with T2D cohort with k = 6.

Supplementary Tables for Chapter III

Supplementary Table 3-1. **Clinical parameters of study subjects.** Data are mean ± SD.

	Lean	**Obese**	**Obese type 2 diabetic**
N (male)	5	4	5
Age [years]	43.8 ± 10.7	37.8 ± 4.5	47.6 ± 2.2
Weight [kg]	79.0 ± 10.2	150.5 ± 23.8	150.3 ± 26.2
Height [cm]	179.8 ± 4.4	184.0 ± 10.6	182.2 ± 4.6
BMI [kg/m^2]	24.4 ± 2.1	44.3 ± 3.5	45.3 ± 7.5
Waist [cm]	93.4 ± 4.1	151.0 ± 14.7	163.3 ± 35.9
Hip [cm]	101.4 ± 4.0	139.7 ± 13.3	132.5 ± 9.3
Waist/Hip [cm]	0.9 ± 0.0	1.1 ± 0.0	1.2 ± 0.2
SBP [mmHg]	120.0 ± 11.0	132.8 ± 8.2	127.3 ± 10.5
DBP [mmHg]	76.8 ± 8.3	91.5 ± 8.0	78.3 ± 12.1
C-Reactive Protein [mg/l]	0.7 ± 0.4	9.8 ± 3.0	10.5 ± 10.1
HbA1c [mmol/mol]	35.0 ± 3.7	37.3 ± 1.7	46.8 ± 4.1
gHbA1c [%]	5.9 ± 0.5	6.3 ± 0.3	7.7 ± 0.6
C-peptide [pmol/l]	584.0 ± 111.6	1458.5 ± 288.0	1789.8 ± 493.7
Cholesterol [mmol/l]	4.5 ± 0.4	5.3 ± 0.5	3.7 ± 1.2
HDL [mmol/l]	1.4 ± 0.4	1.0 ± 0.3	1.0 ± 0.2
LDL [mmol/l]	2.6 ± 0.4	3.7 ± 0.4	2.2 ± 1.1
VLDL [mmol/l]	0.5 ± 0.3	0.6 ± 0.2	0.6 ± 0.2
Triglyceride [mmol/l]	1.1 ± 0.7	1.4 ± 0.5	1.4 ± 0.5
Leukocytes [x 10^9/l]	5.8 ± 1.1	7.6 ± 1.5	6.8 ± 1.2
Insulin [pmol/l]	34 ± 0	183.3 ± 101.2	143.5 ± 73.9
Alcohol/week [units]	6 ± 7.2	0 ± 0	0.3 ± 0.5
Smoking/day [units]	0.6 ± 1.3	0 ± 0	0 ± 0

Supplementary Table 3-2. **RNA-seq study design.**

Individual	Confluence	Day 3	Day 15
Lean 1	x	x	x
Lean 2		x	x
Lean 3	x		x
Lean 4	x		
Lean 5		x	
Obese 1	x		x
Obese 2		x	x
Obese 3	x	x	x
Obese 4	x	x	
Obese T2D 1	x	x	x
Obese T2D 2	x		x
Obese T2D 3		x	
Obese T2D 4		x	x
Obese T2D 5	x		

Supplementary Table 3-3**. Differential splicing results for each cohort and time point comparison.** Available as an excel file online.

Supplementary Table 3-4**. Cluster assignments for each exon and cohort.** Available as an excel file online.

Supplementary Table 3-5**. SNP-exon pairs from GWAS intersection analyses.** Available as an excel file online.

Chapter IV: Discussion

Principal findings

In this work, I expand our understanding of the role of both environmental and genetic risk factors in obesity and metabolic disorder. I identify genes and exons that are responsive in adipose tissue to changes in dietary macronutrient composition, and cluster each set of gene regulatory responses into five main dynamic patterns across the diets. These data emphasize the importance of splicing in the response of a metabolic tissue to diet, and reveal complex response patterns that different functional groups of genes and exons have to macronutrient composition. I also demonstrate the power of this approach to identify possibly novel associations with diet-induced obesity, identifying changes in the expression of genes encoding the BBSome and other BBS-associated genes in response to dietary fat content. In parallel, I also systematically characterize the role of alternative splicing in adipocyte differentiation in humans across three different metabolic states. This study provides further support for the importance of alternative splicing in adipocyte biology, and links splicing events across adipocyte differentiation to GWAS variants associated with BMI and T2D. Taken together, these studies provide a compelling body of data for the importance of splicing in adipocyte biology and obesity etiology and identify avenues for future exploration of specific splicing events.

Future directions

In Chapter 2, I found that there was a strong gene regulatory response to dietary macronutrient composition in mice that affected metabolically relevant genes and pathways. However, it is important to note that this study was performed exclusively in young males of a single strain. Weight and body fat distribution are sexually dimorphic traits,[121,161,162] and it will

therefore be important to replicate these findings in female mice. Extending these results to include mice of both sexes would provide important evidence for the universality of the gene regulatory responses we observed and potentially provide insight into sex-specific responses in a subset of genes and pathways. Even a simplified experimental design with a smaller number of diets (such as a high, medium, and low fat content diet) would help establish whether the majority of observed gene regulatory signals are shared or divergent, and shed light on the degree of sexual dimorphism in the response to dietary macronutrient composition.

