UCLA
UCLA Electronic Theses and Dissertations

Title
Regulatory RNA: a novel checkpoint in cardiometabolic disease

Permalink
https://escholarship.org/uc/item/5hb8p7rb

Author
Salisbury, David Alex

Publication Date
2021

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Regulatory RNA: a novel checkpoint in cardiometabolic disease

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy

in Molecular Biology

by

David Alex Salisbury

2021

© Copyright by

David Alex Salisbury

2021

ABSTRACT OF THE DISSERTATION

Regulatory RNA: a novel checkpoint

in cardiometabolic disease

by

David Alex Salisbury

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2021

Professor Tamer I. Sallam, Co-Chair

Professor Peter John Tontonoz, Co-Chair

Complexity of living organisms increases as a function of evolution. According to the central dogma, proteins represent the primary form of genetic output in living organisms and therefore all biological functions and increases in complexity must have arisen from proteins. Yet, in direct contradiction to this fundamental dogma , the proportion of protein-coding genes actually decreases as a function of organismal complexity. How can this paradigm be reconciled? How else can genetic information be stored and transmitted in higher organisms without involving proteins? Interestingly, the increase in organismal complexity does in fact coincide with a vast expansion of the noncoding genome, suggesting the regulatory RNAs encoded by this portion of

the genome have the potential to act as potent regulators in diverse biological processes. Regulatory RNA may therefore represent the missing piece in our understanding of how genetic information is stored and transmitted in higher eukaryotes and how these organisms are able to bypass the upper limits of biological complexity that would otherwise be imposed by a regulatory network operating solely on proteins. Thus, a better understanding of the physiological roles and mechanisms of regulatory RNAs in health and disease is likely to have significant therapeutic implications and will undoubtedly shed new light on what the basic notion of a gene truly represents.

The work presented here will focus on two types of regulatory RNA: m^6A-modified RNAs and lncRNAs .In chapter 2, we demonstrate that the m^6A RNA modification is essential for proper regulation of hepatic lipid metabolism. *In-vivo* studies showed that both diet and sex alter the global landscape of m^6A. Under control diet feeding, m^6A is dynamically enriched at 3' UTR regions on lipogenic mRNAs but enrichment at these sites decreases dramatically when mice are fed a lipid-rich diet. Furthermore, deletion of the m^6A installing enzyme *Mettl14* increases hepatic lipogenesis and triglyceride accumulation while also diminishing hepatic sexual dimorphism. Our study outlines a protective role for RNA modifications against the development of fatty liver disease and also fills a gap in our understanding of molecular mechanisms underlying sex-specific differences in hepatic lipid traits.

In chapter 3, we report findings from a preclinical investigation of *Mexis* therapy in atherosclerosis. We previously demonstrated the macrophage-specific lncRNA *MeXis* enhances *Abca1* expression thereby boosting cholesterol efflux capacity of macrophages. To test the therapeutic effects of *MeXis* overexpression within lesions, we used a novel genetic model that

allows spatial control of *Mexis* expression from the endogenous locus. We show that *Mexis* overexpression is associated with increased *Abca1* mRNA and protein levels, enhanced cholesterol efflux to an *ApoA1* acceptor, and reduced foam cell formation *in vitro* and *in vivo.* However, when we performed a bone-marrow transplant (BMT) and infused *LDLR*$^{-/-}$ mice with Cre^{+} and Cre^{-} marrow, we did not observe major differences in plaque burden between groups. Surprisingly, *Mexis* overexpression in the atherosclerosis model did however lead to a substantial elevation in systemic inflammatory markers. Our findings have important implications for proposed strategies that aim to enhance lncRNA expression for chronic disease treatment and highlights the importance of carefully interrogating lncRNA effects in multiple contexts.

In chapter 4, we demonstrate the macrophage-specific lncRNA *AK165607* encodes a novel micropeptide termed *ORF5.* We further demonstrate *ORF5* interacts with aggregation-prone proteins while they are being transported to the plasma membrane in secretory lysosomes for extracellular release. Loss of *ORF5* triggers widespread rupturing of lysosome and results in the intracellular accumulation of toxic protein aggregates. *In-vivo* experiments revealed loss of *ORF5* suppresses macrophage inflammatory response and impairs extracellular secretion of *S100A8.*

The dissertation of David Alex Salisbury is approved.

Steven J. Bensinger

Karen Reue

Tamer I. Sallam, Committee Co-Chair

Peter John Tontonoz, Committee Co-Chair

University of California, Los Angeles

2021

Table of Contents

ACKNOWLEDGEMENTS

When I first arrived at UCLA over five years ago, I had no way of knowing that I would grow over the next five years, both personally and as a scientist, more than I had grown during perhaps any other period of my life. Without a doubt, the person most responsible for this growth is my mentor, Tamer Sallam. Leading by example, Tamer has shown me how to be the most efficient and effective scientist possible. Following his example, I have learned how to be an equally effective communicator. Tamer has shown me how to approach science and overcome whatever seemingly insurmountable obstacles that will surely arise along the way. By observing Tamer's own example, I've learned that hard work really does pay off in the end. He has shown me the value of making every day count and the rewards that await those who never give up. He has shown me that hard work, rather than prior experience, is the true predictor of whether or not one will be successful as a scientist, which has given me the courage and perseverance to take on projects and experiments I would have otherwise thought I was incapable of even attempting. Perhaps most importantly, however, it was through Tamer's steadfast belief and confidence in my abilities as a scientist that allowed me to recognize my full potential. Even when I doubted my own ability to do a particular experiment and was ready to just accept the fact that I wasn't smart or technically savvy enough to get it to work, Tamer never once lost confidence in my ability to ultimately tackle the problem and find a solution. It was his steadfast belief and confidence that I could become a successful scientist, despite my complete lack of prior experience with even the most basic molecular techniques, that gave me the courage and self-confidence to keep trying and find a solution. By doing so, I came to realize I truly can do anything I put my mind to, which is perhaps the most valuable lesson one can learn during their time as a trainee. Tamer has had a tremendous impact on my life and has helped me grow exponentially, both professionally but also as a person, and I am truly grateful and honored to have spent the last five years training under his supervision.

I would also like to thank the other members of my committee: Peter Tontonoz, Steve Bensinger, and Karen Reue. Each of these individuals has invested hours of their incredibly valuable time in making helping to make sure I become a successful scientist and have provided exceptional guidance and mentorship to me during my time at UCLA; and I would like to extend a heartfelt thank you to each of these individuals. In particular, I would like to thank Peter Tontonoz. Like Tamer, Peter believed in me even before I believed in myself. No matter how busy he was or what he had going on that week, Peter was always willing to make time to hear my new results and offer whatever brilliant insight or suggestion he had. Seeing the way he approaches science, designs experiments, interprets results, and especially the way he never manages to lose track of the big picture, have dramatically increased by effectiveness as a scientist and as a communicator and for this I am truly grateful.

Dedication

To my parents, Al and Kymberle Salisbury.

The example they set for me through their own actions and the countless sacrifices they have made for me and my brother are the reasons I have become both the person and the scientist I am today.

CURRICULUM VITAE

David Alex Salisbury

Graduate Student Researcher
Department of Cardiology
University of California, Los Angeles
Charles E. Young Dr. South. MRL 6-619
Los Angeles, CA 90095
Telephone: (910) 340-3988
Email: asalisbury1@ucla.edu

Education

September 2016-	**anticipated Ph.D.**	Molecular Biology, University of California, Los Angeles **Mentor:** Dr. Tamer Sallam **GPA:** 3.43 **Thesis:** RNA modifications in metabolic control. Ph.D Thesis. Department of Cardiology, University of California, Los Angeles, Los Angeles, CA
June 2016	**M.S.**	Biology, California State University, Bakersfield. **Advisor:** Dr. Kathleen Szick **GPA:** 3.9 **Thesis:** Investigating cutaneous bacteria as a source of tetrodotoxin in the rough-skinned newt, *Taricha granulosa.* M.S. Thesis. Department of Biology, California State University, Bakersfield, CA.
June 2014	**B.S.**	Biology (Premed); Spanish minor. University of South Carolina, Columbia, SC **GPA:** 3.63

Professional Experience

2016 -	**Graduate Student Researcher**, Department of Cardiology, University of California, Los Angeles
2018 - 2019	**Teaching Assistant**, Cellular, Molecular and Developmental Biology, University of California, Los Angeles
2014 - 2016	**Research Assistant**, Department of Biology, California State University, Bakersfield
2015 - 2016	**Teaching Associate**, Department of Biology, California State University, Bakersfield
2015	**Graduate Assistant**, Research Experience Vitalizing Sciences University Program (REVS-UP), Chevron Corporation and School of Natural Sciences, Mathematics, and Engineering (NSME), California State University, Bakersfield

Other Research Experience

2015 - 2016	Dr. Paul Smith. Microbiology of the Oil Fly, *Helaeomyia petrolei.*

Academic Honors and Awards

2020 - 2021	UCLA Whitcome Pre-Doctoral Fellowship Awardee
2018 - 2020	Ruth L. Kirschstein National Research Service Award (T32HL069766) recipient
2016	UCLA Cota-Robles Scholarship recipient
2015	CSU Graduate Equity Fellowship Awardee
2015	Student Research Scholar, California State University, Bakersfield.

2015	President's Associates Scholarship Award Recipient
2015	CSU Research Council of the University (RCU) Grant, $1,500
2010 - 2014	McKissick Scholarship Award Recipient, University of South Carolina
2010 - 2014	Capstone Scholar, University of South Carolina

Publications

Salisbury, D., Casero, D., Zhang, Z., Wang, D., Kim, J., Wu, X., Vergnes, L., Williams, K.J., Heurtas-Velzquez, A., Reue, K., Chen, J., and Sallam, T. 2021. Transcriptional regulation of N^6-methyladenosine orchestrates sex-dimorphic metabolic traits. *Nature Metabolism, 3*, pp. 940-953

Zhang, Z., Feng, A. C., **Salisbury, D**., Liu, X., Wu, X., Kim, J., and Pan, C. 2020. Collaborative interactions of heterogenous ribonucleoproteins contribute to transcriptional regulation of sterol metabolism in mice. *Nature communications*, *11*(1), 1-9.

Zhang, Z., **Salisbury, D**., & Sallam, T. 2018. Long noncoding RNAs in atherosclerosis: JACC review topic of the week. *Journal of the American College of Cardiology*, *72*(19), 2380-2390.

Sallam, T., Jones, M., Thomas, B.J., Wu, X., Gilliland, T., Qian, K., Eskin, A., Casero, D., Zhang, Z., Sandhu, J., **Salisbury, D**., Rajbhandari, P., Civelek, M., Hong, C., Ito, A., Liu, X., Daniel, B., Lusis, J., Whiteledge, J., Nagy, L., Castrillo, A., Smale, S., and Tontonoz, P. 2018. Transcriptional regulation of macrophage cholesterol efflux and atherogenesis by a long noncoding RNA. *Nature medicine*, *24*(3), p.304.

Tontonoz, P., Wu, X., Jones, M., Zhang, Z., **Salisbury, D.** and Sallam, T., 2017. Long Noncoding RNA Facilitated Gene Therapy Reduces Atherosclerosis in a Murine Model of Familial Hypercholesterolemia. *Circulation*, *136*(8), pp.776-778.

Professional Presentations

D. Salisbury, Z. Zhang, J. Wu, J. Kim, D. Casero, P. Leon, A. Huertas-Vazques, J. Chen, and T. Sallam. Dynamic regulation of *de novo* lipogenesis by internal RNA modifications. ASMB DEUEL Conference on Lipids. Coronado, CA. March 3-6, 2020. **Poster.**

D. Salisbury, Z. Zhang, J. Wu, and T. Sallam. Dynamic modulation of hepatic epitranscriptome in response to diet. David Geffen School of Medicine Research Day. Los Angeles, CA. October 12, 2019. **Poster**

D. Salisbury, Z. Zhang, J. Wu, and T. Sallam. Dissection of diet-induced changes in hepatic transcriptional networks. David Geffen School of Medicine Research Day. Los Angeles, CA. September 29, 2018. **Poster**

Sallam, T., M. Jones, J. Wu, Z. Zhang, **D. Salisbury,** and P. Tontonoz. Long Non-coding RNA Facilitated Gene Therapy Reduces Atherosclerosis in a Murine Model of Familial Hypercholesterolemia. David Geffen School of Medicine Research Day. Los Angeles, CA. October 9, 2017. **Poster.**

D. Salisbury. Investigating cutaneous bacteria as a source of TTX in the rough skin newt, *Taricha granulosa*. 18th Annual Conference of the Student Research Scholars Program. Bakersfield, CA. April 22, 2016. **Poster.**

Bron-Susbilla, C., **D. Salisbury**, A. Stokes, and K. Szick. The effects of tetrodotoxin concentration on cutaneous bacterial communities in the rough skinned newt (Taricha granulosa). Research Experience Vitalizing Sciences University Program (REVS-UP) Annual Symposium. Bakersfield, CA. August 8, 2015. **Poster.**

Chapter 1: Introduction

Background: Regulatory RNA in Cardiometabolic Disease

Watson and Crick's infamous discovery of the structure and function of DNA completely revolutionized our understanding of molecular biology; but it also had a major unintended consequence in that it propagated a reductionist view of molecular biology wherein all important biological functions were ascribed to proteins. RNA, on the other hand, was disregarded as merely a genetic "intermediate", transcriptional "noise" and even "junk" DNA that was, by itself, incapable of performing any type of useful or relevant biological function. Over the next several decades this misconception persisted. In fact, it became nearly universal amongst the scientific community despite subtle clues that RNA could function as much more than an intermediate messenger. Finally, upon completion of the human genome project in 2003, scientist were forced to address this longstanding conundrum when it was revealed that only 2% of the human genome encodes protein while greater than 75% is transcribed to RNA (Katamaya, 2005; ENCODE, 2007; Wilusz, 2009).

In the years since, it has become abundantly clear that these transcribed regions represent much more than transcriptional noise (reviewed by Morris and Mattick, 2014). Instead, transcription of these RNAs increases organismal complexity and represents an entirely new layer in regulation of gene expression. Regulatory RNAs have now emerged as potent regulators of fundamental biological processes and shown to function through diverse mechanisms(Costa, 2005; Rinn, 2007). Consequently, regulatory RNA dysfunction has been implicated in many disease pathologies including obesity, cancer, and atherosclerosis (Wapinski and Chang, 2011). Furthermore, regulatory RNA is currently being exploited in the development of novel therapeutics for treating a variety of human; which would have been inconceivable just a decade ago.

The work presented here will focus on two types of regulatory RNA: m^6A-modified RNAs and lncRNAs. The molecular mechanisms linking each type of regulatory RNA with cardiometabolic disease will be discussed in chapters 2-4. The relevant background information is introduced in the remaining portion of this chapter.

I: Non-alcoholic Fatty Liver Disease (NAFLD)

Epidemiology and disease burden

Non-alcoholic fatty liver disease (NAFLD) is the most common chronic liver disease in the world. NAFLD affects up to 30% of the adult population but is even more prevalent in specific subpopulations, such as the severely obese and T2D patients, where the prevalence of NAFLD reaches 90% and 75%, respectively (Younossi, 2016). In lean individuals, however, NAFLD is far less common with prevalence rate of only 15%. (Watacheril, 2016). Regional differences in the prevalence of NAFLD have also been reported. Prevalence of NAFLD is highest in the Middle East (32%) and South America (30%) but low in Africa (13%) (Hojland, 2016). NAFLD is associated with increased mortality and increased risk of other adverse health outcomes including type 2 diabetes, cardiovascular disease and end-stage kidney disease (Chalasani, 2012; Musso, 2014). Unfortunately, there are currently no approved pharmacological therapies for NASH (Loombda and Chalasani, 2015).

NAFLD encompasses a broad histological spectrum ultimately resulting from abnormal lipid accumulation in the liver. Accumulation of TG in an otherwise healthy liver represents the first step in disease progression, simple steatosis. Simple steatosis plus inflammation leads to the second stage in disease progression: non-alcoholic steatohepatitis (NASH). NASH plus fibrosis can then lead to cirrhosis and ultimately hepatocellular carcinoma (HCC) if left unresolved. Consequently,

NASH has become the second leading cause of liver disease among adults currently awaiting liver transplantation in the United States and is predicted to become the most common cause for liver transplantation by 2030 (Wong, 2015).

The pathogenesis of NAFLD is still hotly debated as two major schools of thought currently exist. The first, known as the two-hit hypothesis, considers hepatic steatosis as the first "hit" (Day and James, 1998). As the first "hit", steatosis then primes the liver for and increases susceptibility to a subsequent second "hit", which then initiates actual tissue injury in the form of increased inflammation and fibrosis. Recently however, a second view, termed the "multiple parallel hits hypothesis", has become the more favored view (Birkenfield and Shulman, 2014; Kirpich, 2015). According to this hypothesis, the different pathological events leading to liver injury occur in parallel, not consecutively as suggested by the traditional two-hit hypothesis. What constitutes a "hit" and what combinations of "hits" have been linked to NAFLD? The next section will explore these questions and address each of the main types of "hits" associated with development of NAFLD.

Molecular basis of NAFLD

The liver represents a critical organ in lipid metabolism and has long been considered a central regulator of lipid homeostasis. As such, the liver's primary role is to maintain the proper balance between lipid import and lipid export. Five major pathways have been shown to be essential for maintaining tight control of hepatic lipid homeostasis: uptake of circulating fatty acids, de novo lipogenesis (DNL), fatty acid oxidation (FAO), lipolysis of hepatic lipid droplets, and hepatic export of lipids in very low-density lipoproteins (VLDL). Consequently, NAFLD, which is characterized by increased accumulation of lipid in the liver, can result from dysregulation of one or more of these pathways.

Increased uptake of FFAs

NAFLD is characterized by the increased accumulation of lipid in the liver, which can be caused by multiple factors. Increased uptake of free fatty acids (FFA) is one mechanism leading to the abnormal accumulation of fat in the liver (Donnelly, 2005). *CD36, FATP2* and *FATP5* are the main receptors involved in the uptake of circulating fatty acids in the liver and have each been implicating in mediating the increased uptake of circulating lipids observed in NAFLD. The mammalian genome encodes 6 *FATP* isoforms but only two of these isoforms, *FATP2* and *FATP5*, are expressed in the liver (Koo, 2013). Multiple lines of evidence have implicated these two isoforms in the pathogenesis of NAFLD. Murine knockdown of *FATP2* decreased uptake of fatty acids and prevented hepatic steatosis induced by a high fat diet (Falcon, 2010). Furthermore, fatty acid uptake by hepatocytes and hepatic triglyceride content are both reduced in *FATP5* deficient mice (Doege, 2006; Doege, 2008). Additional evidence implicating *FATP*-mediated lipid uptake as a key driver of hepatic steatosis is the expression profiles of *FATP2* and 5, both of which are known to be increased in adolescents with NASH ($n = 27$) compared to normal controls ($n = 6$) (Zhu, 2011). Finally, the *FATP5* promotor polymorphism (rs56225452), which represents a gain-of-function mutation in the *FATP5* promotor, was shown to correlate with BMI-dependent hepatic steatosis in males with biopsy proven NAFLD ($n = 103$) (Auinger, 2010).

Increased *CD36*-mediated lipid uptake has also been identified as an important driver of hepatic steatosis. Induction of steatosis in mice fed a high fat diet is associated with increased mRNA and protein levels of *CD36* (Koonen, 2007; Wilson, 2016). Adenovirus-mediated overexpression of *CD36* significantly increased both hepatic fatty acid uptake and hepatic fat accumulation (Koonen, 2007) whereas liver-specific knockout of *CD36* in mice decreased hepatic lipid content (Wilson, 2016). Additional evidence for a causal role of *CD36* in steatosis comes from a recent study that

examined localization of the protein in both healthy and sick individuals (Miquilena-Colina, 2011). Immunohistochemistry of liver sections showed *CD36* localized to the hepatocyte plasma membrane in patients with steatosis and NASH whereas a weak *CD36* expression confined to the cytoplasm was observed in hepatocytes in normal livers, potentially indicating the translocation of *CD36* from cytoplasm to the plasma membrane as a key event triggering the progression of NAFLD.

