Regulation of gene expression during adipogenesis via mRNA processing

A thesis submitted by

Salwa Mohd Mostafa

in partial fulfillment of the requirements for the degree of

Ph.D

In

Genetics

Tufts University

Graduate School of Biomedical Sciences

March 25, 2024

Advisor: Claire Moore, PhD

Abstract

Obesity is characterized by dysregulated adipogenesis leading to increased number and/or size of adipocytes. Understanding the molecular mechanisms governing regulation of adipogenesis is therefore key to designing therapeutic interventions against obesity. In this thesis, we combined computational analysis of sequencing datasets generated from human primary preadipocytes and mature adipocytes, as well as benchwork experiments with the mouse 3T3-L1 cell line, to understand regulation of adipogenesis by long non-coding RNAs, alternative splicing, alternative polyadenylation, and polyadenylation factors. We used analysis of our 3'-end sequencing data and RNA-sequencing data from publicly available databases, to propose mechanisms of molecular regulation of adipogenesis. We also used the 3T3-L1 cell line to study the changes in levels of cleavage and polyadenylation factors and tested if manipulating the levels of one differentially expressed factor, CPSF73, regulates 3T3-L1 differentiation. Using computational analysis, we discovered lncRNAs that have not been previously characterized but may be key regulators of both brown and white adipogenesis in humans. We also demonstrated possible mechanisms of direct or indirect regulation of adipogenesis through alternative splicing. Furthermore, we show that usage of alternative polyadenylation sites of key adipogenesis genes leads to isoform diversity, which can have significant biological consequences on differentiation efficiency. Finally, data from our work with 3T3-L1 cells show that levels of some cleavage and polyadenylation factors such as CPSF73 change during adipogenesis, but the overexpression of CPSF73 was not sufficient to slow adipogenesis. Therefore, our research reveals potential therapeutic avenues for obesity

through manipulation of long non-coding RNA levels, alternative splicing, usage of alternative polyadenylation sites and levels of cleavage and polyadenylation factors.

Dedication

To my parents - my biggest source of strength, joy, and peace - who taught me to live life with purpose and gratitude.

Acknowledgements

I thank God for blessing me with the people and the circumstances that have enabled me to pursue my aspirations and experience everything that life has to offer. I am grateful for all the privileges I have had and to everyone who made that possible.

I am deeply indebted to my mentor, Dr Claire Moore, for her guidance and support throughout my PhD. Claire has given me the opportunity and the platform to learn new skills and enhance my understanding of scientific research. She has been an inspiration not just in science, but also when it comes to enjoying and valuing life experiences. I have always appreciated the stories, photos and treats she would bring back from her trips.

I want to thank all the past (Kasia, Huiyun, Daniel, Juan and Susan) and present (Srimoyee, Atish, Marzieh, Ankita and Maria) members of the Moore lab for their help and encouragement during my time in the lab. I especially want to thank Kasia, who was my rotation mentor and the person with whom I have shared my bay in the early years in the lab, for patiently teaching me new techniques, helping me understand the project we were working on, and for looking out for my best interests, even when I did not know what they were. I also want to thank Huiyun for all her advice and help in and outside the lab, and for being a great friend.

I also want to thank my thesis committee members Dr Phil Hinds, Dr Karl Munger, Dr Charlotte Kuperwasser and previous member Dr Marta Gaglia. Their guidance, support, and advice helped me be a better researcher, and encouraged me to persist.

I am incredibly lucky to have met some amazing people at Tufts. I started my journey in the Genetics Program surrounded by brilliant, passionate students who were always eager to help. I particularly want to thank Lea and Machika for being the people I could always talk to since I joined the program. They have had great insight and advice for me - both personal and professional. I especially want to thank Machika for helping me remain calm and steadfast in the later years. I also want to thank Maya for being so compassionate and kind to me, and for having me at all her awesome parties.

I have had an immense support network outside of Tufts as well. I want to thank Dr Khademul Islam, my very first research mentor in Bangladesh who has encouraged me to pursue a research career. I have numerous friends from my time in Saudi Arabia, Bangladesh and then in the US, who have continued to encourage me and uplift my spirits. My childhood friends, some of whom I have known since kindergarten, have played a huge part in keeping me focused and entertained at the same time. I especially thank my friend, Nushin, who coincidentally ended up in Boston around the same time I started my PhD and was the person I could always count on. She has made my life in Boston so much more fulfilling and enjoyable and has been a major emotional support during my PhD.

I thank my family for giving me the courage and support to dive into uncertainty and to continue learning. I want to thank my aunt and uncle - Khalamoni and Khalu - and my dearest cousins for loving me as I am and taking care of my parents and sister when I could not. I also want to thank my late grandparents, Nana and Nani, for showering me with love in every possible way and in every possible moment that they could. I thank my sister for teaching me what unconditional love is, for wanting to do better and be better, and for getting me interested in science just by being who she is. To my brother, thank you for being my best friend and the best older sibling one can hope for. From helping me with schoolwork to advising me on electronics, you have had a profound influence on my life, and I am so grateful we grew up together.

Finally, to my parents - Ammu and Abbu - I thank you for raising me with so much love and care, for putting me first, for giving me a safe space to make mistakes and grow, and for being my biggest emotional support. I thank you for instilling values in me that remind me to be as motivated, resilient, kind, and respectful as I can be. I will be eternally grateful to you both.

Table of contents

List of tables

List of figures

List of Copyrighted Materials Used

Cristancho, A. G. & Lazar, M. A. Forming functional fat: a growing understanding of adipocyte differentiation. Nat Rev Mol Cell Biol 12, 722–734 (2011).
Rey, F. et al. Role of long non-coding RNAs in adipogenesis: State of the art and implications in obesity and obesity-associated diseases. Obesity Reviews 22, e13203 (2021).
Mitschka, S. & Mayr, C. Context-specific regulation and function of mRNA alternative polyadenylation. *Nat Rev Mol Cell Biol* **23**, 779–796 (2022).
Song, X. *et al.* Alternative splicing events and function in the tumor microenvironment: New opportunities and challenges. *Int Immunopharmacol* **123**, 110718 (2023).

List of Abbreviations

White adipose tissue	WAT
Brown adipose tissue	BAT
Mesenchymal stem cell	MSC
Alkaline phosphatase	ALP
Adipose-Derived Stem Cells	ASCs
MicroRNAs	miRNAs
Long non-coding RNAs	lncRNAs
Messenger RNA	mRNA
Alternative polyadenylation	APA
Alternative splicing	AS
Piwi-interacting RNA	piRNA
Polyadenylation site	pA site
RNA-binding proteins	RBPs
3'- untranslated region	3'-UTR
Skipped exons	SE
Intron retention	RI
Mutually exclusive exons	MXE
Alternative 5′ splice sites	A5SS
Alternative 3′ splice sites	A3SS
Percent spliced in	PSI
Single-cell RNA sequencing	scRNA-seq
Cleavage and polyadenylation specificity factor	CPSF
Cleavage stimulation factor)	CSTF
Mammalian cleavage factor I	CFIm
Mammalian cleavage factor II	CFIIm

Chapter 1: Introduction

1.1 Adipogenesis and obesity

Adipose tissue is a complex organ, with its most important role being the master regulator of energy balance and nutritional homeostasis. Adipocytes are mainly categorized into two types: unilocular white adipocytes that regulate insulin sensitivity, lipid metabolism and satiety; and multilocular brown adipocytes containing large numbers of mitochondria, that are highly specialized cells for dissipation of stored chemical energy into heat. White adipose tissue (WAT) is distributed throughout the body in subcutaneous regions and surrounding visceral organs whereas brown adipose tissue (BAT) is restricted to the paravertebral, supraclavicular and periadrenal regions. Both types of adipocytes develop from mesenchymal stem cells (MSCs) through a process called adipogenesis. Adipogenesis is a two-phase process consisting of early determination and terminal differentiation. In the first phase, mesenchymal stem cells (MSCs) commit to becoming preadipocytes. During the second phase, a transcriptional cascade is activated that contributes to and maintains the maturation of functional adipocytes (Figure 1.1).[2,3]

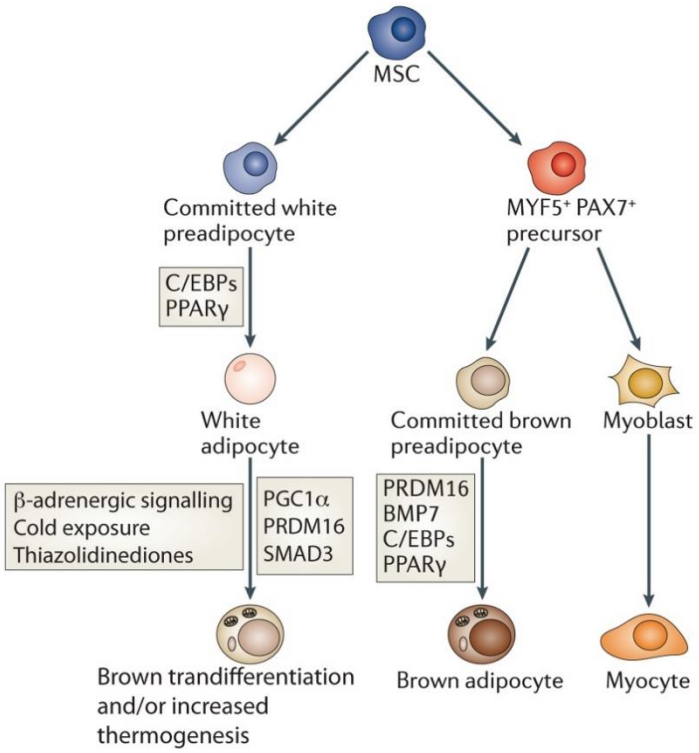


Figure 1.1: Differentiation of mesenchymal stem cells to white and brown adipocytes. Reprinted with permission from Cristancho, A. G. & Lazar, M. A. Forming functional fat: a growing understanding of adipocyte differentiation. *Nat Rev Mol Cell Biol* **12**, 722–734 (2011).

Murine cells lines, such as 3T3-L1, 3T3-F422A and OP9, are commonly used in adipogenesis research,[4] with the most popular model being 3T3-L1 cells. These cells were derived as a subclone of the 3T3 mouse embryonic fibroblast cell line by selecting cells from confluent cultures that displayed spontaneous lipid accumulation.[5] 3T3-L1 cells are easier and less costly to use than primary adipocytes, can be passaged for longer, and have a homogenous response to treatments and experiments.[4] However, these murine cells cannot completely recapitulate the process in human cells due to several differences. For example, 3T3-L1 and 3T3-F442A cell lines express the adipocyte-secreted factor leptin at much lower levels than primary adipocytes.[6] The response to drugs may also be different, for example, levamisole, an inhibitor of alkaline phosphatase (ALP) activity,

decrease ALP activity in 3T3-L1 cells but increases ALP activity in human preadipocytes.[7] Human cell lines have also been used to study adipogenesis. PAZ6, the first human preadipocyte cell line, was isolated from infant BAT transformed with a vector containing the SV40 T and T-antigen coding genes and has been used both as white and brown preadipocyte models. The human SGBS cell line was isolated from the stromal-vascular fraction of an infant with SGBS (Simpson-Golabi-Behemil syndrome) and is widely used for adipogenesis studies (In vitro and ex vivo models of adipocytes). Human Adipose-Derived Stem Cells (ASCs) and human preadipocytes, although being available in small amounts and having limited renewal capacity, are excellent models because they are better able to recapitulate characteristics of adipose tissue, mainly because of retaining depot-specific properties.[8]

An excessive accumulation of WAT, which occurs as a result of increased adipocyte volume (hypertrophy), number of cells (hyperplasia), or both can lead to obesity, a complex and multifaceted disease. Obesity is also associated with an increased risk for other metabolic, cardiovascular and chronic inflammatory diseases. Therefore, understanding the regulation of white adipogenesis is a possible therapeutic approach for obesity.[8,9]

Novel drugs against obesity that are currently in development or in use target the hormones leptin and ghrelin, mitochondrial uncouplers, and growth differentiation factor 15 (GDF15). Incretins (intestinal peptide that stimulates insulin release insulin) were primarily targeted for treatment against diabetes but could also lower body weight.[10]

Treatment strategies targeting molecular regulators of adipogenesis mainly consider signaling proteins, transcriptional factors, and non-coding RNAs such as microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) as biomarkers or drug targets.[11,12]

In our study, we investigated the regulation of gene expression during adipogenesis by lncRNAs and messenger RNA (mRNA) processing mechanisms – alternative polyadenylation (APA) and alternative splicing (AS). We identified genes that are regulated via these modes of regulation, proposed mechanisms of their regulation, identified potential novel regulators of adipogenesis, and finally, demonstrated the probable application of our findings in the therapeutic avenues of browning of WAT[13] and adipocyte dedifferentiation.[14]

1.2 Regulation of gene expression

LncRNAs are a class of non-coding RNAs that are longer than 200 nucleotides, generally do not code for functional proteins and are poorly conserved. LncRNAs have a 5'cap and a 3'poly A tail like mRNA-encoding proteins. However, unlike mRNAs, lncRNAs have strong tissue specificity and a lower abundance. They exist in the nucleus, cytoplasm, and organelles with a greater abundance in the nucleus. LncRNAs play various roles in cell proliferation, differentiation, and apoptosis, as well as the development of tissues and organs. In the nucleus, lncRNAs can regulate transcription, alternative splicing, and subcellular localization of mRNA. In the cytoplasm, lncRNAs can affect the stability and translation of mRNA and "sponge" proteins and miRNAs. LncRNAs were also shown to

function as precursor molecules of piwi-interacting RNA (piRNA) and miRNA and their excretion outside the cell can lead to protein transport out of the cells (Figure 1.2).[15]

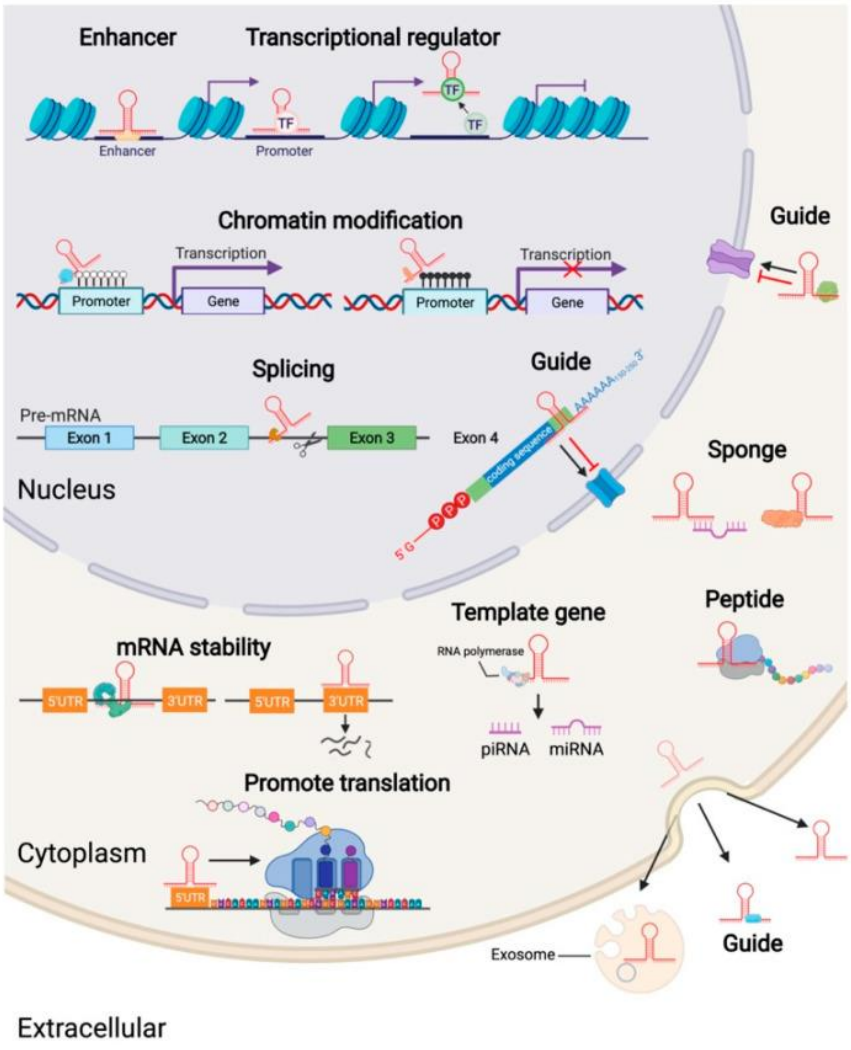


Figure 1.2: Mechanisms of functions of long non-coding RNAs. Reprinted with permission from Rey, F. et al. Role of long non-coding RNAs in adipogenesis: State of the art and implications in obesity and obesity-associated diseases. Obesity Reviews 22, e13203 (2021).

LncRNAs have been shown to regulate adipogenesis, for example by regulating of early and late adipogenesis master regulators, and therefore have implications in obesity. Screening of the whole transcriptome by RNA-sequencing and microarray analysis of adipose tissue of obese and non-obese patients led to identification of adipose-derived

lncRNAs and lncRNAs specifically associated with obesity. Numerous lncRNAs also correlate with obesity-associated inflammatory diseases.[16] However, only one study in 2015 examined global changes in lncRNA expression during differentiation of human adipose-derived stromal cells.[17] Today, improvements in sequencing strategies now enable identification of novel adipogenesis-related lncRNAs that may have been missed in earlier studies. Therefore, our study of lncRNA expression changes in white adipogenesis can reveal novel lncRNA regulators and provide a basis for comparison with future datasets.

Cleavage and polyadenylation is the final step of mature mRNA synthesis in eukaryotes. It involves enzymatic cleavage of pre-mRNA and the addition multiple adenosines called the polyA tail. The process occurs co-transcriptionally, through recognition of specific polyadenylation site (pA site) motifs in the transcript by the cleavage and polyadenylation (C/P) machinery.[18] There are 4 major protein complexes in the C/P machinery - CPSF (cleavage and polyadenylation specificity factor), CSTF (cleavage stimulation factor), CFIm (mammalian cleavage factor I) and CFIIm (mammalian cleavage factor II) - and additional core factors - symplekin, polyA polymerase (PAP), nuclear polyA binding protein (PABPN1) and the C-terminal domain of the largest subunit of RNA polymerase II (Figure 1.3).[19] Manipulating levels of C/P factors can affect differentiation and change APA profiles.[19–22]

Genes containing more than one pA site can undergo alternative polyadenylation (APA) which leads to production of alternate mRNA isoforms. Depending on which pA site is used, APA can change the encoded protein and/or the 3'-UTR (3' untranslated region)

length (Figure 1.3). Use of an intronic pA site can change both the sequence of the encoded protein as well as the 3'-UTR length, resulting in non-functional proteins or proteins with a different function. If APA occurs in the 3'-UTR region, then the resulting mRNA isoforms can encode the same protein but have different 3'-UTR lengths. Cis-regulatory sequences in the 3'-UTR regulate translation, stability, and localization of mRNAs through the action of microRNAs and RNA-binding proteins (RBPs). The major regulators of APA are RBPs and RNA modifications, expression levels and post-translational modifications of C/P subunits, transcriptional dynamics, transcription factor activity, chromatin conformation, and DNA modifications. APA has been identified in multiple organisms, tissues and biological processes and linked to diseases like cancer, diabetes, and muscular dystrophy.[18]

During differentiation, APA can lead to both 3'-UTR lengthening – like in mouse brains, D*rosophila* central nervous system, and human differentiating cardiomyocytes - and 3'-UTR shortening - like during *Drosophila* and mammalian spermatogenesis.[18] A previous study analyzed 3'-reads and showed an overall lengthening during 3T3-L1 differentiation.[23] Another study used RNA sequencing to investigate APA in the 3'-UTR during adipogenesis and identified a modest trend towards longer 3'-UTRs.[24] However, these studies have not focused on particular genes undergoing APA or identified regulators of APA. Given how variations in 3'-UTR length of specific genes [25] as well as modulation of APA regulators [18] can affect the process of differentiation, we aimed to identify key genes in adipogenesis undergoing APA, potential RBPs that regulate APA in those genes, and the levels of C/P factors that could affect global changes in APA during adipogenesis.

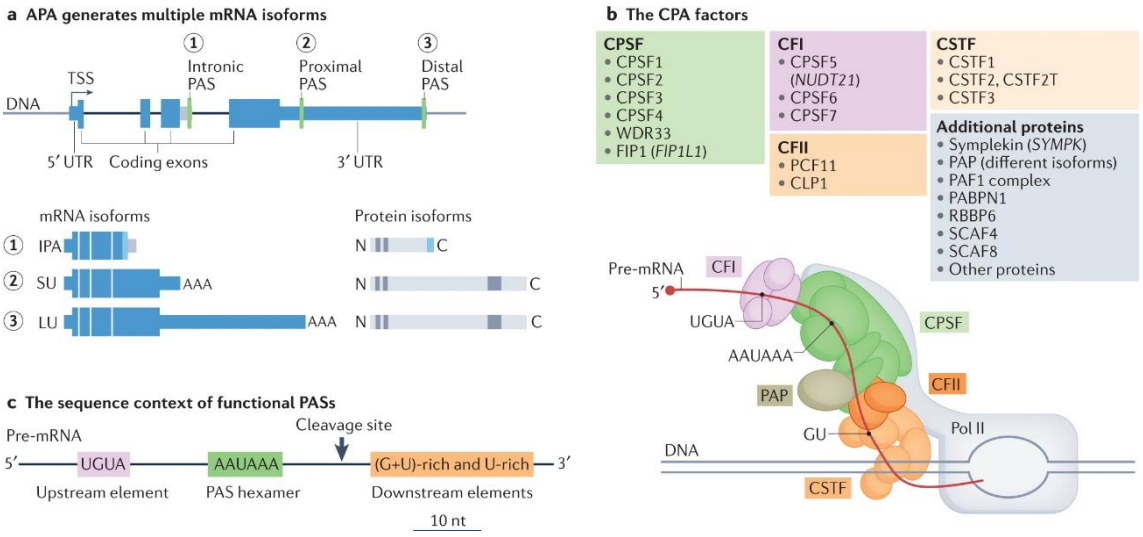


Figure 1.3: Overview of alternative polyadenylation. (A) Locations of polyA sites and consequences of their usage. (B) Cleavage and polyadenylation factors involved in alternative polyadenylation. (C) Regulatory RNA sequence elements of alternative polyadenylation. Reprinted with permission from Mitschka, S. & Mayr, C. Context-specific regulation and function of mRNA alternative polyadenylation. *Nat Rev Mol Cell Biol* **23**, 779–796 (2022).

Alternative splicing (AS) is another co-transcriptional mRNA processing mechanism that leads to transcriptomic and proteomic diversity in eukaryotes. From a single protein-coding gene, this process removes introns from pre-mRNA and combines the exons to form multiple forms of mature mRNA varying in coding potential, UTRs, or RNA stability. There are five different types of alternative splicing: SE (skipped exons), RI (intron retention), MXE (mutually exclusive exons), A5SS (alternative 5′ splice sites), and A3SS (alternative 3′ splice sites) (Figure 1.4). Although about 95% of human multi-exon genes can undergo splicing, most alternatively spliced variants may not be translated into proteins, and therefore, cannot be detected in large-scale proteomics studies. AS is pervasive, occurring during early embryonic development, cell fate determination and differentiation, organogenesis, and tissue development. AS is regulated by various factors, such as the spliceosome complex, global interactions between cis-

acting RNA sequences and the surrounding enhancers or silencers, epigenetics, and transcriptional dynamics.[26] One of the primary regulators of AS is RBPs – multiple RBPs controls each AS event and their combined action produces a distribution of alternatively spliced products in a given cell type.[27]

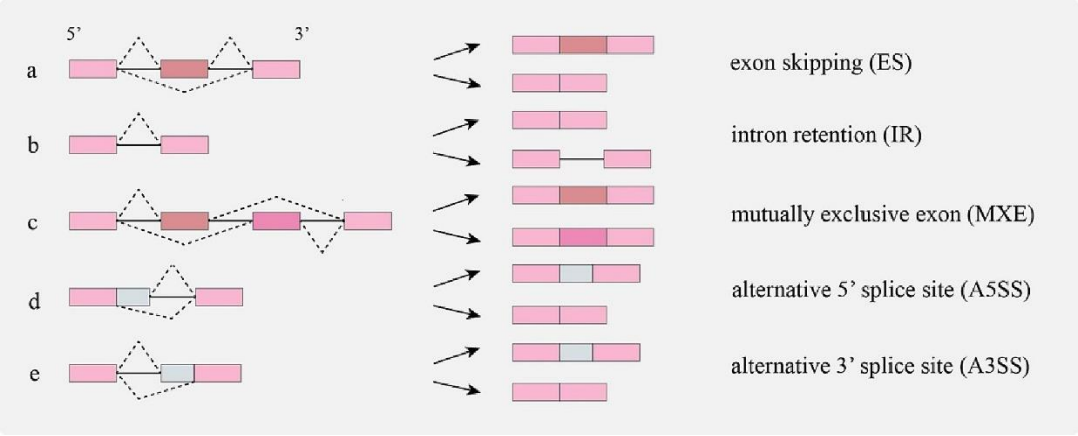


Figure 1.4: Five different types of alternative splicing events. Reprinted with permission from Song, X. et al. Alternative splicing events and function in the tumor microenvironment: New opportunities and challenges. Int Immunopharmacol 123, 110718 (2023).

AS has been detected in both white and brown adipogenesis, with most studies done in 3T3-L1 cells and mouse animal models.[28] Some studies have characterized AS in human adipogenesis[29,30] by looking at specific targets. However, only one paper addressed global alternative splicing during induction of bone marrow-derived mesenchymal stem cells into adipocytes.[31]

Both APA and AS are mRNA processing events that have evolved as a means of fine-tuning the regulation of gene expression during various cellular processes and in response to different stressors.

Polyadenylation has been studied in organisms ranging from archaea to the more complex eukaryotic organisms and in organelles like the chloroplast and mitochondria.[32] The comparison of protein factors responsible for polyadenylation of mRNAs showed that they are conserved across many evolutionary divergent species, but the isoforms of some factors with a species-specific function are not evolutionary conserved. Comparison between human and yeast polyadenylation factors show that yeast does not contain homologs for the CFIm complex and CSTF1. However, yeast-specific proteins like Hrp1p are functionally similar to the mammalian CSTF1 protein and CFIm complex.[33] APA conservation is higher in broadly and highly expressed genes or those with low variation across tissues. The polyA site is defined by surrounding sequence motifs which are variable across different phylogenic groups, for example, yeasts have fewer and more degenerate motifs compared to mammals. Comparison of polyA sites between human, mouse and rat genomes, showed that single polyA sites were more conserved in mammals than alternative polyA sites, underscoring their importance in transcription termination. Conservation of APA is also tissue-dependent, for example, APA is selected for in brain-specific genes but is selected against in liver and kidney, where gene expression may be mainly regulated at the transcriptional level rather than during mRNA processing.[34]

AS is widespread across eukaryotes and the type and extent of AS varies considerably between eukaryotic lineages. For example, the most common AS event in metazoans is exon skipping, whereas in plants and fungi, intron retention is more frequent. AS can also lead to gene-specific changes in cis but can affect widespread changes in splicing if a trans component is involved. Novel splice patterns arising from mutations affecting splice

site of a gene can lead to duplication of existing exons or conversion of constitutive exon to an alternative exon, an intronic sequence into an exon or exonic sequence becoming part of an intron. Mutations in core splicing or regulatory gene can change its coding sequence and/or expression and lead to widespread downstream changes in splicing. AS leads to proteomic diversity but non-functional AS events could also potentially lead to evolution of novel functions. The prevalence of AS has strong correlation with organism complexity, with the highest prevalence in vertebrates. The differences in splicing profiles between lineages could also lead to phenotypic changes that reduce gene flow and result in rapid species divergence.[35] Therefore, with increasing complexity of organisms, evolution of mRNA processing has contributed to the evolution of adaptive traits.

1.3 Sequencing methods and tools

Many lncRNAs that have been characterized in adipogenesis were identified with microarray analysis. However, RNA-sequencing has less technical variability and can detect and quantify novel transcripts.[36] Therefore, RNA-seq datasets are an excellent resource in the identification of novel lncRNAs. While DESeq2 is a widely used tool to accurately identify differentially expressed transcripts, a combination of different methods of analyzing differential gene expression, based on different approaches to trimming, aligning, counting, and normalization of data, can produce more reliable results.[36]

There are multiple bioinformatics tools that analyze RNA-seq data for detection of alternative polyadenylation. These tools identify and quantify APA based on changes in

the mRNA read coverage using either annotated pA sites, which cannot identify novel sites, or perform de novo identification of pA sites.[37] Compared to standard RNA-seq, 3'-end sequencing (3'-Seq) enriches for reads covering the 3′ end of genes and therefore overcomes the limitations of correct estimation of pA site usage by analysis of standard RNA-seq datasets. Indeed, benchmarking analysis based on pA site identification and differential usage of pA sites show that analysis of 3'-seq data is more reliable in identifying and quantifying the usage of pA sites, including intronic pA site usage.[38] There has been research into generation of more reliable 3'-seq data[39,40], but there are limited tools to analyze this data.[41,42] The PolyA-miner tool is especially interesting as it uses "iterative consensus clustering" to detect APA. With most tools expressing APA relative to the most distal pA site, PolyA-miner can detect all APA changes from the most proximal to the most distal, and changes in intermediate pA sites as well. Calling a gene as undergoing APA considers the changes in each of these sites before an overall "shortening" or "lengthening" is called.[43]

Both standard RNA-seq and 3'-seq suffer from biases, such as due to PCR amplification steps during library preparation or when sequencing repetitive regions of the genome. These biases are minimized when using long-read sequencing.[38] Since long-read sequencing allows capture of full-length transcripts, including polyA tails, it removes the need for transcript reconstruction and can therefore also accurately profile APA[38], especially with the use of targeted long-read sequencing, where specific genes are enriched by probe-based cDNA pulldown strategy before sequencing on the Nanopore platform.[44] Single-cell RNA sequencing (scRNA-seq) has also been gaining traction for analyzing global APA patterns in different cell types and would be ideal in comparing

APA profiles in healthy and diseased tissues, such as in tumors.[45] The 3′ tag-based scRNA-seq methods that enrich for mRNA 3′ ends using poly(A) priming can potentially profile APA at single-cell resolution, especially with the development of scRNA-specialized tools such as scDaPars.[46]

Similar to APA analysis, many bioinformatics tools also use RNA-seq data for detection of alternatively spliced isoforms. Many AS tools define a splicing event type, calculate PSI (percent spliced in) values - which is a ratio of normalized counts for inclusion of a transcript element over total normalized reads for that event - and subjects it to statistical evaluation.[47] Therefore, these tools quantify AS, identify types of splicing event, and identify new events. Tools like rMATS and SUPPA only identify events found in annotated transcripts whereas tools like ASPLI and SplAdder additionally detect novel splice events.[48] Benchmarking analysis of AS detection tools showed discrepancies among the 21 different tools that were tested and found SUPPA, DARTS, rMATS and LeafCutter outperform other tools.[47] Short read sequencing technologies are cheap, accurate and provide high coverage reads with good differential expression analysis power but are also limited in their ability to resolve and quantify full-length transcripts. Therefore, long-read sequencing technologies, by sequencing the entire RNA in a single read, can detect changes in spliced isoforms with consequences in RNA regulation or protein function. Additionally, coupling long-read sequencing with single cell sequencing, targeted sequencing and spatial transcriptomics can achieve even higher resolution and informative AS profiling.[49,50]

Therefore, in the next two chapters, we have employed current and novel bioinformatics methods to analyze both standard and 3'-end focused RNA-seq datasets and studied changes in C/P factors to investigate regulation of gene expression during human and mouse adipogenesis, respectively.