In addition to extending these results to investigate sex-specific effects, future studies could expand on this work by investigating gene by environment interactions. Obesity and metabolic disorders are complex diseases with both environmental and genetic risk factors, but the contribution of interactions between the two to disease risk remains less well quantified.[163,164] Although ideally we would like to investigate gene by environment interactions in humans, there are a number of factors that complicate such studies. Accurate data on life-long environmental exposures in humans is difficult to come by, and even short-term dietary studies in humans face issues with compliance and dropout.[165–167] To circumvent these challenges, we can begin to investigate the role of gene by environment interactions in response to diet in mice. Specifically, by replicating the dietary study presented here in mice of different strains and of crosses between strains we can answer questions like whether genetic background impacts response to diet, the degree of sharing of gene regulatory and phenotypic responses between different genetic backgrounds, and whether there are specific variants associated with changes in gene regulatory and phenotypic response to diet. Quantifying the impact of gene by environment interactions on response to diet in mice and identifying particular genes and pathways that are most impacted can help inform future studies in humans and provide candidate genes for further study.

In parallel with extending the results shown here to female mice and mice with different genetic backgrounds, we can also gain additional insight by expanding this study to include more tissues. Obesity and metabolic disorder are complex phenotypes involved many different tissues and body systems, of which adipose tissue is just one part.[67,68] Here, we have established that differences in dietary macronutrient composition lead to extensive splicing and expression changes in adipose tissue, and it seems likely that other metabolic tissues would also have a significant gene regulatory response to dietary macronutrient composition. Efforts are now underway in our group to expand this work beyond adipose tissue to other relevant metabolic tissues such as liver, muscle, and brain. These investigations will compliment and extend the work detailed here by identifying shared pathways and gene regulatory patterns across metabolic tissues, as well as characterizing unique gene regulatory signals to each tissue. By combining data across multiple metabolically relevant tissues, we can get a more complete picture of the gene regulatory response to diet and continue to unravel the precise underpinnings of the observed phenotypic responses.

To conclusively link gene regulatory changes to specific phenotypic responses, more work will need to be done to identify precise causal links. The work presented here has already identified some candidate gene regulatory changes such as the observed changes in BBS gene expression or *IGF1* splicing. Expanding this work to additional tissues will aid in identifying and prioritizing more candidates, but functional characterization is needed to establish a causal link. To investigate the direct downstream consequences of observed gene regulatory changes, candidate genes or exons can be perturbed in mouse adipose tissue or in human mesenchymal stem cells differentiated into adipocytes and the direct phenotypic consequences can be observed. By establishing a clear causal link between specific gene regulatory changes and

phenotypes, this would help provide possible therapeutic targets for modulating the response to diet in humans and could improve our ability to treat obesity and provide actionable dietary guidelines with measurable metabolic outcomes.

In addition to following up on the gene regulatory response to dietary change as outlined here, future studies could also delve deeper into the role of alternative splicing in adipocyte differentiation and function. One particularly impactful avenue of research that would greatly improve our ability to prioritize GWAS variants and provide insight into disease etiology is to expand on our knowledge of what splicing events are most impactful in adipocyte biology and development. In the work presented here, I identify thousands of differential splicing events across adipocyte differentiation and group them into shared splicing patterns and functional categories. This work provides a starting point to prioritize splicing events for functional validation in human preadipocyte lines, which would provide confirmation of the role of specific splicing events in adipocyte development and function. By perturbing individual splicing events, splicing patterns, or splicing regulators in human preadipocytes we can identify their precise impact on adipocyte differentiation and more accurately annotate the role of specific splicing events in particular cellular contexts and states.

In this work we also investigate the intersection between exons that are differentially spliced in adipocyte differentiation and variants associated with BMI and T2D in GWAS. Unfortunately, our sample size was not large enough to perform sQTL mapping to formally identify variants associated with splicing changes across adipocyte differentiation. Extending the work presented here to include a larger number of individuals with complete genotyping data to enable sQTL mapping across adipocyte differentiation would improve our ability to predict which variants may be involved in splicing and which splicing events may be most important for

disease risk. Previous work focusing on cardiomyocyte differentiation has demonstrated that performing QTL mapping across differentiation time can identify novel QTLs not found in the starting or final cell type alone and can elucidate possible disease mechanisms that would have otherwise been missed.[152] Therefore, performing sQTL mapping across adipocyte differentiation would likely contribute novel findings to existing catalogs of sQTLs and help link GWAS variants to specific cellular functions.

Once high-certainty variant-splicing event pairs have been identified, studies should focus on confirming their role in disease risk. To investigate variants implicated in adipocyte differentiation and function, we can use CRISPR to create human mesenchymal stem cells that contain the variant of interest and observe the direct impact of that variant on splicing and adipocyte phenotypes such as differentiation efficiency and fat storage. Detailed functional characterization of variants prioritized by available genomics data will help dissect the precise role of GWAS variants in disease, providing insight into how they confer risk and what the mechanisms of disease progression are. As CRISPR technology continues to progress and more high-throughput techniques such as CRISPR screens are fine-tuned, it will likely soon become feasible to perform these types of validation in a high or medium throughput manner, allowing us to access the full potential of GWAS to elucidate disease risk.