Increased lipogenesis

Enhanced *SREBP-1c*-mediated de novo lipogenesis is another key feature of NAFLD contributing significantly to abnormal accumulation of lipids (Postic, 2008). *De-novo lipogenesis* enables the liver to synthesize new fatty acids from acetyl-CoA and this process is transcriptionally regulated by the transcription factor *Srebp-1c*. *Srebp-1c* was reported to be one of the predominant regulators of *DNL* in NAFLD and multiple lines of evidence implicate elevated *SREBP-1c* and increased expression of downstream targets as a defining feature of NAFLD (Higuchi, 2008).

SREBP-1c expression is enhanced in patients with NAFLD and hepatic triglyceride levels are higher in mice overexpressing *SREBP-1c* (Shimano, 1997; Kohjima, 2007) whereas *SREBP-1c* knockout mice display decreased expression of lipogenic enzymes (Liang, 2002). In addition to elevated *SREBP-1c*, the expression of downstream targets *ACC* and *FASN* was also increased in both patients and animal models with NAFLD (Postic and Girard, 2008; Dorn, 2010). Liver-specific depletion of *ACC1* decreased hepatic lipid accumulation and DNL in mice hepatocytes (Mao, 2006). Altered levels and activity of another *SREBP-1c* target, *DGAT2*, which catalyzes the conversion of diglycerides to triglycerides, has also been associated with NAFLD. Overexpression of *DGAT2* in mice increased hepatic triglyceride content fivefold without affecting insulin sensitivity (Monetti, 2007). Conversely, inhibition of *DGAT2* using antisense oligonucleotides in *db/db* mice fed a methionine and choline-deficient diet-reduced hepatic steatosis (Yamaguchi, 2007).

Multiple other studies have clearly demonstrated the important role of enhanced *DNL* in NAFLD. A small study reported increased *DNL* in NAFLD patients ($n = 5$) compared to controls ($n = 6$) (Diraison, 2003)[38], which was further supported by another study showing enhanced *DNL* in overweight/obese subjects with high liver fat (18.3 ± 3.6%, $n = 13$) compared to those with lower liver fat (3.1 ± 2.7%, $n = 11$) (Lambert, 2014). Furthermore, *DNL* was independently associated with intrahepatic triglyceride levels in this study, and failure to suppress *DNL* during fasting could be a key feature in NAFLD patients. In support of this notion, a separate study found that in obese patients with NAFLD, approximately 26% of hepatic triglycerides were derived from *DNL* and these patients were unable to regulate *DNL* when transitioning from fasting to fed state (Donelly, 2005).

Impaired fatty-acid oxidation (FAO)

Impaired fatty acid beta oxidation (FAO) leads to the accumulation of fatty acids in the liver and can therefore promote increased lipid accumulation in the setting of NAFLD (Fromenty, 2004). Cellular oxidation of fatty acids provides a source of energy to generate ATP (Reddy and Hashimoto, 2001; Rao and Reddy, 2001). FAO occurs primarily in the mitochondria and is transcriptionally regulated by *PPARα* (Begriche, 2013). Activation of *PPARα* induces the transcription of a range of genes related to FAO in the mitochondria, peroxisomes, and cytochromes, thereby reducing hepatic lipid levels (Reddy and Hashimoto, 2001; Kersten, 2017). Consequently, knockout of *PPARα* in *ob/ob* mice induces hepatic steatosis (Lee, 1995). *PPARα* knockout animals were found to exhibit decreased induction of key *PPARα*-target genes, which coincided with the development of hepatic steatosis (Gao, 2015). Additional evidence emphasizes the critical role of *PPARα* in promoting FAO and preventing hepatic lipid accumulation. Although hepatic *PPARα* levels did not differ between patients with steatosis ($n = 16$) and healthy controls ($n = 8$) (Westerbacka, 2007), *PPARα* was downregulated in patients with NASH compared to

patients with steatosis and healthy controls (Gao, 2015). Furthermore, *PPARα* expression decreased with increasing NAFLD activity score and fibrosis stage. Finally, a longitudinal analysis after a 1-year follow-up demonstrated increased *PPARα* was associated with histological improvements in NASH (Francque, 2015).

Reduced Lipid Export

Reduction in lipid clearance can also promote the development of NASH (Fabbrini, 2008). Apart from FAO, export of triglycerides is the only way to prevent hepatic lipid overload (Perry, 2014). Excess lipid is removed from the liver by first being packaged into water-soluble VLDL particles containing apolipoprotein b100 (*Apob*) which are then secreted by hepatocytes (Fabbrini, 2008). Production of these VLDL particles occurs at the endoplasmic reticulum, where the lipidation *ApoB100*-containing particles is catalyzed by the enzyme microsomal triglyceride transfer protein (*MTTP*). The newly formed VLDL particle is then transferred to the Golgi apparatus, where further lipidation takes place until the mature VLDL particle is ultimately formed (Kawano and Cohen, 2013).

Thus, *ApoB-100* and *MTTP* constitute key components in hepatic VLDL secretion and the maintenance of hepatic lipid homeostasis. Consequently, hepatic steatosis, due to impaired triglyceride export, is common in patients with genetic defects in the *ApoB* or *MTTP* gene (hypobetalipoproteinemia and abetaproteinemia, respectively)(Tanoli, 2004; Berriot-Varoqueaux, 2000). Although the mRNA levels of *ApoB* and *MTTP* were found to be higher in patients with NAFLD compared to controls (Higuchi, 2011), *ApoB* synthesis rates were in fact lower in patients with NASH ($n = 7$) compared to lean ($n = 7$) or BMI-matched obese ($n = 7$) controls without NASH (Charlton, 2002). Additionally, hepatic mRNA levels of *apoB100* and *MTTP*, as well as serum VLDL-TG levels, were higher in patients with steatosis ($n = 51$) compared to patients with NASH

(*n* = 53). Additional evidence further implicates deteriorating VLDL assembly and reduced export as key events in the progression from steatosis to NASH. A study reported NAFLD patients that displayed advanced steatosis had decreased *MTTP* levels compared to healthy controls (Fujita, 2009). Furthermore, overexpression of *MTTP* in the Fatty Liver Shionogi mice decreased liver triglycerides and improved NASH (Shindo, 2010).

Decreased Lipolysis

Hepatic triglycerides are primarily stored in the form of lipid droplets. Several lipolytic enzymes, which degrade or break down lipid droplets, are highly expressed in the liver and localize to lipid droplets. Failure to properly degrade these lipid stores can promote abnormal hepatic lipid accumulation and genetic studies have suggested a critical role for lipid droplet proteins in the development of NAFLD. The most strongly associated genetic variant with NASH is a single-nucleotide polymorphism (I148M) in the patatin-like phospholipase domain-containing protein 3 (*PNPLA3*) gene, which encodes a lipolytic enzyme localized to lipid droplets (Romeo, 2008). The I148M variant of *PNPLA3* is resistant to degradation, which allows it to accumulate on lipid droplets and is sufficient to induce steatosis (BasuRay, 2019). Another lipid droplet protein, Hydroxysteroid 17β-dehydrogenase 13 (*HSD17B13*), is also associated with NASH. *HSD17B13* is up-regulated in NASH and several variants of *HSD17B13* are associated with increased steatosis in NAFLD (Ma, 2019). Although genetic studies point to an important role for lipid droplet proteins in regulating features of NAFLD, other studies have suggested lipid droplet proteins may be deleterious; thus additional mechanistic studies are clearly warranted.

Sexual Dimorphism in the Liver

As a sex steroid responsive organ and a major target of GH and T3 action, sexual dimorphism is a major physiological characteristic of the liver (Yokoyama, 2005). Males and female livers exhibit considerable dimorphism in gene expression. In a 2008 proteomics study, gender was found to

affect the expression of more proteins in liver than nutritional status (Valle, 2008). This study identified 23 proteins that were differentially expressed between the sexes and the majority of these proteins were involved in substrate utilization and lipid metabolism. In total, over 1000 genes have now been shown to exhibit sex specific differences in hepatic expression profiles (Lorbek, 2013). A significant portion of these genes are involved in various aspects of lipid metabolism and were previously shown to contribute to sex differences in cardiovascular disease risk (Trapnell, 2012), fatty liver disease (Yokoyama, 2005), and hepatocellular carcinoma (Alnouti and Klassen, 2011). Females, for example, are known to be more susceptible to acute liver failure, autoimmune hepatitis, and toxin-mediated hepatotoxicity (Guy and Peters, 2013), while males are known to have increased susceptibility to liver fibrosis progression (Poynard, 1997) and hepatocellular carcinoma (Clocchiatta, 2017).

Males and females exhibit striking differences in the prevalence of many additional metabolic traits, especially with regard to hepatic TG accumulation and steatosis. A number of studies have demonstrated females have higher basal fasting triglyceride levels than males (Horton, 1999; Zhang, 2006; Link and Reue, 2017). Furthermore, males and females also exhibit considerable dimorphism in key metabolic pathways that help maintain tight control of lipid homeostasis in the liver. For example, gender differences in serum *ApoB* and hepatic triglyceride secretion rates have been previously reported; female rats were demonstrated to have a higher VLDL secretory rate than male rats (Linden, 2001). β-oxidation of FA's, however, seems to be more active in male livers responding to lipid overload compared to females. β-oxidation of FAs is regulated by the transcription factors *PPARα* and *RXRα. PPARα* mRNA/protein *higher in males* and PPRE-dependent promoter activity following treatment with the *PPARα* agonist WY14643 was also higher in males (Pawlak, 2015). In further support of this finding, a comparison of male and female *RXRα* targets revealed male-specific *RXRα* targets were highly enriched for genes involved in β-

oxidation (Kosters, 2013). Sexual dimorphism has also been observed at the level of individual gene expression. Expression of *SCD1*, a key enzyme involved in *DNL*, was found to be differentially expressed between male and female livers (Lee, 1996).

Although the existence of sex-specific differences in hepatic lipid traits has been well documented, the molecular basis for these differences and major determinants remain largely unknown. Growth hormone is widely considered to be the major hormonal determinant of liver sex differences due to its sex-dependent pituitary secretory pattern. In males GH is secreted in a pulsatile fashion whereas females continuously secrete GH. Ablation of circulating GH by hypophysectomy abolishes liver sex differences globally (Wauthier, 2008; Wauthier, 2010). Furthermore, exogenous GH pulses restore male-biased gene expression in females (Holloway, 2006) whereas continuous GH infusion in males induces female-biased genes while repressing male-biased gene expression in the liver (Waxman, 1991). Thus, GH is clearly one factor mediating sex-differences in the liver. However, it has become increasingly clear that this is not the only determinant of liver sex differences. In particular differences in hepatic TG content between males and females do not appear to be the result of sex-specific patterns of GH secretion. Hence, the molecular basis for the observed discrepancy in hepatic lipid content remains largely unknown and hints at the existence of an unknown post-transcriptional mechanism in the liver.

II: Atherosclerosis

Epidemiology and disease burden

The concentrations of blood cholesterol prevalent in most contemporary human societies far exceeds the biological needs of the organism, leading to the development of atherosclerosis (Hopstock, 2017; Schreiner, 2016). Atherosclerosis refers to the accumulation of lipid in the innermost layer of arteries, known as the intima. Atherosclerosis is derived from the Greek word

for ‘gruel’ or ‘porridge’, which refers to the lipid material found in the core of a typical atherosclerotic plaque. Atherosclerosis can cause acute coronary syndromes like myocardial infarction or chronic conditions such as stable angina pectoris. Atherosclerosis also causes many ischaemic strokes and can lead to the formation of aneurysms including those that form in the abdominal aorta (rev. by Libby, 2021).

Atherosclerotic cardiovascular disease (CVD), which includes coronary heart disease, hypertension and stroke, remains a leading cause of vascular disease worldwide. Heart disease resulting from atherosclerotic disease of the coronary arteries and stroke represemt the two leading causes of death in the world (Mozaffarrian, 2017). Over 17 million people died from CVD in 2015, which accounts for 31% of all deaths globally (WHO, 2017). Of these, an estimated 7.4 million occurred as the result of coronary heart disease while 6.7 million were attributed to stroke.. In the US, 37.4% of men and 35.9% of women over 20 years old have some form of CVD. In men with CVD, 37.7% are non-Hispanic white, 46.0% black and 31.3% Hispanic wheras in women, these figures are 35.1%, 47.7% and 33.3%, respectively (Mozaffarrian, 2017). As a result, heart disease is the first cause of death in the US while stroke represents the fifth (Mozaffarrian, 2017).

Large-scale cohort investigations, such as the Framingham study, have revealed many additional risk factors for atherosclerosis that we now regard as ‘traditional’ (Andersson, 2019). Examples of traditional risk factors include tobacco use, unhealthy diet, obesity, physical inactivity, hyperlipidaemia, hypertension and high alcohol use. Recently, though, other non-traditional drivers of atherosclerosis, including disturbed sleep, physical inactivity, the microbiome, air pollution and environmental stress, have gained increasing attention (Drager, 2017; Brook, 2018).

Molecular mechanisms

LDL particles cause atherosclerosis and cumulative exposure of an artery to LDL-C over years remains a principal determinant of atherosclerotic cardiovascular disease (Ference, 2017). In support of a causal role of LDL in atherosclerosis, patients with familial hypercholesterolaemia who accumulate significant quantities of LDL-C at early ages also develop premature atherosclerotic CVD (Nordestgaard, 2013). Furthermore, individuals who have life-long low LDL-C concentrations due to a loss-function-mutation in proprotein convertase subtilisin/kexin type 9 (*PCSK9*), showed a greater reduction of coronary events than that afforded by statin treatment alone (Cohen, 2006). How excessive LDL-C causes atherosclerosis remains controversial. The sequence of events linking LDL-C to development of atherosclerosis is summarized in the oxidized LDL particle hypothesis, which suggests constituents of oxidized LDL particles taken up by macrophages generate ligands that promote foam cell formation and also likely activate the inflammatory response by furnishing neo-epitopes that stimulate humoral and adaptive immunity (Gistera and Hansson, 2017).

Recently however, this view of excess and highly oxidized LDL-C as the main driver of atherosclerotic development has begun to change. Rather than elevations in low-density lipoprotein (LDL) cholesterol, elevations in triglyceride-rich lipoproteins (TGRL) and low high-density lipoprotein (HDL) were recently shown to be the major pattern of lipid abnormality in a large proportion of patients who are treated for atherosclerotic cardiovascular disease (Nordestgaard and Varbo, 2014). There is now substantial human genetic evidence to support a causal role for triglyceride-rich lipoproteins in atherosclerosis (Varbo, 2013) such as the strong observational relationship between levels of lipoprotein(a) (Lp(a)) and atherosclerotic risk (Burgess, 2018). Genetic evidence supporting a protective role for HDL-C against atherosclerosis, on the other hand, is lacking (Musunuru, 2016). Although HDL-C concentrations are consistently found to be inversely associated with the risk of atherosclerotic events in epidemiological studies,

numerous therapies that raise HDL-C have been unable to improve cardiovascular outcomes, which is likely due to the fact that HDL-C tracks inversely with triglyceride concentrations (Libby, 2015).

Role of Liver-X-Receptors (LXRs)

Unable to catabolize cholesterol, mammalian cells must eliminate it through cholesterol efflux. Cholesterol efflux constitutes the first step of reverse cholesterol transport whereby cholesterol is transferred via HDL to the liver for bile acid synthesis and excretion (Lippincourt-Schwartz and Phair, 2010). Transcriptional regulation of macrophage cholesterol efflux is largely mediated through the Liver-X-Receptor (LXR) signaling pathway. LXR is a ligand-activated nuclear receptor that senses cellular cholesterol levels and is considered a master regulator of lipid homeostasis in mammals (Hong and Tontonoz, 2014). In the unliganded state, LXR acts a repressor through interactions with nuclear corepressors. Upon ligand binding, LXRs exchange corepressor for coactivator complexes which induces an array of genes that lower cholesterol levels by promoting cholesterol efflux and inhibiting cholesterol uptake (Calkin and Tontonoz 2010).

Arguably the most important LXR target gene in macrophages is the ATP-binding cassette transporter member 1 (*ABCA1*). *ABCA1* activity is essential for preventing cholesterol accumulation and foam cell formation in atherosclerotic lesions, a key determinant of disease progression. Binding of free *ApoA1* to *Abca1* triggers the transfer of cholesterol from the macrophage plasma membrane to HDL thereby mediating release of cholesterol from the macrophage (Ikonen, 2008). *ABCA1* is mutated in Tangier's disease, which is characterized by the total absence of HDL and premature coronary artery disease Bodzioch, 1999). Deletion of *ABCA1* in immune cells significantly impacts the development and progression of atherosclerotic lesions (Haghpassand, 2001); further highlighting the critical role of this protein in preventing excess cholesterol accumulation in macrophages.

The lncRNA *MeXis,* which is also strongly induced by treatment with the LXR agonist GW3965, was recently shown to augment the expression of *Abca1* in response to LXR activation (Sallam, 2018). This finding expands the repertoire of noncoding gene activation and demonstrates how lncRNAs can function to augment the effects of LXR signaling. The therapeutic potential of sustained overexpression of *MeXis* in the setting of atherosclerosis will be discussed further in chapter 3.

III: Epitranscriptomics

Overview and historical context

The existence of cap and tail modifications on mRNA have been well known for years and demonstrated to have key roles in regulation of gene expression in eukaryotes. Shortly after the discovery of these modifications, however, additional internal modifications of both mRNA and ncRNA began to be discovered. The first, pseudouridine, was discovered in 1960 when the fist biological RNA was sequenced (Cohn 1960). In the years since, over 100 different types of chemical modification to RNAs have been documented in eukaryotes (Boccaletto, 2018; Xuan, 2018). Collectively, these modifications constitute the epitranscriptome and represent an entirely new layer of genetic information. In eukaryotes, tRNAs contain the largest number of modifications with approximately 13 modifications per molecule. However, chemical modifications have been identified on virtually all types of eukaryotic RNA including rRNA, mRNAs, valut RNAs, and lncRNAs (Adams and Cory, 1975; Desrosiers, 1974). Essentially, internal RNA modifications function to regulate protein production by altering splicing, translation, or decay rates of the modified transcript. Furthermore, internal RNA modifications superimpose an additional layer of regulation over an already complex regulatory RNA network, substantially increasing the upper limits of biological complexity imposed by the regulatory RNA network alone. Thus, analogous to

the diverse chemical marks on histone tails, epitranscriptomic modifications encode an additional layer of regulation and can alter fundamental properties of RNA metabolism, which can have incredibly profound biochemical and cellular consequences.

Internal modifications to eukaryotic mRNA can be classified into four groups. Adenosine methylations represent the first group. The most common type of adenosine methylation is N^6-methyladenosine or m^6A, which will be discussed in detail in the following section. Another type of adenosine modification is N^1-methyladenosine or m^1A. Despite being less abundant than m^6A in both human and mouse tissues, the m^1A modification can dramatically alter RNA secondary structure and thus disrupt RNA:Protein interactions because, unlike m^6A, m^1A carries a positive charge (Dominissini, 2016). m^1A is enriched near translation start sites as well as the first splice site of coding transcripts and predominantly exerts its effects on RNA metabolism by modulating translational efficiency (Li, 2016). Under various types of cellular stress, m^1A can be removed by *ALKBH3* (Dominissini, 2016; Li, 2016).