Chapter 2: Transcriptomic analysis reveals regulation of adipogenesis via long non-coding RNA, alternative splicing, and alternative polyadenylation.

Salwa Mohd Mostafa, Luyang Wang, Bin Tian, Joel Graber, and Claire Moore, Submitted to *Scientific Reports*, 1/27/2024.

2.1 Introduction

Obesity is a major global public health burden. It is characterized by excessive accumulation of fat mass in white adipose tissue, which occurs due to increased adipocyte volume, increased number of cells, or both.[8,51] Adipocytes develop from mesenchymal stem cells (MSCs), which become committed preadipocytes and then fully differentiated adipocytes. There are two main types of adipocytes - white and brown adipocytes. White adipose tissue (WAT) functions in fat storage and is characterized by adipocytes containing large unilocular lipid droplets. It is an active endocrine organ that regulates insulin sensitivity, lipid metabolism and satiety. It is distributed throughout the body in subcutaneous regions and surrounds visceral organs. The main function of brown adipose tissue (BAT) is thermogenesis. It is composed of multiloculated adipocytes that contain large numbers of mitochondria. Its location is restricted to the paravertebral, supraclavicular and periadrenal regions.[2] Our focus is on white adipogenesis since obesity is caused by WAT outgrowth.[52]

Murine cells lines, such as 3T3-L1, 3T3-F422A and OP9, are commonly used in adipogenesis research[8], with the most popular model being 3T3-L1 cells. However, these murine cells cannot completely recapitulate the process in human cells due to several differences. For example, 3T3-L1 and 3T3-F442A cell lines express the adipocyte-secreted factor leptin at much lower levels than primary adipocytes. Recently, an increasing number of studies are using human Adipose-Derived Stem Cells (ASCs) and human preadipocytes to better capture characteristics of adipose tissue when addressing questions of adipose-related biology.[8] However, in depth sequencing analysis of these

human cells is limited but critical to dissecting the molecular regulatory mechanisms that characterize human adipogenesis. In our study, we examine gene expression regulation during differentiation of human primary preadipocytes into white adipocytes by using high-depth sequencing libraries and advanced bioinformatic tools, and in this way, expand our understanding of the role of long non-coding RNAs (lncRNAs), alternative splicing, and alternative polyadenylation (APA).

LncRNAs are small RNA transcripts that are longer than 200 nucleotides, generally have a 5'-cap and 3'-polyA tail, and do not code functional proteins. They regulate various processes and some play important roles in adipogenesis based on studies using cell lines and primary cells.[15,53] Since lncRNAs have low conservation between human and mouse[54], studies using primary human cells are of greater relevance to translational research. Studies using primary human adipocytes have characterized individual lncRNAs identified from microarray or RNA-sequencing[55,56] or studied expression in cells from obese and lean individuals.[57] One study in 2015 examined global changes in lncRNA expression during differentiation of human adipose-derived stromal cells.[17] However, the improvement of sequencing strategies may now allow the identification of novel lncRNAs with important roles in adipogenesis that may have been missed in the earlier studies. In our study, we sought to use global profiling to identify such lncRNAs and provide a resource for comparative analysis with future datasets.

APA during adipogenesis has been less extensively studied. APA is a pre-mRNA processing mechanism that generates different mRNA isoforms based on usage of alternative polyA (pA) sites. APA often results in isoforms with the same coding

sequence but different 3'-UTR lengths, with downstream effects on mRNA stability, translatability, localization, and RNA-seeded protein/protein interactions, while intronic APA can alter both the coding sequence and 3'-UTR sequences.[18,58–60] Use of pA sites located in introns affects the coding sequence, leading to transcript degradation or expression of different protein isoforms.[58–60] Previously, analysis of 3'-end reads showed an overall lengthening during 3T3-L1 differentiation.[61] Also, in 3T3-L1 cells, the shorter transcript of heme oxygenase 1 (HO1) was demonstrated to have a stronger inhibitory effect on differentiation of preadipocytes.[25] In 2013, a different group used RNA sequencing to investigate APA in the 3'-UTR and found that differentiated human adipocytes have a modest trend towards longer 3'-UTRs compared to undifferentiated adipocytes.[62] However, a comprehensive analysis of APA at a genome-wide level with standard RNA-seq data is difficult as few reads are localized in the 3'-UTR. A goal of our study was to use 3'-focused sequencing data to accurately and quantitatively characterize alternative pA site usage and location and identify specific white adipogenesis-related genes undergoing such regulation.

Alternative splicing is another pre-mRNA processing mechanism that diversifies the transcriptome and regulates gene expression.[28] It is carried out by the spliceosome and various RNA-binding proteins (RBPs) that activate or repress splicing of regulated exons by binding to enhancer or silencer elements, respectively, in the pre-mRNA.[28] Spliced isoforms and splicing factors relevant to adipogenesis have for the most part been identified in 3T3-L1 cells and mouse animal models.[28] Some studies have characterized specific spliced isoforms in human adipogenesis[30,63], but to our knowledge, only one paper addressed global alternative splicing during the process by inducing mesenchymal

stem cells from bone marrow to differentiate into adipocytes.[31] In the current study, we sought to take advantage of a recent dataset[64] on differentiation of stem cells derived from subcutaneous adipose tissue and analyze global alternative splicing with a commonly used tool, rMATS.[65]

By utilizing multiple sequencing datasets and newly available bioinformatics tools, combined with a thorough investigation into available experimental evidence, we have identified potentially new mechanisms of regulation of white adipocyte differentiation at the level of lncRNAs, alternative splicing and APA.

2.2 Methods

2.2.1 RNA samples and sequencing: Commercially available total RNA of human preadipocytes (Day 0) and their respective differentiated adipocytes (Day 14) were purchased from Zen-Bio Inc (Durham, NC, USA). The preadipocytes were isolated from human subcutaneous adipose tissue, plated and then differentiated into mature adipocytes for 14 days using the company's Adipocyte Differentiation Medium (catalog number DM-2), yielding cells that were rounded with large lipid droplets apparent in the cytoplasm. Details of the samples are provided in Table 5.1. These samples were sent to Admera Health LLC (NJ, USA) for 3'-end sequencing using the Quant-Seq FWD library kit. The generated cDNA libraries were sequenced as 150 bp paired-end reads with read depth of 60 million reads on an Illumina® [NovaSeq X Plus](Illumina, California, USA) instrument.

2.2.2 Differential gene expression analysis: The raw sequence reads of the samples were processed to remove adapters using Cutadapt[66] (v2.8) and the quality of these trimmed reads was checked using FastQC[67] (v0.11.8) and MultiQC[68] (v1.7.0). R1 reads were aligned using STAR[69] (2.6.1d) to Human hg38 genome. Aligned reads were then quantified using featurecounts[70] (v1.6.3) and log2-fold changes were calculated with DESeq2[71](1.40.2) using default parameters and the UCSC annotation file (created on 2015-08-14). A fold change of 1 and an adjusted p-value of less than 0.05 were applied to select differentially expressed genes. The enrichment of Gene Ontology Biological Processes (GOBP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathways for upregulated and downregulated genes were carried out using the DAVID Functional Annotation tools[72] and the default settings. Using GeneCards[73] as reference, we filtered genes for lncRNAs, and identified differentially expressed lncRNAs using the same parameters.

For analysis of differential gene expression of lncRNAs during brown adipogenesis, we used the publicly available GEO dataset GSE168387[74] which contained three replicates each of brown preadipocytes (Day 0) and differentiated adipocytes (Day 14) from the human immortalized brown adipocyte Paz6 cell line. Paired-end reads of samples (GSM5137833, GSM5137834, GSM5137835, GSM5137836, GSM5137837, GSM5137838) were aligned to the hg38 genome, quantified, and the log2-fold changes were calculated as above. Differentially expressed lncRNAs were selected based on adjusted p-value<0.05 and log2 fold change >1 or <−1.

2.2.3 Alternative polyadenylation analysis: APA analysis was performed using a polyA_DB-based APA analysis. Adapter and poly(T) sequences from QuantSeq FWD raw read 2 data were trimmed. Trimmed reads were then aligned to a human (hg19) polyA_DB[75] reference (-100nt, polyA_DB-annotated-PolyA Site (PAS), +25 nt) using STAR-2.7.7a. The last aligned position of each mapped read was compared to polyA_DB-annotated-PASs, allowing ± 24 nt flexibility. Matched reads were considered pA site-supporting (PASS) reads, which were used for further APA analysis. UCSC genome browser plots shown in this paper are based on PASS reads. For 3'-UTR APA analysis, two pA sites with the highest usage in the 3'-UTR were compared, and for intronic APA analysis, all intronic isoforms were combined and compared to all isoforms using pA sites in the terminal exon. A RED value (Relative Expression Difference) was calculated as difference in log2-ratio of isoform abundance (isoforms from use of distal pA sites vs those from proximal pA site) between undifferentiated and differentiated samples. A "shortening" or "lengthening" of gene was called when RED $< -\log_2(1.2)$ or $> \log_2(1.2)$, respectively, and BH (Benjamini and Hochberg)-adjusted P-value < 0.05 (Fisher's exact test).

The enrichment of GOBP and KEGG Pathways for shortened and lengthened genes were carried out using the DAVID Functional Annotation tools[72] and the default settings.

The presence of miRNA-binding sites was identified using the "Custom prediction tool" in the mirDB[76,77] database, and the presence of RBP-binding sites was identified through RBPmap[78,79] (v1.2), using the hg19 database assembly, all 223 human/mouse motifs included in RBPmap, the stringency level set to high and default settings for other parameters. For analysis of RBP-binding, we overlapped results from RBPmap[79] with

CLIP-seq data from the POSTAR3[80] database to identify RBP-binding to the 200 bp upstream and downstream of proximal, distal and intronic polyadenylation sites.

2.2.4 Alternative splicing analysis: We used the GEO dataset GSE176020[64] for alternative splicing analysis. In this dataset, adipose tissue-derived stem cells (ASCs) were induced to differentiate *in vitro*. We analyzed data from undifferentiated (Day 0) and differentiated (Day 9) samples derived from abdominal subcutaneous adipose tissue (GSM5352869, GSM5352872, GSM5352874, GSM5352877, GSM5352879, GSM5352882). FASTQ files from two runs of each sample were merged, followed by alignment of the paired-end reads to human hg38 genome using STAR[69] (v2.6.1d). The BAM (alignment) files generated were used as input to the rMATS[65] (v4.1.2) program, using all default parameters and UCSC annotation file (created on 2015-08-14), to detect splicing events between undifferentiated and differentiated adipocyte samples. The "JCEC" output files containing both junction and exon counts were used to detect significant differentially spliced events. To increase confidence in the rMATS calls, counts of each sample in both conditions were only considered if greater than 25. Differentially spliced events were called if dPsi [IncLevelDifference, or psi (percent-spliced-in) values between undifferentiated and differentiated samples] is greater than 0.05 or less than -0.05, and if the FDR value is less than 0.05. The enrichment of GOBP and KEGG Pathways for differentially spliced genes were carried out using the DAVID Functional Annotation tools[72] and the default settings. Sashimi plots for specific genes were generated using rmats2sashimiplot (v3.0.0) (https://github.com/Xinglab/rmats2sashimiplot).

The splicing output files from rMATS was input into "Motif Map" analysis of rMAPS2[81,82] (http://rmaps.cecsresearch.org/MTool/) using default parameters. This analysis involves scanning for occurrences of binding motifs of known RBPs in the intronic sequences 250 bp upstream or downstream of the target exon or flanking exons, and the first 50 bp of the 5'- or 3'-end of exonic sequences of included, skipped and non-regulated or background exons. A motif score (density) within a 50 bp sliding window is calculated, along with a P-value given in comparison between regulated and control exons.[81,82] Motif maps generated by rMPAPS2 were edited in Inkscape (www.inkscape.org).

2.2.5 Image generation: Volcano plots and bar graphs were generated using Rstudio (2023.09.0) using packages dplyr (v1.1.4)[83], ggplot2 (v3.4.3)[84], Enhanced volcano (v1.18.0)[85] and smplot2 (v0.1.0)[86]. Pie charts were generated using GraphPad Prism version 10.0.2 for Windows, GraphPad Software, Boston, Massachusetts USA, www.graphpad.com.

2.3 Results

2.3.1 Long non-coding RNAs LINC00312, LINC00607 and TYMSOS may be key regulators of white adipogenesis.

We sequenced polyadenylated RNAs from three different sets of preadipocytes isolated from human subcutaneous adipose tissue and compared these 3' end datasets to those from white adipocytes derived from these precursors. Differential gene expression analysis shows upregulation of previously established adipogenesis markers FABP4,

ADIPOQ, LPL and PPARG[87] (Figure 2.1) and confirm efficient differentiation of our samples. GOBP and KEGG enrichment analysis of up-regulated genes shows significantly enriched terms related to adipogenesis, such as fatty acid metabolism, PPAR signaling, and lipid metabolic process, whereas that of down-regulated genes show enriched terms related to cell cycle and cell adhesion (Figure 2.1).

LncRNAs have been reported to have deregulated expression in obese individuals, and some are known to play a role in the differentiation of both brown and white adipocytes[15]. Analysis of our dataset detected upregulation and downregulation of 158 and 77 lncRNAs, respectively, during differentiation into white adipocytes (Figure 2.2 and Table 5.2). Sixteen of the 158 upregulated lncRNAs, and 5 of the downregulated 77 lncRNAs, were previously reported, based on integrative co-expression analysis, as enriched in adipocyte and adipose progenitor cells, respectively.[88] The others have not been identified before as enriched in any particular tissue. Among the top 5 down-regulated and top 5 up-regulated lncRNAs (Figure 2.2A), only ADIPOQ-AS1 is known to function in adipogenesis through the regulation of murine adiponectin.[89] The large change in the expression of the nine novel lncRNAs warrants further investigation into their role in human adipogenesis. Most of these lncRNAs have deregulated expression in cancer, affecting properties like proliferation, migration, and invasion of cancer cells through their target miRNA/gene axis.[90–96] Considering that cell cycle and proliferation genes are downregulated during early adipogenesis[97,98], these lncRNAs may promote adipogenesis by targeting cell proliferation.

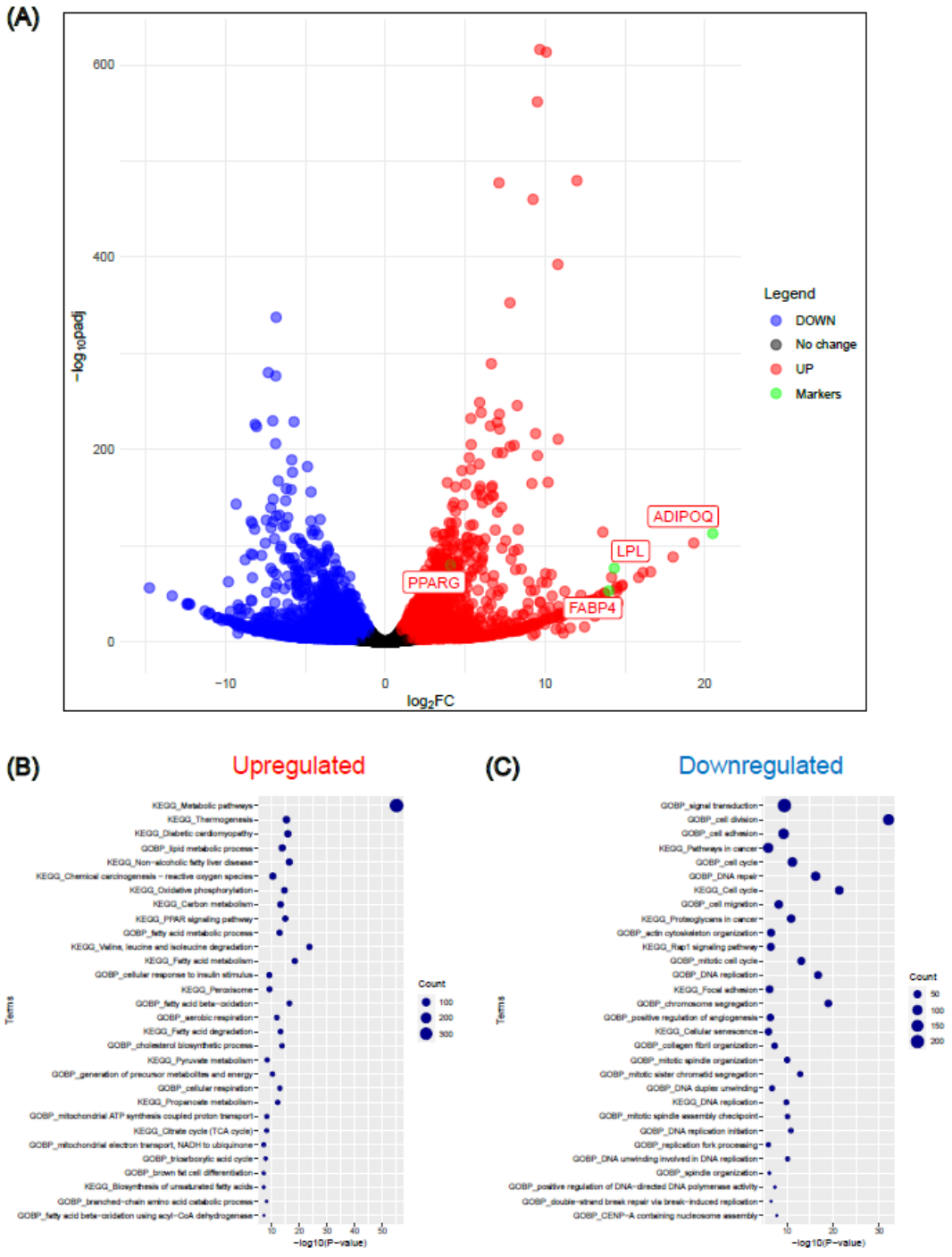


Figure 2.1: Differential gene expression confirms efficient differentiation of preadipocytes. (A) Volcano plot shows the distribution of differential gene expression of all genes at Day 0 vs Day 14 of differentiation. Red represents upregulated genes and blue represents downregulated genes that have p-adjusted values less than 0.05 and log2-fold change of more than 1, or less than -1, respectively. Some adipogenesis marker genes are highlighted. (B) and (C) shows the 30 most significantly enriched (P-value < 0.1) GOBP and KEGG terms for upregulated and downregulated genes, respectively.

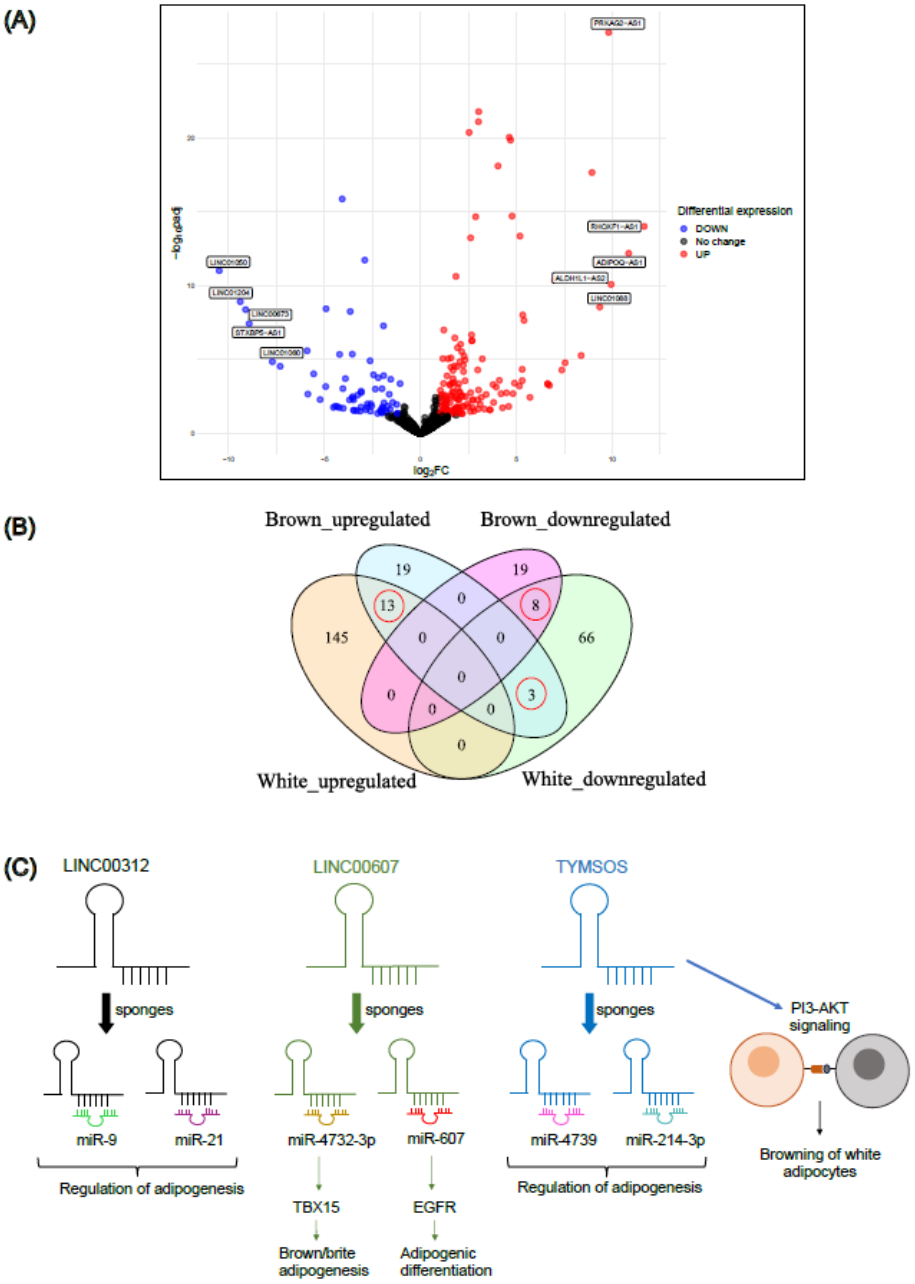


Figure 2.2: Multiple lncRNAs are differentially expressed during white adipogenesis. (A) Volcano plot depicts differential gene expression of lncRNAs at Day 0 vs Day 14 of differentiation. Red represents upregulated genes and blue represents downregulated genes that have p-adjusted values less than 0.05 and log2-fold change of more than or equal to 1, or less than or equal to -1, respectively. The top 5 upregulated and top 5 downregulated lncRNAs are highlighted. (B) Venn diagram shows overlap analysis of lncRNAs differentially expressed during white and brown adipogenesis. (C) Possible mechanisms of lncRNAs LINC00312, LINC00607 and TYMSOS in the regulation of adipogenesis.

To identify lncRNAs differentially expressed in the same or opposite manner during white and brown adipogenesis, we carried out differential expression analysis of a previously published human brown adipogenesis dataset.[99] Overlap analysis with our white adipogenesis dataset (Figure 2.2B) showed upregulation of 13 and downregulation of 8 lncRNAs during differentiation of both brown and white adipocytes (Table 2.1).

Both - UP	**Both - DOWN**	**Brown - UP, White - DOWN**
FOXD2-AS1	LINC00673	LINC00312
IQCH-AS1	LINC01605	LINC00607
LINC00663	LINC01116	TYMSOS
LINC00968	FRMD6-AS1	
LINC01003	SENCR	
LIPE-AS1	SLC8A1-AS1	
LOC100289495	LINC01085	
NIFK-AS1	LINC01119	
PRKAG2-AS1		
RAMP2-AS1		
TPRG1-AS1		
TRHDE-AS1		
LINC01140		

Table 2.1: List of lncRNAs that are differentially expressed during brown and white adipogenesis either in same or opposite direction. Shown are lncRNAs that are upregulated (BOTH - UP) or downregulated (BOTH – DOWN) during white and brown adipogenesis and lncRNAs that are upregulated during brown but downregulated during white adipogenesis (Brown - UP, White – DOWN)

Among the lncRNAs upregulated or downregulated during both brown and white adipogenesis, only LIPE-AS1 and LINC01119 have been shown to have a functional role and are differentially expressed during either murine or human adipogenesis.[100,101] LncRNAs FOXD2-AS1, LINC00968, LINC01116 and SENCR regulate glioma stem cell[102], osteogenic[103], neuronal[104] and endothelial[105] differentiation, respectively, and may play similar roles in adipogenic differentiation via regulation of target genes and miRNAs. The other lncRNAs (LINC01085, LINC01003, LINC00663, LINC01140,

RAMP2-AS1, NIFK-AS1, LINC00673, FRMD6-AS1, SLC8A1-AS1) have been associated with cancer, playing diverse roles in cancer proliferation and migration, cell cycle and signaling pathways.[106–115]

No lncRNAs were upregulated during white adipogenesis but downregulated during brown adipogenesis. Three lncRNAs, LINC00312, LINC00607 and TYMSOS, were upregulated during brown adipogenesis but downregulated during white adipogenesis. None of these have been characterized in adipogenesis, but we hypothesized their regulatory mechanisms based on literature review (Figure 2.2C).

LINC00312 may regulate adipogenesis by two potential mechanisms.

(1) LINC00312 directly binds to and negatively regulates expression of miR-9 in breast cancer.[116] Also, miR-9 is upregulated during 3T3-L1 adipogenesis, but surprisingly, its overexpression inhibits 3T3-L1 differentiation.[117] LINC00312 potentially regulates adipogenesis by regulating availability of miR-9.

(2) LINC00312 directly binds to and negatively regulates expression of miR-21 in colorectal cancer cells.[118] miR-21 is reported to positively regulate adipogenesis and is upregulated in the adipose tissue of obese individuals.[119] miR-21 in 3T3-L1 cells decreased expression of adipogenic markers and increased expression of genes involved in thermogenesis and browning.[120] Therefore, miR-21 is a possible link between LINC00312 and adipogenesis.

LINC00607 may also regulate adipogenesis through two miRNAs.

(1) LINC00607 possibly acts as a sponge for miR-4732-3p in hepatocellular carcinoma.[121] This miRNA negatively targets TBX15[122], a gene essential for differentiation of brown and brite adipocytes, but not white adipocytes.[123]

(2) LINC00607 acts as a sponge of miR-607 in osteosarcoma.[124] Mir-607 was shown to downregulate EGFR in triple-negative breast cancer cells[125] – a gene required for white adipogenic differentiation.[126]

TYMSOS may regulate adipogenesis via three mechanisms.

(1) TYMSOS may regulate adipogenesis via PI3K-AKT signaling, which is important for the differentiation and browning of white preadipocytes.[127] TYMSOS was shown to activate the PI3K/AKT signaling pathway in thyroid cancer cells.[128]

(2) TYMSOS sponges miR-4739 in gastric cancer cells[129], and miR-4739 was shown to promote adipogenic differentiation of hBMSC (human bone marrow stromal cells).[130]

(3) TYMSOS was also shown to act as a sponge for miR-214-3p in NSCLSC (non–small cell lung cancer) cells[131] and miR-214-3p was shown to promote adipogenesis of 3T3-L1 cells.[132]

In summary, our analysis suggests that LINC00312, LINC00607 and TYMSOS may regulate adipogenesis by acting as sponges to limit the availability of specific miRNAs.

2.3.2 Alternative polyadenylation could regulate gene expression during adipogenesis.

APA is a co-transcriptional process that can generate multiple isoforms through differential usage of alternative pA sites in the 3'-UTR as well as the coding region.[18] We analyzed our 3'-end RNA-seq data of preadipocytes (Day 0) and adipocytes (Day 14) to identify differential usage of pA sites during adipogenesis. Our analysis of 4244 genes for significant 3'-UTR APA identified 70 shortened genes and 78 lengthened genes (Figure 2.3A). Genes undergoing 3'-UTR APA did not show a correlation with gene expression (Figure 2.3B). GOBP and KEGG enrichment analysis (Figure 2.4) showed that genes

undergoing 3'-UTR APA are involved in multiple pathways and processes, among which are adipogenesis-related terms such as "fatty acid metabolism," "PPAR signaling pathway", and "regulation of insulin secretion involved in cellular response to glucose stimulus." The lengthened genes encompassed more general terms, such as "collagen fibril organization".