Beyond the specific avenues of research proposed here, there is a general need for better catalogs of splicing changes in different cellular states and in response to different perturbations, as well as better annotations of the impact of splicing events on cellular function. These data would improve our ability to study the role of splicing in metabolic disease, and in tissue function in general. The studies presented here are one step towards a more complete catalog, but more work is needed. In particular, it would be of great benefit to future splicing work to

improve existing annotation databases to include more precise annotations of what individual splicing events do and where and when they have been observed. Genomics work that focuses on gene expression is often aided by annotation data such as gene ontology databases, and creating similar bodies of data at the level of exons or splicing events would greatly speed up efforts to prioritize splicing events from high-throughput methods for follow-up in more low-throughput methods.

Finally, a greater body of work measuring differential splicing in different conditions would also help elucidate whether the phenomenon we observed in both the gene regulatory response to dietary macronutrient change and to adipocyte differentiation of low overlap between splicing and expression changes is a general rule or is specific to adipose tissue or the conditions we measured. It will be important to know if splicing and expression changes are largely orthogonal in many different conditions and tissues, as this has important implications for experimental design and our understanding of gene regulatory mechanisms. If it is a general rule that splicing and expression changes occur in largely separate gene sets, it becomes even more important to not focus solely on gene expression when measuring gene regulatory changes and to investigate splicing in future studies, as the proteome is significantly impacted by both gene expression and splicing. The work presented here suggests that we are likely missing a great deal of impactful proteome changes by looking at gene expression alone, and future studies should prioritize testing both gene expression and splicing changes whenever possible when measuring gene regulatory landscapes.

Concluding remarks

Overall, in this thesis I provide insight into both genetic and environmental contributions to obesity and metabolic disorder, with a particular focus on adipocyte biology and the role of

alternative splicing. In Chapter 2, I characterize the gene expression and splicing response to a meaningful dietary perturbation, and identify BBS genes as candidate genes responding to dietary fat content. In Chapter 3, I provide a resource to help dissect previously uncharacterized GWAS associations with BMI and T2D and to gain insight into adipocyte development and function. These studies illustrate the importance of considering splicing alongside expression to capture a more complete picture of gene regulation in cellular states or in response to environmental change. The work presented here improves our understanding of the etiology of obesity and the role of splicing in human disease, and opens up many avenues for future discovery.

References

1. Ng, M. *et al.* Global, regional and national prevalence of overweight and obesity in children and adults 1980-2013: A systematic analysis. *Lancet Lond. Engl.* **384**, 766–781 (2014).
2. Ortega, F. B., Lavie, C. J. & Blair, S. N. Obesity and Cardiovascular Disease. *Circ. Res.* **118**, 1752–1770 (2016).
3. Abdelaal, M., le Roux, C. W. & Docherty, N. G. Morbidity and mortality associated with obesity. *Ann. Transl. Med.* **5**, 161 (2017).
4. Shultz, S. P., Byrne, N. M. & Hills, A. P. Musculoskeletal Function and Obesity: Implications for Physical Activity. *Curr. Obes. Rep.* **3**, 355–360 (2014).
5. Chooi, Y. C., Ding, C. & Magkos, F. The epidemiology of obesity. *Metabolism* **92**, 6–10 (2019).
6. Aronne, L. J., Nelinson, D. S. & Lillo, J. L. Obesity as a disease state: A new paradigm for diagnosis and treatment. *Clin. Cornerstone* **9**, 9–29 (2009).
7. Correia, F. *et al.* Obstacles in dietary treatment of obesity. *Obes. Metab.* **5**, 107–113 (2009).
8. Romieu, I. *et al.* Energy balance and obesity: what are the main drivers? *Cancer Causes Control* **28**, 247–258 (2017).
9. Hill, J. O., Wyatt, H. R. & Peters, J. C. Energy Balance and Obesity. *Circulation* **126**, 126–132 (2012).
10. Spiegelman, B. M. & Flier, J. S. Obesity and the Regulation of Energy Balance. *Cell* **104**, 531–543 (2001).
11. Poti, J. M., Braga, B. & Qin, B. Ultra-processed Food Intake and Obesity: What Really Matters for Health—Processing or Nutrient Content? *Curr. Obes. Rep.* **6**, 420–431 (2017).

12. Nardocci, M. *et al.* Consumption of ultra-processed foods and obesity in Canada. *Can. J. Public Health.* **110**, 4–14 (2019).

13. Dicken, S. J. & Batterham, R. L. The Role of Diet Quality in Mediating the Association between Ultra-Processed Food Intake, Obesity and Health-Related Outcomes: A Review of Prospective Cohort Studies. *Nutrients* **14**, 23 (2022).

14. Patikorn, C. *et al.* Intermittent Fasting and Obesity-Related Health Outcomes: An Umbrella Review of Meta-analyses of Randomized Clinical Trials. *JAMA Netw. Open* **4**, e2139558 (2021).