Cytosine modifications represent the second group of internal RNA modifications in eukaryotes. Examples include 5-metylcytosine (m^5C) and 5-hydroxymethylcytosine (hm^5C). m^5C was first identified on mRNA and lncRNA species, where it was shown to be enriched in UTRs, particularly around binding sites of *AGO* proteins (Squires, 2010). M^5C is deposited by tRNA methyltransferase NSUN2 (Hussain, 2013; Khoddami and Cairns, 2013) and recognized by the RNA export adaptor protein *ALYREF* (Yang, 2017), potentially implicating the modification in regulation of RNA export. M^5C on RNA can be oxidized to hm^5C. In *Drosophila*, oxidation of m^5C is carried out by the Tet-family enzymes (Fu, 2012), is present on over 1500 mRNAs, and seems to be enriched for exonic and intronic regions protein-coding transcripts (Delatte, 2016). Interestingly, hm^5C has not been detected in mammals, although derivatives of this modification

are likely to exist.

Isomerization of Uridine base (Ψ) is the third class of RNA modifications in eukaryotes. Ψ was originally identified in tRNA and rRNA (Cohn, 1960). However, establishment of the PseudoU-seq protocol in 2014 revealed the modification was far more prevalent than previously thought as more than 300 mRNAs were found to contain the modification in humans (Schwartz, 2014). However, development of a novel pulldown method in 2015 known as CeU-seq revealed the modification was even more widespread and was actually present at more than 2,000 sites in human mRNA (Li, 2015). Like many RNA modifications, Ψ is dynamic. Upon stress conditions like heat shock, the PUS family of enzymes have been shown to catalyze the isomerization of uridine (Fernandez, 2013). In terms of its function, Ψ was found to alter stop codon readthrough (Karijolich and Wu, 2011), although the full biological significance of this modification remains to be determined.

Ribose modification is the final class of eukaryotic RNA modifications. 2'-OMe, or N_m, is highly abundant in tRNAs and rRNAs but has also been detected on mRNA (Schibler and Perry, 1977). 2'-OMe sites have only been mapped in highly abundant RNAs but there are currently no methods to identify potential 2'-OMe sites in low-abundance transcripts (Lichinchi, 2016). However, 2'-OMe sites in rRNA provide a remarkable example of the potent regulatory effects of internal RNA modifications, as rRNA biogenesis is completely blocked in the absence of 2'-OMe. Thus, regardless of whether they are found on coding or noncoding RNAs, dynamic RNA modifications clearly constitute a previously overlooked layer of gene regulation in eukaryotes.

m^6A Modification

Of all the internal RNA modifications, N^6-methyladenosine (m^6A) is by far the most common. Originally discovered in the 1970s, m^6A was shown to accelerate pre-mRNA processing and

mRNA transport in mammalian cells (Perry, 1975; Camper, 1984). However, nearly four decades would pass before advances in analytical chemistry and high-throughput sequencing finally allowed the true biological significance of this modification to fully emerge. Finally, in 2012, two independent studies offered the first view of the m^6A epitranscriptome (Dominissini, 2012; Meyer, 2012). Using an m^6A-specific antibody to capture all m^6A-modified RNA species, over 12,000 sites were identified in humans. M^6A was shown to exhibit a nonrandom distribution wherein m^6A sites were found to contain a consensus RRACH motif and were enriched in terminal exons and 3' UTRs of both coding and ncRNAs. Interestingly, this nonrandom distribution was conserved in mice, suggesting m^6A was an evolutionarily conserved phenomena with the potential to exert profound regulatory effects in fundamental biological processes. Furthermore, despite the prevalence of RRACH sequences in the transcriptome, only 1-5% of these sites are actually methylated *in-vivo;* strongly implying that m^6A deposition is highly regulated and therefore functionally important. Like other modifications, m^6A alters RNA secondary structure and thereby disrupts potential RNA:protein interactions, enabling this modification to potently regulate essentially all aspects of RNA metabolism such as stability, translation efficiency, splicing, and nuclear export.

m^6A is co-transcriptionally deposited by a multicompetent m^6a methyltransferase complex, the core of which is a 200 kDa heterodimer of *Mettl3* and *Mettl14* (Liu, 2014; Ping, 2014). Structural studies, however, indicate *Mettl3* is the only catalytically active subunit while *Mettl14* is essential for stabilizing the conformation and RNA-binding ability of *Mettl3* (Wang, 2016; Sledz, 2016). Besides *Mettl3* and *Mettl14*, other regulatory subunits of the multicompetent methyltransferase complex have been identified, including *WTAP* and its cofactors *VIRMA, ZC3H13,* and *RBM15* (Ping, 2014; Schwartz 2014; Patil, 2016; Wen, 2018). Two additional m^6A writers, *Mettl16* and *ZCCHC4,* have recently been identified and shown to catalyze m6A deposition on snRNAs and

rRNAs, respectively (Brown, 2016; Ma, 2019).

m^6A is a dynamic modification, meaning it is reversible and can be removed from m^6A-modified transcripts under certain conditions. The major RNA demethylases or “erasers” are fat and obesity associated protein (*FTO*) and AlkB family member 5 (*ALKBH5*)(Jia, 2011; Zheng, 2013). While *FTO* can demethylate multiple types of methylated RNAs such as m^3T, m^3U, m^6A_m and m^1A-modified RNAs (Jia, 2008; Wei, 2018), *ALKBH5* appears to be specific for m^6A-modified RNAs (Zheng 2013; Mauer, 2017).

Ultimately, the fate of an m^6A-modified transcript is dictated by the specific “reader” protein that recognizes and binds to the modified adenosine. “Readers” are a group of evolutionarily conserved RNA-binding proteins that recognize the m^6A modification either in the nucleus or in the cytoplasm. Nuclear readers such as *YTHDC1, hnRNPC, hnRNPG*, and *hnRNPA2B1* have been shown to regulate alternative splicing (Alarcon, 2015; Liu, 2015; Xiao, 2016). Additionally, *YTHDC1* can also affect the nuclear export of m^6A-modified transcripts (Roundtree, 2017). Reader proteins are also present and able to recognize m^6A-modified transcripts in the cytoplasm. Recognition of m^6A-containing transcripts by the cytoplasmic readers *IGFBP1/2/3, FMRP*, and *PRRC2A* is associated with increased mRNA stability (Shi, 2017; Hsu, 2017). Conversely, recognition by *YTHDF2, YTHDF3,* and *YTHDC2* sequesters modified transcripts in cytoplasmic P-bodies for degradation and are therefore have been associated with reduced mRNA stability (Lin, 2016; Huang, 2018). Surprisingly, binding of these same readers to m^6A-containing transcripts has also been associated with increased translation efficiency (Wang, 2015; Li, 2017); indicating recognition of m^6A may be combinatorial and the same reader can have different outcomes in specific contexts.

m^{6}A in Health and Disease

m^{6}A methylation has been hypothesized to represent a key mechanism enabling rapid transcriptome turnover during cell differentiation (Roundtree, 2017). Consequently, depletion of *Mettl3* has been shown to inhibit differentiation of mouse ESCs into downstream lineages (Batista, 2014). Furthermore, globally deficient *Mettl3*$^{-/-}$ mice are embryonic lethal, as cells derived from early embryos are unable to resolve their naïve pluripotency (Geula, 2015). Additionally, *ALKBH5* is required for spermatogenesis in mice (Zheng, 2013). In *Drosophila,* m^{6}A methylation of RNA is critical for sex determination and neuronal functions (Haussman, 2016; Lence, 2016). Thus, m^{6}A methylation appears to be critical during embryonic development and differentiation.

Given its well recognized role in regulating numerous fundamental biological processes, it perhaps comes as no surprise that dysregulation of m^{6}A is a common feature of many diseases. Abnormal expression or mutations in *Mettl3* and/or *Mettl14* are frequently associated with disease in humans, particulary cancers (Vu, 2017; Wang, 2017). Levels of *FTO* and *ALKBH5* are also altered in various cancers, triggering widespread deregulation of their downstream targets and contributing significantly to disease progression (Cui, 2017; Li, 2017; Zhang, 2016). Interestingly, inhibition of these erasers with small molecule inhibitors has actually been shown to reduce cancer progression (Chou, 2011; Brat, 2015). Deregulation of m^{6}A is also a common feature of several metabolic disorders. *FTO* was originally identified by GWAS studies for predisposition to diabetes, as *FTO* is required for proper splicing during adipogenesis (Frayling, 2007; Zhang, 2015; Wu, 2017). Another study found m^{6}A was essential for mediating repression of *DNL* during fasting (Salisbury, 2021). Consequently, liver-specific deletion of *Mettl14* resulted in increased hepatic TG accumulation and steatosis in fasted male mice that had been fed chow diet (Salisbury, 2021). Lastly, deregulation of m^{6}A has also been associated with hypertension and cardiovascular disease. Recently, m^{6}A was found to play an important role in blood pressure regulation (Mo,

2019) and cardiac gene expression (Kmietczyk, 2019). Another recent study identified m^6A as a novel mechanism of human CD34+ stem cell-derived exosomes during cardiac repair (Mathiyalagan, 2017). Furthermore, dysregulation of the m^6A modification altered cardiomyocyte Ca^{2+} dynamics and cardiac function (Mathiyalagan, 2017).

IV: long noncoding RNAs (lncRNAs)

Discovery and historical context

Perhaps the first scientist to truly recognize the potential of regulatory RNA was John Mattick in the early 1990s. Based on his observation that the density and size of introns increased with developmental complexity, Mattick hypothesized some genes must have evolved to only express intronic or exonic (non-protein coding) regulatory RNAs, which he further surmised was likely the critical prerequisite that enabled the emergence of developmentally complex organisms (Mattick, 1994). Nearly a decade would pass before any evidence supporting Mattick's hypothesis emerged. Finally, upon completion of the human genome project, Mattick's initial hypothesis was confirmed (Cheng, 2005; Carninci, 2005; Katamaya, 2005) and further corroborated by the ENCODE project (ENCODE, 2007; Dunham, 2012), all of which clearly indicated the vast majority (approximately 75%) of the human and mouse genomes was differentially transcribed despite the fact that less than 2% actually encoded a protein. Furthermore, 95% of human transcription initiation sites were found to be mainly associated with transcription of non-polyadenylated non-coding RNAs (ncRNAs)(Venters, 2013), not with mRNA transcription as many had initially believed. Thus, a new era in genetics had arrived: the rise of regulatory RNA.

Long noncoding RNAs (lncRNAs) represent one such class of regulatory RNA. Long ncRNAs (lncRNAs) are defined as non-protein-coding RNAs greater than 200 nucleotides in length, an arbitrary cutoff that excluded all known classes of small RNAs like snoRNAs, miRNAs, and

piRNAs (Mercer, 2009). In order to be classified as noncoding, an RNA must lack open reading frames encoding proteins greaer than 100 amino acids as well as evidence of codon conservation. However, although most annotated lncRNAs do not encode proteins; a small subset was recently shown to encode “micropeptides” (Frith, 2006; Banfai, 2012). Furthermore, lncRNAs can be placed into one of three subclasses based on their genomic origin: intronic, antisense or intergenic (Katamaya, 2005). However, there is no evidence of any intrinsic difference with regard to function between RNAs based on subclass (Morris and Mattick, 2014).

Largely due to a lack of evolutionary conservation and low expression levels, lncRNAs were initially dismissed as transcriptional “noise” and widely considered to have no biological importance. However, some lncRNAs do actually exhibit evolutionary conservation of the primary sequence, including some ultraconserved lncRNAs (Calin, 2007), as well as primate-specific ones (Necsulea, 2014). Furthermore, it has become increasingly clear that lack of primary sequence conservation in lncRNAs does not indicate lack of function (Pang, 2006), as many lncRNAs show evidence of structural conservation (Smith, 2013). Genes that encode lncRNAs also exhibit many of the same features observed for protein-coding genes, further suggesting they possess important biological functions and represent far more than transcriptional noise. Just like protein-coding genes, lncRNA genes often undergo alternative splicing (Johnsson, 2013), exhibit conservation of promoters (Carninci, 2005), show evidence of regulation by conventional transcription factors (Cawley, 2004), and have similar half-lives as mRNAs that encode proteins (Clark, 2012). Additionally, many lncRNAs exhibit highly context and cell-type dependent expression profiles, implying their expression can be tightly regulated as is the case for protein-coding genes (Carninci, 2005). Dynamic expression of lncRNAs that is mediated by conventional transcription factors in a highly tissue or context-specific fashion has been demonstrated in a range of biological systems, including embryonic stem cells (Dinger, 2008), muscles (Sunwoo,

2009), T cells (Pang, 2009), breast tissues (Hu, 2011), the erythroid system (Hu, 2011), and neurons (Mercer, 2010). In fact, lncRNAs were found to have, on average, higher cell specificity than proteins (Djebali, 2012); consistent with their proposed role as regulatory RNAs mediating the higher order assembly and functions of large protein complexes.

Mechanisms of lncRNAs

lncRNAs have been shown to function through diverse mechanisms. Nuclear functional lncRNAs can modulate gene expression both transcriptionally and epigenetically by acting as guides or scaffolds. In this context, lncRNAs can alter chromatin accessibility at target loci by interacting with components of chromatin-modifying complexes and guiding these complexes to specific genomic sites and scaffold their assembly at these sites. For example several lncRNAs can bind Polycomb repressive complexes (PRCs) (Rinn, 2007), trithorax chromatin-activating complexes (Mohammad, 2008) and DNA methyltransferases (Johnsson, 2013); guiding these epigenetic regulatory complexes to target loci, which alters chromatin accessibility to either promote or inhibit gene expression. Perhaps the most famous naturally occurring lncRNA that acts as a guide and regulates epigenetic processes *in vivo* is the lncRNA *Xist* that mediates X chromosome dosage compensation (Brockdorff, 2002; Brown, 2002). By acting as cellular "guides" lncRNAs can also modulate the splicing of the targeted mRNA (Barry, 2014). lncRNAs can bind complementary sequences on target mRNAs and recruit proteins such as *QKI* and serine/arginine-rich splicing factor 1 (*SRSF1*), thereby affecting splicing. Some lncRNAs lncRNAs function to augment the expression of adjacent protein-coding genes by recruiting transcriptional activators or repressors to promoters of their protein-coding neighbors. For example, the lncRNA *MeXis* recruits the transcriptional activator *DDX17* to the promoter of its protein-coding neighbor *ABCA1* in the setting of cholesterol overload (Sallam, 2018). By doing so, *MeXis* enhances transcription of *ABCA1*,

which promotes cholesterol efflux and prevents accumulation of excess cholesterol in macrophages. Other lncRNAs function as enhancers that can augment the expression of adjacent protein-coding genes through their own transcription (Orom, 2010). By recruiting transcriptional machinery to its own promoter to facilitate its own transcription, these cis-regulatory lncRNAs have also brought transcriptional machinery in close proximity to the promoter of the protein-coding gene, thereby enhancing its expression as well. Enhancer lncRNAs have also been shown to guide the physical looping that occurs between enhancers and facilitate re-positioning of nucleosomes to precisely regulate transcription in specific contexts (Andersson, 2009; Nahkuri, 2009).

lncRNAs can also function as molecular decoys. In this context, lncRNAs can sequester key proteins, such as transcriptional activators, repressors, or even ribosomal protiens. As decoys, lncRNAs can directly affect transcription by binding to proteins such as *DNMT1* and sequestering them from their sites of action, leading to a loss of maintenance of DNA methylation and gene activation (Di-Ruscio, 2013). Decoy lncRNAs can also bind complementary sequences of target mRNAs and sequester the transcript itself, which can either promote or inhibit its translation (Johnsson, 2013). Additionally, some lncRNAs contain multiple miRNA target sites and function as molecular "sponges" that sequester specific miRNAs, which prevents them from binding to and inducing degradation of their target mRNAs (Hennessy, 2019). Thus, by acting as molecular decoys, lncRNAs have the potential to regulate protein localization (Willingham, 2005), mRNA translation (Carrieri, 2012), and mRNA stability (Gong and Maquat, 2011).

lncRNAs in Health and Disease

GWAS studies frequently identify sites located outside of protein-coding genes as causal variants associated with a particular disease. Interestingly, many of these sites map to lncRNA genes,

suggesting these regulatory RNA genes may indeed have functional roles and contribute to the onset and progression of various human diseases. Furthermore, these SNPs could disrupt binding of lncRNAs to their protein partners, which would likely have important physiological consequences. Thus, perhaps not surprisingly, differential expression of lncRNAs has now been documented in cancer and many other diseases (Khaitan, 2011).' Examples of such lncRNAs include *ANRIL, PTCSC3, HOTAIR*, and *MALAT1* which are known to be differentially expressed in T2D, papillary thyroid carcinoma, breast cancer, and metastatic lung cancer, respectively (Takeda, 1998; Mourtada, 2009; Gupta, 2010; Askarian-Amiri, 2011). Furthermore, recent advances in understanding different mechanisms by which lncRNAs function has highlighted their potential to serve as novel therapeutic targets. For example, tissue-specific expression patterns, which have been shown for the majority of lncRNAs, dramatically reduces potential unintended side-effects which makes these lncRNAs ideal drug targets. Furthermore, the fact that many lncRNAs are exclusively expressed in a specific cell type or tumor type highlights their potential to serve as circulating biomarkers that could be used to facilitate early detection of various cancers and other human diseases. In addition to their use as potential biomarkers, other approaches targeting lncRNA-dependent mechanisms for therapeutic purposes include ASOs and vector-mediated gene therapy, which will be discussed further in chapter 3.

V: lncRNA-encoded micropeptides

Overview and evolutionary origin

Despite being widely dismissed for years as transcriptional noise or 'junk' DNA, lncRNAs are now recognized as critical players in diverse biological processes. As a critical regulatory role for these molecules has finally begun to emerge, another novel and completely unexpected role for lncRNAs has also become apparent in recent years; some genes currently annotated as lncRNAs may actually encode functional micropeptides. In fact, recent advances in genomics, ribosome

footprinting, and proteomics has revealed thousands of previously unannotated ORFs present in evolutionarily distant organisms that appear to be translated and encode micropeptides (Bazzini, 2014; Andrews, 2014). Micropeptides are generally classified as proteins less than 100 amino acids and vary in size from 5 to 20kDa (Aspden, 2014). Many localize to the surface of membrane-bound organelles, such as the ER and mitochondria (Makarewich, 2015), where they are primarily thought to function in the assembly, stabilization, and regulation of larger membrane protein complexes (Zickerman, 2010). In this capacity, micropeptides may function as “bridge molecules” that efficiently anchor protein complexes to membranes or mediate interactions between subunits in large multisubunit protein complexes. Micropeptides can also function as chaperones that remain attached to and transiently shield the hydrophobic subunits of multimeric complexes, which can become exposed under stress, and thereby protect against protein aggregation. Interestingly, the functional specificity of micropeptides seems to be primarily encoded through structural patterns, which allows for substantial sequence variation as long as the structural pattern is preserved. As a result, micropeptides often exhibit overall limited sequence conservation but retain a simple common topology, defined by a hydrophobic α-helical transmembrane region separating two hydrophilic N- and the C-terminal domains (Brandt, 2006; Pogoheva, 2018). In this section, we will first discuss different strategies currently used for the identification and detection of putative micropeptide-encoding genes, as well as their evolutionary origin. Next, we will highlight examples of functional micropeptides that participate in diverse biological processes including metabolism, muscle performance, inflammation, and cancer. Lastly, the therapeutic potential of micropeptides, particularly for the development of novel antiviral and anticancer therapies, will be discussed.

The majority of micropeptides identified to date function as accessory subunits that mediate the assembly, stabilization, and regulation of larger membrane protein complexes (Zickerman, 2010).