Several key genes in adipogenesis undergo 3'-UTR APA during differentiation (Figure 2.3C). Shortened genes *SCD* and *ACSL5*, which are involved in fatty acid metabolism[133,134], are upregulated, whereas *CDK16*, a gene involved in cell cycle with known roles in differentiation[135], is unchanged in expression. Lengthened genes *WWTR1* and *E2F1* play important roles in adipogenesis[136,137] and are upregulated and downregulated, respectively, in gene expression. *HADHB*, another lengthened gene, is involved in fatty acid oxidation[138] and is upregulated during adipogenesis. This suggests that adipogenesis is directly affected by APA-mediated regulation of genes.

To assess the gain or loss of miRNA sites due to changes in 3'-UTR length of genes in Figure 2.3C, we used miRDB[76] to investigate the presence of miRNA-binding sites in the 3'-UTR region between the differentially used pA sites. We then checked if any of the predicted miRNAs were known to play a role in adipogenesis based on literature review[119,139,140], and if they are differentially expressed during adipogenesis[141] (Table 2.2).

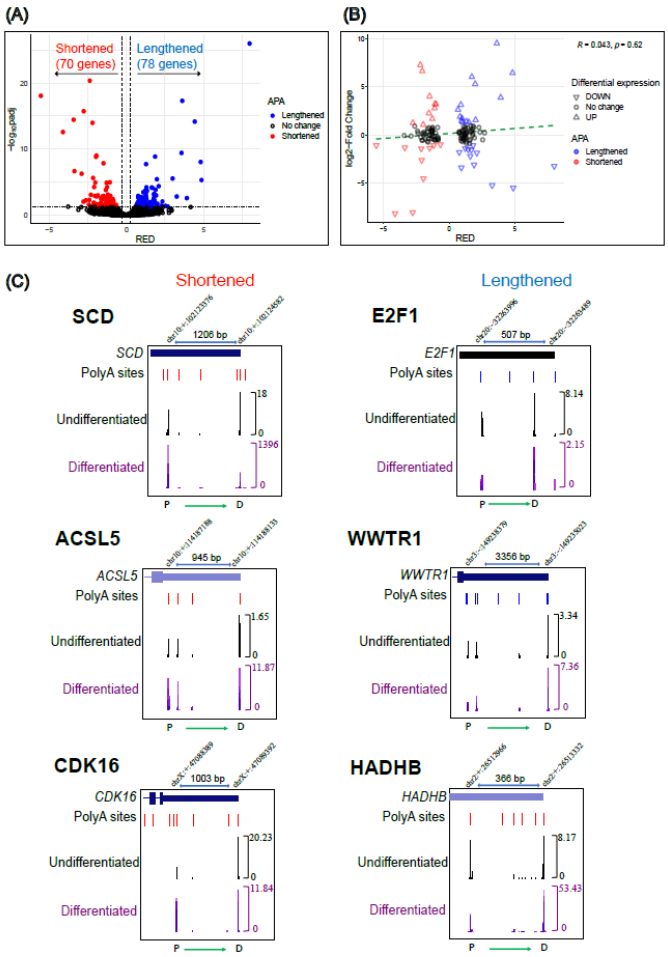


Figure 2.3: Alternative polyadenylation in the 3'-UTR region is regulated during adipogenesis. (A) Volcano plot depicts shortened (red) and lengthened (blue) genes (p-adjusted value <0.05, RED values >= 0.26 (lengthening) or <= -0.26 (shortening). The clear circles are the 4,096 genes that fell below these thresholds. The P-adjusted value and RED value thresholds are shown by dashed lines. (B) Scatter plot shows both RED and Log2-Fold expression change of shortened and lengthened genes. Genes with p-adjusted value <0.05 and log2-Fold change >1 (upregulated) are shown by triangles pointed upward and those with p-adjusted value <0.05 and log2-Fold change < -1 (downregulated) are shown by triangles pointed downward. Genes unchanged in expression are shown by black circles. The color of triangle outline represents shortened (red) and lengthened (blue) genes. The correlation between log2-fold changes and RED values is represented by the green dashed line of best fit, with the Spearman's correlation value and associated p-value shown in the top right of the graph. (C) UCSC genome browser plots depict shortened and lengthened genes. The genes are oriented as 5' (left) to 3' (right). The tracks show gene, annotated polyA-sites (blue indicates RNAs are transcribed from the minus whereas red are transcribed from the plus strand), and peaks (or reads) of merged triplicates of undifferentiated (black) and differentiated (purple) cells. Ranges of normalized coverage signals (reads per million) are given on the right. The gene direction is represented by the green arrow and the exact proximal (P) and distal (D) pA site positions are shown.

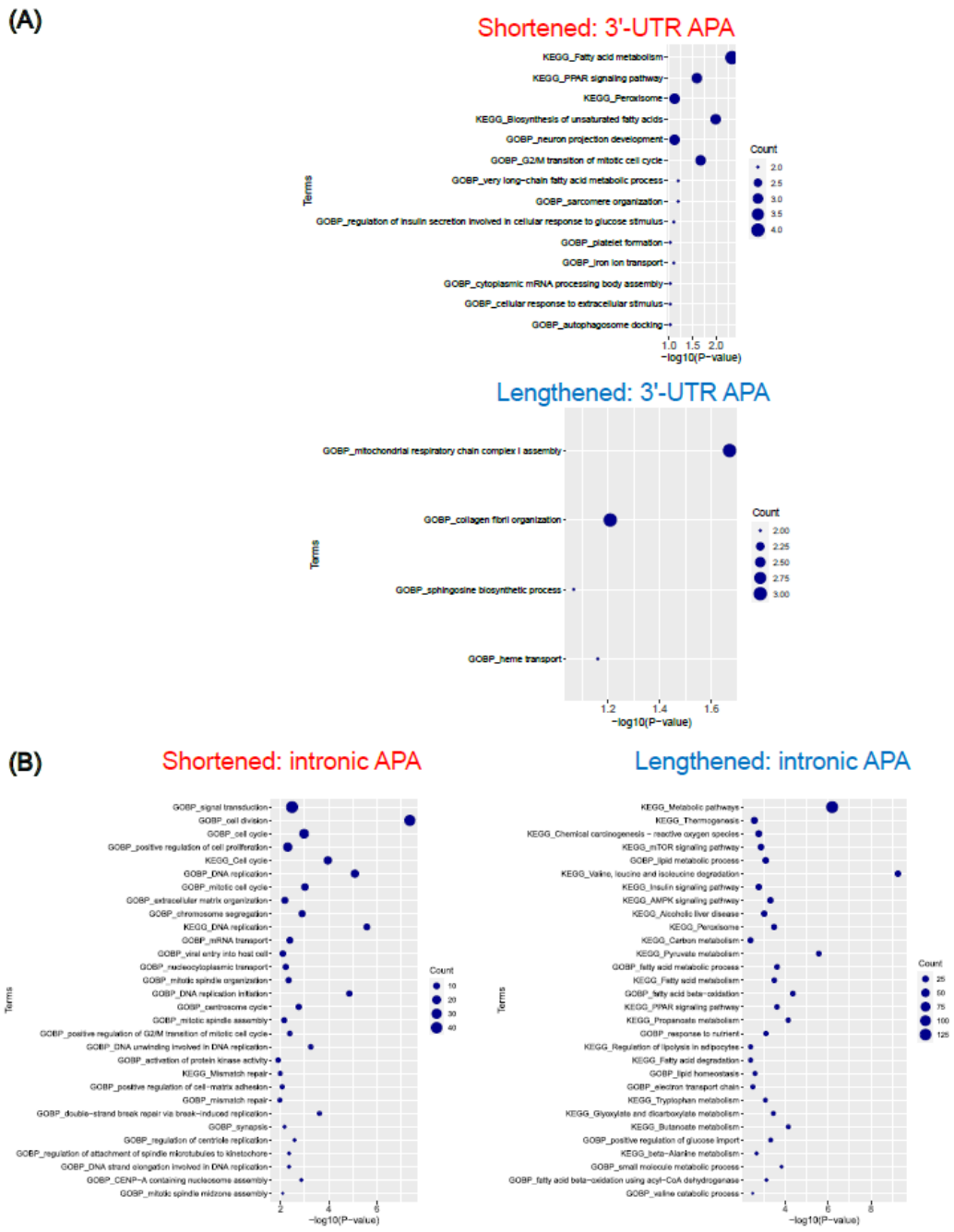


Figure 2.4: GOBP and KEGG enrichment analysis of genes undergoing (A) Shortening or lengthening due to UTR-APA and (B) Shortening or lengthening due to intronic APA (30 most significantly enriched terms are shown)

Gene	miRNAs with known functions in adipogenesis	miRNAs differentially expressed during adipogenesis
SCD	miR-181a-5p, miR-16-1-3p, miR-125b-2-3p	miR-124-3p, miR-125b-2-3p, miR-132-5p, miR-16-1-3p, miR-181a-5p, miR-181b-5p, miR-181c-5p
ACSL5		miR-134-3p, miR-135a-5p, miR-135b-5p, miR-218-1-3p, miR-450a-1-3p
CDK16	miR-23a-5p, miR-23b-5p, miR-125a-5p, miR-125b-5p	miR-23a-5p, miR-23b-5p, miR-125a-5p, miR-125b-5p, miR-495-3p, miR-185-5p
E2F1		
HADHB	miR-30a-3p	miR-30a-3p, miR-30d-3p, miR-30e-3p, miR-376a-5p, miR-450b-5p
WWTR1	miR-340-5p, miR-451b, miR-9-3p, miR-17-3p, miR-10b-3p, miR-582-5p	let-7f-2-3p, miR-130b-5p, miR-142-3p, miR-215-3p, miR-224-3p, miR-335-3p, miR-340-5p, miR-374a-5p, miR-374b-5p, miR-376a-3p, miR-376b-3p, miR-382-5p, miR-451b, miR-495-3p, miR-513a-3p, miR-513b-5p, miR-543, miR-627-5p, miR-9-3p

Table 2.2: List of miRNAs that have binding sites in a 3'-UTR region gained or lost during APA lengthening or shortening, respectively. Shown are the genes undergoing UTR-APA and the miRNAs with binding sites in the UTR that are known to regulate adipogenesis and ones that are differentially expressed during adipogenesis.

Based on Table 2.2 and literature review, we propose below how APA can modulate miRNA-mediated regulation of adipogenesis:

1. The shortening of *SCD* leads to loss of binding sites for miR-181a-5p, miR-16-1-3p, and miR-125b-2-3p. Both miR-181a-5p and miR-16-1-3p promote 3T3-L1 differentiation[142,143] whereas miR-125b-2 inhibits lipogenesis and negatively regulates expression of *SCD* in mice fed with high-fat diet.[144] Therefore, the shortening of *SCD* and consequent avoidance of miRNA-mediated regulation,

may contribute to increased expression of *SCD* and its correct timing, during adipogenesis.

2. Shortening of *CDK16* leads to loss of binding sites for miR-23a-5p, miR-125a-5p, and miR-125b-5p, which are all negative regulators of adipogenesis[145–147], and for miR-23b-5p, which inhibits the thermogenic program of brown adipocytes.[148]
3. Shortened gene *ACSL5* lost binding sites for miRNAs known to be differentially expressed during adipogenesis.
4. *HADHB* lengthening led to gain of binding sites for miR-30a-3p, which triggers anti-inflammatory responses in adipose tissue and increases insulin sensitivity in diet-induced obesity mice.[119] Overexpression of miR-30a can also activate the beige fat transcriptional program.[149] Extended 3'-UTRs can act as sponges to trap miRNAs and prevent their action at other mRNAs.[58] Therefore, lengthening of *HADHB* may not subject it to miR-30a-mediated regulation since *HADHB* is upregulated during adipogenesis, but it could competitively bind to miR-30a-3p to prevent browning of differentiated white adipocytes.
5. Lengthening of *WWTR1* leads to gain of binding sites for miR-340-5p, miR-451b, miR-9-3p, miR-17-3p and miR-10b-3p. Among these miRNAs, miR-17-3p[150] and miR-582-5p[151] may promote adipogenesis, while the others negatively regulate adipogenesis.[152–155] *WWTR1* negatively regulates adipogenesis in mice by down regulating PPARγ activity[136,156] and its expression is upregulated in our dataset. Therefore, the presence of binding sites of these miRNAs in the lengthened 3'-UTR suggests that it may be subjected to miRNA-mediated regulation that affects

its protein levels. Alternatively, the extended *WWTR1* sequence could competitively bind to miRNAs that inhibit adipogenesis.

6. Lengthened gene *E2F1* did not gain any miRNA-binding sites of adipogenesis-related or differentially expressed miRNAs.

We also assessed which RBPs may regulate APA of genes shown in Figure 2.3C. To do so, we checked for RNA-binding motifs by RBPmap[79] and evidence of binding from CLIP-Seq data in POSTAR3[80] in the regions 200 bp upstream and downstream of proximal and distal pA sites. Table 2.3 lists RBPs satisfying both criteria that are known to function in adipogenesis[26,28,157,158], as well as other RBPs that are not reported to have roles in adipogenesis.

Among the "known RBPs", IGF2BP2 was predicted to bind around the proximal pA site of *CDK16*, the distal pA site of *HADHB*, and around both proximal and distal pA sites of *SCD* and *WWTR1*. IGF2BP1 motifs were found around the proximal sites of *SCD* and *HADHB*, and SRSF1 bound around both the distal and proximal pA sites of *E2F1* and only the distal site of *CDK16*. While the "other RBPs" have not been shown to play a role in adipogenesis, they have been shown to regulate APA in other systems[159] and may therefore regulate adipogenesis through APA. For example, both U2AF2 and SRSF7 regulate 3'-UTR length[159], and in our adipogenesis dataset, U2AF2 bound around proximal and distal pA sites of multiple genes while SRSF7 bound only around the proximal pA site of the shortened *CDK16* gene.

RBPs predicted to bind around the polyA sites of multiple genes undergoing 3'-UTR APA.

Gene	SCD		ACSL5		CDK16		E2F1		HADHB		WWTR1	
PolyA site	Prox	Dist	Prox	Dist	Prox	Dist	Prox	Dist	Prox	Dist	Prox	Dist
Known RBPs												
IGF2BP2												
IGF2BP1												
SRSF1												
Other RBPs												
IGF2BP3												
TIA1												
RBM47												
TARDBP												
RBFOX1												
HNRNPA1												
PABPC4												
FUBP3												
HNRNPC												
LIN28A												
RBM15B												
U2AF2												
PCBP2												
FMR1												
PCBP1												

RBPs only predicted to bind around the polyA sites of one of the examined genes undergoing 3'-UTR APA.

Gene		SCD	ACSL5	CDK16	E2F1	HADHB	WWTR1
PolyA site	**Prox**	HNRNPF EWSR1		HNRNPK SRSF7	EIF4G2 FUS HNRNPM PUM1 RBFOX2		
	Dist	HNRNPU PUM2 QKI			FXR2 PABPN1	RBM47	HNRNPD KHSRP TAF15

RBPs predicted to bind around intronic polyA sites showing altered use during adipogenesis

Gene	NCAPG2	ORC1	CDCA2	NDUFB3	MLYCD	PIK3R1
Known RBPs						
SRSF1						
Other RBPs						
HNRNPA1						
HNRNPC						
Unique RBPs						
	TAF15, U2AF2					TARDBP

Table 2.3: Predicted binding of RNA-binding proteins around polyA sites of genes undergoing APA. The table is divided into three sections. The upper section shows RBPs that could bind around the proximal (Prox) or distal (Dist) polyA sites of genes undergoing 3'-UTR APA and are known to function in adipogenesis (Known RBPs), and RBPs that bind around the polyA site but do not yet have defined roles in adipogenesis (Other RBPs). Shaded boxes represent RBP binding around proximal (lighter shade) and distal (darker shade) polyA sites of particular genes. The middle section shows RBPs that are predicted to bind at polyA sites of only one of the examined genes undergoing 3'-UTR APA. The bottom section shows Known and Other RBPs that bind around intronic polyA sites showing altered use during adipogenesis, and RBPs that are predicted to bind at intronic polyA site of only one of the examined genes undergoing intronic APA (Unique RBPs). Shaded boxes in the bottom section represent RBP binding around the intronic site.

In addition to 3'-UTR APA, multiple genes, out of a total of 8505 genes, were identified as undergoing intronic APA, which led to shortening (increased intronic pA site usage) for 843 genes or lengthening (decreased intronic pA site usage) for 1428 genes (Figure 2.5A). Genes undergoing intronic APA showed a strong correlation with expression (read counts), whereby genes with suppressed usage of intronic pA sites (increased RED values) were upregulated, while those with increased usage (decreased RED values) were downregulated (Figure 2.5B). GOBP and KEGG enrichment analysis shows that most genes with increased intronic APA are related to cell cycle and DNA replication, while those with suppressed intronic APA are enriched in categories involved in lipid and fatty acid metabolism and insulin signaling, consistent with the increased need of adipocytes for these gene products (Figure 2.4B).

Examples of genes with changes in intronic APA are shown in Figure 2.5C. Usage of intronic pA sites increases in the *NCAPG2*, *ORC1* and *CDCA2* transcripts during adipogenesis. NCAPG2 and CDCA2 are involved in cell cycle regulation[160,161] and ORC1 in DNA replication[162], and all are downregulated in expression during adipogenesis. Intronic pA site usage is suppressed in *NDUFB3*, *MLYCD*, and *PIK3R1*.

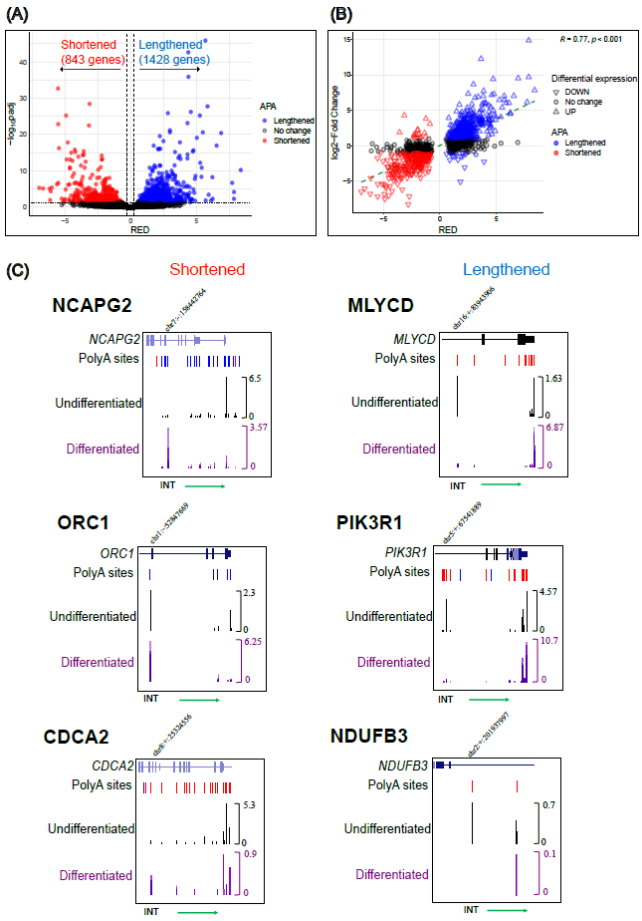


Figure 2.5: Alternative polyadenylation in introns is regulated during adipogenesis (A) Volcano plot depicts shortened (red) and lengthened (blue) genes (p-adjusted value <0.05, RED values >= 0.26 (lengthening) or <= -0.26 (shortening). The clear circles are genes that undergo no APA. The P-adjusted value and RED value thresholds are shown by dashed lines. (B) Scatter plot shows both RED and Log2-Fold expression change of shortened and lengthened genes. Genes with p-adjusted value <0.05 and log2-Fold change >1 (upregulated) are shown by triangles pointed upward and those with p-adjusted value <0.05 and log2-Fold change < -1 (downregulated) are shown by triangles pointed downward. Genes unchanged in expression are shown by black circles. The color of triangle outline represents shortened (red) and lengthened (blue) genes. The correlation between log2-fold changes and RED values is represented by the green dotted line of best fit, with the Spearman's correlation value and associated p-value shown in the top right of graph. (C) UCSC genome browser plots depict shortened and lengthened genes. The genes are oriented as 5' (left) to 3' (right). The tracks show gene, annotated pA-sites (blue indicates RNAs are transcribed from the minus whereas red are transcribed from the plus strand), and peaks (or reads) of merged triplicates of undifferentiated (black) and differentiated (purple) cells. Ranges of normalized coverage signals (reads per million) are given on the right. The gene direction is represented by the green arrow and the exact intronic pA site (INT) position is shown.

NDUFB3 has been implicated in brown adipogenesis[163], while MLYCD is needed for fatty acid biosynthesis[164] and PIK3R1 in the PI3K/AKT pathway is important for insulin signaling in adipocytes.[165] Both *MLYCD* and *PIK3R1* increase in expression during adipogenesis. *NDUFB3* expression is unchanged but the proportion of transcripts ending in the 3' UTR greatly increases.

We analyzed the presence of RBP motifs and binding as described earlier by considering the sequence 200 bp upstream and 200 bp downstream of the target intronic pA site of genes in Figure 2.5C (Table 2.3). For most intronic pA sites, we did not detect binding of an RBP. However, among the RBPs known to function in adipogenesis, SRSF1 bound around intronic pA sites of *NCAPG2* and *CDCA2*. The binding sites of RBPs HNRNPA1 and HNRNPC were found around the intronic pA sites of *CDCA2, NCAPG2, NDUFB3*, and *PIK3R1* and both RBPs are known to function in APA.[166,167]

2.3.3 Alternative splicing regulates key genes in adipogenesis.

By changing which exons are included in the final mRNA, alternative splicing can regulate gene expression by generating more than one mRNA isoform per gene, hence contributing to transcriptomic and proteomic diversity. It happens during adipogenesis, and obese and lean individuals have different alternative splicing profiles.[26] To analyze global changes in alternative splicing during adipogenesis, we used the program rMATS[65] on a recently published RNA-seq dataset of the differentiation of human primary preadipocytes.[64] We identified 100 alternative splicing events, with the most common types of splicing being “skipped exon” (SE) events (Figures 2.6A and 2.6B). One gene can undergo multiple events of the same event type as well as multiple event types.

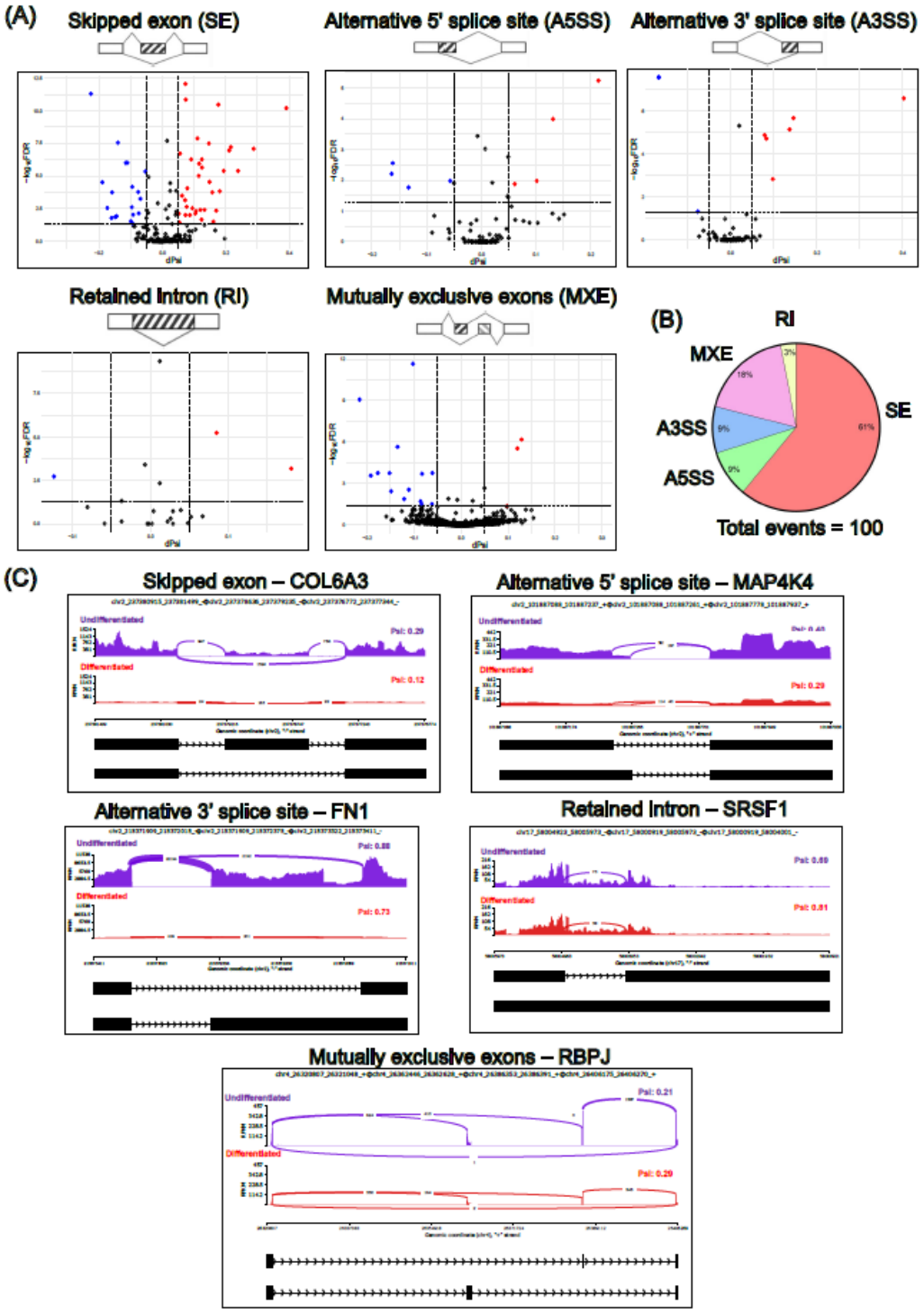


Figure 2.6: **Analysis of the global splicing profile reveals important genes being regulated by splicing during adipogenesis** (A) Volcano plots of distribution of significant splicing events (FDR less than 0.05 and dPsi value greater than or less than 0.05 or -0.05, respectively) in each type of splicing event. Red denotes splicing events where psi value is greater than 0.05 and the target exon/region is included, whereas blue denotes splicing events where dPsi value is less than -0.05 and the target exon/region is

skipped. For the graphical representation of splicing events, blank rectangles represent constitutive exons and shaded exons or introns represent alternatively spliced regions. (B) Percentages of splicing events distributed across the 5 types of splicing events. (C) Sashimi plots depict splicing of genes during adipogenesis. Each plot is labelled with the splicing event type and gene symbol above. Genomic coordinates of the exons in question are shown on top of plot. Tracks show junction counts and the psi value with top (purple) being the undifferentiated sample and the bottom (red) being differentiated. At the bottom, exons are shown as black boxes and introns as lines with arrows showing directions of genes.

To understand the roles the alternatively spliced genes might play in adipogenesis, we carried out GOBP and KEGG enrichment analyses (Figure 2.7). Although the enriched terms are not directly associated with adipogenesis, the spliced genes are involved in various processes important for adipogenesis such as "positive regulation of GTPase activity" and "ECM-receptor interaction".[168,169]

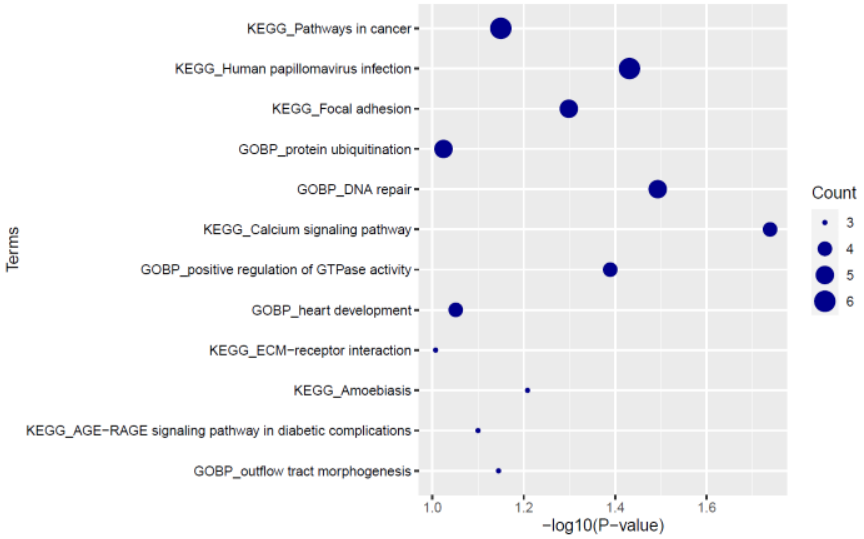


Figure 2.7: GOBP and KEGG enrichment analysis of spliced genes.

We analyzed some of the spliced genes in greater detail (Figure 2.6C):

(1) **Skipping of exon 6 of COL6A3**: *COL6A3* encodes the alpha-3 chain of type VI collagen, which is highly enriched in adipose tissue. The expression of *COL6A3*

expression in abdominal subcutaneous adipose tissue positively correlated with BMI and total body fat mass.[170] Exon 6 is skipped less frequently in preadipocytes compared to adipocytes. Studies have shown that the inclusion of exon 6 of *COL6A3* is correlated with upregulation in cancer.[171,172] This suggests that the skipping of exon 6 may be responsible for the downregulation that we observe in adipocytes.

(2) **Alternative 5' splice site usage in MAP4K4**: In mice, MAP4K4 negatively regulates adipogenesis and insulin sensitivity.[173] Also, during brown adipogenesis in mice, alternatively spliced *MAP4K4* isoforms had different effects on the phosphorylation of JNK, which in turn was correlated with the differentiation and metabolic signature of brown adipocytes.[174] MAP4K4 was not studied in human adipogenesis, but its silencing in myotubes was shown to prevent TNF-α-mediated insulin resistance and increase glucose uptake.[175] In our analysis, the expression of *MAP4K4* decreases during adipogenesis, and alternative 5'- splice site usage in the target exon of *MAP4K4* may regulate adipogenesis through changes in *MAP4K4* expression and/or through signaling via JNK phosphorylation.