15. Morales-Suarez-Varela, M., Collado Sánchez, E., Peraita-Costa, I., Llopis-Morales, A. & Soriano, J. M. Intermittent Fasting and the Possible Benefits in Obesity, Diabetes, and Multiple Sclerosis: A Systematic Review of Randomized Clinical Trials. *Nutrients* **13**, 3179 (2021).

16. Harris, L. *et al.* Intermittent fasting interventions for treatment of overweight and obesity in adults: a systematic review and meta-analysis. *JBI Evid. Synth.* **16**, 507 (2018).

17. Bouchard, C. Genetics of Obesity: What We Have Learned Over Decades of Research. *Obesity* **29**, 802–820 (2021).

18. Elks, C. E. *et al.* Variability in the Heritability of Body Mass Index: A Systematic Review and Meta-Regression. *Front. Endocrinol.* **3**, 29 (2012).

19. 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).

20. 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).

21. Uffelmann, E. *et al.* Genome-wide association studies. *Nat. Rev. Methods Primer* **1**, 1–21 (2021).

22. Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **20**, 467–484 (2019).

23. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).

24. Loos, R. J. F. & Yeo, G. S. H. The genetics of obesity: from discovery to biology. *Nat. Rev. Genet.* **23**, 120–133 (2022).

25. Rao, S., Yao, Y. & Bauer, D. E. Editing GWAS: experimental approaches to dissect and exploit disease-associated genetic variation. *Genome Med.* **13**, 41 (2021).

26. Joslin, A. C. *et al.* A functional genomics pipeline identifies pleiotropy and cross-tissue effects within obesity-associated GWAS loci. *Nat. Commun.* **12**, 5253 (2021).

27. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018).

28. Broekema, R. V., Bakker, O. B. & Jonkers, I. H. A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. *Open Biol.* **10**, 190221 (2020).

29. Chen, W., McDonnell, S. K., Thibodeau, S. N., Tillmans, L. S. & Schaid, D. J. Incorporating Functional Annotations for Fine-Mapping Causal Variants in a Bayesian Framework Using Summary Statistics. *Genetics* **204**, 933–958 (2016).

30. Kichaev, G. & Pasaniuc, B. Leveraging Functional-Annotation Data in Trans-ethnic Fine-Mapping Studies. *Am. J. Hum. Genet.* **97**, 260–271 (2015).

31. Su, C. *et al.* Mapping effector genes at lupus GWAS loci using promoter Capture-C in follicular helper T cells. *Nat. Commun.* **11**, 3294 (2020).

32. Torres, J. M. *et al.* Inferring causal genes at type 2 diabetes GWAS loci through chromosome interactions in islet cells. *Wellcome Open Res.* **8**, 165 (2023).

33. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).

34. Tung, Y. C. L., Yeo, G. S. H., O'Rahilly, S. & Coll, A. P. Obesity and FTO: Changing Focus at a Complex Locus. *Cell Metab.* **20**, 710–718 (2014).

35. Loos, R. J. F. & Yeo, G. S. H. The bigger picture of FTO—the first GWAS-identified obesity gene. *Nat. Rev. Endocrinol.* **10**, 51–61 (2014).

36. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371–375 (2014).

37. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N. Engl. J. Med.* **373**, 895–907 (2015).

38. Sobreira, D. R. *et al.* Extensive pleiotropism and allelic heterogeneity mediate metabolic effects of IRX3 and IRX5. *Science* (2021).

39. Zeng, B. *et al.* Comprehensive Multiple eQTL Detection and Its Application to GWAS Interpretation. *Genetics* **212**, 905–918 (2019).

40. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).

41. Pavlides, J. M. W. *et al.* Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. *Genome Med.* **8**, 84 (2016).

42. Sun, W. & Hu, Y. eQTL Mapping Using RNA-seq Data. *Stat. Biosci.* **5**, 198–219 (2013).

43. Gilad, Y., Rifkin, S. A. & Pritchard, J. K. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* **24**, 408–415 (2008).

44. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* **99**, 1245–1260 (2016).

45. Westra, H.-J. & Franke, L. From genome to function by studying eQTLs. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **1842**, 1896–1902 (2014).

46. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).

47. Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437–451 (2017).

48. Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* **10**, 741–754 (2009).

49. Kalsotra, A. & Cooper, T. A. Functional consequences of developmentally regulated alternative splicing. *Nat. Rev. Genet.* **12**, 715–729 (2011).

50. Kim, H. K., Pham, M. H. C., Ko, K. S., Rhee, B. D. & Han, J. Alternative splicing isoforms in health and disease. *Pflüg. Arch. - Eur. J. Physiol.* **470**, 995–1016 (2018).

51. Tazi, J., Bakkour, N. & Stamm, S. Alternative splicing and disease. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **1792**, 14–26 (2009).

52. THE GTEX CONSORTIUM. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

53. Brotman, S. M. *et al.* Subcutaneous adipose tissue splice quantitative trait loci reveal differences in isoform usage associated with cardiometabolic traits. *Am. J. Hum. Genet.* **109**, 66–80 (2022).

54. Atla, G. *et al.* Genetic regulation of RNA splicing in human pancreatic islets. *Genome Biol.* **23**, 196 (2022).

55. Vitolins, M. Z. & Case, T. L. What Makes Nutrition Research So Difficult to Conduct and Interpret? *Diabetes Spectr. Publ. Am. Diabetes Assoc.* **33**, 113–117 (2020).