Unlike the central subunits of these large membrane protein complexes that harbor core bioenergetic function and are conserved between eukaryotes and prokaryotes, the micropeptide accessory subunits are not present in prokaryotes and show a variable distribution and degree of conservation among eukaryotes (Zickerman, 2010; Almen, 2009). Although sequence conservation is low, a significant proportion of micropeptide accessory subunits are characterized by a common single transmembrane (STMD) topology, which is defined by a hydrophobic α-helical transmembrane region separating two hydrophilic N- and the C-terminal domains. Importantly, because the functional specificity of these peptides is primarily encoded by structural patterns, substantial variation of the primary sequence is possible as long as the structural pattern is preserved. For example, the sequence of an 8.5 kDa subunit from the yeast cytochrome bc_1 complex revealed no homology to any known proteins (Brandt, 1994). When the yeast subunit was aligned to the 6.4 kDa Bovine homolog, only 28% identity and 51% similarity were observed. However, the predicted structure of the two subunits was found to be highly conserved; both proteins contained 19% charged and 11% aromatic residues and a hydrophobic helix was located in the middle of both proteins. Thus, structure rather than primary sequence may be a better criteria for determining whether or not a particular micropeptide is evolutionarily conserved.

Interestingly, the common STMD structural signature is found in accessory subunits of other membrane protein complexes, particularly those attached to ROS-generating organelles like the ER, mitochondria, and choloroplasts (Makarewich, 2020). In fact, proteins with a single transmembrane (TM) α-helix are the most abundant and functionally diverse group of membrane proteins and are present in the genomes of evolutionarily distant organisms (Krogh, 2001; Almen, 2009; Pogozheva, 2018). When comparative analysis of 6039 STMD proteins from six evolutionarily distant organisms was performed using data from the publicly-available Membranome database (Lomize, 2017), proteins with single TM α-helices were found to constitute

20–30% of all proteins encoded by sequenced genomes (Pogozheva, 2018). Furthermore, the fraction of single-spanning TM α-helix proteins among all helical TM proteins increases from around 15 to 25% in bacteria to greater than 40% in mammals (Liu, 2001; Worch, 2010), underscoring the functional and evolutionary importance of STMD micropeptides. Interestingly, further analysis revealed STMD microproteins did not originate from the mitochondrial genome; they are encoded entirely by the nuclear genome and evolved only after the first mitochondria had developed (Gross, 2009). Thus, nuclear encoded STMD subunits or micropeptides are thought to be a key step during organelle evolution that allowed the host cell to gain control over the assembly and activity of larger membrane protein complexes, and the increasing number of accessory subunits in eukaryotes is proposed to correlate with their increasing demand for regulation of activity (Ludwig, 2001; Little, 2010). Furthermore, the occurrence of STMD microproteins in rapidly evolving regions of the genome, such as ncRNA genes, supports the notion they serve as novel adapter domains essential for the assembly of dynamic multimeric complexes that are constantly evolving new functions.

Detection strategies

Why has the functional role of micropeptides been so elusive and difficult to decipher? The use of stringent bioinformatic criteria when processing next-generation sequencing (NGS) data has inadvertently constrained the identification of important transcripts that exhibit non-canonical protein-coding potential (Saghatelian and Couso, 2015). For example, excluding reads less than 300 nucleotides, reads with non-AUG start codons, single exon reads, as well as reads lacking primary sequence conservation amongst species have all strongly biased against the discovery and identification of micropeptide-encoding genes (Bazzini, 2014; Andrews, 2014). Despite this, Ribo-seq has remained one of the most powerful and commonly used methods in the discovery and annotation of micropeptide-encoding genes (Ingolia, 2011; Chew, 2013). Translational

efficiency (Ingolia, 2011) and ribosome release score (RRS) analyses (Guttman, 2013) can also be utilized to assess the coding potential of putative ORFs. Translational efficiency can be calculated by dividing the number of ribosomal reads covering exons by the number of RNA reads covering exons and must be normalized for length. RRS analysis exploits the known phenomenon of ribosome release from a transcript directly after recognition of an ORFs stop codon by enumerating the number of ribosome profiling reads before and after the in frame stop codon of a putative ORF. Once the RRS value for a putative ORF has been determined, it can be compared against the RRS values of known protein-coding genes, which can be useful for determining the relative likelihood that a putative ORF undergoes active translation. Other strategies for identifying non-canonical ORFs encoded by lncRNAs involve the use of bioinformatics-based tools that can identify hidden ORFs in user supplied DNA sequences, such as sORF finder (Hanada, 2010) and MiPepid (Zhu and Gribskov, 2019). Another bioinformatics-based tool that has proven useful for the identification of micropeptide-encoding genes is PhyloCSF, a comparative genomics method that identifies the evolutionarily conserved protein coding potential of genomic sequences (Lin, 2011). Although each method has its limitations, the use of Ribo-seq in conjunction with other techniques such as publicly-available bioinformatics-based tools has proven to be an incredibly powerful tool for identifying cryptic micropeptide hidden in genes that have been misannotated as lncRNAs.

Many lncRNAs contain putative ORFs that occur by chance but are not actually translated into a functional peptides. Therefore, once a putative micropeptide-encoding gene has been identified, several additional techniques are commonly used to validate micropeptide production. *In-vitro* translation assays, which involve incorporation of radio-labeled methionines into putative translation products, are a common but essential tool for validating production of a micropeptide from a gene of interest. Transfection of tagged constructs, in which a FLAG or other epitope tag

has been inserted in-frame of a potential ORF, have also proven useful for demonstrating production of a micropeptide by western blot analysis or immunofluorescence-based detection (Anderson, 2015; Makarewich, 2018). Although powerful for an initial demonstration of micropeptide production, these methods have inherent limitations, and the results obtained can be somewhat artificial. To overcome this, other strategies are often utilized in conjunction with *in-vitro* translation assays and transfection of tagged ORF constructs. Generation of endogenous knock-in cell lines, where a FLAG or other epitope tag has been inserted in-frame into the endogenous ORF locus, has successfully been used to demonstrate production of a micropeptide by various labs (Anderson, 2015; Jackson, 2018). Not only does this method avoid the inherent artifacts of forced overexpression, it can provide additional insight into regulatory mechanisms governing expression of the micropeptide. CRISPR-based loss-of-function experiments are perhaps the most powerful tool currently available for validating production of a putative micropeptide and have successfully been used to demonstrate essential physiological roles for micropeptides (Bi, 2017). In fact, global screens that look for growth defects or other phenotypes that occur in response to CRISPR/Cas9 perturbations of putative ORFs can be used to simultaneously identify and validate production of functional micropeptides. Other CRISPR/Cas9 perturbations, such insertion of a pre-mature stop codon into the coding sequence of a putative ORF, have also been successfully used to validate micropeptide production from a gene of interest (Anderson, 2015; Bi, 2017). Lastly, custom antibodies have also been successfully employed for validating production of an endogenous micropeptide (Bi, 2017; Makarewich, 2018). However, due to low molecular weight and substantial hydrophobicity of most micropeptides, generation of custom antibodies may not always be a feasible option.

In conclusion, many tools exist for successfully validating production of a micropeptide from a gene of interest, but genome-wide experimental and computational identification of these

functional ORFs remains challenging.

Micropeptides in metabolism

Mitochondria, which generate the bulk of cellular ATP, are composed of over 1,000 proteins that assemble into distinct complexes residing on the inner and outer mitochondrial membranes (IMMs and OMMs) (Kuhlbrandt, 2015). For example, mitochondrial oxidative phosphorylation (OXPHOS) machinery consists of approximately 100 different subunits distributed across five major complexes, with the majority of these subunits deriving from nuclear-encoded genes (Kuhlbrandt, 2015; Calvo, 2016). In fact, mitochondrial proteins are highly enriched among the hundreds of currently annotated proteins with <100 amino acids, and the majority of these small mitochondria-localized peptides are thought to serve as assembly factors that facilitate the precise construction and arrangement of larger mitochondrial protein complexes (Calvo, 2016). However, given the large number of unsolved clinical cases involving mitochondrial deficiency, many additional assembly factors likely remain to be discovered. Interestingly, several groups have identified lncRNA-derived microproteins that regulate mitochondrial functions by modulating the assembly and stability of larger, mitochondria-bound supercomplexes involved in fatty acid β-oxidation, the TCA cycle, and the electron transport chain. Micropeptide regulator of β-oxidation, or *MOXI*, is one such example (Makarewich, 2018). *MOXI* is composed of 56 amino acids and is encoded by *1500011K16Rik*, a muscle-enriched gene previously annotated as a lncRNA. The N-terminus of the *MOXI* protein is hydrophobic and predicted to encode a transmembrane domain, whereas the C-terminal region is highly enriched in basic residues. *MOXI* was initially identified in a bioinformatics screen for putative micropeptides encoded by muscle-enriched lncRNAs using PhyloCSF, a comparative genomics method that identifies the evolutionarily conserved protein coding potential of genomic sequences (Lin, 2011). Translation of *MOXI* was validated using a combination of approaches which included *in-vitro* translation assays, western blot and

immunofluorescence-based detection of epitope-tagged ORF constructs, mass spectrometry, and direct detection with a custom antibody. The authors showed that *MOXI* localizes to the inner mitochondrial membrane (IMM) where it specifically interacts with mitochondrial trifunctional protein (*MTP*), an important subunit of a larger enzymatic complex involved in fatty acid β-oxidation (Uchida, 1992; Jackson, 1992). Isolated heart and skeletal muscle mitochondria from *MOXI* knockout mice exhibit a diminished ability to metabolize fatty acids, while transgenic *MOXI* overexpression led to enhanced β-oxidation. *MOXI* knockout mice also exhibit a profound reduction in exercise capacity, clearly demonstrating a role for this micropeptides in regulation of mitochondrial metabolism and energy homeostasis.

Another group simultaneously reported the production of this 56-amino acid peptide containing a single transmembrane domain from the lncRNA*1500011k16Rik*, which they termed *Mitoregulin* or *Mtln* (Stein, 2018). Several characteristics led these authors to initially hypothesize the lncRNA *1500011k16Rik* may encode a micropeptide. Analysis of nucleotide conservation of the genes homologous to *1500011k16Rik* revealed that a region of the putative ORF was highly conserved among 60 vertebrate species, and analysis of aggregated ribosome profiling data revealed substantial coverage of the putative ORF encoding *Mtln.* Alignment of the putative translation products from other species revealed a high degree of conservation at the amino acid level, as well as a high ratio of synonymous over nonsynonymous codon substitutions and the complete absence of premature in-frame stop codons. Production of a micropeptide from lncRNA *1500011k16Rik* was then confirmed by these authors through mass spectrometry and immunofluorescence-based detection of a *Mtln*-mCherry fusion protein. Consistent with the findings of the other group, these authors also determined *Mtln* localizes to mitochondria. By immunoprecipitating the *Mtln* peptide from an NIH 3T3 cell line ectopically expressing HA-*Mtln*, the authors identified NADH cytochrome b5 reductase 3 (*Cyb5r3*) as an interacting protein. The two

main processes that require the membrane-bound form of *Cyb5r3* are Δ9 fatty acid desaturation (Keyes and Cinti, 1980) and cholesterol biosynthesis (Reddy, 1977), which led the authors to perform LC-MS and lipidome analysis to determine if loss of *Mtln* altered cellular lipid composition. Interestingly, *Mtln* depletion altered the levels of phospholipids and triglycerides, suggesting *Mtln* may be important for proper maintenance of lipid homeostasis. Deletion of *Mtln* was also shown to markedly decrease the activity of mitochondrial respiratory chain complex I. Through a series of additional experiments, the authors went on to show the interaction of *Mtln* with *Cyb5r3* stimulates complex I functioning by providing a favorable lipid composition of the membrane, thereby demonstrating another important role for this micropeptide in cell metabolism.

In addition to mitochondria-localized micropeptides, essential roles for other micropeptides in metabolism have also been described. One such example is lipid drop-associated micropeptide 1 *(LDAMP1*) (Huang, 2020). Lipid droplets (LDs) are an important organelle, highly conserved from bacteria to humans, that serve key functions in cellular lipid homeostasis and metabolic disorders (Farese, 2009; Zhang and Liu, 2019). To identify potential microproteins that may be associated with lipid droplets, the authors isolated all proteins of low molecular weignt from the LDs of myoblasts. In so doing, they successfully a identified a total of 15 LD-associated micropeptides. One of these, *LDAMP1*, was selected for further analysis. Endogenous expression of *LDAMP1* was confirmed by FLAG tag knock-in. The authors also confirmed *LDAMP1* was localized to LDs through a combination of confocal microscopy, cellular fractionation, and immunogold labeling experiments. Furthermore, oleate-induced triglyceride accumulation was found to be significantly reduced in myoblasts overexpressing *LDAMP1*, potentially implicating micropeptides in LD dynamics and related metabolic functions. Other examples of micropeptides with roles in metabolism are beta-cell glucose regulated micropeptide 1 (*BGRM1*) and *ORF60* (Cao, 2019; Gupta, 2019). BGRM1 was initially identified through a bioinformatics screen that analyzed the

coding potential of lncRNAs expressed exclusively in the human pancreas (Cao, 2019). *BGRM1* is a 65-amino acid micropeptide that localizes to the ER and is encoded by a conserved lncRNA whose expression is altered in islets from obese or diabetic humans and mice. Overexpression of *BGRM1* elevated glucose-stimulated insulin production in INS-1 cells as well as islets from humans and mice, and also promoted ER calcium efflux. The authors subsequently demonstrated that *BGRM1* protects against ER stress and also mediates insulin production through modulation of intracellular calcium homeostasis. Thus, micropeptides can function as essential regulators of calcium homoeostasis and insulin secretion in beta cells. *ORF60* represents yet another micropeptide that is encoded by a lncRNA, contains a single transmembrane domain, and is well conserved across species (Gupta, 2019). *ORF60* expression increases in response to HFD feeding. By using CRISPR/Cas9 to generate *ORF60* KO mice, the authors successfully demonstrated an essential physiological role for this novel micropeptide-encoding gene. Under both chow and HFD-feeding, *ORF60* KO mice exhibited impaired glucose tolerance and impaired response to insulin compared to WT counterparts, suggesting loss of *ORF60* may be associated with insulin resistance. BAT and WAT harvested from *ORF60* KO mice also showed evidence of insulin signaling attenuation, suggesting a novel role for this conserved micropeptide in regulation of insulin sensitivity, although the molecular mechanism by which *ORF60* influences insulin signaling pathways remains unclear.

In conclusion, numerous micropeptides have been found to associate with metabolically important organelles where they serve key function in diverse metabolic pathways such as β-oxidation of fatty acids and insulin signaling, underscoring the regulatory potential of micropeptides in maintenance of metabolic homeostasis.

Micropeptides in inflammation

Functional peptides encoded by short open reading frames hidden in genes misannotated as lncRNAs are also emerging as important mediators of inflammation. Although, by definition, lncRNAs are unable to code for proteins, 35% of highly expressed macrophage lncRNAs were recently shown to interact with ribosomes during bacterial infection, suggesting small peptides generated from non-canonical ORFs hidden in lncRNAs are potentially widespread and may play crucial roles during the immune response (Jackson, 2018). Virus infection of human lung cancer cells was also shown to induce 19 novel ORFs hidden in noncoding RNA genes (Razooky, 2017), further implicating micropeptides as important mediators of the immune response. One of the first reports to demonstrate micropeptides encoded by noncannocial ORFs could function as key regulators of the immune response was the identification of a small peptide encoded by the lncRNA *AW112010,* which was later shown to play an essential role in the orchestration of mucosal immunity during both bacterial infection and colitis (Jackson, 2018). To identify *AW112010* and other potentially important protein-coding lncRNAs in the innate immune response, the authors first identified genes that were markedly altered after LPS stimulation. Ribosome association with these candidate lncRNAs was then confirmed using RiboTag RNA-seq, a modified form of Ribo-seq (Haimon, 2018). In RiboTag RNA-seq, CRISPR/Cas9 is used to insert an in-frame HA tag in the *RPL22* gene, which codes for a subunit of the ribosome. After immunoprecipitating tagged ribosomes with an HA-antibody, ribosome-bound mRNAs can then be eluted and analyzed by qPCR or RNA-seq. RiboTag^{LysM} BMDMs stimulated with LPS or infected with *Salmonella* revealed a significant increase in ribosome-association for the lncRNA *Aw112010.* The authors also performed translational efficiency and ribosome release score analysis on lncRNA *AW112010*, a strategy previously used to challenge the notion that some lncRNAs encode translated proteins (Guttman, 2013). Interestingly, the values obtained for *AW112010* were similar to the values obtained for known protein-coding

genes, as were the values obtained for 96 additional lncRNAs whose expression was markedly altered by LPS stimulation. Because generation of a custom antibody for the micropeptide encoded by lncRNA *AW112010* was not feasible, the authors used CRISPR-Cas9 technology to introduce a C-terminal HA-epitope tag into the *Aw112010* gene in mice. A band corresponding to the predicted size of the *AW112010*-encoded micropeptide could be detected by western blot analysis using lysates from LPS-treated *Aw112010*^{HA} mice. Furthermore, when the same CRISPR-based strategy was used to insert a pre-mature STOP codon into the *AW112010* ORF locus, this band disappeared. MS analysis further confirmed that *Aw112010* is a *bona fide* protein-coding lncRNA gene that is translated during the innate immune response to bacterial infection. Because the insertion of a premature stop codon in the coding region of *AW112010* was found to be sufficient for abrogation of its protein-coding potential, *Aw112010*^{Stop} knock-in mice were subsequently used to determine whether translation of the non-canonical ORF within *Aw112010* was physiologically relevant during the immune response. The caecum of *Aw112010*^{stop} mice showed increased bacterial load and these mice presented with higher bacterial burden and dissemination to the liver and spleen compared wild-type littermates when infected with *Salmonella. Aw112010*^{Stop} mice also became moribund with bacterial infection significantly quicker than wild-type mice when infected with 1×10^2 CFUs of *S. Typhimurium* via oral gavage. Interestingly*, Aw112010*^{Stop} mice were significantly protected from colitis compared to WT counterparts, as measured by weight loss and colonic shortening. Mechanistically, loss of the the *AW112010*-encoded micropeptide caused a major defect in *IL-12p40* production, which hampered the ability of mice to combat *Salmonella* infection and undergo mucosal inflammation but also conferred protection against colitis.

Another micropeptide that functions as an important mediator of the immune response is mitochondrial micropeptide-47, or *Mm47* (Bhatta, 2020). *Mm47* is composed of 47 amino acids

and encoded by the lncRNA *1810058I24Rik*. *1810058I24Rik* is downregulated in both human and murine myeloid cells exposed to LPS, as well as other TLR ligands and inflammatory cytokines. Interestingly, analysis of *1810058I24Rik* subcellular localization revealed the transcript localized primarily to the cytosol, suggesting it may have coding potential. By performing *in-vitro* translation assays with ^{35}S-labeled methionine in combination with microscopy and subcellular fractionation experiments in macrophages, the authors successfully demonstrated endogenous expression of the micropeptide and showed that it localized to mitochondria. CRISPR/Cas9–mediated deletion of *Mm47*, as well as siRNA knockdown studies conducted in mice primary macrophages, revealed the transcriptional response downstream of TLR4 remained intact in cells lacking *Mm47*, however these cells were severely compromised for *NLRP3* inflammasome responses. Although the precise mechanism by which *Mm47* on the mitochondria contributes to activation of the *NLRP3* inflammasome is still an open question, the results of this study clearly show that a novel 47-amino acid micropeptide is required for this process.