(3) Alternative 3' splice site usage in FN1: FN1 is an extracellular matrix (ECM) protein with roles in cell adhesion, migration, and fibrosis. *FN1* is also known to undergo alternative splicing that results in proteins with different tissue localization.[176] It is dysregulated in obese adipose tissue[177], and we also observed that its expression is downregulated during adipogenesis.

(4) **Mutually exclusive exons of RBPJ:** RBPJ has been labelled as "mutually exclusive exons" event by rMATS. It is involved in both transcriptional activation and repression of genes and is a downstream effector of the Notch signaling pathway.[178] Though RBPJ is not directly associated with adipogenesis, Notch signaling is a known regulator of the process.[179,180] Therefore, alternatively spliced transcripts of *RBPJ* may regulate adipogenesis through Notch signaling.

(5) **Intron retention of SRSF1**: SRSF1 is a splicing factor that negatively regulates adipogenesis[28], but there have been no studies on the expression of its variants in the adipocyte system. Intron retention in the 3'-UTR region of SRSF1 is higher in differentiated cells in our analysis. With intron retention, the transcript was shown to be more stable and produce more protein.[181,182] Since SRSF1 is a negative regulator of adipogenesis, splicing out of the intron may be required for undifferentiated cells to undergo differentiation, but after differentiation, SRSF1 may have other functions in mature adipocytes.

To identify splicing factors that could contribute to global alternative splicing changes, we used rMAPS2[81,82] (RNA map analysis and plotting server). The program rMAPS2 carries out motif enrichment analysis (based on known consensus RBP motifs from the literature) on various regions of alternatively spliced exons – specifically, intronic sequences 250 bp upstream or downstream of target and flanking exons, as well as the first 50 bp of the 5'- or 3'-end of the exons.[81] Since SE events had the highest number of regulated exons, we limited our analysis to enrichment of RBP motifs in SE events. Among the RBPs identified by rMAPS2, we observed nine RBPs that are already known to influence adipogenesis.[26,28,157,158,183] Predictions of where these known RBPs could bind

(along with the motifs) to regulate global changes in splicing during adipogenesis are shown in Figure 2.8 and Figure 2.9 and summarized in Figure 2.8D. For example, SRSF1 could mainly promote exon skipping by binding in the intronic sequence upstream of the target exon. Sam68 (KHDRBS1), on the other hand, could lead to exon inclusion if it binds just downstream of the 3'-end of the upstream flanking exon, but can also cause a skipped exon event if it binds just downstream of 3'-end of the target exon. IGF2BP2 could stimulate inclusion of exons by binding in the intronic sequence downstream of the target exon or promote skipping if it binds upstream of the target exon (Figure 2.8B). We also identified six other RBPs that are not known to have roles in adipogenesis but could potentially be major regulators of alternative splicing (Figure 2.8C and 2.8D). Both ESRP1 and PABPC1 could cause skipping of exons. LIN28A and KHPRBS2 could mainly promote inclusion of exons by binding to intronic regions downstream and upstream of the target exon, respectively. CPEB4 could lead to skipping or inclusion of the target exon by binding upstream or downstream of the exon, respectively. HNRNPA1 is predicted to mainly bind to regions downstream of the target exon and promote inclusion and would lead to skipping only if it binds to a region within 125 bp upstream of the downstream flanking exon.

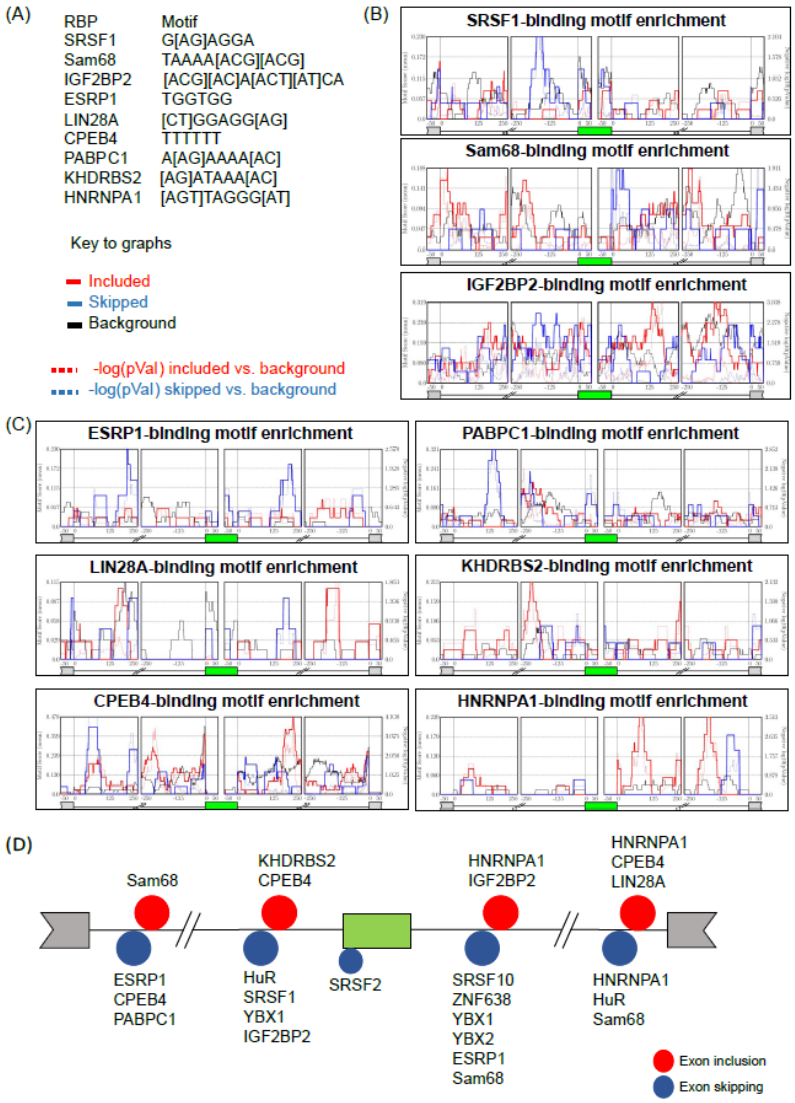


Figure 2.8: Specific splicing factors may regulate global "skipped exon" events. (A) Specific motif sequences used to search for binding sites in the vicinity of the skipped exon for the splicing regulators shown in panels (B) and (C) and key to the binding profiles of RBPs traced in the graphs. The solid lines are motif scores (Y axis, left side), calculated as the overall percentage of nucleotides covered by the motif in a 50 bp window, with red denoting the motif density score for exons with increased inclusion (38 events), blue for exons with increased skipping (22 events), and black for background (non-regulated) exons (70 events). The dotted lines are –log10(P-value) (Y-axis, right side) based on comparison of motif scores between regulated exons against background exons. Only -log10(P-value) greater than 1.3 are considered. The green exon is the target exon (that is skipped or included) and the flanking exons are shown in grey. (B) Maps of the RBPs SRSF1, Sam68 (KHDRBS1), and IGF2BP2, which are known to play a role in adipogenesis. (C) Maps of novel RBPs ESRP1, LIN28A, CPEB4, PABPC1, KHDRBS2, and HNRNPA1 that could alter the global alternative splicing profile during adipogenesis. (D) Summary of binding profiles of the RBPs, with proteins promoting inclusion indicated as red circles and those promoting skipping as blue circles.

(A)

RBP	Motif
SRSF10	AGAGA[AG][AG]
ZNF638	[CGT]GTT[GC][GT]T
HuR	TT[AT]GTTT
SRSF2	GGAG[AT][AGT]
YBX1	AACATC
YBX2	AACA[AT]C[AGT]

Key to graphs

- Included
- Skipped
- Background

-log(pVal) included vs. background
-log(pVal) skipped vs. background

(B)

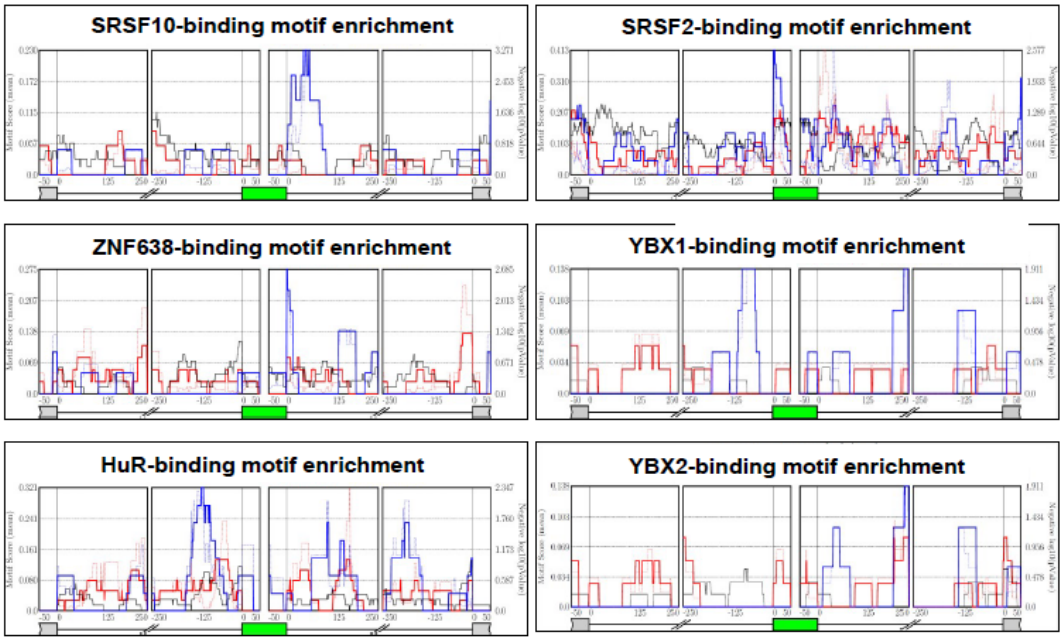


Figure 2.9: Splicing factors known to play a role in adipogenesis may regulate global "skipped exon" events. (A) Specific motif sequences used to search for binding sites in the vicinity of the skipped exon for the splicing regulators shown in panel (B) and key to the binding profiles of RBPs traced in the graphs. The solid lines are motif scores (Y axis, left side), calculated as the overall percentage of nucleotides covered by the motif in a 50 bp window, with red denoting the motif density score for exons with increased inclusion (38 events), blue for exons with increased skipping (22 events), and black for background (non-regulated) exons (70 events). The dotted lines are –log10(P-value) (Y-axis, right side) based on comparison of motif scores between regulated exons against background exons. Only values greater than 1.3 are considered (P-value < 0.05). The green exon is the target exon (that is skipped or included) and the flanking exons are shown in grey. (B) Maps of more RBPs - SRSF10, SRSF2, ZNF638, HuR, RBM4, YBX1 and YBX2 - which are known to play a role in adipogenesis.

2.4 Discussion

Obesity is an increasing global health problem. It is therefore important to understand the basic mechanisms which regulate human adipogenesis because its dysregulation can lead to obesity and obesity-related disorders.[2] In this study, we have used multiple sequencing datasets to identify potentially important regulators of human adipogenesis at the level of lncRNAs, alternative splicing and alternative polyadenylation.

2.4.1 Long non-coding RNAs

Using the sequencing dataset that we generated from human primary preadipocytes, we first explored lncRNA-mediated regulation. We reasoned that the most important lncRNAs were likely to be those with high change in expression during white adipogenesis, and similarities or differences in expression when compared to their changes during brown adipogenesis. The changes in lncRNAs levels may drive adipogenesis or facilitate function of the differentiated adipocytes. It will be important to characterize their role in adipogenic systems as they may be crucial in maintaining the differentiated state or in determining susceptibility to develop obesity or related disorders and may therefore be important therapeutic targets. Additionally, adiposity has been associated with certain types of cancer[52], and because some of the lncRNAs highlighted in our study have been associated with cancer, they may affect the susceptibility to developing cancer. LncRNAs that undergo a similar change in expression in both white and brown adipogenesis may function in the commitment of mesenchymal stem cells into the adipocyte lineage or the maintenance of the differentiated cell state. The lncRNAs that change in opposite directions during white and brown adipogenesis are especially

interesting, since browning of white adipose tissue has received increased attention as a therapeutic approach to obesity and its related disorders.[184,185] Therefore, the lncRNAs LINC00312, LINC00607 and TYMSOS, that increase during brown adipogenesis but decrease during white adipogenesis, could be important therapeutic candidates.

Our analysis of lncRNAs is limited by lack of time-course data. For example, we did not see a change in expression of HOTAIR between preadipocytes at Day 0 versus mature adipocytes at Day 14, but this finding is consistent with the previous report that HOTAIR increases during early adipogenesis and decreases at later stages.[64] There are also depot-specific[87] as well as sex-specific[186] differences in adipose tissue gene expression. For example, HOTAIR was shown to vary in expression in gluteal versus abdominal subcutaneous tissue.[187] Since our samples are not depot-specific or sex-specific and only examined the terminal differentiated state, our experimental design does not allow discrimination of these reported differences.

2.4.2 Alternative polyadenylation (APA)

APA during adipogenesis has not been extensively investigated. The two studies on global APA changes were published ten years ago and mainly focused on identifying a trend to longer 3' UTRs during adipogenesis[61,62], and the one examining human cells used total RNA-seq data and did not characterize intronic APA. Our study is the first to generate 3'-end sequencing data of human preadipocytes and adipocytes, which provides much more accurate and quantitative characterization of pA site usage. We observed only a slight trend towards 3'-UTR lengthening during human adipogenesis, consistent with the modest tendency towards longer 3'-UTRs that was previously reported[62], but unlike

the more notable trend seen during mouse 3T3-L1 adipogenesis.[61] This divergence may be due to differences in models used to study the process, as well as differences in experimental design and analysis. Interestingly, increased or decreased use of intronic pA sites was much more prevalent in our data set than changes in the 3'-UTR at the gene's end. Suppression of an intronic pA site was most common, and a prime example is *PIK3R1*, a gene that regulates adipogenesis and insulin sensitivity.[165] This APA event is expected to yield more transcripts extending to the end of the gene and being translated into full-length protein. The strong enrichment in this group of genes involved in lipid and fatty acid metabolism and insulin signaling would be consistent with adipocytes needing more of these types of proteins.

Changes in 3'-UTR length could subject genes to different degrees of miRNA-mediated regulation. We found that several miRNAs with known functions in adipogenesis had potential binding sites in 3'-UTR sequences of the shortened or lengthened transcripts that we examined, suggesting miRNA-mediated regulation of these genes. We also identified several other miRNAs that are differentially expressed during adipogenesis and may play important, yet undefined roles in this transition.

To identify RBPs that may function in both 3'-UTR APA and intronic APA, we used motif enrichment analysis and global protein/RNA crosslinking data to identify binding sites in the regions 200 bp upstream and downstream of pA sites. We found that RBPs with known roles in adipogenesis bind around alternate pA sites of seven of the twelve genes that we examined, and also identified several other RBPs that could be regulating ten of these twelve genes. Interestingly, IGF2BP2, SRSF1, and HNRNPA1 are possible

regulators of skipped exon events during adipogenesis, and thus might be regulating both alternative splicing and APA.

For *ORC1* and *MLYCD* that are undergoing intronic APA, no RBPs were detected to bind around the intronic pA site. Besides RBPs, other factors such as transcriptional dynamics and expression levels of subunits of the cleavage and polyadenylation complex can regulate APA[18], and may function in regulating intronic APA of these genes during adipogenesis. Regulation may also occur by suppressing recognition of the splice sites of the flanking exons.[60] It is important to note that the CLIP-seq evidence for our RBP-binding predictions was based on experiments with non-adipogenic cell lines, such as HEK293, and further experiments are needed to demonstrate a direct function of the various RBPs on regulating APA during adipogenesis.

2.4.3 Alternative splicing

Spliced isoforms have been shown to play roles in adipogenesis[28,30,63], and our analysis of a publicly available RNA-seq dataset revealed 100 alternative splicing events occurring during human adipogenesis. Like our study, the global analysis of alternative splicing in human adipogenesis published by Yi *et al*[31], also identified exon skipping as the most prevalent type of alternative splicing. However, they analyzed cells derived from bone marrow whereas the cells in our study originated from subcutaneous adipose tissue. The use of primary human preadipocytes, the stringent thresholds that we have used, and the application of a widely used and reliable alternative splicing analysis tool, rMATS[65] makes our analysis a valuable resource for future studies to dissect the importance of alternative splicing to adipogenesis.

Our study uncovered possible mechanisms of adipogenesis regulation by specific splicing events. Of note is splicing of *MAP4K4* transcripts, which encode a potential therapeutic target that is alternatively spliced during brown adipogenesis in mice.[174] Alternative splicing of *RBPJ*, a gene involved in the Notch signaling pathway[178], is interesting as this pathway is associated with adipocyte dedifferentiation, a therapeutic potential in the field of obesity.[188] SRSF1, which shows increased intron retention in mature adipocytes, is an important regulator of alternative splicing in adipogenesis[28], and in our analysis, it is a potential mediator of both exon skipping and APA. Additionally, by motif enrichment analysis, we demonstrated how RBPs previously characterized in adipogenesis can potentially regulate alternative splicing of multiple genes during differentiation, thus extending their roles in adipogenesis beyond the few targets examined in previous studies. We have also identified novel RBPs that could regulate the global alternative splicing profile during adipogenesis but are as yet uncharacterized.

To the best of our knowledge, our study presents the first comprehensive analysis of different modes of gene regulation during human adipogenesis. We generated 3'-sequencing datasets from matched preadipocyte and adipocyte RNA samples, as well as used publicly available datasets for analysis with specialized tools, and with these resources, identified multiple potential mechanisms of regulation that can be addressed in future research. In addition to illustrating how lncRNAs, alternative polyadenylation and alternative splicing may regulate adipogenesis, we also identified novel lncRNA, miRNA, and RBP regulators that may play a crucial role in fine-tuning adipocyte

differentiation. Understanding the basic mechanisms of gene regulation during adipogenesis will aid in designing therapeutics to combat obesity and its related disorders.

Chapter 3: Cleavage and polyadenylation factors are potential regulators of adipogenesis.

Salwa Mohd Mostafa and Claire Moore, To be submitted to *BMC Research Notes*.

3.1 Introduction

Adipogenesis is a finely regulated process for differentiation of preadipocytes to adipocytes. Dysregulation of this process can lead to health complications, such as the development of obesity and obesity-related disorders.[2] Therefore, understanding the regulation of adipogenesis at the molecular level would contribute to designing therapeutics. Cleavage and polyadenylation (C/P) is a mRNA processing step in which transcripts are cleaved at a specific site, followed by addition of a poly(A) tail that is important for mRNA stability, export, and translation. C/P is brought about by four major complexes – CPSF (cleavage and polyadenylation specificity factor), CSTF (cleavage stimulation factor), CFIm (mammalian cleavage factor I) and CFIIm (mammalian cleavage factor II) - and additional core factors - symplekin, polyA polymerase (PAP), nuclear polyA binding protein (PABPN1) and the C-terminal domain (CTD) of the largest subunit of RNA polymerase II.[19] These proteins are imperative to the transcription termination step and play important roles in alternative polyadenylation (APA) – a regulatory step that leads to isoform diversity due to differential usage of polyadenylation (pA) sites.[19] Previously, Hoque et al. found that there is an overall lengthening of 3'-UTR due to APA during 3T3-L1 differentiation.[61] However, changes in the expression of C/P factors during adipogenesis are still unknown. Our group has successfully shown earlier that protein levels of C/P factors change during macrophage differentiation and that CSTF64, a protein in the CSTF complex with known roles in APA, directly regulates the process.[189] Therefore, we aimed to study the changes of C/P factors during adipogenesis using 3T3-L1 as the model system.

3.2 Methods

3.2.1 Cell culture and Oil Red O staining: 3T3-L1 cells were cultured in maintenance media (Dulbecco's Modified Eagle Medium (DMEM - high glucose) containing 10% fetal bovine serum (FBS) and antibiotics (100 U/ml penicillin, 100 ug/ml streptomycin)). Cells were grown at 37°C with 5% CO_2. Cells were passaged every 3 days to prevent cell confluency over 80%. For differentiation, cells were plated at a density of 1.0 x 10^4 cells/cm^2 and grown till they reached 100% confluency. Cells were then maintained for two additional days for contact inhibition. On day 0, the cells were induced to differentiate using differentiation media (maintenance media supplemented with 0.5 mM 3-isobutyl-1-methylxanthine, 1 uM dexamethasone, 5 ug/mL insulin, and 1 uM rosiglitazone). On day 3, the differentiation media was replaced with insulin media (maintenance media containing 5 ug/mL insulin). On day 5, insulin media was replaced with maintenance media and cultured for 2 additional days before being harvested for protein or RNA.

3T3-L1 cells induced with control (3T3-L1-C) and CPSF73-overexpressing (3T3-L1-OE) vectors were generated previously in the lab.[190] Cells were plated at 1.0 x 10^4 cells/cm^2 and grown till they were 100% confluent and reached contact inhibition. Cells were grown and differentiated in maintenance media, differentiation media and insulin media as described earlier, except with the addition of 1.5 ug/ml of doxycycline (Dox) when the cells were plated for the differentiation experiment. The Dox-containing media were replenished every other day. Cells on Day 0 were subjected to differentiation media for 2 days, insulin media for 2 days, and cultured in maintenance media for 4 days before being harvested for protein and RNA and stained with Oil Red O staining.

For Oil Red O staining of Dox-induced 3T3-L1-C and 3T3-L1-OE, cells at Day 0 and Day 8 were washed twice with phosphate-buffered saline (PBS) and then fixed with 4% formaldehyde for one hour at room temperature. Then, the cells were washed once with PBS, once with 60% isopropanol, and dried. A 0.5% stock solution of Oil Red O (Sigma–Aldrich, Sigma O0625) in isopropanol was prepared and filtered through a 0.2-um filter. A fresh working solution was prepared by mixing the stock solution with distilled water in a 6:4 ratio, incubating it for 20 minutes at room temperature, and re-filtering through a 0.2-um filter. The working solution was then added to the fixed cells and incubated at room temperature for 1 hour. The cells were then washed extensively with distilled water, dried and photographed.

3.2.2 RNA extraction and RT-qPCR: Cells were lysed with Trizol (Thermo Fisher Scientific, 15-596-018) and RNA was extracted using the Zymo Direct-Zol RNA Miniprep kit (Zymo Research, R2050. One ug of RNA was then reverse transcribed to cDNA using NEB LunaScript® RT SuperMix kit (New England Biolabs, M3010L), and qPCR was carried out using NEB Luna® Universal qPCR Master Mix (New England Biolabs, M3003L) and primers listed in Table 3.1. Expression of adipogenesis markers were normalized to that of mouse *Rpl13a* and the qPCR results were quantified using the ddCt method.

Primers	**Sequence**
Fabp4 Forward	AAGGTGAAGAGCATCATAACCCT
Fabp4 Reverse	TCACGCCTTTCATAACACATTCC
Adipoq Forward	TGTTCCTCTTAATCCTGCCCA
Adipoq Reverse	CCAACCTGCACAAGTTCCCTT
Cfd Forward	CATGCTCGGCCCTACATGG
Cfd Reverse	CACAGAGTCGTCATCCGTCAC
Pparg Forward	CTCCAAGAATACCAAAGTGCGA
Pparg Reverse	GCCTGATGCTTTATCCCCACA
Rpl13a Forward	CTGCTCTCAAGGTTGTTCGGCT
Rpl13a Reverse	CCTTCCGTTTCTCCTCCAGAGT

Table 3.1: List of primers

3.2.3 Western blotting: Cells were lysed with RIPA buffer (150 mM sodium chloride, 50 mM Tris-HCl (pH 7.4), 1% NP-40, 0.1% SDS, 5mM EDTA, 0.1% sodium deoxycholate, 1mM Dithiothreitol, 10mM sodium fluoride, 200uM sodium orthovanadate, 1mM sodium pyrophosphate, 10mM B-glycerophosphate) with protease inhibitor cocktail (Bimake, B14001; Thermo Fisher Scientific, PI78425). Cells in RIPA buffer were incubated in ice for 15 min and then homogenized by vortexing. The lysates were cleared by centrifugation at 12,000 x g for 15 min at 4°C. Protein quantification was done by the BCA assay (Thermo Fisher Scientific, A53226). The lysates were prepared for western blotting with the addition of 4X SDS loading buffer (+355 mM β-mercaptoethanol) and heated to 95°C for 5 min. Then, 25-50 mg of lysate was resolved on a 10% Bis-Tris gel

and transferred to PVDF membranes. Total protein staining was done using Revert Total Protein Stain (LI-COR, 926-11021). Membranes were then blocked for 5 minutes using EveryBlot Blocking Buffer (Bio-Rad, 12,010,020) or for 30 minutes using 5% milk in 1x TBST (Tris Buffered Saline Buffer with Tween 20). Primary antibodies (Table 3.2) were diluted in 1% milk in 1x TBST, EveryBlot or only 1x TBST and incubated with membrane overnight at 4 °C. Membranes were washed three times in 1x TBST and incubated with HRP (HorseRadish Peroxidase)-labeled secondary antibodies for 1 hour at room temperature. A Syngene imager was used to capture the chemiluminescence signal of proteins and ImageJ was used to quantify the band intensities. A LI-COR Odyssey CLx was used to capture the total protein stain which was then quantified using Image Studio (version 5.5).

3.2.4 Differential gene expression analysis: For analysis of differential gene expression of during 3T3-L1 differentiation, we used the undifferentiated Day 0 (GSM3728574, GSM3728575, GSM3728576) and differentiated Day 8 (GSM3728580, GSM3728581, GSM3728582) samples from the publicly available GEO dataset GSE129957.[191] The raw sequence reads of the samples were processed to remove adapters using Cutadapt[66] (v2.8) and the quality of these trimmed reads was checked using FastQC[67] (v0.11.8) and MultiQC[68] (v1.7.0). The reads were then aligned using STAR[69] (2.6.1d) to mouse mm10 genome. Aligned reads were then quantified using featurecounts[70] (v1.6.3) and log2-fold changes were calculated with DESeq2[71](1.40.2) using default parameters. A fold change of 2 and an adjusted p-value of less than 0.05 were applied to select differentially expressed genes.

Protein	Antibody source	Blocking buffer
CPSF73	Santa Cruz Biotechnology (SCBT), sc-393001	5% milk in 1x TBST
CPSF30	Novus Biologicals, NB100-79827	EveryBlot Blocking Buffer
CPSF100	SCBT, sc-165983	5% milk in 1x TBST
CPSF160	SCBT, sc-166281	5% milk in 1x TBST
CSTF64	Bethyl Laboratories, A301-092A	5% milk in 1x TBST
CSTF77	SCBT, sc-376575	5% milk in 1x TBST
CFIm25	Bethyl Laboratories	5% milk in 1x TBST
CFIm59	SCBT, sc-393880	5% milk in 1x TBST
PCF11	SCBT, sc-514158	5% milk in 1x TBST
CLP1	Proteintech, 14746-1-AP	5% milk in 1x TBST
SYMPK	SCBT, sc-398897	5% milk in 1x TBST
PABPN1	Abclonal, A6041	5% milk in 1x TBST

Table 3.2: List of antibodies

3.3 Results

3.3.1 Levels of some cleavage and polyadenylation factors change during 3T3-L1 differentiation.

We analyzed differential gene expression of representative subunits of the C/P complex between undifferentiated and differentiated 3T3-L1 cells using a publicly available RNA-

seq dataset. None of the tested C/P factors are differentially expressed at the level of mRNA (Figure 3.1A), but some are changed at the level of proteins (Figure 3.1B). Multiple proteins decrease during differentiation, especially all of the tested CPSF proteins, which show the strongest decrease. CTSF77, CFIm59 and CFIIm (PCF11 and CLP1) also decrease in protein levels during 3T3-L1 differentiation. CSTF64, CFIm25, SYMPK and PABPN1 do not have remarkable differences in protein levels between undifferentiated and differentiated cells.

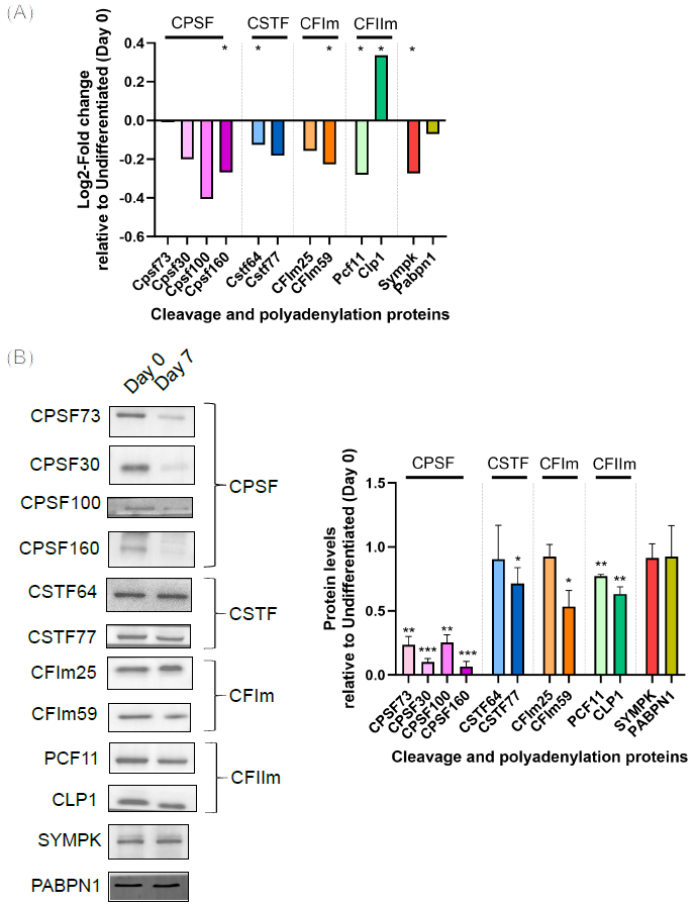


Figure 3.1: Protein levels of cleavage and polyadenylation factors change during 3T3-L1 differentiation. (A) There was no major difference in expression of mRNA of any of the C/P factors (log2-fold change of 1 and p-adjusted value < 0.05). P-adjusted values less than 0.05 are labelled with an asterisk. (B) Representative cropped blots of protein levels of representative C/P factors from each complex are shown, along with quantifications (n=3, mean ± SD) of protein levels at Day 7 relative to Day 0, where values above 1 indicate increased levels and values below 1 indicate decreased levels

during differentiation. The protein levels were normalized to total protein levels. Significance testing was done using student's unpaired two-tailed t-test with Welch's correction, p-value less than 0.05 are significant and denoted with an asterisk. Full-length blots are presented in Figure 3.2.