56. Mozaffarian, D., Rosenberg, I. & Uauy, R. History of modern nutrition science—implications for current research, dietary guidelines, and food policy. *BMJ* **361**, k2392 (2018).

57. Delgado-Lista, J. *et al.* Long-term secondary prevention of cardiovascular disease with a Mediterranean diet and a low-fat diet (CORDIOPREV): a randomised controlled trial. *The Lancet* **399**, 1876–1885 (2022).

58. Hall, K. D. *et al.* Effect of a plant-based, low-fat diet versus an animal-based, ketogenic diet on ad libitum energy intake. *Nat. Med.* **27**, 344–353 (2021).

59. Speakman, J. R. Use of high-fat diets to study rodent obesity as a model of human obesity. *Int. J. Obes.* **43**, 1491–1492 (2019).

60. de Moura e Dias, M. *et al.* Diet-induced obesity in animal models: points to consider and influence on metabolic markers. *Diabetol. Metab. Syndr.* **13**, 32 (2021).

61. Simpson, S. J., Raubenheimer, D. & Bone, Q. A multi-level analysis of feeding behaviour: the geometry of nutritional decisions. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **342**, 381–402 (1997).

62. Simpson, S. J. *et al.* The Geometric Framework for Nutrition as a tool in precision medicine. *Nutr. Healthy Aging* **4**, 217–226.

63. Ruth Archer, C., Royle, N., South, S., Selman, C. & Hunt, J. Nutritional Geometry Provides Food for Thought. *J. Gerontol. A. Biol. Sci. Med. Sci.* **64A**, 956–959 (2009).

64. Simpson, S. J. *et al.* Dietary protein, aging and nutritional geometry. *Ageing Res. Rev.* **39**, 78–86 (2017).

65. Solon-Biet, S. M. *et al.* The Ratio of Macronutrients, Not Caloric Intake, Dictates Cardiometabolic Health, Aging, and Longevity in Ad Libitum-Fed Mice. *Cell Metab.* **19**, 418–430 (2014).

66. Solon-Biet, S. M. *et al.* Macronutrient balance, reproductive function, and lifespan in aging mice. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3481–3486 (2015).

67. Das, U. N. Obesity: Genes, brain, gut, and environment. *Nutrition* **26**, 459–473 (2010).

68. Kim, J. B. Dynamic cross talk between metabolic organs in obesity and metabolic diseases. *Exp. Mol. Med.* **48**, e214–e214 (2016).

69. Booth, A., Magnuson, A., Fouts, J. & Foster, M. T. Adipose tissue: an endocrine organ playing a role in metabolic regulation. *Horm. Mol. Biol. Clin. Investig.* **26**, 25–42 (2016).

70. Havel, P. J. Role of adipose tissue in body-weight regulation: mechanisms regulating leptin production and energy balance. *Proc. Nutr. Soc.* **59**, 359–371 (2000).

71. Blüher, M. Adipose tissue dysfunction contributes to obesity related metabolic diseases. *Best Pract. Res. Clin. Endocrinol. Metab.* **27**, 163–177 (2013).

72. Ahmed, B., Sultana, R. & Greene, M. W. Adipose tissue and insulin resistance in obese. *Biomed. Pharmacother. Biomedecine Pharmacother.* **137**, 111315 (2021).

73. De Fano, M. *et al.* Adipose Tissue Plasticity in Response to Pathophysiological Cues: A Connecting Link between Obesity and Its Associated Comorbidities. *Int. J. Mol. Sci.* **23**, 5511 (2022).

74. Vernia, S. *et al.* An alternative splicing program promotes adipose tissue thermogenesis. *eLife* **5**, e17672 (2016).

75. Kaminska, D. & Pihlajamäki, J. Regulation of alternative splicing in obesity and weight loss. *Adipocyte* **2**, 143–147 (2013).

76. Chao, Y. *et al.* Regulatory roles and mechanisms of alternative RNA splicing in adipogenesis and human metabolic health. *Cell Biosci.* **11**, 66 (2021).

77. Emont, M. P. *et al.* A single cell atlas of human and mouse white adipose tissue. *Nature* **603**, 926 (2022).

78. Isakson, P., Hammarstedt, A., Gustafson, B. & Smith, U. Impaired Preadipocyte Differentiation in Human Abdominal Obesity: Role of Wnt, Tumor Necrosis Factor-α, and Inflammation. *Diabetes* **58**, 1550–1557 (2009).

79. Andersen, E. *et al.* Preadipocytes from obese humans with type 2 diabetes are epigenetically reprogrammed at genes controlling adipose tissue function. *Int. J. Obes.* **43**, 306–318 (2019).

80. Longo, M. *et al.* Adipose Tissue Dysfunction as Determinant of Obesity-Associated Metabolic Complications. *Int. J. Mol. Sci.* **20**, 2358 (2019).

81. Farris, K. *et al. Impact of Dietary Macronutrient Composition on Gene Regulation in Adipose Tissue*. https://www.researchsquare.com/article/rs-2973236/v1 (2023) doi:10.21203/rs.3.rs-2973236/v1.