Another study (Razooky, 2017) has also identified novel peptide-based regulators of the antiviral response previously hidden amongst the >10,000 lncRNA-encoding genes in the mammalian genome. Using influenza as a viral infection model, they performed RNA-seq and ribosome profiling to track expression and translation of putative lncRNA genes that could potentially encode for micropeptides. They successfully identified 17 candidates using this approach. Four of these micropeptides, which are encoded by the lncRNAs *MMP24-AS1, ZFAS, RP11622K12.1, MIR22HG* were found to be highly conserved amongst vertebrates, and the authors were able to validate their endogenous expression using custom antibodies. Interestingly, when the endogenous protein levels of the peptide encoded by lncRNA *MIR22HG* was probed with a custom antibody, the upregulation of the protein that observed upon influenza infection was comparable to the degree of upregulation predicted by ribosome profiling techniques.

Furthermore, the peptide encoded by the lncRNA *MIR22HG* was shown to be stably expressed even in the absence of infection, suggesting it may also function in other contexts besides inflammation.

In conclusion, many annotated lncRNAs contain short open reading frames (sORFs) that undergo active protein translation, and the translation of these functional but non-canonical ORFs represents a critical event during the innate immune response to infection and inflammation that likely has major implications for human health and disease.

Micropeptides as therapeutic targets

The use of small peptides as drugs has a long and successful history in the treatment of human disease (Pu, 2019). In fact, the unique properties of micropeptides make them ideal candidates for the development of novel therapeutics that could be clinically used for the treatment of various human diseases. In particular, micropeptides may prove especially useful for the development of novel antiviral and anticancer therapies. Reasons why micropeptides may possess the unique ability to be therapeutically exploited specifically in these two contexts are discussed further below.

Just as micropeptides can promote the attachment and assembly of large multisubunit complexes to cellular membranes, they can also directly block this process. Thus, certain micropeptides have the inherent capacity to function as membrane fusion inhibitors, a characteristic that makes them ideal candidates for the development novel antiviral therapeutics. Attachment of viral particles to prospective host cells is the critical first step of the viral infection process. Thus, micropeptides that block fusion of the viral envelope with the host cell plasma membrane or with its target receptor embedded therein could potentially be exploited as novel antivirals that protect against

infections by HIV-1, Covid-19, and other viruses by interfering with this critical first step of a viral infection. Furthermore, given that many of the currently administered vaccines do not recognize all possible epitopes of a particular virus and thus fail to provide complete protection, the development of novel, small peptide-based membrane fusion inhibitors may represent a more effective and cost-efficient solution for combating certain types of viral infection in the years to come. In particular, exploiting the ability of some small peptides to function as natural membrane fusion inhibitors has proven especially useful in the development of therapeutics targeting HIV-1. For example, *Griffithsin (GRFT),* a 12.7 kDa protein derived from red algae (Mori, 2005), can prevent HIV-1 infection by inhibiting Env-mediated cell-cell fusion, even when administered at the pM level. *CDM49*, a small peptide composed of only 28 amino acids that mimics *CD4*, can bind the conserved CD4 binding pocket on *gp120*; thereby blocking *CD4:gp120* interactions and preventing HIV-1 attachment to host cells (Vita, 1999). Lastly, the 36 amino-acid peptide *enfuvirtide* (Fuzeon™), developed by Roche, was the first peptide-based HIV entry inhibitor approved by the U.S. FDA in 2003 (Lalezari, 2003). Enfuvirtide targets the HIV-1 envelope protein and blocks *CD4* receptor binding, effectively preventing HIV-1 entry into host T-cells. Thus, the inherent capacity of some micropeptides to function as membrane fusion inhibitors highlights the potential of these molecules to serve as novel therapeutic targets and facilitate the development of new types of antiviral medications.

The discovery and proper annotation of micropeptides hidden in genes misannotated as lncRNAs may also aid in the identification of novel anticancer therapeutic targets. Although chemotherapy is the major treatment method for cancer, it has fierce side effects, including the insufficient uptake of drugs in tumor cells, non-specific cytotoxicity of chemotherapeutic drugs and the rapid emergence of drug resistance, which largely constrain its clinical application (Gandhi, 2014; He, 2016). Interestingly, some small peptides exhibit potent anti-tumor properties. One such example

is *mifamurtide*, which has been widely used in clinics and and achieved varying degrees of efficacy (Ando, 2011). Compared with traditional chemotherapy agents, small peptide-based drugs possess many unique advantages, including high specificity and activity, low immunogenicity, and less cytotoxicity to normal tissues (Leader, 2008). The anti-cancer mechanisms of peptides and small proteins are mainly divided into two types (Vaishya, 2015). The first type is related to direct induction of apoptosis in tumor cells via specific pathways, while the second type relates to direct inhibition of tumor cell growth by targeting tumor angiogenesis or stimulating an immune response. An example of the latter is the *HOXB-AS3* peptide, a conserved 53-amino acid peptide encoded by the lncRNA *HOXB-AS3* (Huang, 2017). *In-vitro* studies showed that *HOXB-AS3*, rather than its lncRNA, inhibited proliferation, migration, invasion, and colony formation of colon cancer cells (CRC) by antagonizing *hnRNP A1* protein. The *HOXB-AS3* peptide also clearly impaired the *in-vivo* growth of CRC xenografts and decreased both the number and size of lung metastatic nodules, indicating *HOXB-AS3* may have great potential for treating colon cancer. Strategies for rescuing or strengthening the function of putative tumor suppressor micropeptides include vaccination with the synthesized peptide or viral vector vaccines that encode the relevant peptides sequence, both of which are now in development for clinical use as alternative cancer therapies (Radvanyi, 2018). Even if these antitumor peptides have a very short half-life *in-vivo* or exhibit difficulties passing the biofilm barrier, they could theoretically be wrapped in nanomaterials to avoid being quickly metabolized and still get delivered to tumor cells and play an anti-cancer role.

Lastly, peptides derived from small non-cannocial ORFs have also been detected on human leukocyte antigen (HLA) complexes, revealing these small ORFs as a source of antigens (Chong, 2020). Furthermore, recent studies have shown that non-canonical tumor-specific HLA peptides derived from non-coding regions could elicit an anti-tumor immune response (Khurana, 2016),

suggesting micropeptides may also represent a rich and completely untapped pool of novel targetable epitopes that could significantly increase the efficacy of cancer immunotherapies.

Although the repertoire of functional micropeptides has increased substantially in recent years, there are likely many additional micropeptides hidden in genes annotated as lncRNAs that remain to be discovered. Identification of these cryptic peptides will likely have major implications for human health and disease. Thus, lncRNA-encoded micropeptides may represent the next frontier in the development of novel forms of antiviral and anticancer therapies, a substantial step forward for a molecule that, less than 10 years before, the majority of the scientific community still overwhelmingly referred to as ‘junk’ DNA.

References:

Alarcón, C. R., Goodarzi, H., Lee, H., Liu, X., Tavazoie, S., & Tavazoie, S. F. (2015). HNRNPA2B1 is a mediator of m6A-dependent nuclear RNA processing events. *Cell*, *162*(6), 1299-1308.

Almén, Markus Sällman, et al. "Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin." *BMC biology* 7.1 (2009): 50.

Alnouti Y, Klaassen CD. 2011. Mechanisms of gender-specific regulation of mouse sulfotransferases (Sults). *Xenobiotica* **41**:187–197.

Ando, Kosei, et al. "Mifamurtide for the treatment of nonmetastatic osteosarcoma." *Expert opinion on pharmacotherapy* 12.2 (2011): 285-292.

Anderson, Douglas M., et al. "A micropeptide encoded by a putative long noncoding RNA regulates muscle performance." *Cell* 160.4 (2015): 595-606.

Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C., & Komorowski, J. (2009). Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome research*, *19*(10), 1732-1741.

Andersson, C., Johnson, A. D., Benjamin, E. J., Levy, D., & Vasan, R. S. (2019). 70-year legacy of the Framingham Heart Study. *Nature Reviews Cardiology*, *16*(11), 687-698.

Andrews, Shea J., and Joseph A. Rothnagel. "Emerging evidence for functional peptides encoded by short open reading frames." Nature Reviews Genetics 15.3 (2014): 193-204.

Askarian-Amiri, M. E., Crawford, J., French, J. D., Smart, C. E., Smith, M. A., Clark, M. B., ... & Mattick, J.

S. (2011). SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer. *Rna*, *17*(5), 878-891.

Aspden, Julie L., et al. "Extensive translation of small open reading frames revealed by Poly-Ribo-Seq." *Elife* 3 (2014): e03528.

Auinger, A., Valenti, L., Pfeuffer, M., Helwig, U., Herrmann, J., Fracanzani, A. L., ... & Rubin, D. (2010). A promoter polymorphism in the liver-specific fatty acid transport protein 5 is associated with features of the metabolic syndrome and steatosis. *Hormone and metabolic research*, *42*(12), 854-859.

Bánfai, B., Jia, H., Khatun, J., Wood, E., Risk, B., Gundling, W. E., ... & Lipovich, L. (2012). Long noncoding RNAs are rarely translated in two human cell lines. *Genome research*, *22*(9), 1646-1657.

Barry, G., Briggs, J. A., Vanichkina, D. P., Poth, E. M., Beveridge, N. J., Ratnu, V. S., ... & Mattick, J. S. (2014). The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. *Molecular psychiatry*, *19*(4), 486-494.

BasuRay, S., Wang, Y., Smagris, E., Cohen, J. C., & Hobbs, H. H. (2019). Accumulation of PNPLA3 on lipid droplets is the basis of associated hepatic steatosis. *Proceedings of the National Academy of Sciences*, *116*(19), 9521-9526.

Batista, P. J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., ... & Chang, H. Y. (2014). m6A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell stem cell*, *15*(6), 707-719.

Bazzini, Ariel A., et al. "Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation." The EMBO journal 33.9 (2014): 981-993.

Begriche K, Massart J, Robin MA, Bonnet F, Fromenty B (2013) Mitochondrial adaptations and dysfunctions in nonalcoholic fatty liver disease. *Hepatology* 58:1497–1507.

Berriot-Varoqueaux N, Aggerbeck LP, Samson-Bouma M, Wetterau JR (2000) The role of the microsomal triglygeride transfer protein in abetalipoproteinemia. *Annu Rev Nutr* 20:663–697.

Bhatta, Ankit, et al. "A Mitochondrial Micropeptide Is Required for Activation of the Nlrp3 Inflammasome." *The Journal of Immunology* 204.2 (2020): 428-437.

Bi, Pengpeng, et al. "Control of muscle formation by the fusogenic micropeptide myomixer." *Science* 356.6335 (2017): 323-327.

Birkenfeld, A. L., & Shulman, G. I. (2014). Nonalcoholic fatty liver disease, hepatic insulin resistance, and type 2 diabetes. *Hepatology*, *59*(2), 713-723.

Boccaletto, P., Machnicka, M. A., Purta, E., Piątkowski, P., Bagiński, B., Wirecki, T. K., ... & Bujnicki, J. M. (2018). MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic acids research*, *46*(D1), D303-D307.

Bodzioch, M., Orso, E., Klucken, J., Langmann, T., Bottcher, A., Diederich, W., Drobnik, W., Barlage, S., Buchler, C., Porsch-Ozcurumez, M., Kaminski, W. E., Hahmann, H. W., Oette, K., Rothe, G., Aslanidis, C., Lackner, K. J., and Schmitz, G. (1999) The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. *Nature genetics22*, 347-351.

Brandt, Ulrich, et al. "Isolation and characterization of QCR10, the nuclear gene encoding the 8.5-kDa subunit 10 of the Saccharomyces cerevisiae cytochrome bc1 complex." *Journal of Biological Chemistry* 269.17 (1994): 12947-12953.

Brandt, Ulrich. "Energy converting NADH: quinone oxidoreductase (complex I)." *Annu. Rev. Biochem.* 75 (2006): 69-92.

Brat, D. J. N., Verhaak, R. G., Aldape, K. D., Yung, W. K., Salama, S. R., Cooper, L. A., ... & Chin, L. (2015). Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. N. Engl. J. Med. *2015*, *2481*, 10.

Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., ... & Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*, *71*(3), 515-526.

Brook, R. D., Newby, D. E., & Rajagopalan, S. (2018). Air pollution and cardiometabolic disease: an update and call for clinical trials. *American journal of hypertension*, *31*(1), 1-10.

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. G., Xing, Y., Lawrence, J., & Willard, H. F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*, *71*(3), 527-542.

Brown, J. A., Kinzig, C. G., DeGregorio, S. J., & Steitz, J. A. (2016). Methyltransferase-like protein 16 binds the 3′-terminal triple helix of MALAT1 long noncoding RNA. *Proceedings of the National Academy of Sciences*, *113*(49), 14013-14018.

Burgess, S., Ference, B. A., Staley, J. R., Freitag, D. F., Mason, A. M., Nielsen, S. F., ... &

Danesh, J. (2018). Association of LPA variants with risk of coronary disease and the implications for lipoprotein (a)-lowering therapies: a Mendelian randomization analysis. *JAMA cardiology*, *3*(7), 619-627.

Calin, G. A., Liu, C. G., Ferracin, M., Hyslop, T., Spizzo, R., Sevignani, C., ... & Croce, C. M. (2007). Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer cell*, *12*(3), 215-229.

Calkin, A. C., & Tontonoz, P. (2010). Liver x receptor signaling pathways and atherosclerosis. *Arteriosclerosis, thrombosis, and vascular biology*, *30*(8), 1513-1518.

Calvo, Sarah E., Karl R. Clauser, and Vamsi K. Mootha. "MitoCarta2. 0: an updated inventory of mammalian mitochondrial proteins." *Nucleic acids research* 44.D1 (2016): D1251-D1257.

Camper, S. A., Albers, R. J., Coward, J. K., & Rottman, F. M. (1984). Effect of undermethylation on mRNA cytoplasmic appearance and half-life. *Molecular and cellular biology*, *4*(3), 538-543.

Cao, Huojun, et al. "Comprehensive identification of micropeptides encoded by long noncoding RNAs in human tissues." *The FASEB Journal* 33.1_supplement (2019): 714-1.

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., ... & Gustincich, S. (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature*, *491*(7424), 454-457.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., ... & Zhu, S. (2005). The transcriptional landscape of the mammalian genome. *science*, *309*(5740), 1559-1563.

Cawley, S., Bekiranov, S., Ng, H. H., Kapranov, P., Sekinger, E. A., Kampa, D., ... & Gingeras, T. R.

(2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell*, *116*(4), 499-509.

Chalasani, N., Younossi, Z., Lavine, J. E., Diehl, A. M., Brunt, E. M., Cusi, K., ... & Sanyal, A. J. (2012). The diagnosis and management of non-alcoholic fatty liver disease: Practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. *Hepatology*, *55*(6), 2005-2023.

Charlton M, Sreekumar R, Rasmussen D, Lindor K, Nair KS (2002) Apolipoprotein synthesis in nonalcoholic steatohepatitis. Hepatology 35:898–904.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., ... & Gingeras, T. R. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *science*, *308*(5725), 1149-1154.

Chew, Guo-Liang, et al. "Ribosome profiling reveals resemblance between long non-coding RNAs and 5′ leaders of coding RNAs." *Development* 140.13 (2013): 2828-2834.

Chong, Chloe, et al. "Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes." *Nature communications* 11.1 (2020): 1-21.

Chou, W. C., Lei, W. C., Ko, B. S., Hou, H. A., Chen, C. Y., Tang, J. L., ... & Tien, H. F. (2011). The prognostic impact and stability of Isocitrate dehydrogenase 2 mutation in adult patients with acute myeloid leukemia. *Leukemia*, *25*(2), 246-253.

Clark, M. B., Johnston, R. L., Inostroza-Ponta, M., Fox, A. H., Fortini, E., Moscato, P., ... & Mattick, J. S. (2012). Genome-wide analysis of long noncoding RNA stability. *Genome research*, *22*(5), 885-898.

Cohn, W. E. (1960). Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics. *Journal of Biological Chemistry*, *235*(5), 1488-1498.

Cohen, J. C., Boerwinkle, E., Mosley Jr, T. H., & Hobbs, H. H. (2006). Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *New England Journal of Medicine*, *354*(12), 1264-1272.

Cory, S., & Adams, J. M. (1975). The modified 5′-terinal sequences in messenger RNA of mouse myeloma cells. *Journal of molecular biology*, *99*(4), 519-547.

Costa, F. F. (2005) Non-coding RNAs: new players in eukaryotic biology. *Gene*: 357, 83-94

Clocchiatti A, Cora E, Zhang Y, Dotto GP. 2016. Sexual dimorphism in cancer. *Nat Rev Cancer* **16**:330–339. doi:10.1038/nrc.2016.30.

Cory, S., & Adams, J. M. (1975). The modified 5′-terinal sequences in messenger RNA of mouse myeloma cells. *Journal of molecular biology*, *99*(4), 519-547.

Cui, Q. I., Shi, H., Ye, P., Li, L., Qu, Q., Sun, G., ... & Shi, Y. (2017). m6A RNA methylation regulates the self-renewal and tumorigenesis of glioblastoma stem cells. *Cell reports*, *18*(11), 2622-2634.

Day, C.P, and James, O.F. (1998) Steatohepatitis: a tale of two “hits”? *Gastroenterology* 114:842–845

Delatte, B., Wang, F., Ngoc, L. V., Collignon, E., Bonvin, E., Deplus, R., ... & Fuks, F. (2016). RNA biochemistry. *Transcriptome-wide distribution and function of RNA hydroxymethylcytosine*, 282-5.

Desrosiers, R., Friderici, K., & Rottman, F. (1974). Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proceedings of the National Academy of Sciences*, *71*(10), 3971-3975.

Dinger, M. E., Amaral, P. P., Mercer, T. R., Pang, K. C., Bruce, S. J., Gardiner, B. B., ... & Mattick, J. S. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome research*, *18*(9), 1433-1445.

Diraison F, Moulin P, Beylot M (2003) Contribution of hepatic de novo lipogenesis and reesterification of plasma non esterified fatty acids to plasma triglyceride synthesis during non-alcoholic fatty liver disease. *Diabetes Metab* 29:478–485.

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., ... & Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, *489*(7414), 101-108.

Di Ruscio, A., Ebralidze, A. K., Benoukraf, T., Amabile, G., Goff, L. A., Terragni, J., ... & Tenen, D. G. (2013). DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature*, *503*(7476), 371-376.

Doege, H., Baillie, R. A., Ortegon, A. M., Tsang, B., Wu, Q., Punreddy, S., ... & Stahl, A. (2006). Targeted deletion of FATP5 reveals multiple functions in liver metabolism: alterations in hepatic lipid homeostasis. *Gastroenterology*, *130*(4), 1245-1258.

Doege, H., Grimm, D., Falcon, A., Tsang, B., Storm, T. A., Xu, H., ... & Stahl, A. (2008). Silencing of hepatic fatty acid transporter protein 5 in vivo reverses diet-induced non-alcoholic fatty liver disease and improves hyperglycemia. *Journal of biological chemistry*, *283*(32), 22186-22192.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., ... & Rechavi, G. (2012). Topology of the human and mouse m 6 A RNA methylomes revealed by m 6 A-seq. *Nature*, *485*(7397), 201-206.

Dominissini, D., Nachtergaele, S., Moshitch-Moshkovitz, S., Peer, E., Kol, N., Ben-Haim, M. S., ... & He, C. (2016). The dynamic N 1-methyladenosine methylome in eukaryotic messenger RNA. *Nature*, *530*(7591), 441-446.

Donnelly, K. L., Smith, C. I., Schwarzenberg, S. J., Jessurun, J., Boldt, M. D., & Parks, E. J. (2005). Sources of fatty acids stored in liver and secreted via lipoproteins in patients with nonalcoholic fatty liver disease. *The Journal of clinical investigation*, *115*(5), 1343-1351.