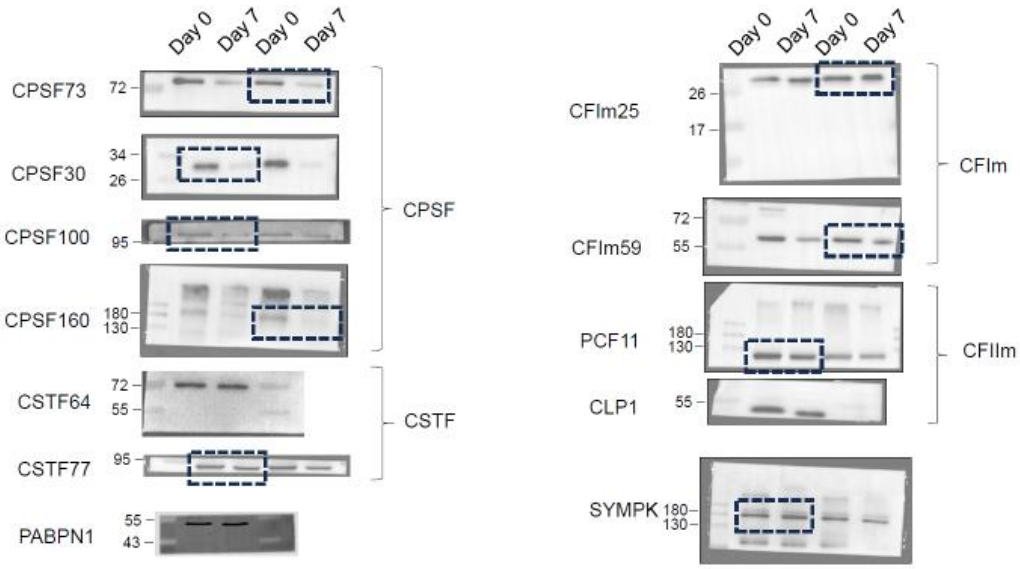


Figure 3.2: Uncropped blots of proteins shown in Figure 3.1.

Considering that we did not see a strong trend towards 3'-UTR lengthening in primary human adipocytes compared to what has been previously observed in 3T3-L1 cells, it would be interesting to compare the changes in levels of human C/P proteins during adipogenesis to that of mouse, as the RNA levels of human C/P proteins also do not change in expression during adipogenesis (Figure 3.3).

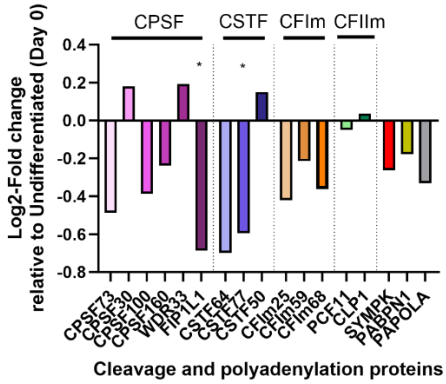


Figure 3.3 RNA levels of cleavage and polyadenylation factors do not change during human primary preadipocyte differentiation. The thresholds are log2-fold change of 1 and p-adjusted value < 0.05. Log2-fold changes are shown with p-adjusted values less than 0.05 are labelled with an asterisk.

3.3.2 Overexpression of CPSF73 alone does not the change efficiency or rate of 3T3-L1 differentiation.

Considering the observed large, significant decrease in protein levels of CPSF73 during adipogenesis and its role in adipogenesis through regulation by UBE3D[190], we investigated whether the differentiation of 3T3-L1 cells is affected when it is overexpressed. We used Control- and CPSF73-overexpressing 3T3-L1 cells previously generated in the lab[190] and induced CPSF73 overexpression when cells were plated for differentiation, so that CPSF73 was overexpressed in both undifferentiated (Day 0) and differentiated (Day 8) cells (Figure 3.4A). Using Oil Red O staining of fixed undifferentiated and differentiated cells, we did not see a remarkable change in differentiation efficiency upon overexpression of CPSF73 (Figure 3.4B). Since the overexpression of CPSF73 could be affecting the rate at which cells achieve the differentiated state without necessarily affecting the overall differentiation efficiency, we tested the levels of adipogenesis markers *Fabp4*, *Ppparg*, Adipsin (*Cfd*) and Adiponectin (*Adipoq*)[190] across five timepoints during differentiation (Days 0, 2, 4, 6 and 8). However, the expression of markers during the process was not affected by CPSF73 overexpression (Figure 3.4C).

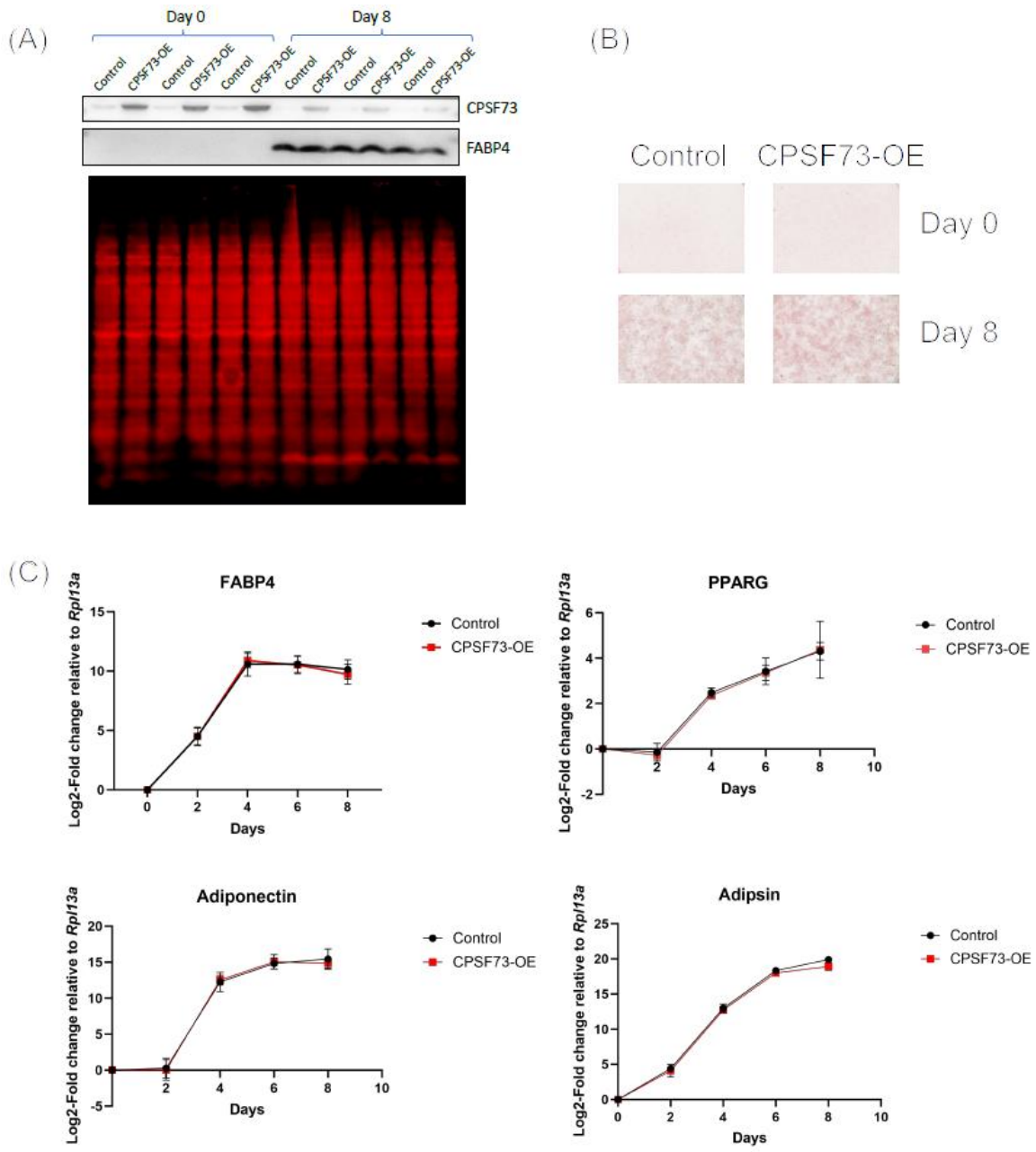


Figure 3.4: Overexpression of CPSF73 alone is not sufficient to affect differentiation efficiency. (A) CPSF73 is overexpressed in both undifferentiated and differentiated adipocytes compared to its control. Three replicates from Day 0 and from Day 8 of differentiation are shown. The FABP4 marker indicates successful differentiation, and the total protein stain serves as loading control. (B) Oil Red O staining of Dox-induced control and CPSF73-ovexpressing 3T3-L1 cells at Day 0 and Day 8 of differentiation does not show a change in differentiation efficiency due to overexpression of CPSF73 (B) Overexpression of CPSF73 does not significantly change rate of 3T3-L1 differentiation. Expression of mRNAs of adipogenesis markers normalized to *Rpl13A* are shown relative to Day 0 of differentiation.

3.4 Discussion

C/P factors play important roles in the 3'-end processing of pre-mRNAs to mRNAs and are implicated in diseases like cancer.[192] Changes in their protein levels can directly regulate APA[19] and therefore, studying their changes during 3T3-L1 differentiation identifies potential mechanisms to regulate gene expression during adipogenesis, for example, in the lengthening of genes due to APA during 3T3-L1 differentiation.[61] We have shown that protein levels of multiple C/P factors change during 3T3-L1 differentiation without changes in RNA expression. This may be due to changes in translational efficiency, protein stability, or the effect of miRNAs blocking protein translation without mRNA degradation[193–195], that affect protein levels without affecting RNA levels.

Of special note is the decrease in protein levels of the tested CPSF factors. The CPSF factors are required for cleavage of the pre-mRNA, and to direct AAUAAA-dependent poly(A) tail addition.[19] Therefore, decrease in the levels of CPSF factors could lead to changes in cleavage efficiency of pre-mRNAs or contribute to APA changes during differentiation, and therefore regulate adipogenesis. We tested the effect of overexpressing CPSF73, the levels of which are substantially decreased in differentiated cells compared to undifferentiated cells, on adipogenesis.

3.5 Limitation

In this study, we have only tested the effect of CPSF73 overexpression on overall efficiency of differentiation and the rate at which cells reach the differentiated state without testing other potential effects on (a) the release of secretory factors like

extracellular matrix proteins during adipogenesis[196], (b) lipid droplet dynamics[197] or fatty acid and protein composition,[198,199] which could have implications in browning of white adipose tissue[200] and the development of lipid disorders[201], and (c) metabolite concentrations at different timepoints during differentiation which are relevant in understanding obesity and related disorders like Type 2 diabetes mellitus.[202]

Moreover, we have not tested the effects of overexpressing other components of the C/P complex individually or in combination, which could have a more substantial effect on 3T3-L1 differentiation. For example, CSTF77 is known to interact with CPSF160 and potentially enable CPSF-CstF cooperative RNA binding during the assembly of the C/P machinery.[19] Additionally, CFIm59 decreases during 3T3-L1 differentiation and may therefore be a potential regulator. The CFIm complex consists of a CFIm25 dimer and a dimer of CFIm59 or CFIm68. While CFIm59 was thought to have a redundant function with CFIm68, it was shown to have a distinct effect on APA of *PTEN* gene in the mouse fibroblast cell line NIH3T3.[203] Although proteins of the CFIIm complex (PCF11 and CLP1) do not undergo as a large decrease in levels, their decrease during adipogenesis might affect pre-mRNA cleavage efficiency or the assembly of C/P machinery.[19,204,205] Interestingly, CSTF64 and PABN1, two well-known APA regulators[19], are not altered in levels.

In summary, this paper reveals important changes in levels of C/P factors during adipogenesis and provides the basis for further experiments to characterize the roles of C/P factors in the process.

Chapter 4: Discussion

Abnormal increase in number or size of white adipocytes due to dysregulated adipogenesis leads to obesity and related disorders. Therefore, understanding the regulation of adipogenesis is important in the discovery of novel therapeutic targets.[12] In Sections 2.4 and 3.4, we have discussed limitations and suggested future experiments pertaining to the overarching theme of the individual chapters. In this Discussion section, we present how findings from Chapters 2 and 3 together advance research in the field of adipogenesis, expand on the limitations of our analysis, and propose follow-up experiments. Our overall objective was to identify potential regulators of adipogenesis by (1) using sequencing datasets of human adipogenesis to study how gene expression is regulated via lncRNAs, alternative splicing (AS) and alternative polyadenylation (APA) and (2) study changes in the RNA and protein levels of cleavage and polyadenylation (C/P) factors in mouse 3T3-L1 cells and the potential of these changes to affect adipogenesis and regulate APA.

We used publicly available whole transcript RNA-seq datasets and generated a 3'-end focused white adipogenesis dataset using RNA samples purchased from Zen-Bio. Principal component analysis (PCA) of differential gene expression data shows two distinct clusters of preadipocytes and their matched differentiated adipocytes (Figure 4.1) which provides additional support to the reliability of our dataset for an adipogenesis study. Only Replicate 1 is a sample from male individual and replicates 2 and 3, which cluster closer together, are from female individuals. This denotes possible sex-specific differences in differential expression of genes during adipogenesis. However, the human

primary preadipocyte and adipocyte RNA samples that we have used are not controlled for sex, age, ethnicity, or source depot. Therefore, they cannot be used to make predictions or conclusions based on differences in those factors.

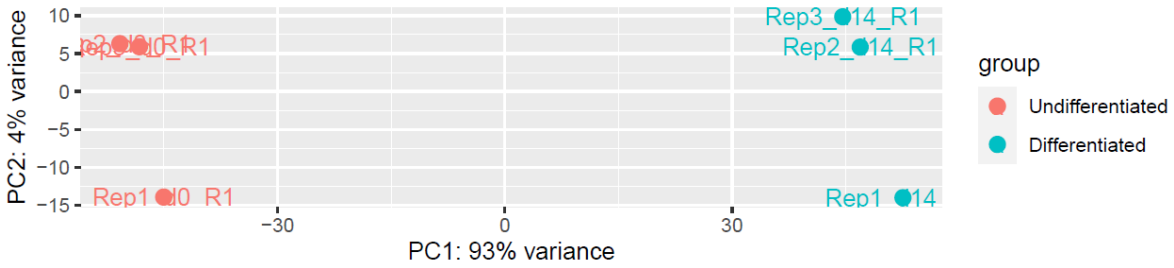


Figure 4.1: Principal component analysis of differentially expressed genes between undifferentiated (Day 0) and differentiated (Day 14) adipocytes. PC1 and PC2 are shown, with 93% and 4% of variance explained, respectively.

4.1 LncRNAs:

In our study, we identified novel lncRNA regulators of adipogenesis by focusing on the lncRNAs that undergo greatest change in expression or are differentially expressed in both brown and white adipogenesis. We were especially interested in the lncRNAs, LINC00312, LINC00607 and TYMSOS, that increase during brown adipogenesis but decrease during white adipogenesis. Browning of white adipose tissue (WAT) is one of the therapeutic approaches for obesity[184,185], but there have been no studies where lncRNAs were targeted to promote browning of WAT. In our study, we proposed mechanisms in which lncRNAs LINC00607 and TYMSOS can regulate browning of WAT through the gene *TBX15* and the PI3K/AKT signaling pathway, respectively, and therefore be studied as potential targets for this therapeutic strategy.The proposed mechanisms for these lncRNAs in adipogenesis involve the sponging of miRNAs which then derepresses the target genes of miRNAs. The upregulation of some lncRNAs during brown adipogenesis, in log2-fold change, is as high as about 6. However, for lncRNAs

LINC00312, LINC00607 and TYMSOS, the log2-fold changes are 1.06, 2.48, and 1.56, respectively. This brings into question whether relatively modest physiological changes in expression of lncRNA is sufficient to effectively sponge the target miRNAs and have a biological consequence, especially if the target miRNAs are highly abundant. However, sponging of miRNAs by lowly expressed lncRNAs that lead to even a modest upregulation of transcription factors can have amplified downstream consequences through signaling to multiple effector targets. Also, effective sponging of miRNAs depends on both spatial and temporal expression of lncRNAs.[206] For example, these lncRNAs may be highly upregulated at earlier stages of adipogenesis to sponge target miRNAs and facilitate the start of the differentiation program, followed by its downregulation so that there is a lesser extent of upregulation when their expression at Day 0 and Day 8 are compared. Additionally, lncRNAs that are predominantly localized in the cytoplasm and closely associate with the RISC complex (RNA-induced silencing complex) will likely have stronger sponging effects.[206,207] Therefore, RNA-FISH experiments can be done to observe changes in expression or subcellular localization of lncRNAs during adipogenesis. Also, the expression of the miRNA's target genes or any downstream effects or changes in 3T3-L1 morphology can be studied to assess functional consequences in response to manipulating the levels of the lncRNA in question. Additionally, the sponging efficacy of these lncRNAs may be more relevant and consequential in cases of aberrant adipogenesis compared to functional adipogenesis. LncRNAs can be used biomarkers for obesity and related disorders.[53,208] These three lncRNAs were not among the lncRNAs differentially expressed between obese and non-obese patients identified by microarray analysis.[209] This may be a limitation of the

microarray analysis which could only examine lncRNAs known at the time, or these lncRNAs may not be relevant biomarkers for obesity. However, they could still be biomarkers for metabolic diseases associated with dysregulated adipogenesis or a means to brown WAT.

It would have been both interesting and informative to have functional enrichment analysis data for differentially expressed lncRNAs as we did for genes undergoing AS and APA. However, due to lack of characterization of many of the lncRNAs differentially expressed during adipogenesis, the analysis could not be done.

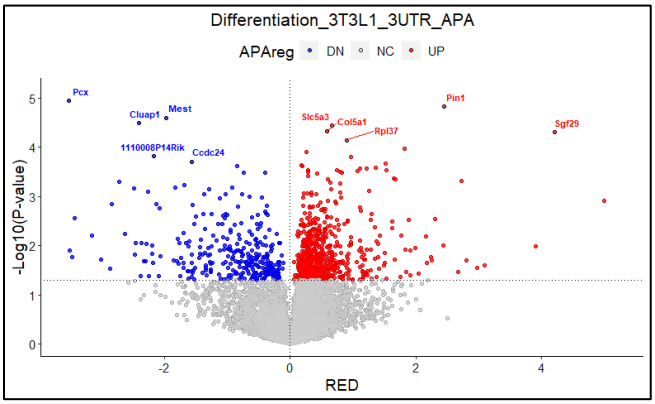
Also, while we used in-depth review of existing literature on the relevant lncRNAs to propose potential mechanisms of their regulation in adipogenesis, their interactions with the respective miRNAs must be validated in adipogenesis systems. In future experiments, the specific functions of the identified lncRNAs can be characterized by using knockdown and overexpression studies to see changes in gene expression and effects on adipogenesis.

4.2 Alternative polyadenylation

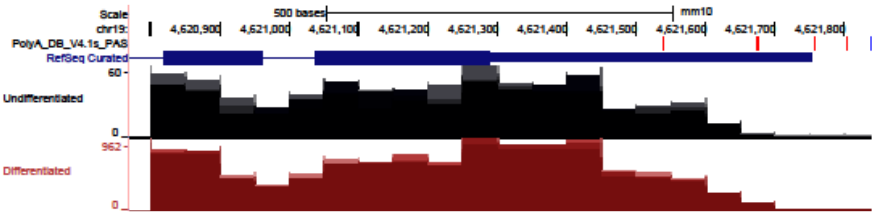
Previous studies on APA have mainly focused on identifying overall changes in 3'-UTR length[61,62] without focus on specific targets or regulators of APA during adipogenesis. Our study is the first to use 3'-end sequencing data of human preadipocytes and adipocytes and mouse 3T3-L1 cells to identify adipogenesis-related genes regulated by APA, factors potentially regulating APA, and C/P factors that may regulate APA and/or differentiation of preadipocytes.

Benchmark analysis showed that the overall performance of tools used to detect APA from RNA-seq data is "moderate", that there are considerable differences in APA prediction results between different tools, and that 3'-seq should be the method of choice for APA analysis.[38,210] I had previously applied APAlyzer[211] to a publicly available RNA-seq data of 3T3-L1 differentiation.[191] Except for a few genes, analysis of UCSC Genome Browser plots of most genes did not reflect the APA identified by APAlyzer (Figure 4.2). This experience raises concerns for the many studies that have now been published using such tools to predict global APA changes and reinforces the use of 3'-end focused RNA-sequencing for APA analysis for a more accurate and quantitative characterization of pA site usage. However, newer tools, such as LABRAT[212] and PolyAMiner-Bulk[213], claim to do a better job of predicting altered pA site use from bulk-RNA sequencing data. Given how there are more publicly available RNA-seq datasets compared to 3'-end sequencing datasets, it is worth developing newer tools that can be used to identify APA with greater precision from RNA-seq data. Additionally, since RNA-seq measures steady-state levels of mRNA, a shift in isoform abundance could be due to changes in stability of one of the isoforms, and not changes in processing. Also, the true magnitude in shift of processing from a distal to a proximal pA site may be masked because of heightened degradation of the longer isoform. Therefore, additional experiments can be done to test mRNA stability using transcriptional inhibitors or the pulse-chase experiment, or to measure transcriptional activity using Nuclear Run-On sequencing that maps nascent transcripts.

(A)



(B)



(C)

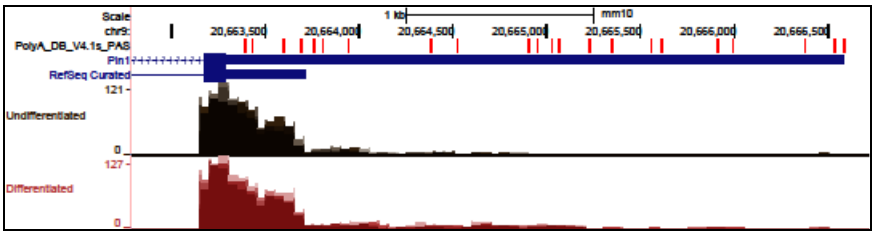


Figure 4.2 APAlyzer analysis of publicly available mouse RNA-seq dataset GSE129957 reveals genes potentially undergoing APA. (A) Volcano plot showing genes undergoing shortening (DN) or lengthening (UP) during 3T3-L1 differentiation. However, UCSC browser plots do not reflect a substantial change in usage of pA sites or show high variation between replicates, for example, (B) shortened gene *Pcx* and (C) lengthened gene *Pin1*. Replicates are merged using "transparent method" in UCSC genome browser. The sequencing depth of the dataset may be too low for accurate identification of APA sites and their differential usage.

To note, application of the tool PolyA-miner[41] and our collaborator's polyA_DB based APA analysis on our 3'-end dataset highlighted differences in the identification of APA genes by different programs. As Figure 4.3 shows, genes with the highest RED values or

lowest p-adjusted values were not identified as APA by PolyA-miner. Combining more than one tool for computational analysis may be a better strategy to prioritize targets for future functional studies but also involves the possibility of excluding genes undergoing the strongest APA.

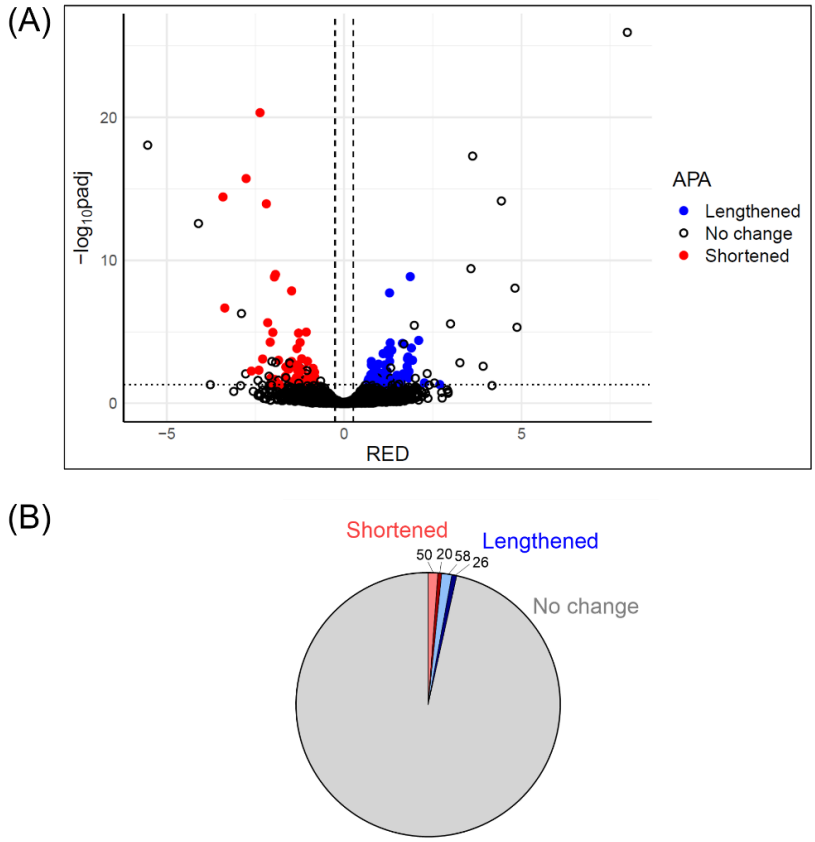


Figure 4.3 More than half of the genes predicted to undergo UTR-APA by polyA_DB based analysis were also predicted by PolyA-miner analysis. (A) Volcano plot showing distribution of shortened and lengthened genes detected by polyA_DB based analysis, where red and blue dots represent shortened and lengthened genes, respectively, as predicted by both PolyA-miner and polyA_DB based analysis (polyA_DB based analysis: p-adjusted values <0.05, RED values >= 0.26 (lengthening) or <= -0.26 (shortening); PolyA-miner: p-adjusted value <0.05, PolyAindex <0 (shortened) or >0 (lengthened)). The clear circles are genes that are predicted to undergo no APA by polyA_DB based analysis or APA predicted only by polyA_DB based analysis. The P-value and RED value thresholds are shown by dashed lines. (B) Pie chart showing number of genes identified as shortened (red) or lengthened (blue) by polyA_DB based analysis. The lighter shades of red and blue denote genes identified as changed by both PolyAMiner and polyA_DB based analysis, whereas darker shades denote genes identified as changed by polyA_DB based analysis only.

While I could validate by RT-qPCR 5 out of 6 genes predicted to undergo 3'-UTR shortening by both PolyA-miner and polyA_DB based analysis, I was not able to validate any of the genes predicted to be lengthened (Figure 4.4). The disparity between RNA-seq analysis and qRT-PCR may be explained by inaccurate prediction of APA due to activation of cryptic promoters between proximal and distal pA sites which lead to misrepresentation of abundance of longer isoforms. Other methods, such as 3'-RACE, Northern blot analysis, or high-depth or targeted long-read sequencing may be better validation approaches. However, we are unsure as to why only the lengthening of genes could not be validated by qRT-PCR.

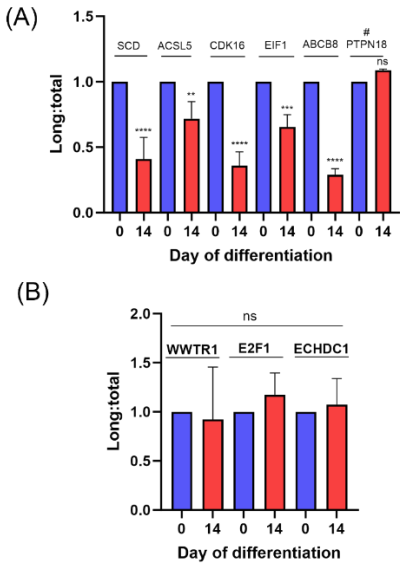


Figure 4.4 Shortening of genes due to UTR-APA during adipogenesis could be validated by qPCR but not lengthening of genes. (A) Validation tests for shortened genes identified by both PolyA-miner and polyA_DB based analysis, except for ABCB8 and PTPN18 which were detected by PolyA-miner only. The validated shortening of ABCB8 highlights differences in APA identification by different programs. (B) Validation tests for lengthened genes identified by both PolyA-miner and polyA_DB based analysis, except for WRD41 and DCUN1D4 which were detected by PolyA-miner only. All significance testing was done using One-way Anova by comparing the means of each pair of undifferentiated and differentiated samples. Replicates n=3 unless otherwise indicated with "#" in which case n=2. Ratios are calculated for longer isoforms to total (long plus short isoforms) where a ratio of less than 1 represents shortening and greater than 1 represents lengthening of gene.