82. Thom, G. & Lean, M. Is There an Optimal Diet for Weight Management and Metabolic Health? *Gastroenterology* **152**, 1739–1751 (2017).

83. Willcox, B. J. *et al.* Caloric Restriction, the Traditional Okinawan Diet, and Healthy Aging. *Ann. N. Y. Acad. Sci.* **1114**, 434–455 (2007).

84. Ludwig, D. S., Willett, W. C., Volek, J. S. & Neuhouser, M. L. Dietary fat: From foe to friend? *Science* **362**, 764–770 (2018).

85. Simpson, S. J. & Raubenheimer, D. Obesity: the protein leverage hypothesis. *Obes. Rev.* **6**, 133–142 (2005).

86. Raubenheimer, D., Simpson, S. J. & Mayntz, D. Nutrition, ecology and nutritional ecology: toward an integrated framework. *Funct. Ecol.* **23**, 4–16 (2009).

87. Raubenheimer, D. & Simpson, S. J. Nutritional Ecology and Human Health. *Annu. Rev. Nutr.* **36**, 603–626 (2016).

88. Raubenheimer, D. *et al.* An integrative approach to dietary balance across the life course. *iScience* **25**, 104315 (2022).

89. Lee, K. P. *et al.* Lifespan and reproduction in Drosophila: New insights from nutritional geometry. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 2498–2503 (2008).

90. Pellegrinelli, V., Carobbio, S. & Vidal-Puig, A. Adipose tissue plasticity: how fat depots respond differently to pathophysiological cues. *Diabetologia* **59**, 1075–1088 (2016).

91. Laclaustra, M., Corella, D. & Ordovas, J. M. Metabolic syndrome pathophysiology: The role of adipose tissue. *Nutr. Metab. Cardiovasc. Dis.* **17**, 125–139 (2007).

92. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).

93. Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nat. Rev. Genet.* **17**, 19–32 (2016).

94. Raubenheimer, D. Toward a quantitative nutritional ecology: the right-angled mixture triangle. *Ecol. Monogr.* **81**, 407–427 (2011).

95. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, (2019).

96. Park, J. *et al.* VEGF-A–Expressing Adipose Tissue Shows Rapid Beiging and Enhanced Survival After Transplantation and Confers IL-4–Independent Metabolic Improvements. *Diabetes* **66**, 1479–1490 (2017).

97. Sun, K. *et al.* Dichotomous effects of VEGF-A on adipose tissue dysfunction. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5874–5879 (2012).

98. Harper, S. J. & Bates, D. O. VEGF-A splicing. *Nat. Rev. Cancer* **8**, 880–887 (2008).

99. Arcondéguy, T., Lacazette, E., Millevoi, S., Prats, H. & Touriol, C. VEGF-A mRNA processing, stability and translation: a paradigm for intricate regulation of gene expression at the post-transcriptional level. *Nucleic Acids Res.* **41**, 7997–8010 (2013).

100. Scavo, L. M., Karas, M., Murray, M. & Leroith, D. Insulin-like growth factor-I stimulates both cell growth and lipogenesis during differentiation of human mesenchymal stem cells into adipocytes. *J. Clin. Endocrinol. Metab.* **89**, 3543–3553 (2004).

101. Smith, P. J., Wise, L. S., Berkowitz, R., Wan, C. & Rubin, C. S. Insulin-like growth factor-I is an essential regulator of the differentiation of 3T3-L1 adipocytes. *J. Biol. Chem.* **263**, 9402–9408 (1988).

102. Chang, H. R., Kim, H. J., Xu, X. & Ferrante, A. W. Macrophage and adipocyte IGF1 maintain adipose tissue homeostasis during metabolic stresses. *Obes. Silver Spring Md* **24**, 172–183 (2016).

103. Philippou, A., Maridaki, M., Pneumaticos, S. & Koutsilieris, M. The Complexity of the IGF1 Gene Splicing, Posttranslational Modification and Bioactivity. *Mol. Med.* **20**, 202–214 (2014).

104. Hede, M. S. *et al.* E-Peptides Control Bioavailability of IGF-1. *PLOS ONE* **7**, e51152 (2012).

105. Cui, H., Yi, Q., Feng, J., Yang, L. & Tang, L. Mechano growth factor E peptide regulates migration and differentiation of BMSCs. *J. Mol. Endocrinol.* **52**, (2013).

106. Kineman, R. D., Rio-Moreno, M. del & Sarmento-Cabral, A. 40 YEARS of IGF1: Understanding the tissue-specific roles of IGF1/IGF1R in regulating metabolism using the Cre/loxP system. *J. Mol. Endocrinol.* **61**, T187–T198 (2018).

107. Hilgendorf, K. I. Primary Cilia Are Critical Regulators of White Adipose Tissue Expansion. *Front. Physiol.* **12**, 769367 (2021).

108. Marion, V. *et al.* Transient ciliogenesis involving Bardet-Biedl syndrome proteins is a fundamental characteristic of adipogenic differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1820–1825 (2009).