Dorn C, Riener MO, Kirovski G, Saugspier M, Steib K, Weiss TS, Gabele E, Kristiansen G, Hartmann A, Hellerbrand C (2010) Expression of fatty acid synthase in nonalcoholic fatty liver disease. *Int J Clin Exp Pathol* 3:505–514.

Drager, L. F., McEvoy, R. D., Barbe, F., Lorenzi-Filho, G., & Redline, S. (2017). Sleep apnea and cardiovascular disease: lessons from recent trials and need for team science. *Circulation*, *136*(19), 1840-1850.

Dunham, I. et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.

ENCODE Project consortium., Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., Andrews, R. M., Flicek, P., Boyle, P. J., Cao, H., Carter, N. P., Clelland, G. K., Davis, S., Day, N., Dhami, P., Dillon, S. C., Dorschner, M. O., Fiegler, H., Giresi, P. G., Goldy, J., Hawrylycz, M.,

Haydock, A., Humbert, R., James, K. D., Johnson, B. E., Johnson, E. M., Frum, T. T., Rosenzweig, E. R., Karnani, N., Lee, K., Lefebvre, G. C., Navas, P. A., Neri, F., Parker, S. C., Sabo, P. J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F. S., Dekker, J., Lieb, J. D., Tullius, T. D., Crawford, G. E., Sunyaev, S., Noble, W. S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I. L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H. A., Sekinger, E. A., Lagarde, J., Abril, J. F., Shahab, A., Flamm, C., Fried, C., Hackermuller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J. S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M. C., Thomas, D. J., Weirauch, M. T., Gilbert, J., Drenkow, J., Bell, I., Zhao, X., Srinivasan, K. G., Sung, W. K., Ooi, H. S., Chiu, K. P., Foissac, S., Alioto, T., Brent, M., Pachter, L., Tress, M. L., Valencia, A., Choo, S. W., Choo, C. Y., Ucla, C., Manzano, C., Wyss, C., Cheung, E., Clark, T. G., Brown, J. B., Ganesh, M., Patel, S., Tammana, H., Chrast, J., Henrichsen, C. N., Kai, C., Kawai, J., Nagalakshmi, U., Wu, J., Lian, Z., Lian, J., Newburger, P., Zhang, X., Bickel, P., Mattick, J. S., Carninci, P., Hayashizaki, Y., Weissman, S., Hubbard, T., Myers, R. M., Rogers, J., Stadler, P. F., Lowe, T. M., Wei, C. L., Ruan, Y., Struhl, K., Gerstein, M., Antonarakis, S. E., Fu, Y., Green, E. D., Karaoz, U., Siepel, A., Taylor, J., Liefer, L. A., 76 Wetterstrand, K. A., Good, P. J., Feingold, E. A., Guyer, M. S., Cooper, G. M., Asimenos, G., Dewey, C. N., Hou, M., Nikolaev, S., Montoya-Burgos, J. I., Loytynoja, A., Whelan, S., Pardi, F., Massingham, T., Huang, H., Zhang, N. R., Holmes, I., Mullikin, J. C., Ureta-Vidal, A., Paten, B., Seringhaus, M., Church, D., Rosenbloom, K., Kent, W. J., Stone, E. A., Program, N. C. S., Baylor College of Medicine Human Genome Sequencing, C., Washington University Genome Sequencing, C., Broad, I., Children's Hospital Oakland Research, I., Batzoglou, S., Goldman, N., Hardison, R. C., Haussler, D., Miller, W., Sidow, A., Trinklein, N. D., Zhang, Z. D., Barrera, L., Stuart, R., King, D. C., Ameur, A., Enroth, S., Bieda, M. C., Kim, J., Bhinge, A. A., Jiang, N., Liu, J., Yao, F., Vega, V. B., Lee, C. W., Ng, P., Shahab, A., Yang, A., Moqtaderi, Z., Zhu, Z., Xu, X., Squazzo, S., Oberley, M. J., Inman, D., Singer, M. A., Richmond, T. A., Munn, K. J., RadaIglesias, A., Wallerman, O., Komorowski, J., Fowler, J. C., Couttet, P., Bruce, A. W., Dovey, O. M., Ellis, P. D., Langford, C. F., Nix, D. A., Euskirchen, G., Hartman, S., Urban, A. E., Kraus, P., Van Calcar, S.,

Heintzman, N., Kim, T. H., Wang, K., Qu, C., Hon, G., Luna, R., Glass, C. K., Rosenfeld, M. G., Aldred, S. F., Cooper, S. J., Halees, A., Lin, J. M., Shulha, H. P., Zhang, X., Xu, M., Haidar, J. N., Yu, Y., Ruan, Y., Iyer, V. R., Green, R. D., Wadelius, C., Farnham, P. J., Ren, B., Harte, R. A., Hinrichs, A. S., Trumbower, H., Clawson, H., HillmanJackson, J., Zweig, A. S., Smith, K., Thakkapallayil, A., Barber, G., Kuhn, R. M., Karolchik, D., Armengol, L., Bird, C. P., de Bakker, P. I., Kern, A. D., LopezBigas, N., Martin, J. D., Stranger, B. E., Woodroffe, A., Davydov, E., Dimas, A., Eyras, E., Hallgrimsdottir, I. B., Huppert, J., Zody, M. C., Abecasis, G. R., Estivill, X., Bouffard, G. G., Guan, X., Hansen, N. F., Idol, J. R., Maduro, V. V., Maskeri, B., McDowell, J. C., Park, M., Thomas, P. J., Young, A. C., Blakesley, R. W., Muzny, D. M., Sodergren, E., Wheeler, D. A., Worley, K. C., Jiang, H., Weinstock, G. M., Gibbs, R. A., Graves, T., Fulton, R., Mardis, E. R., Wilson, R. K., Clamp, M., Cuff, J., Gnerre, S., Jaffe, D. B., Chang, J. L., Lindblad-Toh, K., Lander, E. S., Koriabine, M., Nefedov, M., Osoegawa, K., Yoshinaga, Y., Zhu, B., and de Jong, P. J. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*:447, 799-816 220.

Fabbrini, E., Mohammed, B. S., Magkos, F., Korenblat, K. M., Patterson, B. W., & Klein, S. (2008). Alterations in adipose tissue and hepatic lipid kinetics in obese men and women with nonalcoholic fatty liver disease. *Gastroenterology*, *134*(2), 424-431.

Falcon, A., Doege, H., Fluitt, A., Tsang, B., Watson, N., Kay, M. A., & Stahl, A. (2010). FATP2 is a hepatic fatty acid transporter and peroxisomal very long-chain acyl-CoA synthetase. *American Journal of Physiology-Endocrinology and Metabolism*, *299*(3), E384-E393.

Farese RV, Jr.., Walther TC. Lipid droplets finally get a little R-E-S-P-E-C-T. Cell. 2009;**139**(5):855–60

Ference, B. A., Ginsberg, H. N., Graham, I., Ray, K. K., Packard, C. J., Bruckert, E., ... & Catapano, A. L. (2017). Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus

statement from the European Atherosclerosis Society Consensus Panel. *European heart journal*, *38*(32), 2459-2472.

Fernández, I. S., Ng, C. L., Kelley, A. C., Wu, G., Yu, Y. T., & Ramakrishnan, V. (2013). Unusual base pairing during the decoding of a stop codon by the ribosome. *Nature*, *500*(7460), 107-110.

Francque S, Verrijken A, Caron S, Prawitt J, Paumelle R, Derudas B, Lefebvre P, Taskinen MR, Van Hul W, Mertens I, Hubens G, Van Marck E, Michielsen P, Van Gaal L, Staels B (2015) PPARalpha gene expression correlates with severity and histological treatment response in patients with non-alcoholic steatohepatitis. *J Hepatol* 63:164–173.

Frith, M. C., Forrest, A. R., Nourbakhsh, E., Pang, K. C., Kai, C., Kawai, J., ... & PLoS Genetics EIC Wayne Frankel. (2006). The abundance of short proteins in the mammalian proteome. *PLoS genetics*, *2*(4), e52.

Fromenty, B., Robin, M. A., Igoudjil, A., Mansouri, A., & Pessayre, D. (2004). The ins and outs of mitochondrial dysfunction in NASH. *Diabetes & metabolism*, *30*(2), 121-138.

Fujita K, Nozaki Y, Wada K, Yoneda M, Fujimoto Y, Fujitake M, Endo H, Takahashi H, Inamori M, Kobayashi N, Kirikoshi H, Kubota K, Saito S, Nakajima A (2009) Dysfunctional very-low-density lipoprotein synthesis and release is a key factor in nonalcoholic steatohepatitis pathogenesis. *Hepatology* 50:772–780.

Fu, Y. (2012). *Dynamic regulation of RNA modifications by AlkB family dioxygenases*. The University of Chicago.

Gandhi, N. S., Tekade, R. K., and Chougule, M. B. (2014). Nanocarrier mediated delivery of siRNA/miRNA in combination with chemotherapeutic agents for cancer therapy: current progress and

advances. *J. Control. Release* 194, 238–256. doi: 10.1016/j.jconrel.2014.09.001

Gao Q, Jia Y, Yang G, Zhang X, Boddu PC, Petersen B, Narsingam S, Zhu YJ, Thimmapaya B, Kanwar YS, Reddy JK (2015) PPARalpha-deficient ob/ob obese mice become more obese and manifest severe hepatic steatosis due to decreased fatty acid oxidation. *Am J Pathol* 185:1396–1408.

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A. A., Kol, N., Salmon-Divon, M., ... & Hanna, J. H. (2015). m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science*, *347*(6225), 1002-1006.

Gisterå, A., & Hansson, G. K. (2017). The immunology of atherosclerosis. *Nature Reviews Nephrology*, *13*(6), 368-380.

Gong, C., & Maquat, L. E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature*, *470*(7333), 284-288.

Gross, Jeferson, and Debashish Bhattacharya. "Mitochondrial and plastid evolution in eukaryotes: an outsiders' perspective." *Nature Reviews Genetics* 10.7 (2009): 495-505.

Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., ... & Chang, H. Y. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, *464*(7291), 1071-1076.

Gupta, Nikhil, et al. "272-LB: The Micropeptide ORF60 Modulates Insulin Sensitivity." (2019): 272-LB.

Guttman, Mitchell, et al. "Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins." *Cell* 154.1 (2013): 240-251.

Guy J, Peters MG. 2013. Liver disease in women: the influence of gender on epidemiology, natural history, and patient outcomes. *Gastroenterol Hepatol* **9**:633–639.

Haghpassand, M., Bourassa, P. A., Francone, O. L., and Aiello, R. J. (2001) Monocyte/macrophage expression of ABCA1 has minimal contribution to plasma HDL levels. The Journal of clinical investigation108, 1315-1320.

Haimon, Zhana, et al. "Re-evaluating microglia expression profiles using RiboTag and cell isolation strategies." *Nature immunology* 19.6 (2018): 636-644.

Hanada, Kousuke, et al. "sORF finder: a program package to identify small open reading frames with high coding potential." *Bioinformatics* 26.3 (2010): 399-400.

Haussmann, I. U., Bodi, Z., Sanchez-Moran, E., Mongan, N. P., Archer, N., Fray, R. G., & Soller, M. (2016). m 6 A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature*, *540*(7632), 301-304.

He, C., Tang, Z., Tian, H., and Chen, X. (2016). Co-delivery of chemotherapeutics and proteins for synergistic therapy. *Adv. Drug Deliv. Rev.* 98, 64–76. doi: 10.1016/j.addr.2015.10.021.

Hennessy, E. J., van Solingen, C., Scacalossi, K. R., Ouimet, M., Afonso, M. S., Prins, J., ... & Moore, K. J. (2019). The long noncoding RNA CHROME regulates cholesterol homeostasis in primates. *Nature metabolism*, *1*(1), 98-110.

Higuchi N, Kato M, Shundo Y, Tajiri H, Tanaka M, Yamashita N, Kohjima M, Kotoh K, Nakamuta M, Takayanagi R, Enjoji M (2008) Liver X receptor in cooperation with SREBP-1c is a major lipid synthesis regulator in nonalcoholic fatty liver disease. Hepatol Res 38:1122–

1129.

Higuchi N, Kato M, Tanaka M, Miyazaki M, Takao S, Kohjima M, Kotoh K, Enjoji M, Nakamuta M, Takayanagi R (2011) Effects of insulin resistance and hepatic lipid accumulation on hepatic mRNA expression levels of apoB, MTP and L-FABP in non-alcoholic fatty liver disease. *Exp Ther Med* 2:1077–1081.

Hojland Ipsen D, Tveden-Nyborg P, Lykkesfeldt J (2016) Normal weight dyslipidemia: is it all about the liver? Obesity (Silver Spring) 24:556–567.

Holloway MG, Laz EV, Waxman DJ. 2006. Codependence of growth hormone-responsive, sexually dimorphic hepatic gene expression on signal transducer and activator of transcription 5b and hepatic nuclear factor 4α. Mol. Endocrinol. **20**:647–660.

Hong, C., & Tontonoz, P. (2014). Liver X receptors in lipid metabolism: opportunities for drug discovery. *Nature reviews Drug discovery*, *13*(6), 433-444.

Hopstock, L. A., Bønaa, K. H., Eggen, A. E., Grimsgaard, S., Jacobsen, B. K., Løchen, M. L., ... & Wilsgaard, T. (2017). Longitudinal and secular trends in total cholesterol levels and impact of lipid-lowering drug use among Norwegian women and men born in 1905–1977 in the population-based Tromsø Study 1979–2016. *BMJ open*, *7*(8), e015001.

Horton, J. D., Shimano, H., Hamilton, R. L., Brown, M. S., & Goldstein, J. L. (1999). Disruption of LDL receptor gene in transgenic SREBP-1a mice unmasks hyperlipidemia resulting from production of lipid-rich VLDL. *The Journal of clinical investigation*, *103*(7), 1067-1076.

Hsu, P. J., Zhu, Y., Ma, H., Guo, Y., Shi, X., Liu, Y., ... & He, C. (2017). Ythdc2 is an N 6-

methyladenosine binding protein that regulates mammalian spermatogenesis. *Cell research*, *27*(9), 1115-1127.

Huang, H., Weng, H., Sun, W., Qin, X., Shi, H., Wu, H., ... & Chen, J. (2018). Recognition of RNA N 6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nature cell biology*, *20*(3), 285-295.

Huang, Ting, et al. "Identification of a Micropeptide on Lipid Droplet That Mediates Lipid Storage and Insulin Sensitivity." *bioRxiv* (2020).

Hu, W., Yuan, B., Flygare, J., & Lodish, H. F. (2011). Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation. *Genes & development*, *25*(24), 2573-2578.

Hussain, S., Sajini, A. A., Blanco, S., Dietmann, S., Lombard, P., Sugimoto, Y., ... & Frye, M. (2013). NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell reports*, *4*(2), 255-261.

Ikonen, E. (2008). Cellular cholesterol trafficking and compartmentalization. *Nature reviews Molecular cell biology*, *9*(2), 125-138.

Ingolia, Nicholas T., Liana F. Lareau, and Jonathan S. Weissman. "Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes." *Cell* 147.4 (2011): 789-802.

Jackson, S, et al. Combined enzyme defect of mitochondrial fatty acid oxidation. *J Clin Invest* 1992. **90**:1219-1225.

Jackson, Ruaidhri, et al. "The translation of non-canonical open reading frames controls mucosal

immunity." *Nature* 564.7736 (2018): 434-438.

Jia, G., Yang, C. G., Yang, S., Jian, X., Yi, C., Zhou, Z., & He, C. (2008). Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS letters*, *582*(23-24), 3313-3319.

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., ... & He, C. (2011). N 6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature chemical biology*, *7*(12), 885-887.

Johnsson, P., Lipovich, L., Grandér, D., & Morris, K. V. (2014). Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *1840*(3), 1063-1071.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C. C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K. C., Hallinan, J., Mattick, J., Hume, D. A., Lipovich, L., Batalov, S., Engstrom, P. G., Mizuno, Y., Faghihi, M. A., Sandelin, A., Chalk, A. M., Mottagui-Tabar, S., Liang, Z., Lenhard, B., Wahlestedt, C., Group, R. G. E. R., Genome Science, G., and Consortium, F. (2005) Antisense transcription in the mammalian transcriptome. *Science*:309, 1564-1566.

Kawano Y, Cohen D.E. (2013) Mechanisms of hepatic triglyceride accumulation in non-alcoholic fatty liver disease. *J Gastroenterol* 48:434–441.

Kersten, S., & Stienstra, R. (2017). The role and regulation of the peroxisome proliferator activated receptor alpha in human liver. *Biochimie*, *136*, 75-84.

Keyes, S. R., and D. L. Cinti. "Biochemical properties of cytochrome b5-dependent microsomal fatty acid elongation and identification of products." *Journal of Biological Chemistry* 255.23 (1980): 11357-11364.

Khaitan, D., Dinger, M. E., Mazar, J., Crawford, J., Smith, M. A., Mattick, J. S., & Perera, R. J. (2011). The melanoma-upregulated long noncoding RNA SPRY4-IT1 modulates apoptosis and invasion. *Cancer research*, *71*(11), 3852-3862.

Khoddami, V., & Cairns, B. R. (2013). Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nature biotechnology*, *31*(5), 458-464.

Khurana, E. et al. Role of non-coding sequence variants in cancer. *Nat. Rev. Genet.* **17**, 93–108 (2016)

Kirpich, I. A., Marsano, L. S. & McClain, C. J. (2015). Gut–liver axis, nutrition, and non-alcoholic fatty liver disease. *Clin. Biochem.* **48**, 923–930.

Kmietczyk, V., Riechert, E., Kalinski, L., Boileau, E., Malovrh, E., Malone, B., ... & Völkers, M. (2019). m6A-mRNA methylation regulates cardiac gene expression and cellular growth. *Life science alliance*, *2*(2).

Kohjima M, Enjoji M, Higuchi N, Kato M, Kotoh K, Yoshimoto T, Fujino T, Yada M, Yada R, Harada N, Takayanagi R, Nakamuta M (2007) Re-evaluation of fatty acid metabolism-related gene expression in nonalcoholic fatty liver disease. Int J Mol Med 20:351–358.

Koo, S. H. (2013). Nonalcoholic fatty liver disease: molecular mechanisms for the hepatic steatosis. *Clinical and molecular hepatology*, *19*(3), 210.

Koonen, D. P., Jacobs, R. L., Febbraio, M., Young, M. E., Soltys, C. L. M., Ong, H., ... & Dyck,

J. R. (2007). Increased hepatic CD36 expression contributes to dyslipidemia associated with diet-induced obesity. *diabetes*, *56*(12), 2863-2871.

Kosters, A., Sun, D., Wu, H., Tian, F., Felix, J. C., Li, W., & Karpen, S. J. (2013). Sexually dimorphic genome-wide binding of retinoid X receptor alpha (RXRα) determines male-female differences in the expression of hepatic lipid processing genes in mice. *PloS one*, *8*(8), e71538.

Krogh, Anders, et al. "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes." *Journal of molecular biology* 305.3 (2001): 567-580.

Kühlbrandt, Werner. "Structure and function of mitochondrial membrane protein complexes." *BMC biology* 13.1 (2015): 89.

Lalezari, J.P.; Henry, K.; O'Hearn, M.; Montaner, J.S.; Piliero, P.J.; Trottier, B.; Walmsley, S.; Cohen, C.; Kuritzkes, D.R.; Eron, J.J.; et al. Enfuvirtide, an HIV-1 fusion inhibitor, for drug-resistant HIV infection in North and South America. N. Engl. J. Med. 2003, 348, 2175–2185.