We identified key genes in adipogenesis that undergo APA and some of these genes are also differentially expressed. For other genes, APA may affect mRNA localization without affecting gene expression, though changes in mRNA localization during adipogenesis have not been studied yet. An important facilitator of the biological consequences of 3'-UTR APA is miRNA binding, which could regulate the expression of target mRNAs it binds to, and also regulate competing-endogenous RNA (ceRNA) if the target mRNA undergoes 3'-UTR shortening. For example, during tumorigenesis, the shortening of 3'-UTR of oncogenes prevents sequestering of miRNAs, and the released miRNAs then bind to repress the tumor-suppressor genes, which are the ceRNA partners of the oncogenes.[214]

For the key genes we have identified as undergoing 3'-UTR APA, we proposed potential consequences of APA on gene expression via gain or loss of miRNA-binding sites. This analysis revealed new miRNAs that have not been associated with adipogenesis before but could be important regulators and therefore can have clinical applications for obesity and related metabolic disorders.[215] Our analysis also demonstrates possible miRNA-mediated regulation of gene expression via APA which could be used to inform designing of miRNA-based therapeutics as target genes with shortened 3'-UTRs can escape regulation by miRNAs, whereas non-target genes with lengthened 3'-UTR could also be subjected to miRNA-mediated regulation and therefore lead to off-target effects.
RBPs can also bind in the 3'-UTR region between the two polyA sites, and the interplay between miRNAs and RBPs can determine target gene expression. RBPs can compete for the same binding motif as miRNA or change the secondary structure of the target mRNA

to prevent association of RISC complexes, thereby inhibiting the action of miRNAs. RBPs can also change the 3'-UTR structure to favor miRNA binding which can then enhance miRNA action. Additionally, miRNAs and RBPs also regulate each other's expression.[216] We also proposed potential mechanisms of APA regulation by RBPs which can have implications in drug designing. For example, glucocorticoids, a class of medication used to treat inflammatory and autoimmune diseases, mediated APA in human B-lymphoblastoid cell lines through regulating expression of RBPs, and the RBPs may be responsible for the cell-type specific APA resulting from the treatment.[217] Therefore, in addition to analysis of APA during adipogenesis, it is important to analyze potential RBPs that are regulating APA. While the binding of target genes to the APA must be validated in adipogenesis systems, our findings provide a means to prioritize future research into RBP regulators of adipogenesis. For example, some of the RBPs that were identified as potential regulators of AS and/or APA were also differentially expressed during adipogenesis (Figure 4.5) and therefore may be ideal candidates for further experiments. Furthermore, RBPs with potentially dual modes of regulating gene expression are also interesting. For example, HNRNPA1 is a splicing factor that also regulates APA[166], can recruit CPSF/SYMPK to a regulated exon and stably associates with CPSF/SYMPK in the absence of RNA.[218]

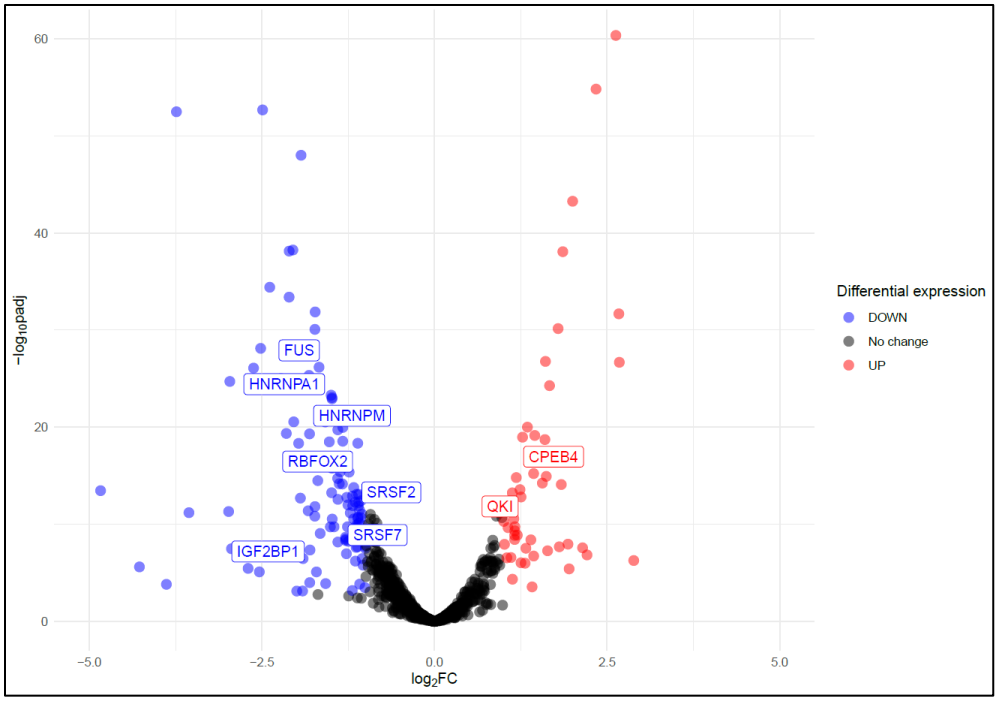


Figure 4.5 Volcano plot of all RNA-binding proteins (RBPs) differentially expressed during adipogenesis (reference list from Sundararaman *et al*[1]). Red represents upregulated genes and blue represents downregulated genes that have p-adjusted values less than 0.05 and log2-fold change of more than 1, or less than -1, respectively. The RBPs identified as potential regulators of splicing and APA during adipogenesis in Chapter 1 and are also differentially expressed are highlighted. The y-axis was limited to 60 and x-axis was limited to between -5 and 5 for display.

We used the mouse 3T3-L1 cell line to study changes in RNA and protein levels of C/P factors. Overall, all CPSF subunits and some other C/P subunits that we tested decreased during adipogenesis, which contrasts with the findings of another member of our lab, who observed an overall increase in C/P proteins during macrophage differentiation.[189] These differences could account for cells or tissues having distinct APA profiles. CPSF73 was one of the C/P factors that was decreased during adipogenesis. However, its overexpression does not lead to changes in rate or efficiency of 3T3-L1 differentiation and raises the possibility of whether CPSF73 produced is non-functional. However, a previous lab member used the same overexpression construct to overexpress CPSF73 in

MDA-MB-231 cells which affected the migration and invasion of the cells.[219] Another lab member overexpressed CPSF73 in 3T3-L1 cells and showed that the overexpression rescued the differentiation defects due to loss of UBE3D.[190] Therefore, it is unlikely that the overexpressed CPSF73 is not functional.

The decrease in protein levels of CPSF73 but not mRNA levels, even when CPSF73 is overexpressed, and the higher protein level of differentiated cells overexpressing CPSF73 versus differentiated control cells, suggests regulation of CPSF73 through changes in protein stability. CPSF73 can be stabilized or degraded through ubiquitin E3 ligases UBE3D[219] and RBBP6[220], and therefore, overexpressing CPSF73 while also inhibiting its degradation by blocking the ubiquitin-proteasome pathway through use of proteasome inhibitors or mediating its ubiquitination through overexpression of UBE3D and knockdown of RBBP6 may be sufficient in affecting differentiation of 3T3-L1 cells.

Downregulation of multiple C/P factors led to APA in C2C12 myoblast cells - for example, knockdown of CFIm68, CFIm25, and PABPN1 led to more shortened events whereas knockdown of FIP1 and PCF11 had the opposite effect.[21] Modulation of C/P factors could also mediate differentiation without changes in APA. For example, downregulation of CFIm25 in H9 human embryonic stem cells affected the proliferation rate and differentiation potential through transcriptional regulation without affecting mRNA 3' processing efficiency and APA profile.[22] Therefore, changes in levels of C/P factors during 3T3-L1 differentiation could provide insight into novel regulators of adipogenesis.

A combined knockdown or overexpression of CPSF factors may not be feasible, and further exploration of this issue may not be possible unless a common regulator of CPSF proteins is identified. However, a time-course analysis of protein levels may be useful in prioritizing factors for modulation studies, or an overexpression screen can be done for all C/P factors that decrease in levels during adipogenesis. Also, C/P factors with dual modes of gene expression regulation that undergo changes during adipogenesis can be prioritized, for example, CPSF30 could potentially regulate adipogenesis through APA as well as through AS.[221]

In Chapter 2, we have mainly identified targets of APA that are protein-coding genes playing a direct role in adipogenesis. However, some genes that may not encode key adipogenesis proteins but undergo strong APA can still be important candidates for further research, for example, phytochemicals can affect adipocyte differentiation through regulation of cell cycle genes.[12] We identified multiple cell cycle and cell proliferation genes that undergo 3'-UTR and intronic APA and can therefore be good therapeutic candidates.

Also, up to 70% of lncRNAs contain multiple polyA sites[45] and may be subject to APA. However, none of the differentially expressed lncRNAs in our dataset were identified as undergoing APA during adipogenesis. This may be due to the stringent threshold of polyA_DB based APA analysis since PolyA-miner did identify some lncRNAs as undergoing APA - for example, LIPE-AS1, an upregulated lncRNA, was identified as shortened by PolyA-miner. In addition to characterizing the role of lncRNAs in adipogenesis, their regulation via APA can also be explored in future experiments.

4.3 Alternative splicing

Spliced isoforms play important roles in adipogenesis[28,30,63], and we identified 100 alternative splicing events in 68 genes during human adipogenesis. Of the 68 genes, 9 genes were upregulated, and 10 genes were downregulated. However, our differential expression analysis is restricted to gene-level expression changes and not isoform-specific expression, which could be affected due to AS during adipogenesis. For example, in mice, there are two spliced isoforms of the *Lpin1* (lipin) gene that shows temporal differences in expression during 3T3-L1 differentiation and also exhibit differences in protein subcellular localization.[222]

An example of drugs targeting AS is the drug Eteplirsen, an anti-sense oligonucleotide (ASO) that targets dystrophin (DMD) pre-mRNA splicing (exon 51 skipping) to treat Duchenne muscular dystrophy.[223] Therefore, in addition to identifying splicing events, we also highlighted how genes, such as *MAP4K4* and *RBPJ*, that undergo AS may be potential targets for drugs that leverage brown adipogenesis or dedifferentiation of adipocytes as therapeutic strategies.

Similar to APA, the use of more than one tool increases confidence in the computational prediction of AS events. While rMATS is a widely used tool for AS[65], a comparison of its predictions to that of another commonly used tool, SUPPA[224], can be used to prioritize targets for further analysis. This is especially true for identification of mutually exclusive exon (MXE) events since rMATS can identify events that are not true MXE events – transcripts may contain either one of the two exons in consideration, but there are also transcripts where both or none of the exons are present.

Additionally, we used motif enrichment analysis to identify splicing factors regulating global skipped exon events. To further study regulation of a specific alternative splicing event by RBPs, data from the POSTAR3[80] database can be used for CLIP-Seq binding evidence or RNA immunoprecipitation can be done to identify RBPs binding to the mRNA precursor transcripts.

Adipose tissue has remarkable plasticity and, in response to physiologic cues, can undergo metabolic, structural, and phenotypic remodeling. White adipocytes respond to changes in temperature by undergoing the reversible process of browning which involves changes to lipid droplets, and upregulation of fuel oxidation and mitochondrial biogenesis. White adipocytes can also undergo reversible dedifferentiation *in vivo* and *in vitro* to proliferative fibroblasts, switch roles between nutrient uptake and nutrient release in response to hormonal and neuronal signals and increase in size and number in response to over-nutrition. Obesity and aging are associated with decline in adipose tissue plasticity, which can lead to fibrosis and inflammation, resulting in insulin resistance and metabolic disease. Therefore, targeting the plasticity of mature white adipocytes is an important therapeutic strategy.[225]

Besides its role in energy storage, adipose tissue play important roles of regulating multiple aspects of physiology such as maintenance of energy levels, insulin sensitivity, body temperature, and immune responses.[225] Given the role of white adipose tissue as an endocrine organ that secretes adipokines to coordinate the systemic metabolic state, such as fibrosis and inflammation in other tissues, understanding gene expression regulation during adipogenesis has implications beyond obesity. Larger adipocytes correlate with a

more pro-inflammatory adipokine profile and undergo hypoxic stress, stimulating collagen deposition and fibrosis of the adipose depots.[226] Adipose tissue also affects the behavior of cells around them - both mature adipocytes and adipose progenitor cells affect the growth and metastasis of cancer cells and as such, there are strong epidemiological associations between fat mass and the incidence and mortality of multiple types of malignancies.[3] Therefore, studying adipogenesis can also inform the strategies to counteract inflammation that leads to development of metabolic disorders and understand the development of obesity-associated cancers.

Overall, our study explored and predicted important regulators of gene expression during adipogenesis, proposed potential mechanisms of regulation and therefore provided a foundation for future functional studies.

Chapter 5: Appendix

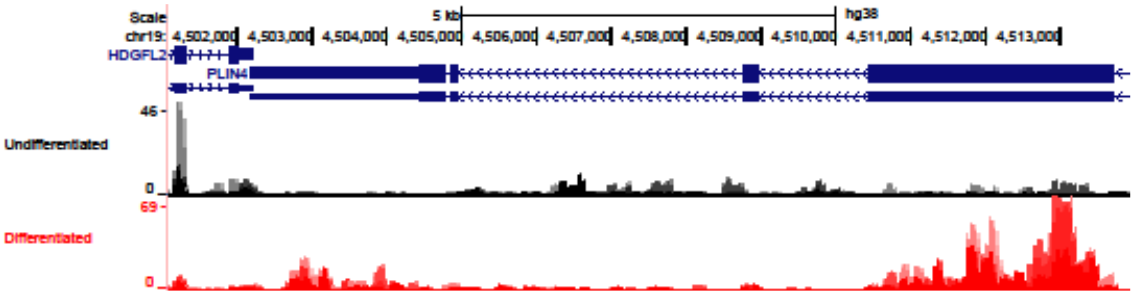


Figure 5.1 DoGFinder analysis of publicly available human adipogenesis dataset GSE176020 did not show interesting changes in transcriptional readthrough during adipogenesis. We only focused on genes that were downregulated and had increased transcriptional readthrough because (1) the expected biological consequence of increased transcriptional readthrough is decrease in expression and (2) From the UCSC genome browser plots of highly upregulated genes, it is difficult to discern between an actual readthrough and overall increased expression throughout the region. Nineteen genes were both downregulated and had increased transcriptional readthrough in differentiated samples compared to undifferentiated samples. In the browser plots, a clear transcriptional readthrough was observed for only a few genes, an example is the gene HDGL2, a hepatoma-derived growth factor (HDGF) family that is an epigenetic regulator of myogenesis.

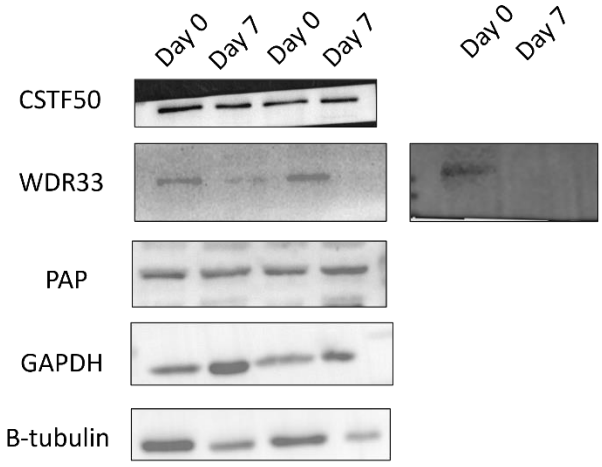


Figure 5.2 Western blot of additional CPA factors from Chapter 3 (Figure 3.1). The blots show protein levels at Day 0 (undifferentiated) and Day 7 (differentiated) 3T3-L1 cells.

In cancer cells, some genes have been shown to undergo APA, with the smaller isoform usually being predominant. Cancer cells also display differences in chromatin and expression of epigenetic factors compared to normal cells. Previous research has shown that usage of pA sites is associated with specific epigenetic marks. Also, a previous post-doc in the lab, Dr Michaels, and I found that deletion of histone H3K4 methyltransferase Set1 and histone H3K36 methyltransferase Set2 in yeast leads to shift in polyadenylation site usage for some genes.[227] **We hypothesized that epigenetics can contribute to cancer phenotypes by affecting the choice of pA site of genes.** Using qPCR, we tested genes known to undergo APA in cancer by comparing triple-negative breast cancer cell line MDA-MB-231 to control cell line MCF10A. We could only validate shortening in cFOS gene.

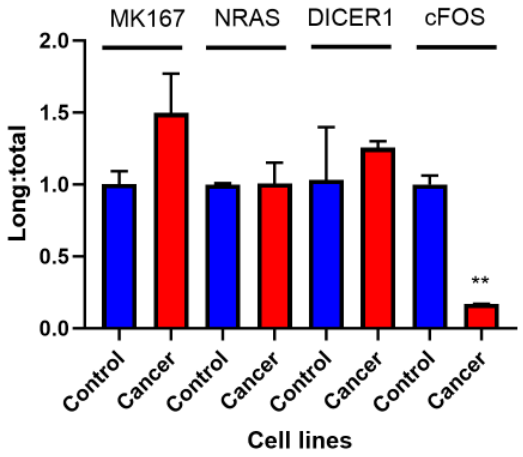


Figure 5.3 Only cFOS undergoes shortening in MDA-MB-231 (cancer) cells relative to MCF10A (control) cells. qPCR-validation of genes known to undergo APA in cancer. Significance testing was done using unpaired t-tests, n=2. Ratios are calculated for longer isoforms to total (long plus short isoforms) where a ratio of less than 1 represents shortening and greater than 1 represents lengthening of gene.

Based on our findings in yeast, we generated MDA-MB-231 cell lines with constitutive knockdown of the enzymes H3K4 methyltransferases SETD7 and SMYD2, and H3K4 demethylase LSD1. Using qPCR, we identified some genes undergoing APA due to knockdown of these enzymes. Knockdown of SETD7 decreases proliferation rate of MDA-MB-231 and knockdown of SMYD2 increases proliferation rate of MDA-MB-231. Knockdown oonLSD1 decreases proliferation rate of MDA-MB-231 at Day 6.

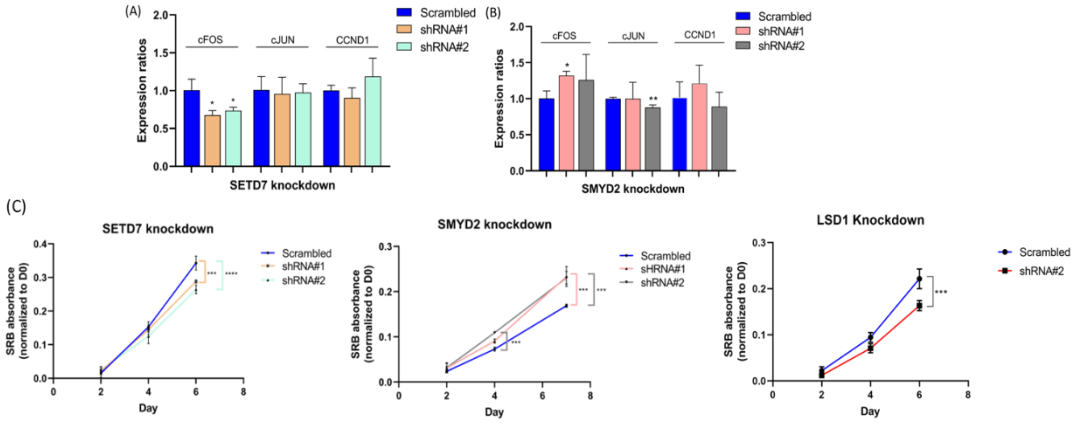


Figure 5.4 Knockdown of SETD7 and SMYD2 in MDA-MB-231 cells cause APA in some genes and knockdown of SETD7, SMYD2 and LSD1 in MDA-MB-231 cells affect proliferation of cells. qPCR-validation of three genes known to undergo APA in cancer for (A) SETD7-KD and (B) SMYD2-KD. Significance testing was done using unpaired t-tests, n=2. The expression ratio (long:total) is calculated as ratio of longer isoforms to total (long plus short isoforms) where a ratio of less than 1 represents shortening and greater than 1 represents lengthening of gene. (C) SRB proliferation assay for SETD7, SMYD2 and LSD1 knockdown in MDA-MB-231 cells using two shRNAs or one shRNA for LSD1 at mentioned timepoints.

To detect global change in APA, and to confirm APA in previously tested genes, we sent samples of Control and SETD7-KD (shRNA # 2), SMYD2-KD (shRNA # 2) and LSD1-KD in the MDA-MB-231 cell line for 3'-end sequencing. Click-Seq technologies carried out library preparation, sequencing, and computational analysis. The genes we tested earlier were not identified as shortened or lengthened with the DPAC tool used by the company. This may be because qPCR, being very sensitive, can pick up small differences in isoforms that are not identified by RNA-Seq. We selected genes based on high significance and high dPDUI values, which is indicative of extent of APA (distal PA usage index) and checked the alignment files in the browser plots. However, most plots did not reflect a shift in pA site usage, especially for LSD1-KD. There were either very small changes which are unlikely to have a significant biological consequence, or showed variability between replicates which decreased our confidence in the calls. Some plots did reflect the APA shift predicted; however, they could not be validated by qPCR. We also tested multiple genes identified as shortened for SETD7-KD, but the results were variable between runs or between replicates, and there were either no signals or multiple peaks in melting curves.

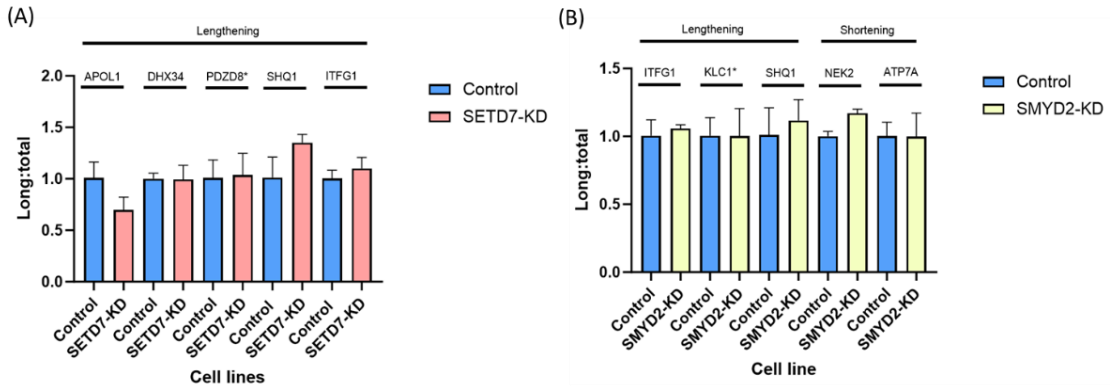


Figure 5.5 APA detected by 3'-end sequencing due to SETD7 and SMYD2 knockdown in MDA-MB-231 cells could not be validated by qPCR. Validation of APA using qPCR for mentioned genes after (A) SETD7 and (B) SMYD2 knockdown in MDA-MB-231 cell lines. Significance testing was done using One-way ANOVA, n=3 or n=2 (genes with asterisk). Ratios are calculated for longer isoforms to total (long plus short isoforms) where a ratio of less than 1 represents shortening and greater than 1 represents lengthening of gene.

We tried a different method - the 3'-RACE technique - to verify the APA changes due to difficulties getting reproducible results with the qPCR, and to visualize presence of longer isoform using same forward primers for short and long isoform in PCR. This overcomes a problem with qPCR validation approach, that is, for genes that show lengthening with qPCR, the reason could be activation of cryptic promoters between the proximal and distal pA rather than actual APA causing lengthening. However, APA could not be validated by 3'-RACE either.

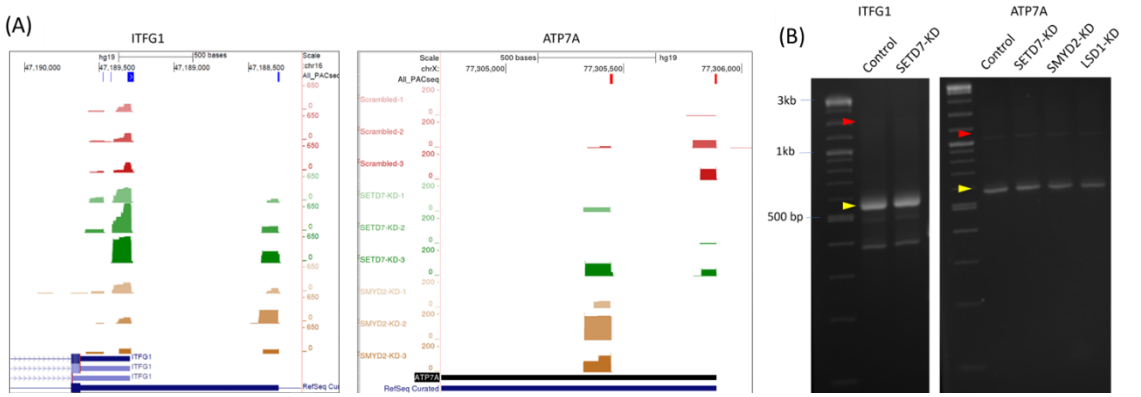


Figure 5.6 Predicted APA in ITFG1 and ATP7A could not be validated by qPCR. (A) Genome Browser plots for ITFG1 and ATP7A identified as lengthened and shortened, respectively, in both SETD7-KD and SMYD2-KD. The directions of both genes are left to right. (B) We tried the 3'-RACE technique for validation instead of qPCR but could not validate APA. Yellow arrows point to the shorter isoform whereas red arrows point to the longer isoform.

In mammalian cells, there are multiple enzymes responsible for H3K4 methylation and therefore other enzymes may be playing a redundant role to SETD7- and SMYD2- and masking its APA effects. Therefore, we decided to shift our focus on the knockdown of protein lysine methyltransferases DOT1L (H3K79me1,2,3), EZH2 (H3K27me3) and EHMT2 (H3K9me) in MDA-MB-231 generated by Dr Michaels. All these proteins are overexpressed in cancer. We sent these new samples to Click-Seq technologies for APA analysis. However, APA could not be validated, and we observed the same variability issue with qPCR was observed as in the SETD7- and SMYD2- samples. We had also sent the DOT1L1-, EZH2-, and EHMT2-KD samples for long-read sequencing (data not shown), and though we did not perform further analysis, the read coverage plots did not show visible or large differences in 3'-UTR lengths between the two conditions for specific genes (CDC25C, TUSC2, UPF2) or the reads were too low to make any deduction. We concluded that there might be limitations in the validation approaches or in the predictions of APA with the DPAC tool, although additional analysis with PolyA-miner did not help in identifying targets. Otherwise, there may be no changes or subtle changes in pA site usage due to knockdown of epigenetic enzymes. Therefore, we discontinued the project.

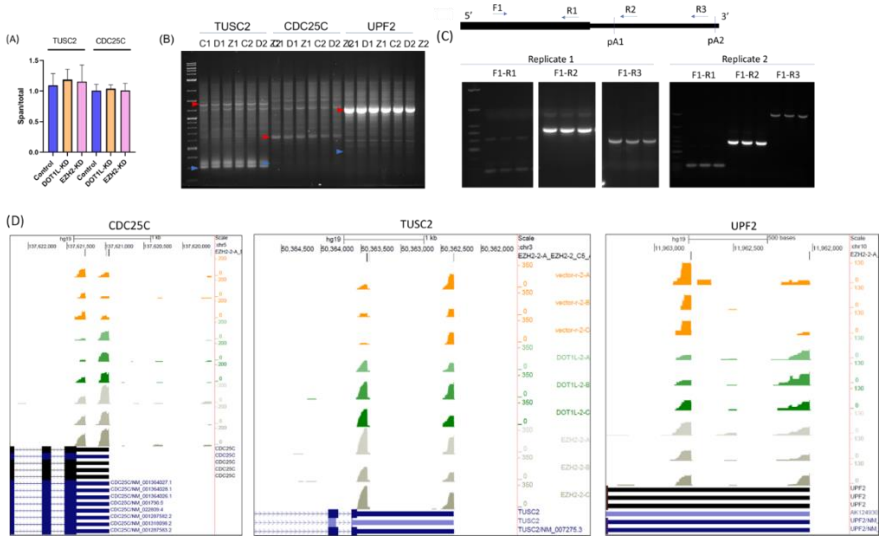


Figure 5.7 APA detected by 3'-end seq due to DOT1L1 and EZH2 knockdown in MDA-MB-231 cell lines could not be validated by qPCR and 3'-RACE. (A) APA analysis by qPCR of mentioned genes after DOT1L and EZH2 knockdown in MDA-MB-231 cell lines. Significance testing was done using t-test, n=4. Span/total (long:total) ratios are calculated for longer isoforms to total (long plus short isoforms) where a ratio of less than 1 represents shortening and greater than 1 represents lengthening of gene. (B) 3'-RACE for the mentioned genes in two replicates of Control (C) DOT1L- (D) and EZH2 (E)-knockdown in MDA-MB-231 cell lines. Blue arrow represents shorter isoform whereas red arrow represents the longer isoform (C) Depiction of location of primers to verify shortening of TUSC2 via PCR and the PCR results of 2 replicates of Control, DOT1L- and EZH2-knockdown using combination of the primers. (D) Genome Browser plots for CDC25C, TUSC2 and UPF2. The directions of all genes are left to right.