109. Forsythe, E. & Beales, P. L. Bardet–Biedl syndrome. *Eur. J. Hum. Genet.* **21**, 8–13 (2013).

110. Khan, S. a. *et al.* Genetics of human Bardet–Biedl syndrome, an updates. *Clin. Genet.* **90**, 3–15 (2016).

111. Forti, E., Aksanov, O. & Birk, R. Z. Temporal expression pattern of Bardet-Biedl syndrome genes in adipogenesis. *Int. J. Biochem. Cell Biol.* **39**, 1055–1062 (2007).

112. Tsoucas, D. *et al.* Accurate estimation of cell-type composition from gene expression data. *Nat. Commun.* **10**, 2975 (2019).

113. Novas, R., Cardenas-Rodriguez, M., Irigoín, F. & Badano, J. L. Bardet–Biedl syndrome: Is it only cilia dysfunction? *FEBS Lett.* **589**, 3479–3491 (2015).

114. Zhou, Z. *et al.* Impaired cooperation between IFT74/BBS22–IFT81 and IFT25–IFT27/BBS19 causes Bardet-Biedl syndrome. *Hum. Mol. Genet.* **31**, 1681–1693 (2022).

115. Seo, S. *et al.* Requirement of Bardet-Biedl syndrome proteins for leptin receptor signaling. *Hum. Mol. Genet.* **18**, 1323–1331 (2009).

116. Cognard, N. *et al.* Comparing the Bbs10 complete knockout phenotype with a specific renal epithelial knockout one highlights the link between renal defects and systemic inactivation in mice. *Cilia* **4**, 10 (2015).

117. Marion, V. *et al.* BBS-Induced Ciliary Defect Enhances Adipogenesis, Causing Paradoxical Higher-Insulin Sensitivity, Glucose Usage, and Decreased Inflammatory Response. *Cell Metab.* **16**, 363–377 (2012).

118. Barroso, I. & McCarthy, M. I. The Genetic Basis of Metabolic Disease. *Cell* **177**, 146–161 (2019).

119. Ghosh, S. & Bouchard, C. Convergence between biological, behavioural and genetic determinants of obesity. *Nat. Rev. Genet.* **18**, 731–748 (2017).

120. Gloudemans, M. J. *et al.* Integration of genetic colocalizations with physiological and pharmacological perturbations identifies cardiometabolic disease genes. *Genome Med.* **14**, 31 (2022).

121. Hansen, G. T. *et al.* Genetics of sexually dimorphic adipose distribution in humans. *Nat. Genet.* **55**, 461–470 (2023).

122. Krentz, N. A. J. & Gloyn, A. L. Insights into pancreatic islet cell dysfunction from type 2 diabetes mellitus genetics. *Nat. Rev. Endocrinol.* **16**, 202–212 (2020).

123. Cheng, M., Huang, X., Zhang, M. & Huang, Q. Computational and functional analyses of T2D GWAS SNPs for transcription factor binding. *Biochem. Biophys. Res. Commun.* **523**, 658–665 (2020).

124. Crean, A. J. *et al.* Male reproductive traits are differentially affected by dietary macronutrient balance but unrelated to adiposity. *Nat. Commun.* **14**, 2566 (2023).

125. Sørensen, A., Mayntz, D., Raubenheimer, D. & Simpson, S. J. Protein-leverage in Mice: The Geometry of Macronutrient Balancing and Consequences for Fat Deposition. *Obesity* **16**, 566–571 (2008).

126. Piper, M. D. W. *et al.* Matching Dietary Amino Acid Balance to the In Silico-Translated Exome Optimizes Growth and Reproduction without Cost to Lifespan. *Cell Metab.* **25**, 610–621 (2017).

127. Lawson, J. & Willden, C. Mixture Experiments in R Using mixexp. *J. Stat. Softw.* **72**, 1–20 (2016).

128. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

129. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

130. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**, 2008–2017 (2012).

131. Van Hauwaert, E. L. *et al.* Isolation of nuclei from mouse white adipose tissues for single-nucleus genomics. *STAR Protoc.* **2**, 100612 (2021).

132. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* 1–12 (2023) doi:10.1038/s41587-023-01767-y.

133. Huot, M.-É. *et al.* The Sam68 STAR RNA-Binding Protein Regulates mTOR Alternative Splicing during Adipogenesis. *Mol. Cell* **46**, 187–199 (2012).

134. Song, J. & Richard, S. Sam68 Regulates S6K1 Alternative Splicing during Adipogenesis. *Mol. Cell. Biol.* **35**, 1926–1939 (2015).

135. Aprile, M. *et al.* PPARγΔ5, a Naturally Occurring Dominant-Negative Splice Isoform, Impairs PPARγ Function and Adipocyte Differentiation. *Cell Rep.* **25**, 1577-1592.e6 (2018).

136. Li, H. *et al.* SRSF10 Regulates Alternative Splicing and Is Required for Adipocyte Differentiation. *Mol. Cell. Biol.* **34**, 2198–2207 (2014).

137. Patel, R. S., Carter, G., Cooper, D. R., Apostolatos, H. & Patel, N. A. Transformer 2β homolog (Drosophila) (TRA2B) Regulates Protein Kinase C δI (PKCδI) Splice Variant Expression during 3T3L1 Preadipocyte Cell Cycle. *J. Biol. Chem.* **289**, 31662–31672 (2014).