Leader, B., Baca, Q. J., and Golan, D. E. (2008). Protein therapeutics: a summary and pharmacological classification. *Nat. Rev. Drug Discov.* 7, 21–39. doi: 10.1038/nrd2399

Liang G, Yang J, Horton JD, Hammer RE, Goldstein JL, Brown MS (2002) Diminished hepatic response to fasting/refeeding and liver X receptor agonists in mice with selective deficiency of sterol regulatory element-binding protein-1c. *J Biol Chem* 277:9520–9528.

Lambert JE, Ramos-Roman MA, Browning JD, Parks EJ (2014) Increased de novo lipogenesis is a distinct characteristic of individuals with nonalcoholic fatty liver disease.

Gastroenterology 146:726–735.

Lee SS, Pineau T, Drago J, Lee EJ, Owens JW, Kroetz DL, Fernandez-Salguero PM, Westphal H, Gonzalez FJ (1995) Targeted disruption of the alpha isoform of the peroxisome proliferator-activated receptor gene in mice results in abolishment of the pleiotropic effects of peroxisome proliferators. *Mol Cell Biol* 15:3012–3022.

Lee, K. N., Pariza, M. W., & Ntambi, J. M. (1996). Differential expression of hepatic stearoyl-CoA desaturase gene 1 in male and female mice. *Biochimica Et Biophysica Acta (BBA)-Lipids and Lipid Metabolism*, *1304*(2), 85-88.

Lence, T., Akhtar, J., Bayer, M., Schmid, K., Spindler, L., Ho, C. H., ... & Roignant, J. Y. (2016). m 6 A modulates neuronal functions and sex determination in Drosophila. *Nature*, *540*(7632), 242-247.

Li, X., Zhu, P., Ma, S., Song, J., Bai, J., Sun, F., & Yi, C. (2015). Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nature chemical biology*, *11*(8), 592-597.

Li, X., Yang, J., Zhu, Y., Liu, Y., Shi, X. E., & Yang, G. (2016). Mouse maternal high-fat intake dynamically programmed mRNA m6A modifications in adipose and skeletal muscle tissues in offspring. *International journal of molecular sciences*, *17*(8), 1336.

Li, A., Chen, Y. S., Ping, X. L., Yang, X., Xiao, W., Yang, Y., ... & Yang, Y. G. (2017). Cytoplasmic m 6 A reader YTHDF3 promotes mRNA translation. *Cell research*, *27*(3), 444-447.

Libby, P. (2021). The changing landscape of atherosclerosis. *Nature*, *592*(7855), 524-533.

Libby, P. Triglycerides on the rise: should we swap seats on the seesaw? (2015). Eur. Heart J. 36, 774–776.

Lichinchi, G., Zhao, B. S., Wu, Y., Lu, Z., Qin, Y., He, C., & Rana, T. M. (2016). Dynamics of human and viral RNA methylation during Zika virus infection. *Cell host & microbe*, *20*(5), 666-673.

Lin, Michael F., Irwin Jungreis, and Manolis Kellis. "PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions." *Bioinformatics* 27.13 (2011): i275-i282.

Lin, S., Choe, J., Du, P., Triboulet, R., & Gregory, R. I. (2016). The m6A methyltransferase METTL3 promotes translation in human cancer cells. *Molecular cell*, *62*(3), 335-345.

Lippincott-Schwartz, J., & Phair, R. D. (2010). Lipids and cholesterol as regulators of traffic in the endomembrane system. *Annual review of biophysics*, *39*, 559-578.

Lindén, D., Alsterholm, M., Wennbo, H., & Oscarsson, J. (2001). PPARα deficiency increases secretion and serum levels of apolipoprotein B-containing lipoproteins. *Journal of Lipid Research*, *42*(11), 1831-1840.

Link, J. C., & Reue, K. (2017). Genetic basis for sex differences in obesity and lipid metabolism. *Annual review of nutrition*, *37*, 225-245.

Little, Alex George, et al. "Evolution of the nuclear-encoded cytochrome oxidase subunits in vertebrates." *Physiological genomics* 42.1 (2010): 76-84.

Liu, Jinfeng, and Burkhard Rost. "Comparing function and structure between entire proteomes." *Protein Science* 10.10 (2001): 1970-1979.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., ... & He, C. (2014). A METTL3–METTL14 complex mediates mammalian nuclear RNA N 6-adenosine methylation. *Nature chemical biology*, *10*(2),

93-95.

Lomize, Andrei L., et al. "Membranome: a database for proteome-wide analysis of single-pass membrane proteins." *Nucleic acids research* 45.D1 (2017): D250-D255.

Lorbek G, Perse M, Horvat S, Bjorkhem I, Rozman D. 2013. Sex differences in the hepatic cholesterol sensing mechanisms in mice. Molecules **18**:11067–11085.

Ludwig, Bernd, et al. "Cytochrome c oxidase and the regulation of oxidative phosphorylation." *Chembiochem* 2.6 (2001): 392-403.

Ma, Y., Belyaeva, O. V., Brown, P. M., Fujita, K., Valles, K., Karki, S., ... & Rotman, Y. (2019). 17-beta hydroxysteroid dehydrogenase 13 is a hepatic retinol dehydrogenase associated with histological features of nonalcoholic fatty liver disease. *Hepatology*, *69*(4), 1504-1519.

Ma, H., Wang, X., Cai, J., Dai, Q., Natchiar, S. K., Lv, R., ... & He, C. (2019). N 6-Methyladenosine methyltransferase ZCCHC4 mediates ribosomal RNA methylation. *Nature chemical biology*, *15*(1), 88-94.

Makarewich, Catherine A. "The hidden world of membrane microproteins." Experimental Cell Research 388.2 (2020): 111853.

Makarewich, Catherine A., et al. "MOXI is a mitochondrial micropeptide that enhances fatty acid β-oxidation." *Cell reports* 23.13 (2018): 3701-3709.

Mao J, DeMayo FJ, Li H, Abu-Elheiga L, Gu Z, Shaikenov TE, Kordari P, Chirala SS, Heird WC, Wakil SJ (2006) Liver-specific deletion of acetyl-CoA carboxylase 1 reduces hepatic

triglyceride accumulation without affecting glucose homeostasis. Proc Natl Acad Sci USA 103:8552–8557.

Mathiyalagan, P., Liang, Y., Adamiak, M., Sassi, Y., Agarwal, N., Ishikawa, K., ... & Sahoo, S. (2017). Modulation of N6-Methyladenosine (m6A) as a novel mechanism of human CD34+ stem cell-derived exosomes in cardiac repair. *Circulation*, *136*(suppl_1), A18434-A18434.

Mattick, J. S. (1994). Introns: evolution and function. *Current opinion in genetics & development*, *4*(6), 823-831.

Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A. V., Patil, D. P., ... & Jaffrey, S. R. (2017). Reversible methylation of m 6 A m in the 5′ cap controls mRNA stability. *Nature*, *541*(7637), 371-375.

Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nature reviews genetics*, *10*(3), 155-159.

Mercer, T. R., Qureshi, I. A., Gokhan, S., Dinger, M. E., Li, G., Mattick, J. S., & Mehler, M. F. (2010). Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation. *BMC neuroscience*, *11*(1), 1-15.

Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., & Jaffrey, S. R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell*, *149*(7), 1635-1646.

Miquilena-Colina ME, Lima-Cabello E, Sanchez-Campos S, Garcia-Mediavilla MV, Fernandez-Bermejo M, Lozano-Rodriguez T, Vargas-Castrillon J, Buque X, Ochoa B, Aspichueta P, Gonzalez-Gallego J, Garcia-Monzon C (2011) Hepatic fatty acid translocase CD36 upregulation is associated with insulin resistance, hyperinsulinaemia and increased

steatosis in non-alcoholic steatohepatitis and chronic hepatitis C. Gut 60:1394–1402.

Mo, X. B., Lei, S. F., Zhang, Y. H., & Zhang, H. (2019). Examination of the associations between m 6 A-associated single-nucleotide polymorphisms and blood pressure. *Hypertension Research*, *42*(10), 1582-1589.

Mohammad, F., Pandey, R. R., Nagano, T., Chakalova, L., Mondal, T., Fraser, P., & Kanduri, C. (2008). Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. *Molecular and cellular biology*, *28*(11), 3713-3728.

Monetti M, Levin MC, Watt MJ, Sajan MP, Marmor S, Hubbard BK, Stevens RD, Bain JR, Newgard CB, Farese RV Sr, Hevener AL, Farese RV Jr (2007) Dissociation of hepatic steatosis and insulin resistance in mice overexpressing DGAT in the liver. *Cell Metab* 6:69–78.

Mori, T.; O'Keefe, B.R.; Sowder, R.C., 2nd; Bringans, S.; Gardella, R.; Berg, S.; Cochran, P.; Turpin, J.A.; Buckheit, R.W., Jr.; McMahon, J.B.; et al. Isolation and characterization of griffithsin, a novel HIV-inactivating protein, from the red alga Griffithsia sp. J. Biol. Chem. 2005, 280, 9345–9353

Morris, K. V., & Mattick, J. S. (2014). The rise of regulatory RNA. *Nature Reviews Genetics*, *15*(6), 423-437.

Mourtada-Maarabouni, M., Pickard, M. R., Hedge, V. L., Farzaneh, F., & Williams, G. T. (2009). GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene*, *28*(2), 195-208.

Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., ... &

Turner, M. B. (2015). Heart disease and stroke statistics—2015 update: a report from the American Heart Association. *circulation*, *131*(4), e29-e322.

Musunuru, K. & Kathiresan, S. Surprises from genetic analyses of lipid risk factors for atherosclerosis. (2016). A window into novel aspects of lipids and atherosclerosis emerging from contemporary human genetics. *Circ. Res.* 118, 579–585

Musso, G., Gambino, R., Tabibian, J. H., Ekstedt, M., Kechagias, S., Hamaguchi, M., ... & Cassader, M. (2014). Association of non-alcoholic fatty liver disease with chronic kidney disease: a systematic review and meta-analysis. *PLoS medicine*, *11*(7), e1001680.

Nahkuri, S., Taft, R. J., & Mattick, J. S. (2009). Nucleosomes are preferentially positioned at exons in somatic and sperm cells. *Cell cycle*, *8*(20), 3420-3424.

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., ... & Kaessmann, H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*, *505*(7485), 635-640.

Nordestgaard, B. G., Chapman, M. J., Humphries, S. E., Ginsberg, H. N., Masana, L., Descamps, O. S., ... & European Atherosclerosis Society Consensus Panel. (2013). Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society. *European heart journal*, *34*(45), 3478-3490.

Nordestgaard, B. G., & Varbo, A. (2014). Triglycerides and cardiovascular disease. *The Lancet*, *384*(9943), 626-635.

Ørom, U. A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., ... & Shiekhattar, R. (2010). Long noncoding RNAs with enhancer-like function in human cells. *Cell*, *143*(1), 46-58.

Pang, K. C., Frith, M. C., & Mattick, J. S. (2006). Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in genetics*, *22*(1), 1-5.

Pang, K. C., Dinger, M. E., Mercer, T. R., Malquori, L., Grimmond, S. M., Chen, W., & Mattick, J. S. (2009). Genome-wide identification of long noncoding RNAs in CD8+ T cells. *The Journal of Immunology*, *182*(12), 7738-7748.

Parthasarathy, G., Revelo, X., & Malhi, H. (2020). Pathogenesis of nonalcoholic steatohepatitis: an overview. *Hepatology communications*, *4*(4), 478-492.

Patil, D. P., Chen, C. K., Pickering, B. F., Chow, A., Jackson, C., Guttman, M., & Jaffrey, S. R. (2016). m 6 A RNA methylation promotes XIST-mediated transcriptional repression. *Nature*, *537*(7620), 369-373.

Pawlak, M., Lefebvre, P., & Staels, B. (2015). Molecular mechanism of PPARα action and its impact on lipid metabolism, inflammation and fibrosis in non-alcoholic fatty liver disease. *Journal of hepatology*, *62*(3), 720-733.

Perry, R. P., Kelley, D. E., Friderici, K., & Rottman, F. (1975). The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5′ terminus. *Cell*, *4*(4), 387-394.

Perry RJ, Samuel VT, Petersen KF, Shulman GI (2014) The role of hepatic lipids in hepatic insulin resistance and type 2 diabetes. *Nature* 510:84–91.

Ping, X. L., Sun, B. F., Wang, L., Xiao, W., Yang, X., Wang, W. J., ... & Yang, Y. G. (2014). Mammalian

WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell research*, *24*(2), 177-189.Postic, C., & Girard, J. (2008). Contribution of de novo fatty acid synthesis to hepatic steatosis and insulin resistance: lessons from genetically engineered mice. *The Journal of clinical investigation*, *118*(3), 829-838.

Pogozheva, Irina D., and Andrei L. Lomize. "Evolution and adaptation of single-pass transmembrane proteins." *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1860.2 (2018): 364-377.

Poynard T, Bedossa P, Opolon P. 1997. Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. Lancet **349**:825–832.

Pu, Jing, et al. "Development of Protein-and Peptide-Based HIV Entry Inhibitors Targeting gp120 or gp41." *Viruses* 11.8 (2019): 705.

Radvanyi, L. G. (2018). Targeting the cancer mutanome of breast cancer. *Nat. Med.* 24, 703–704. doi: 10.1038/s41591-018-0065-z

Rao MS, Reddy JK (2001) Peroxisomal beta-oxidation and steatohepatitis. *Semin Liver Dis* 21:43–55.

Razooky, Brandon S., et al. "Viral infection identifies micropeptides differentially regulated in smORF-containing lncRNAs." *Genes* 8.8 (2017): 206.

Reddy JK, Hashimoto T (2001) Peroxisomal beta-oxidation and peroxisome proliferator-activated receptor alpha: an adaptive metabolic system. *Annu Rev Nutr* 21:193–230.

Reddy, V. V., D. A. V. I. D. Kupfer, and E. L. I. A. H. U. Caspi. "Mechanism of C-5 double bond

introduction in the biosynthesis of cholesterol by rat liver microsomes." *Journal of Biological Chemistry* 252.9 (1977): 2797-2801.

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A.,

Farnham, P. J., Segal, E., and Chang, H. Y. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cel:l*129, 1311-1323 223.

Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nature Genetics* **40:** 1461-1465.

Roundtree, I. A., Luo, G. Z., Zhang, Z., Wang, X., Zhou, T., Cui, Y., ... & He, C. (2017). YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. *Elife*, *6*, e31311.

Saghatelian, Alan, and Juan Pablo Couso. "Discovery and characterization of smORF-encoded bioactive polypeptides." *Nature chemical biology* 11.12 (2015): 909.

Sallam, T., Jones, M., Thomas, B. J., Wu, X., Gilliland, T., Qian, K., ... & Tontonoz, P. (2018). Transcriptional regulation of macrophage cholesterol efflux and atherogenesis by a long noncoding RNA. *Nature medicine*, *24*(3), 304-312.

Salisbury, D. A., Casero, D., Zhang, Z., Wang, D., Kim, J., Wu, X., ... & Sallam, T. (2021). Transcriptional regulation of N6-methyladenosine orchestrates sex-dimorphic metabolic traits. *Nature Metabolism*, *3*(7), 940-953.

Schibler, U., & Perry, R. P. (1977). The 5′-termini of heterogeneous nuclear RNA: a comparison among molecules of different sizes and ages. *Nucleic acids research*, *4*(12), 4133-4150.

Schreiner, P. J., Jacobs Jr, D. R., Wong, N. D., & Kiefe, C. I. (2016). Twenty-five year secular trends in lipids and modifiable risk factors in a population-based biracial cohort: the Coronary Artery Risk Development in Young Adults (CARDIA) study, 1985–2011. *Journal of the American Heart Association*, *5*(7), e003384.

Schwartz, S., Bernstein, D. A., Mumbach, M. R., Jovanovic, M., Herbst, R. H., León-Ricardo, B. X., ... & Regev, A. (2014). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell*, *159*(1), 148-162.

Shi, H., Wang, X., Lu, Z., Zhao, B. S., Ma, H., Hsu, P. J., ... & He, C. (2017). YTHDF3 facilitates translation and decay of N 6-methyladenosine-modified RNA. *Cell research*, *27*(3), 315-328.

Shimano H, Horton JD, Shimomura I, Hammer RE, Brown MS, Goldstein JL (1997) Isoform 1c of sterol regulatory element binding protein is less active than isoform 1a in livers of transgenic mice and in cultured cells. *J Clin Investig* 99:846–854.

Shindo N, Fujisawa T, Sugimoto K, Nojima K, Oze-Fukai A, Yoshikawa Y, Wang X, Yasuda O, Ikegami H, Rakugi H (2010) Involvement of microsomal triglyceride transfer protein in nonalcoholic steatohepatitis in novel spontaneous mouse model. J Hepatol 52:903–912.

Śledź, P., & Jinek, M. (2016). Structural insights into the molecular mechanism of the m6A writer complex. *elife*, *5*, e18434.

Smith, M. A., Gesell, T., Stadler, P. F., & Mattick, J. S. (2013). Widespread purifying selection on RNA structure in mammals. *Nucleic acids research*, *41*(17), 8220-8236.

Squires, J. E., & Preiss, T. (2010). Function and detection of 5-methylcytosine in eukaryotic

RNA. *Epigenomics*, *2*(5), 709-715.

Stein, Colleen S., et al. "Mitoregulin: a lncRNA-encoded microprotein that supports mitochondrial supercomplexes and respiratory efficiency." *Cell reports* 23.13 (2018): 3710-3720.

Sunwoo, H., Dinger, M. E., Wilusz, J. E., Amaral, P. P., Mattick, J. S., & Spector, D. L. (2009). MEN ε/β nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome research*, *19*(3), 347-359.

Takeda, K., Ichijo, H., Fujii, M., Mochida, Y., Saitoh, M., Nishitoh, H., ... & Miyazono, K. (1998). Identification of a novel bone morphogenetic protein-responsive gene that may function as a noncoding RNA. *Journal of Biological Chemistry*, *273*(27), 17079-17085.

Tanoli T, Yue P, Yablonskiy D, Schonfeld G (2004) Fatty liver in familial hypobetalipoproteinemia: roles of the APOB defects, intra-abdominal adipose tissue, and insulin sensitivity. *J Lipid Res* 45:941–947.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**:562–578.

Uchida, Y, Izai, K, Orii, T, Hashimoto, T. Novel fatty acid β-oxidation enzymes in rat liver mitochondria. II. Purification and properties of enoyl-coenzyme A (CoA) hydratase/3-hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase trifunctional protein. *J Biol Chem* 1992. **267**:1034-1041.

Varbo, A., Benn, M., Tybjærg-Hansen, A., & Nordestgaard, B. G. (2013). Elevated remnant cholesterol causes both low-grade inflammation and ischemic heart disease, whereas

elevated low-density lipoprotein cholesterol causes ischemic heart disease without inflammation. *Circulation*, *128*(12), 1298-1309.

Vita, C.; Drakopoulou, E.; Vizzavona, J.; Rochette, S.; Martin, L.; Menez, A.; Roumestand, C.; Yang, Y.S.; Ylisastigui, L.; Benjouad, A.; et al. Rational engineering of a miniprotein that reproduces the core of the CD4 site interacting with HIV-1 envelope glycoprotein. Proc. Natl. Acad. Sci. USA 1999, 96, 13091–13096

Venters, B. J., & Pugh, B. F. (2013). Genomic organization of human transcription initiation complexes. *Nature*, *502*(7469), 53-58.

Wapinski, O., & Chang, H. Y. (2011). Long noncoding RNAs and human disease. *Trends in cell biology*, *21*(6), 354-361.

Wattacheril J, Sanyal AJ (2016) Lean NAFLD: an underrecognized outlier. Curr Hepatol Rep 15:134–139.