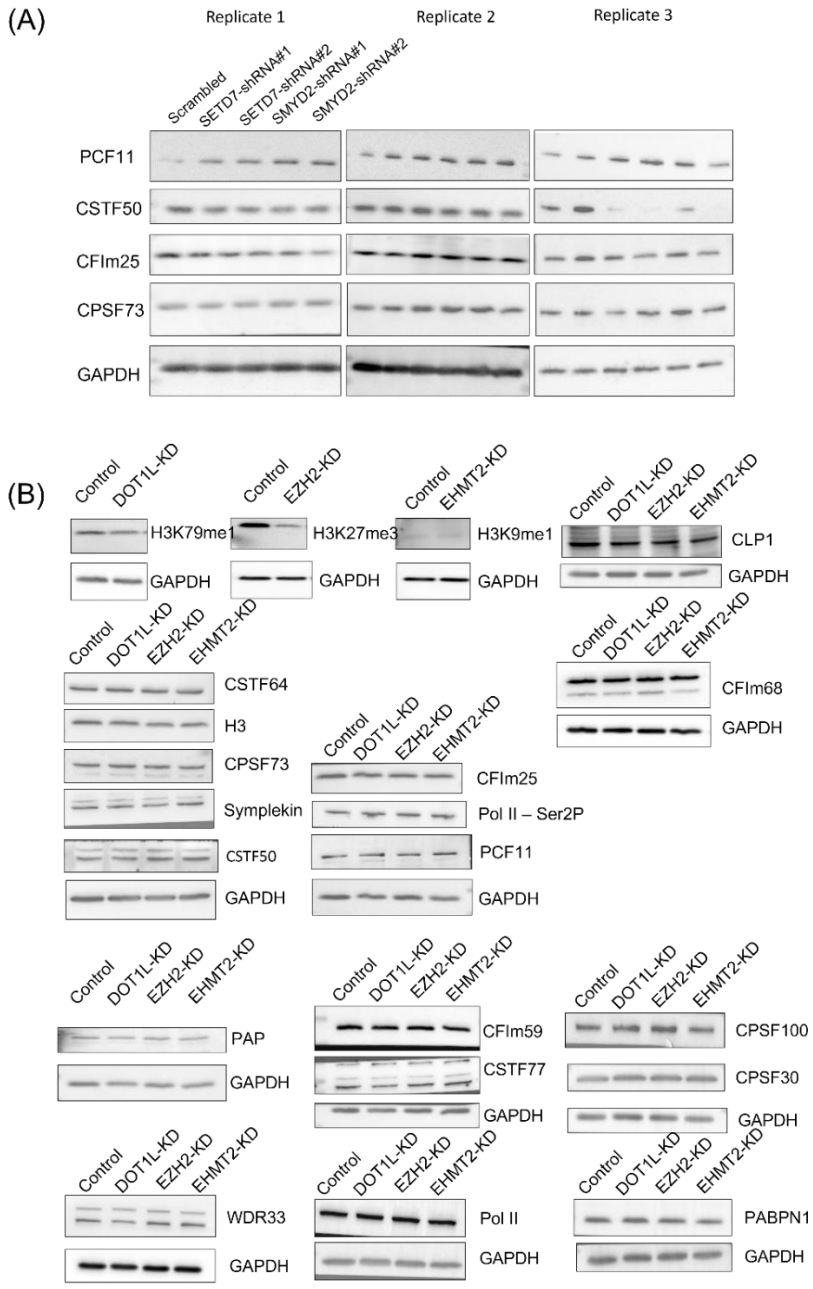


Figure 5.8 Knockdown of SETD7 and SMYD2 PCF11 in MDA-MB-231 cells decreases only PCF11 levels whereas for decrease in when SETD7 and SMYD2 is knocked down of DOT1L-, EZH2-, and EHMT2-KD does not affect levels of C/P factors. A) Levels of subunits of cleavage and polyadenylation factors in control versus SETD7-, SMYD2- and LSD1-knockdown in MDA-MB-231 cells. The lanes of replicates 2 and 3 follow the same sequence as replicate 1. (B) Representative western blots of one replicate depicting levels of various cleavage and polyadenylation factors in DOT1L-, EZH2-, and EHMT2-KD in MDA-MB-231 cells. The factors we tested were unchanged due to manipulation of epigenetic factors, except for the increase in PCF11 levels due to SETD7- and SMYD2-KD.

Sample	Catalog Number	Lot#	Description	Number of donors	Donor Ages	Sex	Ethnicity	BMI	Diabetic	Depot
1	RNAmi-P10-2	L012502	Preadipocyte Total RNA (+miRNA)	1	40	M	Unknown	25.3	N	Abdomen
2	RNAmi-A10-2	L012502	Adipocyte Total RNA (+miRNA)	1	40	M	Unknown	25.3	N	Abdomen
3	RNAmi-P10-3	LM050709B	Preadipocyte Total RNA (+miRNA)	1	42	F	Caucasian	34.8	N	Abdomen/ Hips
4	RNAmi-A10-3	LM050709B	Adipocyte Total RNA (+miRNA)	1	42	F	Caucasian	34.8	N	Abdomen/ Hips
5	RNAmi-P10-SL	SL0046	Preadipocyte Total RNA (+miRNA)	6	50.17	F	Caucasian	26.93	N	Abdomen/ Thighs/Hips/ Knee
6	RNAmi-A10-SL	SL0046	Adipocyte Total RNA (+miRNA)	6	50.17	F	Caucasian	26.93	N	Abdomen/ Thighs/Hips/ Knee

Table 5.1 Details of donors from Zen-Bio Inc (Durham, NC, USA)

UP_Symbol	log2FoldChange	padj	DOWN_Symbol	log2FoldChange	padj
LOC100288069	3.428496183	0.002289	SNHG3	-1.57217	1.21E-08
ZNF436-AS1	1.799982248	0.003716	RNU5F-1	-5.00366	0.000375
RNU11	2.436674367	0.001487	DLEU2L	-1.91686	0.000123
LINC01144	3.338357532	0.022339	ROR1-AS1	-2.54507	0.01394
FOXD2-AS1	2.720941554	0.005151	LINC01133	-1.9259	5.32E-08
LOC100507634	1.887863704	0.004098	DNM3OS	-1.95601	0.020194
NFIA-AS1	1.080339447	0.043085	C1orf220	-2.85045	0.025215
LINC01140	1.116514348	0.041653	LINC00842	-3.91279	0.000194
PKN2-AS1	3.175386209	0.003651	ZNF503-AS2	-1.44784	0.007948
HIPK1-AS1	3.112092107	0.010817	LINC00856	-4.93786	0.000678
LINC00623	1.658502948	0.000391	LINC00595	-5.90609	2.57E-06
LINC01138	1.996769849	0.000432	EMX2OS	-1.06108	0.000427
LINC00869	1.930714652	1.59E-06	KIRREL3-AS2	-4.02358	0.019806
LOC100506023	2.015621039	0.031836	SENCR	-5.2272	0.005112
LINC00467	1.698918204	0.000194	LINC00941	-4.21353	4.46E-06
TAF1A-AS1	2.986837831	5.07E-05	TMPO-AS1	-3.55274	4.34E-06
LINC01354	5.186377445	4.53E-14	TBX5-AS1	-4.04215	0.000925
LOC101928272	1.78071661	0.040599	FZD10-AS1	-4.53684	0.017256
LINC00999	2.664101816	2.17E-07	LINC01048	-7.30976	2.89E-05
NUTM2B-AS1	1.206401491	0.000419	LINC01050	-10.4883	9.95E-12
LOC642361	1.529776535	0.00047	FRMD6-AS1	-2.44597	0.000107
AGAP11	4.699114286	1.50E-20	MEG3	-2.02654	0.023044
OLMALINC	4.043751861	8.27E-19	MEG9	-2.05108	0.031683
LOC101928053	2.18637496	0.043863	LOXL1-AS1	-3.65902	5.65E-09
ZBED5-AS1	1.312124658	0.006717	LINC00597	-2.58706	0.009904
BDNF-AS	1.544063533	0.008256	ATP2A1-AS1	-1.5777	0.022365

LINC01252	1.903136883	0.019316	PYCARD-AS1	-5.86709	0.002167
LOH12CR2	1.310883111	0.017252	FTO-IT1	-4.22268	0.017698
RASSF8-AS1	1.446736002	8.92E-06	VPS9D1-AS1	-5.56598	9.35E-05
ARNTL2-AS1	2.792762217	0.03012	MGC12916	-1.98673	0.016267
LINC00938	1.593359103	7.77E-06	SCARNA20	-6.69952	0.000211
HOXC-AS1	1.167570045	0.02317	LINC00673	-9.11296	4.25E-09
LOC100240735	1.287997188	0.019869	TYMSOS	-1.20619	0.044575
TRHDE-AS1	5.30642703	0.000266	DLGAP1-AS1	-1.23915	0.010993
MAPKAPK5-AS1	1.042352359	0.003006	DLGAP1-AS2	-2.6034	0.031256
PXN-AS1	1.337628572	0.021967	GATA6-AS1	-3.08881	0.001377
THRIL	1.137184374	0.032137	LOC101927571	-3.45359	0.027215
PAN3-AS1	1.807720216	0.021836	SNAR-G2	-2.62387	0.003784
LINC00548	6.725523012	0.000572	SNAR-G1	-3.60446	3.98E-05
LINC00598	1.693984209	0.031319	SNAR-F	-1.31147	0.049638
SMIM2-AS1	2.492532153	0.014998	LINC00211	-4.33179	0.015446
CPB2-AS1	1.890367349	0.026973	SLC8A1-AS1	-3.48033	0.005365
RBM26-AS1	1.670108676	0.001463	LINC01119	-2.79115	0.016641
UBAC2-AS1	3.224737335	9.06E-06	ALMS1-IT1	-2.02156	0.040429
FGF14-AS2	1.32147879	0.003222	LOC284950	-3.08511	0.001751
LINC01070	4.597623487	0.014889	LINC00342	-2.90153	1.91E-12
RPPH1	1.762211978	2.73E-05	LOC101929532	-4.43425	0.014328
DHRS4-AS1	1.045809223	0.001215	HAGLROS	-2.23931	0.043901
ARHGAP5-AS1	1.963007095	0.018205	LINC01116	-1.5474	0.000218
SLC25A21-AS1	1.203948651	0.032549	LINC00607	-1.8973	0.008443
FOXN3-AS1	2.325038533	1.06E-05	MCM8-AS1	-2.09843	0.029732
DICER1-AS1	1.981516165	0.01751	LINC01260	-1.7413	0.040408
OIP5-AS1	1.218077118	1.02E-07	LINC01436	-2.76488	0.010075

RNU5B-1	2.016395958	0.001741	LINC00312	-1.62494	0.002218
IQCH-AS1	2.538742822	4.48E-21	GHRLOS	-1.79239	0.032501
ADPGK-AS1	7.369814092	5.12E-05	LINC00852	-3.69822	0.005372
ST20-AS1	1.585245645	3.33E-05	KIF9-AS1	-3.4928	0.026796
CPEB1-AS1	2.077931502	0.040471	SCARNA22	-2.11374	7.56E-07
SNHG19	1.029820494	2.76E-05	LINC01085	-3.10079	0.016321
LINC01569	2.252788467	0.015549	LINC01094	-3.27044	0.002522
LOC730183	2.23591148	2.42E-05	LINC01060	-7.72029	1.40E-05
SNAI3-AS1	2.695366663	5.26E-07	STARD4-AS1	-4.0798	1.41E-16
PITPNA-AS1	1.943472981	0.000822	P4HA2-AS1	-3.53769	0.003269
LINC00974	3.812685602	0.008017	VTRNA1-3	-1.57656	0.003604
RAMP2-AS1	6.987106502	4.48E-11	SAP30L-AS1	-2.19744	0.000171
NBR2	1.339240764	0.004714	HCG4	-3.31511	0.02861
HOXB-AS1	2.67409329	5.50E-07	LOC730101	-2.62437	1.22E-05
TOB1-AS1	5.401043901	2.26E-08	STXBP5-AS1	-8.92667	3.73E-08
C1QTNF1-AS1	1.713673473	0.047893	AGPAT4-IT1	-3.54256	0.00399
TTC39C-AS1	3.02833777	0.001155	SND1-IT1	-1.18587	0.043569
SCARNA17	4.007061466	8.00E-13	LINC01605	-4.91989	3.71E-09
RAB11B-AS1	1.792209089	3.46E-07	LOC101929217	-3.13599	0.03063
LINC00663	2.20788981	0.003079	ASAP1-IT1	-2.35056	0.000995
LIPE-AS1	1.316728208	0.034339	FAM225A	-2.00406	0.000943
ZNF667-AS1	1.113691723	0.013105	DBH-AS1	-2.00925	0.038247
DNAJC27-AS1	1.803816652	0.02265	PPP1R26-AS1	-1.70954	0.014328
STARD7-AS1	2.271539503	5.82E-06	LINC01204	-9.39267	1.23E-09
NIFK-AS1	1.285592726	0.002615			
ZEB2-AS1	4.113321062	0.000256			
IDH1-AS1	1.72794159	0.001661			

LINC00652	5.70852514	0.003592			
SNHG11	1.147722646	0.004003			
LINC01270	1.578209285	0.02762			
URB1-AS1	1.919768641	3.05E-05			
AATBC	8.376183193	5.35E-06			
LOC101928796	2.120141825	0.002744			
THAP7-AS1	1.550630717	0.000541			
MORC2-AS1	1.005144908	0.045684			
LINC01521	2.877437808	2.24E-15			
LINC01315	4.202159719	0.002046			
PRRT3-AS1	2.056579986	1.38E-05			
THRB-AS1	4.904107384	0.000394			
LINC00693	5.159800645	0.000614			
ENTPD3-AS1	1.757496058	0.022965			
PRKAR2A-AS1	2.615367419	0.001744			
ARIH2OS	1.334500522	0.0307			
SEMA3B-AS1	2.975769353	0.005336			
RASSF1-AS1	5.305442815	4.62E-05			
ADAMTS9-AS1	3.786138562	0.000489			
ZBTB11-AS1	1.330338927	0.011271			
ALDH1L1-AS2	9.938770576	8.42E-11			
LOC100507389	3.6261281	0.025175			
LOC100289361	1.987006798	0.005042			
TIPARP-AS1	2.728489257	0.018401			
LINC00886	1.316806252	0.028344			
SCARNA7	2.678394812	0.003779			
ADIPOQ-AS1	10.85560145	6.65E-13			

TPRG1-AS1	4.821971494	0.002032			
TMEM44-AS1	2.245336215	6.50E-05			
NOP14-AS1	1.117374441	0.002271			
LINC00504	2.454397615	0.009915			
LINC01088	9.346888515	2.82E-09			
LOC100507053	1.90259112	0.03669			
FLJ20021	1.695525159	5.48E-05			
LINC01061	1.389636838	0.025874			
LINC01612	6.6310626	0.000453			
LINC01093	7.541771461	1.66E-05			
EXOC3-AS1	1.205506798	0.009865			
OXCT1-AS1	4.482650377	0.001789			
NNT-AS1	1.169269287	8.68E-06			
ZBED3-AS1	8.939933798	2.29E-18			
SCAMP1-AS1	1.849146304	2.43E-11			
LINC01554	2.739416252	0.005687			
LINC01023	1.755825105	0.000363			
LOC100652758	1.361294366	0.024619			
HEIH	3.023435369	8.46E-22			
LINC00847	4.26906441	1.53E-49			
CTC-338M12.4	2.609379579	0.032435			
TRIM52-AS1	1.386638007	0.000127			
FOXCUT	3.014624972	0.022554			
LINC01011	1.43130168	0.037252			
LOC100506207	3.946869533	0.000726			
ZSCAN16-AS1	3.035382848	1.76E-22			
FOXP4-AS1	2.176470412	0.013986			

LOC101927189	4.280200752	0.018202			
LINC00472	2.093927816	9.22E-07			
CAHM	2.057206004	0.0016			
LOC100289495	2.35610017	0.003207			
PSMG3-AS1	1.386357881	0.006741			
LOC646762	1.120699561	0.00758			
LINC00957	2.150173694	2.93E-06			
LINC01393	3.621449117	0.025086			
PRKAG2-AS1	9.811871047	7.78E-28			
LINC01003	2.61790338	5.98E-14			
RBPMS-AS1	4.776797504	2.03E-15			
LOC102723701	5.332527337	9.75E-09			
LINC00968	1.628891144	0.003146			
LINC01301	1.991941659	0.005278			
RDH10-AS1	2.162230883	0.045799			
CA3-AS1	6.629926107	0.000437			
HAS2-AS1	2.039298621	0.001706			
LINC01239	2.325017344	0.000273			
B4GALT1-AS1	1.586268007	0.002619			
LINC00963	1.899075659	0.015275			
SNHG7	1.01466137	0.00264			
RHOXF1-AS1	11.66068142	9.98E-15			
LINC00632	4.634862648	9.51E-21			
CDR1	5.325326077	1.83E-53			

Table 5.2 List of upregulated (UP_Symbol) and downregulated (DOWN_Symbol) lncRNAs during white adipogenesis.

Chapter 6: Bibliography

1. Sundararaman, B. *et al.* Resources for the Comprehensive Discovery of Functional RNA Elements. *Mol Cell* **61**, 903–913 (2016).

2. Cristancho, A. G. & Lazar, M. A. Forming functional fat: a growing understanding of adipocyte differentiation. *Nat Rev Mol Cell Biol* **12**, 722–734 (2011).

3. Rosen, E. D. & Spiegelman, B. M. What we talk about when we talk about fat. *Cell* **156**, 20–44 (2014).

4. Ruiz-Ojeda, F. J., Rupérez, A. I., Gomez-Llorente, C., Gil, A. & Aguilera, C. M. Cell Models and Their Application for Studying Adipogenic Differentiation in Relation to Obesity: A Review. *Int J Mol Sci* **17**, (2016).

5. Sadie-Van Gijsen, H. Adipocyte biology: It is time to upgrade to a new model. *J Cell Physiol* **234**, 2399–2425 (2019).

6. Cristancho, A. G. & Lazar, M. A. Forming functional fat: a growing understanding of adipocyte differentiation. (2011) doi:10.1038/nrm3198.

7. Cave, E. & Crowther, N. J. The Use of 3T3-L1 Murine Preadipocytes as a Model of Adipogenesis. *Methods in Molecular Biology* **1916**, 263–272 (2019).

8. Ruiz-Ojeda, F. J., Rupérez, A. I., Gomez-Llorente, C., Gil, A. & Aguilera, C. M. Cell models and their application for studying adipogenic differentiation in relation to obesity: a review. *Int J Mol Sci* **17**, 1040 (2016).

9. Kusminski, C. M., Bickel, P. E. & Scherer, P. E. Targeting adipose tissue in the treatment of obesity-associated diabetes. *Nature Reviews Drug Discovery 2016 15:9* **15**, 639–660 (2016).

10. Müller, T. D., Blüher, M., Tschöp, M. H. & DiMarchi, R. D. Anti-obesity drug discovery: advances and challenges. *Nature Reviews Drug Discovery 2021 21:3* **21**, 201–223 (2021).

11. Haider, N. & Larose, L. Harnessing adipogenesis to prevent obesity. *Adipocyte* **8**, 98–104 (2019).

12. Jakab, J. *et al.* Adipogenesis as a potential anti-obesity target: a review of pharmacological treatment and natural products. *Diabetes, Metabolic Syndrome and Obesity* **14**, 67–83 (2021).

13. Schirinzi, V., Poli, C., Berteotti, C. & Leone, A. Browning of Adipocytes: A Potential Therapeutic Approach to Obesity. *Nutrients 2023, Vol. 15, Page 2229* **15**, 2229 (2023).

14. Song, T. & Kuang, S. Adipocyte dedifferentiation in health and diseases. *Clin Sci* **133**, 2107–2119 (2019).

15. Zhang, P. *et al.* LncRNA-mediated adipogenesis in different adipocytes. *Int J Mol Sci* **23**, 7488 (2022).

16. Rey, F. *et al.* Role of long non-coding RNAs in adipogenesis: State of the art and implications in obesity and obesity-associated diseases. *Obesity Reviews* **22**, e13203 (2021).

17. Luan, A. *et al.* RNA sequencing for identification of differentially expressed non-coding transcripts during adipogenic differentiation of adipose-derived stromal cells. *Plast Reconstr Surg* **136**, 752 (2015).

18. Gallicchio, L., Olivares, G. H., Berry, C. W. & Fuller, M. T. Regulation and function of alternative polyadenylation in development and differentiation. *RNA Biol* **20**, 908–925 (2023).

19. Neve, J., Patel, R., Wang, Z., Louey, A. & Furger, A. M. Cleavage and polyadenylation: ending the message expands gene regulation. *RNA Biol* **14**, 865–890 (2017).

20. Ogorodnikov, A. *et al.* Transcriptome 3′end organization by PCF11 links alternative polyadenylation to formation and neuronal differentiation of neuroblastoma. *Nature Communications 2018 9:1* **9**, 1–16 (2018).

21. Li, W. *et al.* Systematic profiling of poly(A)+ transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. *PLoS Genet* **11**, e1005166 (2015).

22. Ran, Y. *et al.* CFIm25 regulates human stem cell function independently of its role in mRNA alternative polyadenylation. *RNA Biol* **19**, 686–702 (2022).

23. Hoque, M. *et al.* Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. *Nat Methods* **10**, 133–139 (2013).

24. Spangenberg, L., Correa, A., Dallagiovanna, B. & Naya, H. Role of Alternative Polyadenylation during Adipogenic Differentiation: An In Silico Approach. *PLoS One* **8**, (2013).

25. Cui, J. *et al.* Shortening of HO1 3′UTRs by alternative polyadenylation suppresses adipogenesis in 3T3-L1. *J. Agric. Food Chem* **69**, 8049 (2021).

26. Chao, Y. *et al.* Regulatory roles and mechanisms of alternative RNA splicing in adipogenesis and human metabolic health. *Cell & Bioscience 2021 11:1* **11**, 1–16 (2021).

27. Fu, X. D. & Ares, M. Context-dependent control of alternative splicing by RNA-binding proteins. *Nature Reviews Genetics 2014 15:10* **15**, 689–701 (2014).

28. Naing, Y. T. & Sun, L. The role of splicing factors in adipogenesis and thermogenesis. *Mol Cells* **46**, 268 (2023).

29. Aprile, M. *et al.* PPARgD5, a Naturally Occurring Dominant-Negative Splice Isoform, Impairs PPARg Function and Adipocyte Differentiation. doi:10.1016/j.celrep.2018.10.035.

30. Carter, G. *et al.* Dysregulated alternative splicing pattern of PKCδ during differentiation of human preadipocytes represents distinct differences between lean and obese adipocytes. *ISRN Obes* **2013**, 1–9 (2013).

31. Yi, X., Yang, Y., Wu, P., Xu, X. & Li, W. Alternative splicing events during adipogenesis from hMSCs. *J Cell Physiol* **235**, 304–316 (2020).

32. Ho, E. S., Gunderson, S. I. & Duffy, S. A multispecies polyadenylation site model. *BMC Bioinformatics* **14**, 1–10 (2013).

33. Darmon, S. K. & Lutz, C. S. MRNA 3′ end processing factors: A phylogenetic comparison. *Comp Funct Genomics* **2012**, (2012).

34. Wang, R., Zheng, D., Yehia, G. & Tian, B. A compendium of conserved cleavage and polyadenylation events in mammalian genes. *Genome Res* **28**, 1427–1441 (2018).

35. Wright, C. J., Smith, C. W. J. & Jiggins, C. D. Alternative splicing as a source of phenotypic diversity. *Nature Reviews Genetics 2022 23:11* **23**, 697–710 (2022).

36. Costa-Silva, J., Domingues, D. & Lopes, F. M. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS One* **12**, e0190152 (2017).

37. Bryce-Smith, S. *et al.* Extensible benchmarking of methods that identify and quantify polyadenylation sites from RNA-seq data. *RNA* **29**, 1839–1855 (2023).

38. Shah, A., Mittleman, B. E., Gilad, Y. & Li, Y. I. Benchmarking sequencing methods and tools that facilitate the study of alternative polyadenylation. *Genome Biol* **22**, 1–21 (2021).

39. Zheng, D., Liu, X. & Tian, B. 3′READS+, a sensitive and accurate method for 3′ end sequencing of polyadenylated RNA. *RNA* **22**, 1631–1639 (2016).

40. Routh, A. *et al.* Poly(A)-ClickSeq: click-chemistry for next-generation 3′-end sequencing without RNA enrichment or fragmentation. *Nucleic Acids Res* **45**, e112–e112 (2017).

41. Yalamanchili, H. K. *et al.* PolyA-miner: accurate assessment of differential alternative poly-adenylation from 3 Seq data using vector projections and non-negative matrix factorization. *Nucleic Acids Res* **48**, (2020).

42. Routh, A. DPAC: A Tool for Differential Poly(A)–Cluster Usage from Poly(A)–Targeted RNAseq Data. *G3 Genes|Genomes|Genetics* **9**, 1825–1830 (2019).

43. Yalamanchili, H. K. *et al.* A computational pipeline to infer alternative poly-adenylation from 3′ sequencing data. *Methods Enzymol* **655**, 185–204 (2021).

44. Zhang, Z., Bae, B., Cuddleston, W. H. & Miura, P. Coordination of Alternative Splicing and Alternative Polyadenylation revealed by Targeted Long-Read Sequencing. *bioRxiv* 2023.03.23.533999 (2023).

45. Xu, S. M., Curry-Hyde, A., Sytnyk, V. & Janitz, M. RNA polyadenylation patterns in the human transcriptome. *Gene* **816**, 146133 (2022).

46. Ye, W., Lian, Q., Ye, C. & Wu, X. A Survey on Methods for Predicting Polyadenylation Sites from DNA Sequences, Bulk RNA-seq, and Single-cell RNA-seq. *Genomics Proteomics Bioinformatics* **21**, 67–83 (2023).

47. Jiang, M., Zhang, S., Yin, H., Zhuo, Z. & Meng, G. A comprehensive benchmarking of differential splicing tools for RNA-seq analysis at the event level. *Brief Bioinform* **24**, 1–13 (2023).

48. Halperin, R. F. *et al.* Improved methods for RNAseq-based alternative splicing analysis. *Scientific Reports 2021 11:1* **11**, 1–15 (2021).

49. Wright, D. J. *et al.* Long read sequencing reveals novel isoforms and insights into splicing regulation during cell state changes. *BMC Genomics* **23**, 1–12 (2022).

50. De Paoli-Iseppi, R., Gleeson, J. & Clark, M. B. Isoform Age - Splice Isoform Profiling Using Long-Read Technologies. *Front Mol Biosci* **8**, 711733 (2021).

51. Lin, X. & Li, H. Obesity: epidemiology, pathophysiology, and therapeutics. *Front Endocrinol (Lausanne)* **12**, 706978 (2021).

52. Evseeva, M. N., Balashova, M. S., Kulebyakin, K. Y. & Rubtsov, Y. P. Adipocyte biology from the perspective of in vivo research: review of key transcription factors. *Int J Mol Sci* **23**, 322 (2021).

53. Sufianov, A. *et al.* The role of long non-coding RNAs in the development of adipose cells. *Noncoding RNA Res* **8**, 255–262 (2023).

54. Ding, C. *et al.* De novo reconstruction of human adipose transcriptome reveals conserved lncRNAs as regulators of brown adipogenesis. *Nat Commun* **9**, 1–14 (2018).

55. Xiao, T. *et al.* Long noncoding RNA ADINR regulates adipogenesis by transcriptionally activating C/EBPα. *Stem Cell Reports* **16**, 1006 (2021).

56. Fan, L., Xu, H., Li, D., Li, H. & Lu, D. A novel long noncoding RNA, AC092834.1, regulates the adipogenic differentiation of human adipose-derived mesenchymal stem cells via the DKK1/Wnt/β-catenin signaling pathway. *Biochem Biophys Res Commun* **525**, 747–754 (2020).

57. Rey, F. *et al.* Transcriptional characterization of subcutaneous adipose tissue in obesity affected women highlights metabolic dysfunction and implications for lncRNAs. *Genomics* **113**, 3919–3934 (2021).

58. Zhang, Q. & Tian, B. The emerging theme of 3'UTR mRNA isoform regulation in reprogramming of cell metabolism. *Biochem Soc Trans* **51**, 1111–1119 (2023).

59. Kamieniarz-Gdula, K. & Proudfoot, N. J. Transcriptional control by premature termination: a forgotten mechanism. *Trends in Genetics* **35**, 553 (2019).

60. Mitschka, S. & Mayr, C. Context-specific regulation and function of mRNA alternative polyadenylation. *Nat Rev Mol Cell Biol* **23**, 779–796 (2022).

61. Hoque, M. *et al.* Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. *Nat Methods* **10**, 133–139 (2013).

62. Spangenberg, L., Correa, A., Dallagiovanna, B. & Naya, H. Role of alternative polyadenylation during adipogenic differentiation: an in silico approach. *PLoS One* **8**, e75578 (2013).

63. 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).

64. Potolitsyna, E., Hazell Pickering, S., Germier, T., Collas, P. & Briand, N. Long non-coding RNA HOTAIR regulates cytoskeleton remodeling and lipid storage capacity during adipogenesis. *Sci Rep* **12**, 10157 (2022).

65. Shen, S. *et al.* rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A* **111**, E5593–E5601 (2014).

66. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**, 10–12 (2011).

67. Andrews, S. FastQC: a quality control tool for high throughput sequence data [Online]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).

68. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

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

70. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

71. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

72. Sherman, B. T. *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* **50**, W216–W221 (2022).

73. Safran, M. *et al.* The GeneCards Suite. *Practical Guide to Life Science Databases* 27–56 (2022) doi:10.1007/978-981-16-5812-9_2/FIGURES/8.

74. Castellá, M. *et al.* Adipose tissue plasticity in pheochromocytoma patients suggests a role of the splicing machinery in human adipose browning. *iScience* **26**, 106847 (2023).

75. Wang, R., Nambiar, R., Zheng, D. & Tian, B. PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Res* **46**, D315–D319 (2018).

76. Chen, Y. & Wang, X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res* **48**, 127–131 (2019).

77. Liu, W. & Wang, X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol* **20**, 1–10 (2019).

78. Paz, I., Argoetti, A., Cohen, N., Even, N. & Mandel-Gutfreund, Y. RBPmap: a tool for mapping and predicting the binding sites of rna-binding proteins considering the motif environment. *Methods in Molecular Biology* **2404**, 53–65 (2022).

79. Paz, I., Kosti, I., Ares, M., Cline, M. & Mandel-Gutfreund, Y. RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res* **42**, W361–W367 (2014).

80. Zhao, W. *et al.* POSTAR3: an updated platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. *Nucleic Acids Res* **50**, D287–D294 (2022).

81. Park, J. W., Jung, S., Rouchka, E. C., Tseng, Y.-T. & Xing, Y. rMAPS: RNA map analysis and plotting server for alternative exon regulation. *Nucleic Acids Res* **44**, 333–338 (2016).

82. Hwang, J. Y. *et al.* rMAPS2: an update of the RNA map analysis and plotting server for alternative splicing regulation. *Nucleic Acids Res* **48**, W300–W306 (2020).