138. Mikoluk, C., Nagengast, A. A. & DiAngelo, J. R. The splicing factor transformer2 (tra2) functions in the Drosophila fat body to regulate lipid storage. *Biochem. Biophys. Res. Commun.* **495**, 1528–1533 (2018).

139. Splicing across adipocyte differentiation is highly dynamic and impacted by metabolic phenotype. https://www.researchsquare.com (2023) doi:10.21203/rs.3.rs-3487148/v1.

140. Chew, N. W. S. *et al.* The global burden of metabolic disease: Data from 2000 to 2019. *Cell Metab.* **35**, 414-428.e3 (2023).

141. Naing, Y. T. & Sun, L. The Role of Splicing Factors in Adipogenesis and Thermogenesis. *Mol. Cells* **46**, 268–277 (2023).

142. Péterfy, M., Phan, J., Xu, P. & Reue, K. Lipodystrophy in the fld mouse results from mutation of a new gene encoding a nuclear protein, lipin. *Nat. Genet.* **27**, 121–124 (2001).

143. Péterfy, M., Phan, J. & Reue, K. Alternatively Spliced Lipin Isoforms Exhibit Distinct Expression Pattern, Subcellular Localization, and Role in Adipogenesis *. *J. Biol. Chem.* **280**, 32883–32889 (2005).

144. Mei, B., Zhao, L., Chen, L. & Sul, H. S. Only the large soluble form of preadipocyte factor-1 (Pref-1), but not the small soluble and membrane forms, inhibits adipocyte differentiation: role of alternative splicing. *Biochem. J.* **364**, 137–144 (2002).

145. Kaminska, D. *et al.* Adipose tissue INSR splicing in humans associates with fasting insulin level and is regulated by weight loss. *Diabetologia* **57**, 347–351 (2014).

146. Urbut, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019).

147. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

148. Dominguez, D. *et al.* An extensive program of periodic alternative splicing linked to cell cycle progression. *eLife* **5**, e10288 (2016).

149. Stilling, R. M. *et al.* De-regulation of gene expression and alternative splicing affects distinct cellular pathways in the aging hippocampus. *Front. Cell. Neurosci.* **8**, 373 (2014).

150. Solovyeva, E. M. *et al.* New insights into molecular changes in skeletal muscle aging and disease: Differential alternative splicing and senescence. *Mech. Ageing Dev.* **197**, 111510 (2021).

151. Li, Y. I. *et al.* RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600–604 (2016).

152. Strober, B. J. *et al.* Dynamic genetic regulation of gene expression during cellular differentiation. *Science* **364**, 1287–1290 (2019).

153. Takata, A., Matsumoto, N. & Kato, T. Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat. Commun.* **8**, 14519 (2017).

154. Zhang, Y. *et al.* Regional Variation of Splicing QTLs in Human Brain. *Am. J. Hum. Genet.* **107**, 196–210 (2020).

155. Riolo, G., Cantara, S. & Ricci, C. What's Wrong in a Jump? Prediction and Validation of Splice Site Variants. *Methods Protoc.* **4**, 62 (2021).

156. Arner, P., Arner, E., Hammarstedt, A. & Smith, U. Genetic Predisposition for Type 2 Diabetes, but Not for Overweight/Obesity, Is Associated with a Restricted Adipogenesis. *PLOS ONE* **6**, e18284 (2011).

157. Nylander, V. *et al.* Ionizing Radiation Potentiates High-Fat Diet–Induced Insulin Resistance and Reprograms Skeletal Muscle and Adipose Progenitor Cells. *Diabetes* **65**, 3573–3584 (2016).

158. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

159. Haeussler, M. *et al.* The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* **47**, D853–D858 (2019).

160. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

161. Grove, K., Fried, S., Greenberg, A., Xiao, X. & Clegg, D. A microarray analysis of sexual dimorphism of adipose tissues in high-fat-diet-induced obese mice. *Int. J. Obes. 2005* **34**, 989–1000 (2010).

162. MacCannell, A. D. V. *et al.* Sexual dimorphism in adipose tissue mitochondrial function and metabolic flexibility in obesity. *Int. J. Obes.* **45**, 1773–1781 (2021).

163. Qi, L. & Cho, Y. A. Gene-environment interaction and obesity. *Nutr. Rev.* **66**, 684–694 (2008).

164. Reddon, H., Guéant, J.-L. & Meyre, D. The importance of gene–environment interactions in human obesity. *Clin. Neurosci. Res.* **130**, 1571–1597 (2016).

165. Thessen, A. E. *et al.* Community Approaches for Integrating Environmental Exposures into Human Models of Disease. *Environ. Health Perspect.* **128**, 125002 (2020).

166. Fabricatore, A. N. *et al.* Attrition from randomized controlled trials of pharmacological weight loss agents: a systematic review and analysis. *Obes. Rev. Off. J. Int. Assoc. Study Obes.* **10**, 333–341 (2009).

167. Crichton, G. E. *et al.* Long-term dietary intervention trials: critical issues and challenges. *Trials* **13**, 111 (2012).