83. Wickham, H., François, R., Henry, L., Müller, K. & Vaughan, D. Dplyr: A Grammar of Data Manipulation. https://dplyr.tidyverse.org (2023).

84. Hadley Wickham. Ggplot2: elegant graphics for data analysis. https://ggplot2.tidyverse.org (2016).

85. Blighe, K., Rana, S. & Lewis, M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. https://bioconductor.org/packages/EnhancedVolcano (2023) doi:10.18129/B9.bioc.EnhancedVolcano.

86. Min, S. H. & Zhou, J. Smplot: An R package for easy and elegant data visualization. *Front Genet* **12**, 802894 (2021).

87. Mathur, N. *et al.* Human visceral and subcutaneous adipose stem and progenitor cells retain depot-specific adipogenic properties during obesity. *Front Cell Dev Biol* **10**, 983899 (2022).

88. Norreen-Thorsen, M. *et al.* A human adipose tissue cell-type transcriptome atlas. *Cell Rep* **40**, 111046 (2022).

89. Cai, R. *et al.* Adiponectin AS lncRNA inhibits adipogenesis by transferring from nucleus to cytoplasm and attenuating Adiponectin mRNA translation. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1863**, 420–432 (2018).

90. Ou, Y. *et al.* Targeting antisense lncRNA PRKAG2-AS1, as a therapeutic target, suppresses malignant behaviors of hepatocellular carcinoma cells. *Front Med (Lausanne)* **8**, 649279 (2021).

91. Zhang, F. *et al.* LINC00673 silencing inhibits cell migration and invasion by suppressing PI3K/AKT signaling in glioma. *Neuroreport* **29**, 718–722 (2018).

92. Ji, Z. *et al.* C-Myc-activated long non-coding RNA LINC01050 promotes gastric cancer growth and metastasis by sponging miR-7161-3p to regulate SPZ1 expression. *Journal of Experimental & Clinical Cancer Research* **40**, 351 (2021).

93. Cen, D., Huang, H., Yang, L., Guo, K. & Zhang, J. Long noncoding RNA STXBP5-AS1 inhibits cell proliferation, migration, and invasion through inhibiting the PI3K/AKT signaling pathway in gastric cancer cells. *Onco Targets Ther* **12**, 1929–1936 (2019).

94. Shao, S., Wang, C., Wang, S., Zhang, H. & Zhang, Y. LncRNA STXBP5-AS1 suppressed cervical cancer progression via targeting miR-96-5p/PTEN axis. *Biomedicine & Pharmacotherapy* **117**, 109082 (2019).

95. Zhang, W. *et al.* LINC01088 inhibits tumorigenesis of ovarian epithelial cells by targeting miR-24-1-5p. *Sci Rep* **8**, 1–10 (2018).

96. Zhang, Q. *et al.* LINC01060 knockdown inhibits osteosarcoma cell malignant behaviors in vitro and tumor growth and metastasis in vivo through the PI3K/Akt signaling. *Cancer Biol Ther* **24**, 2198904 (2023).

97. Marquez, M. P. *et al.* The role of cellular proliferation in adipogenic differentiation of human adipose tissue-derived mesenchymal stem cells. *Stem Cells Dev* **26**, 1578 (2017).

98. Marcon, B. H. *et al.* Cell cycle genes are downregulated after adipogenic triggering in human adipose tissue-derived stem cells by regulation of mRNA abundance. *Sci Rep* **9**, 1–10 (2019).

99. Castellá, M. S. *et al.* Adipose tissue plasticity in pheochromocytoma patients suggests a role of the splicing machinery in human adipose browning. *iScience* **26**, 106847 (2023).

100. Thunen, A., La Placa, D., Zhang, Z. & Shively, J. E. Role of lncRNA LIPE-AS1 in adipogenesis. *Adipocyte* **11**, 11 (2022).

101. Chen, K., Xie, S. & Jin, W. Crucial lncRNAs associated with adipocyte differentiation from human adiposederived stem cells based on co-expression and ceRNA network analyses. *PeerJ* **2019**, e7544 (2019).

102. Wang, Y., Cheng, Y., Yang, Q., Kuang, L. & Liu, G. Overexpression of FOXD2-AS1 enhances proliferation and impairs differentiation of glioma stem cells by activating the NOTCH pathway via TAF-1. *J Cell Mol Med* **26**, 2620 (2022).

103. Liao, C., Zhou, Y., Li, M., Xia, Y. & Peng, W. LINC00968 promotes osteogenic differentiation in vitro and bone formation in vivo via regulation of miR-3658/RUNX2. *Differentiation* **116**, 1–8 (2020).

104. DOUKA, K. *et al.* Cytoplasmic long noncoding rnas are differentially regulated and translated during human neuronal differentiation. *RNA* **27**, 1082–1101 (2021).

105. Boulberdaa, M. *et al.* A role for the long noncoding RNA SENCR in commitment and function of endothelial cells. *Molecular Therapy* **24**, 978 (2016).

106. Pan, M. *et al.* The effect and mechanism of LINC00663 on the biological behavior of glioma. *Neurochem Res* **46**, 1737–1746 (2021).

107. Fei, Y., Li, Y. & Chen, F. LncRNA-IQCH-AS1 sensitizes thyroid cancer cells to doxorubicin via modulating the miR-196a-5p/PPP2R1B signalling pathway. *J Chemother* **35**, 250–258 (2023).

108. Song, H. *et al.* A novel biomarker NIFK-AS1 promotes hepatocellular carcinoma cell cycle progression through interaction with SRSF10. *J Gastrointest Oncol* **13**, 1927–1941 (2022).

109. Xin, Y. *et al.* SLC8A1 antisense RNA 1 suppresses papillary thyroid cancer malignant progression via the FUS RNA binding protein (FUS)/NUMB like endocytic adaptor protein (Numbl) axis. *Bioengineered* **13**, 12572–12582 (2022).

110. Zhang, J. *et al.* Docetaxel resistance-derived LINC01085 contributes to the immunotherapy of hormone-independent prostate cancer by activating the STING/MAVS signaling pathway. *Cancer Lett* **545**, 215829 (2022).

111. Zhu, X. *et al.* m6A-mediated upregulation of LINC01003 regulates cell migration by targeting the CAV1/FAK signaling pathway in glioma. *Biol Direct* **18**, 1–13 (2023).

112. Li, L., Gan, Y. P. & Peng, H. RAMP2-AS1 inhibits CXCL11 expression to suppress malignant phenotype of breast cancer by recruiting DNMT1 and DNMT3B. *Exp Cell Res* **416**, 113139 (2022).

113. Huang, S. K. *et al.* Overexpression of LINC00673 promotes the proliferation of cervical cancer cells. *Front Oncol* **11**, 669739 (2021).

114. Sun, W. *et al.* LncRNA FRMD6-AS1 promotes hepatocellular carcinoma cell migration and stemness by regulating SENP1/HIF-1α axis. *Pathol Res Pract* **243**, 154377 (2023).

115. Peng, G., Yan, J., Shi, P. & Li, H. LINC01140 hinders the development of breast cancer through targeting miR-200b-3p to downregulate DMD. *Cell Transplant* **32**, 1–14 (2023).

116. Chen, Y. *et al.* Long non-coding RNA LINC00312 regulates breast cancer progression through the miR-9/CDH1 axis. *Mol Med Rep* **21**, 1296–1303 (2020).

117. Xu, B. *et al.* Overexpression of microRNA-9 inhibits 3T3-L1 cell adipogenesis by targeting PNPLA3 via activation of AMPK. *Gene* **730**, 144260 (2020).

118. Li, G., Wang, C., Wang, Y., Xu, B. & Zhang, W. LINC00312 represses proliferation and metastasis of colorectal cancer cells by regulation of miR-21. *J Cell Mol Med* **22**, 5565 (2018).

119. Kurylowicz, A. MicroRNAs in human adipose tissue physiology and dysfunction. *Cells* **10**, 3342 (2021).

120. Lhamyani, S. *et al.* MiR-21 mimic blocks obesity in mice: A novel therapeutic option. *Molecular Therapy: Nucleic Acid* **26**, 401–416 (2021).

121. Qi, F. *et al.* Tumor mutation burden-associated LINC00638/miR-4732-3p/ULBP1 axis promotes immune escape via PD-L1 in hepatocellular carcinoma. *Front Oncol* **11**, 729340 (2021).

122. Li, P., Li, Y., Bai, S., Zhang, Y. & Zhao, L. miR-4732-3p prevents lung cancer progression via inhibition of the TBX15/TNFSF11 axis. *Epigenomics* **15**, 195–207 (2023).

123. Gburcik, V., Cawthorn, W. P., Nedergaard, J., Timmons, J. A. & Cannon, B. An essential role for tbx15 in the differentiation of brown and 'brite' but not white adipocytes. *Am J Physiol Endocrinol Metab* **303**, E1053–E1060 (2012).

124. Zheng, Y., Chen, Z., Zhou, Z., Xu, X. & Yang, H. Silencing of long non-coding RNA LINC00607 prevents tumor proliferation of osteosarcoma by acting as a sponge of miR-607 to downregulate E2F6. *Front Oncol* **10**, 584452 (2021).

125. Yang, T. *et al.* Cantharidin induces apoptosis of human triple negative breast cancer cells through mir-607-mediated downregulation of EGFR. *J Transl Med* **21**, 597 (2023).

126. Latorre, J. *et al.* The relevance of EGFR, ErbB receptors and neuregulins in human adipocytes and adipose tissue in obesity. *Biomedicine & Pharmacotherapy* **156**, 113972 (2022).

127. Pan, J., Kothan, S., Moe Moe, A. T. & Huang, K. Dysfunction of insulin-AKT-UCP1 signalling inhibits transdifferentiation of human and mouse white preadipocytes into brown-like adipocytes. *Adipocyte* **11**, 213 (2022).

128. Jia, J. *et al.* LncRNA TYMSOS promotes epithelial-mesenchymal transition and metastasis in thyroid carcinoma through regulating MARCKSL1 and

activating the PI3K/Akt signaling pathway. *Critical Reviews™ in Eukaryotic Gene Expression* **33**, 1–14 (2023).

129. Gu, Y. *et al.* TYMSOS drives the proliferation, migration, and invasion of gastric cancer cells by regulating ZNF703 via sponging miR-4739. *Cell Biol Int* **45**, 1710–1719 (2021).

130. Elsafadi, M. *et al.* MicroRNA-4739 regulates osteogenic and adipocytic differentiation of immortalized human bone marrow stromal cells via targeting LRP3. *Stem Cell Res* **20**, 94–104 (2017).

131. Yuan, Y. *et al.* FOXM1/lncRNA TYMSOS/miR-214-3p–mediated high expression of NCAPG correlates with poor prognosis and cell proliferation in non–small cell lung carcinoma. *Front Mol Biosci* **8**, 785767 (2022).

132. Xi, F. X. *et al.* MicroRNA-214-3p targeting Ctnnb1 promotes 3T3-L1 preadipocyte differentiation by interfering with the Wnt/β-Catenin signaling pathway. *Int J Mol Sci* **20**, 1816 (2019).

133. Flowers, M. T. & Ntambi, J. M. Role of stearoyl-coenzyme A desaturase in regulating lipid metabolism. *Curr Opin Lipidol* **19**, 248–256 (2008).

134. Luo, Q. *et al.* Role of ACSL5 in fatty acid metabolism. *Heliyon* **9**, e13316 (2023).

135. Wang, X. *et al.* The roles, molecular interactions, and therapeutic value of CDK16 in human cancers. *Biomedicine & Pharmacotherapy* **164**, 114929 (2023).

136. El Ouarrat, D. *et al.* TAZ is a negative regulator of PPARγ activity in adipocytes and TAZ deletion improves insulin sensitivity and glucose tolerance. *Cell Metab* **31**, 162 (2020).

137. Chen, J. *et al.* E2F1 regulates adipocyte differentiation and adipogenesis by activating ICAT. *Cells* **9**, 1024 (2020).

138. Sekine, Y. *et al.* HADHB, a fatty acid beta-oxidation enzyme, is a potential prognostic predictor in malignant lymphoma. *Pathology* **54**, 286–293 (2022).

139. Ru, W. *et al.* Non-coding RNAs and adipogenesis. *Int J Mol Sci* **24**, 9978 (2023).

140. Zhang, X., Price, N. L. & Fernández-Hernando, C. Non-coding RNAs in lipid metabolism. *Vascul Pharmacol* **114**, 93–102 (2019).

141. Ortega, F. J. *et al.* MiRNA expression profile of human subcutaneous adipose and during adipocyte differentiation. *PLoS One* **5**, e9022 (2010).

142. Ouyang, D. *et al.* MiR-181a-5p regulates 3T3-L1 cell adipogenesis by targeting Smad7 and Tcf7l2. *Acta Biochim Biophys Sin* **48**, 1034–1041 (2016).

143. Xu, J., Zhang, L., Shu, G. & Wang, B. microRNA-16–5p promotes 3T3-L1 adipocyte differentiation through regulating EPT1. *Biochem Biophys Res Commun* **514**, 1251–1256 (2019).

144. Wei, L. M. *et al.* MiR-125b-2 knockout increases high-fat diet-induced fat accumulation and insulin resistance. *Sci Rep* **10**, 1–10 (2020).

145. Shen, L. *et al.* MicroRNA-23a regulates 3T3-L1 adipocyte differentiation. *Gene* **575**, 761–764 (2016).

146. Rockstroh, D., Löffler, D., Kiess, W., Landgraf, K. & Körner, A. Regulation of human adipogenesis by miR125b-5p. *Adipocyte* **5**, 283 (2016).

147. Du, J. *et al.* MicroRNA-125a-5p affects adipocytes proliferation, differentiation and fatty acid composition of porcine intramuscular fat. *Int J Mol Sci* **19**, 501 (2018).

148. You, L. *et al.* The role of microRNA-23b-5p in regulating brown adipogenesis and thermogenic program. *Endocr Connect* **9**, 457 (2020).

149. Saha, P. K. *et al.* Browning and beiging of adipose tissue: its role in the regulation of energy homeostasis and as a potential target for alleviating metabolic diseases: miR-30a targets gene networks that promote browning of human and mouse adipocytes. *Am J Physiol Endocrinol Metab* **319**, E667 (2020).

150. Wang, Q. *et al.* miR-17-92 cluster accelerates adipocyte differentiation by negatively regulating tumor-suppressor Rb2/p130. *Proc Natl Acad Sci U S A* **105**, 2889–2894 (2008).

151. Wang, P. *et al.* Genome-wide microRNA screening reveals miR-582-5p as a mesenchymal stem cell-specific microRNA in subchondral bone of the human knee joint. *J Cell Physiol* **234**, 21877 (2019).

152. Gan, M. *et al.* Ssc-miR-451 regulates porcine primary adipocyte differentiation by targeting ACACA. *Animals (Basel)* **10**, 1–14 (2020).

153. Li, H. *et al.* MiRNA-10b reciprocally stimulates osteogenesis and inhibits adipogenesis partly through the TGF-β/SMAD2 signaling pathway. *Aging Dis* **9**, 1058 (2018).

154. Xu, B. *et al.* Overexpression of microRNA-9 inhibits 3T3-L1 cell adipogenesis by targeting PNPLA3 via activation of AMPK. *Gene* **730**, 144260 (2020).

155. Liu, H. *et al.* miR-340-5p inhibits sheep adipocyte differentiation by targeting ATF7. *Animal Science Journal* **91**, e13462 (2020).

156. Hong, J. H. *et al.* TAZ, a transcriptional modulator of mesenchymal stem cell differentiation. *Science* **309**, 1074–1078 (2005).

157. Zhang, P. *et al.* RNA-Binding proteins in the regulation of adipogenesis and adipose function. *Cells* **11**, 2357 (2022).

158. Chen, J. *et al.* Integrative analyses of mRNA expression profile reveal the involvement of IGF2BP1 in chicken adipogenesis. *Int J Mol Sci* **20**, 2923 (2019).

159. Tian, B. & Manley, J. L. Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology 2016 18:1* **18**, 18–30 (2016).

160. Zhan, P. *et al.* NCAPG2 promotes tumour proliferation by regulating G2/M phase and associates with poor prognosis in lung adenocarcinoma. *J Cell Mol Med* **21**, 665 (2017).

161. Tao, Q. *et al.* The roles of the cell division cycle-associated gene family in hepatocellular carcinoma. *J Gastrointest Oncol* **12**, 781–794 (2021).

162. DePamphilis, M. L. The 'ORC cycle': A novel pathway for regulating eukaryotic DNA replication. *Gene* **310**, 1–15 (2003).

163. Kim, J. *et al.* Regulation of brown and white adipocyte transcriptome by the transcriptional coactivator NT-PGC-1α. *PLoS One* **11**, 0159990 (2016).

164. Zhou, L. *et al.* Fatty acid oxidation mediated by malonyl-CoA decarboxylase represses renal cell carcinoma progression. *Cancer Res* **83**, 3920–3939 (2023).

165. Tsay, A. & Wang, J. C. The role of PIK3R1 in metabolic function and insulin sensitivity. *Int J Mol Sci* **24**, 12665 (2023).

166. Jia, Q. *et al.* HNRNPA1-mediated 3' UTR length changes of HN1 contributes to cancer- and senescence-associated phenotypes. *Aging* **11**, 4407–4437 (2019).

167. Gruber, A. J. *et al.* A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Res* **26**, 1145–1159 (2016).

168. Ruiz-Ojeda, F. J., Méndez-Gutiérrez, A., Aguilera, C. M. & Plaza-Díaz, J. Extracellular matrix remodeling of adipose tissue in obesity and metabolic diseases. *Int J Mol Sci* **20**, 4888 (2019).

169. Hasegawa, K. *et al.* Regulation of de novo lipid synthesis by the small GTPase Rac1 in the adipogenic differentiation of progenitor cells from mouse white adipose tissue. *Int J Mol Sci* **24**, 4608 (2023).

170. Pasarica, M. *et al.* Adipose tissue collagen VI in obesity. *J Clin Endocrinol Metab* **94**, 5155 (2009).

171. Thorsen, K. *et al.* Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. *Mol Cell Proteomics* **7**, 1214–1224 (2008).

172. Arafat, H. *et al.* Tumor-specific expression and alternative splicing of the COL6A3 gene in pancreatic cancer. *Surgery* **150**, 306–315 (2011).

173. Tang, X. *et al.* An RNA interference-based screen identifies MAP4K4/NIK as a negative regulator of PPARγ, adipogenesis, and insulin-responsive hexose transport. *Proc Natl Acad Sci U S A* **103**, 2087 (2006).

174. Peng, H. Y. *et al.* RBM4A-SRSF3-MAP4K4 splicing cascade constitutes a molecular mechanism for regulating brown adipogenesis. *Int J Mol Sci* **19**, 2646 (2018).

175. Bouzakri, K. & Zierath, J. R. MAP4K4 gene silencing in human skeletal muscle prevents tumor necrosis factor-alpha-induced insulin resistance. *Journal of Biological Chemistry* **282**, 7783–7789 (2007).

176. Spada, S., Tocci, A., Di Modugno, F. & Nisticò, P. Fibronectin as a multiregulatory molecule crucial in tumor matrisome: from structural and functional features to clinical practice in oncology. *Journal of Experimental & Clinical Cancer Research* **40**, 1–14 (2021).

177. Berger, E. *et al.* Pathways commonly dysregulated in mouse and human obese adipose tissue: FAT/CD36 modulates differentiation and lipogenesis. *Adipocyte* **4**, 161 (2015).

178. Lake, R. J., Tsai, P.-F., Choi, I., Won, K.-J. & Fan, H.-Y. RBPJ, the major transcriptional effector of notch signaling, remains associated with chromatin throughout mitosis, suggesting a role in mitotic bookmarking. *PLoS Genet* **10**, 1004204 (2014).

179. Shan, T., Liu, J., Wu, W., Xu, Z. & Wang, Y. Roles of notch signaling in adipocyte progenitor cells and mature adipocytes. *J. Cell. Physiol* **232**, 1258–1261 (2017).

180. Liu, M. C., Logan, H. & Newman, J. J. Distinct roles for Notch1 and Notch3 in human adipose-derived stem/stromal cell adipogenesis. *Mol Biol Rep* **47**, 8439–8450 (2020).

181. Jacob, A. G. & Smith, C. W. J. Intron retention as a component of regulated gene expression programs. *Hum Genet* **136**, 1043 (2017).

182. Valacca, C. *et al.* Sam68 regulates EMT through alternative splicing–activated nonsense-mediated mRNA decay of the SF2/ASF proto-oncogene. *J Cell Biol* **191**, 87 (2010).

183. Lin, J. C. Impacts of alternative splicing events on the differentiation of adipocytes. *Int J Mol Sci* **16**, 22169 (2015).

184. Schirinzi, V., Poli, C., Berteotti, C. & Leone, A. Browning of adipocytes: a potential therapeutic approach to obesity. *Nutrients 2023, Vol. 15, Page 2229* **15**, 2229 (2023).

185. Roth, C. L., Molica, F., Kwak, B. R., Alvarez-Llamas, G. & Martin-Lorenzo, M. Browning of white adipose tissue as a therapeutic tool in the fight against atherosclerosis. *Metabolites 2021, Vol. 11, Page 319* **11**, 319 (2021).

186. Anderson, W. D. *et al.* Sex differences in human adipose tissue gene expression and genetic regulation involve adipogenesis. *Genome Res* **30**, 1379–1392 (2020).

187. Divoux, A. *et al.* Identification of a novel lncRNA in gluteal adipose tissue and evidence for its positive effect on preadipocyte differentiation. *Obesity* **22**, 1781–1785 (2014).

188. Liu, S. *et al.* Potential key factors involved in regulating adipocyte dedifferentiation. *J Cell Physiol* **237**, 1639–1647 (2022).

189. Mukherjee, S., Graber, J. H. & Moore, C. L. Macrophage differentiation is marked by increased abundance of the mRNA 3' end processing machinery, altered poly(A) site usage, and sensitivity to the level of CstF64. *Front Immunol* **14**, 1091403 (2023).

190. Heller-Trulli, D., Liu, H., Mukherjee, S. & Moore, C. L. UBE3D regulates mRNA 3′-end processing and maintains adipogenic potential in 3T3-L1 cells. *Mol Cell Biol* **42**, 174–196 (2022).

191. Sun, W. *et al.* A transcriptomic analysis reveals novel patterns of gene expression during 3T3-L1 adipocyte differentiation. *Front Mol Biosci* **7**, 564339 (2020).

192. Zhang, Y. *et al.* Alternative polyadenylation: methods, mechanism, function, and role in cancer. *Journal of Experimental and Clinical Cancer Research* **40**, 1–19 (2021).

193. Hershey, J. W. B., Sonenberg, N. & Mathews, M. B. Principles of translational control: an overview. *Cold Spring Harb Perspect Biol* **4**, a011528 (2012).

194. Ying, S. Y., Chang, D. C. & Lin, S. L. The microRNA (miRNA): overview of the RNA genes that modulate gene function. *Mol Biotechnol* **38**, 257–268 (2008).

195. Cenik, C. *et al.* Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Res* **25**, 1610–1621 (2015).

196. Ojima, K., Oe, M., Nakajima, I., Muroya, S. & Nishimura, T. Dynamics of protein secretion during adipocyte differentiation. *FEBS Open Bio* **6**, 816–826 (2016).

197. Boschi, F., Rizzatti, V., Zamboni, M. & Sbarbati, A. Simulating the dynamics of lipid droplets in adipocyte differentiation. *Comput Methods Programs Biomed* **138**, 65–71 (2017).

198. Brasaemle, D. L., Dolios, G., Shapiro, L. & Wang, R. Proteomic Analysis of Proteins Associated with Lipid Droplets of Basal and Lipolytically Stimulated 3T3-L1 Adipocytes. *Journal of Biological Chemistry* **279**, 46835–46842 (2004).

199. Arisawa, K., Ichi, I., Yasukawa, Y., Sone, Y. & Fujiwara, Y. Changes in the phospholipid fatty acid composition of the lipid droplet during the differentiation of 3T3-L1 adipocytes. *The Journal of Biochemistry* **154**, 281–289 (2013).

200. Barneda, D., Frontini, A., Cinti, S. & Christian, M. Dynamic changes in lipid droplet-associated proteins in the "browning" of white adipose tissues. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1831**, 924–933 (2013).

201. Krahmer, N., Farese, R. V. & Walther, T. C. Balancing the fat: Lipid droplets and human disease. *EMBO Mol Med* **5**, 973–983 (2013).

202. Roberts, L. D., Virtue, S., Vidal-Puig, A., Nicholls, A. W. & Griffin, J. L. Metabolic phenotyping of a model of adipocyte differentiation. *Physiol Genomics* **39**, 109–119 (2009).

203. Tseng, H. W. *et al.* Distinct, opposing functions for CFIm59 and CFIm68 in mRNA alternative polyadenylation of Pten and in the PI3K/Akt signalling cascade. *Nucleic Acids Res* **50**, 9397–9412 (2022).

204. Haddad, R. *et al.* An essential role for Clp1 in assembly of polyadenylation complex CF IA and Pol II transcription termination. *Nucleic Acids Res* **40**, 1226–1239 (2012).

205. Ghazy, M. A. *et al.* The interaction of Pcf11 and Clp1 is needed for mRNA 3′-end formation and is modulated by amino acids in the ATP-binding site. *Nucleic Acids Res* **40**, 1214–1225 (2012).

206. Thomson, D. W. & Dinger, M. E. Endogenous microRNA sponges: evidence and controversy. *Nature Reviews Genetics 2016 17:5* **17**, 272–283 (2016).

207. Du, Z. *et al.* Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer. *Nature Communications 2016 7:1* **7**, 1–10 (2016).

208. Su, X., Huang, H., Lai, J., Lin, S. & Huang, Y. Long noncoding RNAs as potential diagnostic biomarkers for diabetes mellitus and complications: A systematic review and meta-analysis. *J Diabetes* **16**, e13510 (2024).

209. Sun, J. *et al.* Differentially expressed circulating LncRNAs and mRNA identified by microarray analysis in obese patients. *Scientific Reports 2016 6:1* **6**, 1–10 (2016).

210. Chen, M. *et al.* A survey on identification and quantification of alternative polyadenylation sites from RNA-seq data. *Brief Bioinform* **21**, 1261–1276 (2020).

211. Wang, R. & Tian, B. APAlyzer: A bioinformatics package for analysis of alternative polyadenylation isoforms. *Bioinformatics* **36**, 3907–3909 (2020).

212. Goering, R. *et al.* LABRAT reveals association of alternative polyadenylation with transcript localization, RNA binding protein expression, transcription speed, and cancer survival. *BMC Genomics* **22**, 1–27 (2021).

213. Jonnakuti, V. S., Wagner, E. J., Maletić-Savatić, M., Liu, Z. & Yalamanchili, H. K. PolyAMiner-Bulk: A Machine Learning Based Bioinformatics Algorithm to Infer and Decode Alternative Polyadenylation Dynamics from bulk RNA-seq data. *bioRxiv* 2023.01.23.523471 (2023).

214. Park, H. J. *et al.* 3′ UTR shortening represses tumor-suppressor genes in trans by disrupting ceRNA crosstalk. *Nature Genetics 2018 50:6* **50**, 783–789 (2018).

215. Ma, J., Metzinger-Le Meuth, V. & Metzinger, L. The roles of microRNAs in obesity: emphasizing links with chronic kidney disease and cardiovascular disorders. *Obesities 2023, Vol. 3, Pages 243-252* **3**, 243–252 (2023).

216. Ciafrè, S. A. & Galardi, S. microRNAs and RNA-binding proteins. *RNA Biol* **10**, 934–942 (2013).

217. Nguyen, T. T. L. *et al.* Glucocorticoids mediate transcriptome-wide alternative polyadenylation: Potential mechanistic and clinical implications. *Clin Transl Sci* **15**, 2758–2771 (2022).

218. Misra, A., Ou, J., Zhu, L. J. & Green, M. R. Global Promotion of Alternative Internal Exon Usage by mRNA 3′ End Formation Factors. *Mol Cell* **58**, 819–831 (2015).

219. Liu, H., Heller-Trulli, D. & Moore, C. L. Targeting the mRNA endonuclease CPSF73 inhibits breast cancer cell migration, invasion, and self-renewal. *iScience* **25**, (2022).

220. Lin, P. *et al.* RBBP6 maintains glioblastoma stem cells through CPSF3-dependent alternative polyadenylation. *Cell Discovery 2024 10:1* **10**, 1–19 (2024).

221. Dubois, J. *et al.* The Nonstructural NS1 Protein of Influenza Viruses Modulates TP53 Splicing through Host Factor CPSF4 . *J Virol* **93**, 2168–2186 (2019).

222. Péterfy, M., Phan, J. & Reue, K. Alternatively spliced lipin isoforms exhibit distinct expression pattern, subcellular localization, and role in adipogenesis. *Journal of Biological Chemistry* **280**, 32883–32889 (2005).

223. Winkle, M., El-Daly, S. M., Fabbri, M. & Calin, G. A. Noncoding RNA therapeutics — challenges and potential solutions. *Nature Reviews Drug Discovery 2021 20:8* **20**, 629–651 (2021).

224. Trincado, J. L. *et al.* SUPPA2: Fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol* **19**, 1–11 (2018).

225. Sakers, A., De Siqueira, M. K., Seale, P. & Villanueva, C. J. Adipose-tissue plasticity in health and disease. *Cell* **185**, 419–446 (2022).

226. Ghaben, A. L. & Scherer, P. E. Adipogenesis and metabolic health. *Nature Reviews Molecular Cell Biology 2018 20:4* **20**, 242–258 (2019).

227. Kaczmarek Michaels, K., Mohd Mostafa, S., Ruiz Capella, J. & Moore, C. L. Regulation of alternative polyadenylation in the yeast Saccharomyces cerevisiae by histone H3K4 and H3K36 methyltransferases. *Nucleic Acids Res* (2020) doi:10.1093/nar/gkaa292.