Wang, X., Feng, J., Xue, Y., Guan, Z., Zhang, D., Liu, Z., ... & Yin, P. (2016). Structural basis of N 6-adenosine methylation by the METTL3–METTL14 complex. *Nature*, *534*(7608), 575-578.

Wang, S., Sun, C., Li, J., Zhang, E., Ma, Z., Xu, W., ... & Yin, R. (2017). Roles of RNA methylation by means of N6-methyladenosine (m6A) in human cancers. *Cancer letters*, *408*, 112-120.

Wauthier V, Sugathan A, Meyer RD, Dombkowski AA, Waxman DJ. 2010. Intrinsic sex differences in the early growth hormone responsiveness of sex-specific genes in mouse liver. *Mol. Endocrinol.* **24**:667–678.

Wauthier V, Waxman DJ. 2008. Sex-specific early growth hormone response genes in rat

liver. *Mol. Endocrinol*. **22**:1962–1974

Waxman DJ, Pampori NA, Ram PA, Agrawal AK, Shapiro BH. 1991. Interpulse interval in circulating growth hormone patterns regulates sexually dimorphic expression of hepatic cytochrome P450. Proc. Natl. Acad. Sci. U. S. A. **88**:6868–6872.

Wei, J., Liu, F., Lu, Z., Fei, Q., Ai, Y., He, P. C., ... & He, C. (2018). Differential m6A, m6Am, and m1A demethylation mediated by FTO in the cell nucleus and cytoplasm. *Molecular cell*, *71*(6), 973-985.

Wen, J., Lv, R., Ma, H., Shen, H., He, C., Wang, J., ... & Diao, J. (2018). Zc3h13 regulates nuclear RNA m6A methylation and mouse embryonic stem cell self-renewal. *Molecular cell*, *69*(6), 1028-1038.

Westerbacka J, Kolak M, Kiviluoto T, Arkkila P, Siren J, Hamsten A, Fisher RM, Yki-Jarvinen H (2007) Genes involved in fatty acid partitioning and binding, lipolysis, monocyte/macrophage recruitment, and inflammation are overexpressed in the human fatty liver of insulin-resistant subjects. *Diabetes* 56:2759–2765.

Worch, Remigiusz, et al. "Focus on composition and interaction potential of single-pass transmembrane domains." *Proteomics* 10.23 (2010): 4196-4208.

World Health Organization. Cardiovascular diseases (CVDs) Fact Sheet. 2017.

Willingham, A. T., Orth, A. P., Batalov, S., Peters, E. C., Wen, B. G., Aza-Blanc, P., ... & Schultz, P. G. (2005). A Strategy for Probing the Function of Noncoding RNAs Finds a Repressor of NFAT.

Wilson CG, Tran JL, Erion DM, Vera NB, Febbraio M, Weiss EJ (2016) Hepatocyte-specific disruption of CD36 attenuates fatty liver and improves insulin sensitivity in HFD-fed

mice. Endocrinology 157:570–585.

Wilusz, J. E., Sunwoo, H., and Spector, D. L. (2009) Long noncoding RNAs: functional surprises from the RNA world. Genes & development23, 1494-1504 219.

Wong, R. J., Aguilar, M., Cheung, R., Perumpail, R. B., Harrison, S. A., Younossi, Z. M., & Ahmed, A. (2015). Nonalcoholic steatohepatitis is the second leading etiology of liver disease among adults awaiting liver transplantation in the United States. *Gastroenterology*, *148*(3), 547-555.

Wu, G., Xiao, M., Yang, C., & Yu, Y. T. (2011). U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. *The EMBO journal*, *30*(1), 79-89.

Wu, W., Feng, J., Jiang, D., Zhou, X., Jiang, Q., Cai, M., ... & Wang, Y. (2017). AMPK regulates lipid accumulation in skeletal muscle cells through FTO-dependent demethylation of N 6-methyladenosine. *Scientific reports*, *7*(1), 1-13.

Wu, R., Li, A., Sun, B., Sun, J. G., Zhang, J., Zhang, T., ... & Yuan, Z. (2019). A novel m 6 A reader Prrc2a controls oligodendroglial specification and myelination. *Cell research*, *29*(1), 23-41.

Vaishya, R., Khurana, V., Patel, S., and Mitra, A. K. (2015). Long-term delivery of protein therapeutics. *Expert Opin. Drug Del.* 12, 415–440. doi: 10.1517/17425247.2015.961420

Vu, L. P., Pickering, B. F., Cheng, Y., Zaccara, S., Nguyen, D., Minuesa, G., ... & Kharas, M. G. (2017). The N 6-methyladenosine (m 6 A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nature medicine*, *23*(11), 1369-1376.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y. S., Hao, Y. J., Sun, B. F., ... & Yang, Y. G. (2016). Nuclear

m6A reader YTHDC1 regulates mRNA splicing. *Molecular cell*, *61*(4), 507-519.

Xuan, J. J., Sun, W. J., Lin, P. H., Zhou, K. R., Liu, S., Zheng, L. L., ... & Yang, J. H. (2018). RMBase v2. 0: deciphering the map of RNA modifications from epitranscriptome sequencing data. *Nucleic acids research*, *46*(D1), D327-D334.

Yamaguchi K, Yang L, McCall S, Huang J, Yu XX, Pandey SK, Bhanot S, Monia BP, Li YX, Diehl AM (2007) Inhibiting triglyceride synthesis improves hepatic steatosis but exacerbates liver damage and fibrosis in obese mice with nonalcoholic steatohepatitis. Hepatology 45:1366–1374.

Yang, X., Yang, Y., Sun, B. F., Chen, Y. S., Xu, J. W., Lai, W. Y., ... & Yang, Y. G. (2017). 5-methylcytosine promotes mRNA export—NSUN2 as the methyltransferase and ALYREF as an m 5 C reader. *Cell research*, *27*(5), 606-625.

Yokoyama Y, Nimura Y, Nagino M, Bland KI, Chaudry IH. 2005. Current understanding of gender dimorphism in hepatic pathophysiology. J Surg Res **128**:147–156.

Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M (2016) Global epidemiology of nonalcoholic fatty liver disease—meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology 64:73–84

Zhang, W., Patil, S., Chauhan, B., Guo, S., Powell, D. R., Le, J., ... & Unterman, T. G. (2006). FoxO1 regulates multiple metabolic pathways in the liver: effects on gluconeogenic, glycolytic, and lipogenic gene expression. *Journal of Biological Chemistry*, *281*(15), 10105-10117.

Zhang, M., Zhang, Y., Ma, J., Guo, F., Cao, Q., Zhang, Y., ... & Zhao, R. (2015). The demethylase activity

of FTO (fat mass and obesity associated protein) is required for preadipocyte differentiation. *PLoS One*, *10*(7), e0133788.

Zhang, C., Samanta, D., Lu, H., Bullen, J. W., Zhang, H., Chen, I., ... & Semenza, G. L. (2016). Hypoxia induces the breast cancer stem cell phenotype by HIF-dependent and ALKBH5-mediated m6A-demethylation of NANOG mRNA. *Proceedings of the National Academy of Sciences*, *113*(14), E2047-E2056.

Zhang C, Liu P. The New Face of the Lipid Droplet: Lipid Droplet Proteins. Proteomics. 2019;19(10):e1700223.

Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., ... & He, C. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Molecular cell*, *49*(1), 18-29.

Zhu, L., Baker, S. S., Liu, W., Tao, M. H., Patel, R., Nowak, N. J., & Baker, R. D. (2011). Lipid in the livers of adolescents with nonalcoholic steatohepatitis: combined effects of pathways on steatosis. *Metabolism*, *60*(7), 1001-1011.

Zhu, Mengmeng, and Michael Gribskov. "MiPepid: MicroPeptide identification tool using machine learning." *BMC bioinformatics* 20.1 (2019): 559.

Zickermann, Volker, et al. "Small single transmembrane domain (STMD) proteins organize the hydrophobic subunits of large membrane protein complexes." FEBS letters 584.12 (2010): 2516-2525.

Chapter 2: Transcriptional regulation of N^6-methyladenosine orchestrates sex-dimorphic metabolic traits.

Title: Transcriptional regulation of *N*^{6}-methyladenosine orchestrates sex-dimorphic metabolic traits

Authors: David A Salisbury [1,2,3], David Casero [4], Zhengyi Zhang [1,2], Dan Wang [1,2], Jason Kim [1,2], Xiaohui Wu [1,2], Laurent Vergnes [5], Aashiq Mirza [6], Paola Leon-Mimila[1], Kevin J. Williams [7], Adriana Huertas-Velzquez [1], Samie Jaffrey [6], Karen Reue [2,5], Jianjun Chen [8] and Tamer Sallam [1,2,3*]

Affiliations:

[1] Division of Cardiology, Department of Medicine, University of California, Los Angeles, CA

[2] Molecular Biology Institute, University of California, Los Angeles, CA

[3] Molecular Biology Interdepartmental Program, University of California, Los Angeles, CA

[4] Inflammatory Bowel & Immunobiology Research Institute, Cedars-Sinai, Los Angeles, CA

[5] Department of Human Genetics, University of California, Los Angeles, CA

[6] Department of Pharmacology, Weill Cornell Medicine, Cornell University, New York, NY

[7] Department of Biological Chemistry, University of California, Los Angeles, CA

[8] Department of Systems Biology, The Beckman Research Institute of City of Hope, Duarte, CA

*Corresponding author, Contact Information:
Division of Cardiology, Department of Medicine
David Geffen School of Medicine at UCLA
650 Charles E. Young South Dr. CHS A2-327
Los Angeles, CA 90095-1679
Email: tsallam@mednet.ucla.edu

Abstract

Males and females exhibit striking differences in the prevalence of metabolic traits including hepatic steatosis, a key driver of cardiometabolic morbidity and mortality. Here, we investigate the role of RNA methylation, a widespread regulatory mechanism of transcript turnover, in metabolic sex-differences. We show that presence of the RNA modification N^6-methyladenosine (m^6A) triages lipogenic transcripts for degradation and guards against hepatic triglyceride accumulation. In male but not female mice, this protective checkpoint stalls under lipid-rich or refed conditions, congruent with an on-demand need for fatty acid biosynthesis and lipid storage. Loss of m^6A control in male livers increases hepatic triglyceride stores leading to a more "feminized" hepatic lipid composition. Crucially, liver-specific deletion of the m^6A complex protein *Mettl14* from male and female mice significantly diminishes sex-specific differences in steatosis and sex-biased gene expression. We further surmise that m^6A installing machinery shows discordant regulation between the sexes and is subject to transcriptional control by the sex-responsive BCL6-STAT5 axis in response to dietary conditions. These data show that m^6A is essential for precise and synchronized control of lipogenic enzyme activity and provide insights into the molecular basis for the existence sex-specific differences in hepatic lipid traits.

Introduction

The liver is the central tissue orchestrating rapid adaptations in gene regulation to maintain metabolic homeostasis. Control of mRNA biogenesis is a key target of multiple regulatory pathways in liver and its disruption can result in metabolic disturbances including fatty liver disease. For example, chronic high fat/high carbohydrate feeding stimulates the master lipogenic transcription factor SREBP1C, transcriptional induction of lipogenic genes, and increased hepatic triglyceride stores [1]. Reciprocally, under fasting conditions, activity of SREBP1C and its downstream targets *Scd1*, *Dgat2* and *Fasn* are reduced [1,2]. Fasting is characterized by low insulin, almost absent nuclear SREBP1C and low demand for fatty acid biosynthesis[3]. Mysteriously, despite lower level of lipogenic mRNA and protein compared with fed state in liver, the absolute expression of lipogenic genes during fast is remarkably high (top 0.1-2% by percentile rank). Others have reported sex-specific discordance in lipogenic mRNAs, their protein levels and/or triglyceride content [4,5]. These observations hint the existence of a post-transcriptional axis contributing to lipogenesis; however, the mechanisms underlying synchronized control of lipogenic mRNA metabolism and protein function remain poorly understood.

Recent evidence suggests that selective N^6-methyladenosine (m^6A) methylation of RNA can affect the stability, translation and/or localization of mRNA[6-8]. The installation of m^6A occurs at predictable motifs through the collaborative activities of a 'writer ' complex made up of the methyltransferases METTL14 and METTL3 [9]. Although a group of elegant studies have shown that m^6A plays a critical role in diverse biologic processes [7,8,10], understating the molecular basis of temporospatial variation in m^6A levels and how the activity of m^6A installing machinery may be fine-tuned in response to environmental cues is still emerging. Here, we demonstrate that m^6A modifications potently alter the fate of lipogenic mRNA and hepatic triglyceride stores. In

addition, we show that m^{6}A modifications are dynamically altered in response to dietary conditions and provide mechanistic evidence that m^{6}A installing machinery is subject to tight transcriptional control by sex-dimorphic dietary regulators to sustain metabolic control. These data identify a new pathway for lipid degradation and, at least in part, explain the molecular basis for the existence sex-specific differences in hepatic triglyceride composition.

Materials and Methods:

Reagents, plasmids, and cell transfection

shRNAs targeting *Mettl14* were originally obtained from Horizon Discovery and cloned into pLKO.1 vector. The following shRNA sequences were used: 5′-GCT AAA GGA TGA GTT AAT-3′ (human, TRCN0000015935). The resulting plasmids were transiently transfected using Fugene HD transfection reagent (Promega) or electroporated using the Neon transfection system (Invitrogen) and then tested for their ability to knockdown *Mettl14* at 48 hrs post-transfection. The following parameters were used for electroporation of cells: pulse voltage = 1450V, pulse width = 10ms, pulse number = 3. AAV.TBG.PI.Cre.rBG (Addgene) was used for acute deletion of *Mettl14* and *BCL6* in mouse liver, AAV.TBG.PI.eGFP.WPRE.bGH (Addgene) or AAV.TBG.PI.Null.bGH (Addgene) were used as controls. For overexpression of m^{6}A writers, the following viruses were generated by VectorBuilder: AAV.TBG.mMettl3[NM_019721.2].WPRE, AAV.TBG.mMettl14[NM_201638.2].WPRE, and AAV.TBG.mWtap[NM_001113533.1].WPRE. and AAV.TBG.PI.Null.bGH (Addgene) was used as the control. For targeted m^{6}A-modification of cellular mRNAs, plasmids encoding nuclease-dead *Cas13b* and truncated *Mettl3* fused to a nuclear localization sequence (pCMV-dCas13-M3nls) and gRNA expression plasmid (pU6-PspCas13b-gRNA) were obtained from Addgene (155366 and 155368; respectively). For complete list of antibodies and primers please see Supplementary Table 1 and Supplementary Table 2.

Animals and diets

All animals used in this study were in C57BL/6 background unless otherwise noted. Our study used both male and female mice. Mice were housed in a temperature-controlled room under a 12-h light/12-hr dark cycle and pathogen-free conditions. Mice were fed standard chow diet, Western diet (40% Kcal with 0.15% cholesterol and moderate sucrose, or high fat diet (60%Kcal with moderate sucrose and no cholesterol) (Research Diets). Duration of diet was 4 weeks for most experiments but for male and female comparison WD RNA-seq experiments duration was shorter for 10 days. For m^6A writer overexpression experiments, mice were fed FPC-NASH diet (Envigo) and provided water supplemented with glucose and fructose (23.1g Fructose + 18.9g glucose in 1 L water, Sigma) for 8 weeks as previously described [42]. For RNA-seq, m^6A-seq, and m^6A writer *in vivo* overexpression experiments, age-matched WT mice were purchased from Jackson laboratories. *Mettl14*$^{f/f}$ mice were generated as previously described [13]. Age-matched male and female *bcl6* $^{f/f}$ mice were purchased from Jackson Laboratories. For chronic *Mettl14* deficiency experiments, Albumin-cre transgenics were purchased from Jackson Laboratories and crossed with *Mettl14*$^{f/f}$ mice. To generate liver-specific knockout mice, we treated *Mettl14*$^{f/f}$ and *bcl6* $^{f/f}$ mice with adeno-associated virus with TBG promoter (AAV8.TBG.Cre) or control (AAV8.TBG.GFP or AAV8.TBG.Null) purchased from Penn Vector Core. AAV was administered via intraperitoneal injection at a dose of 5×10^{11} GC per mouse. Four Core Genotypes (FCG) C57BL/6J mice on *Apoe* null background were from a colony maintained at UCLA. XX female mice were mated with XY–(Sry+) male mice to generate XX, XX(Sry+), XY–, and XY– (Sry+) offspring as previously described [28]. To generate mice overexpressing m^6A writers in the liver, we treated age-matched mice purchased from Jackson Laboratories with 2.5×10^{11} GC each of AAV.TBG.mMettl3, AAV.TBG.mMettl14, and AAV.TBG.mWtap or 7.5×10^{11} GC of AAV8.TBG.Null per mouse. NAFLD diet composed of WD supplemented with glucose/fructose water. We picked a short term timepoint of 4 weeks to

better decipher direct effects of RNA modifications. Mice were euthanized 4 weeks after AAV injection and fasted for 4-6 hr prior to euthanization unless otherwise noted. All mouse experiments were approved by the UCLA institutional Animal Care and Research Advisory Committee.

Cell culture

Mouse primary hepatocytes were isolated as previously described [43] [44] and cultured in William's E medium with 5% BSA. AML cells were originally obtained from ATCC and cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. AAV studies were performed as previously described [45]. SCAP$^{-/-}$ HEK293T cells were kindly provided by Dr. Peter Espenshade (Johns Hopkins University School of Medicine, Baltimore, MD, USA), and were cultured in the above-mentioned culture medium additionally supplemented with 5 μg/ml cholesterol, 1 mM sodium mevalonate, and 20 μM oleic acid as described previously [46]. All cells were incubated at 37°C in a humidified incubator containing 5% CO_2.

Gene expression and immunoblot analysis

For gene expression analysis, total RNA was isolated using TRIzol reagent (Invitrogen) and reverse transcribed using a homemade RT, as we previously described [47]. cDNA was quantified by real-time PCR using SYBR Green Master Mix (Diagenode) on BioRad Real-time PCR instrument. Gene expression levels were determined by using a standard curve. Each gene was normalized to the housekeeping gene 36B4, β-actin, or GAPDH. Total cellular protein was isolated from approximately 300mg of frozen liver tissue or cells of a 6-well plate using RIPA lysis buffer with protease inhibitor cocktail (Roche). Protein concentration was determined by BCA assay (Invitrogen), diluted in Nupage loading dye (Invitrogen), heated at 80 °C for 5 min,

and run on 4–12% NuPAGE Bis-Tris Gel (Invitrogen) at 200V for 30min. Proteins were transferred to 0.22μM nitrocellulose membranes (Invitrogen) for 1 hr at 100V on ice and blocked with 5% milk in TBS-T to quench nonspecific protein binding. Samples were blotted with the indicated primary antibodies and detected using Alexa-488, Alexa-647, or IgG HRP-conjugated secondary antibodies. Uncropped gel images with molecular weight markers are provided in supplemental data.

MS-Lipid quantification

Liver tissues were frozen in liquid nitrogen and stored at -80°C, blood was collected by retro-orbital bleeding and plasma further separated by centrifugation. Lipidomics analysis was performed as previously described [48]. In brief, 50–100 mg of frozen liver were homogenized in the Omni Bead Ruptor Elite with 2 mL homogenizer tube system (Omni, 19-628D). Samples are homogenized in cold phosphate-buffered saline (PBS) for 3 cycles of 10 s each at 5 m/s with a 10 s dwell time between cycles. A total of 3–6 mg of homogenized material were applied to a modified Bligh and Dyer extraction. Prior to biphasic extraction, a 13 lipid class Lipidyzer Internal Standard Mix is added to each sample (AB Sciex, 5040156). Following two successive extractions, pooled organic layers are dried down in a Genevac EZ-2 Elite. Lipid samples are resuspended in 1:1 methanol/dichloromethane with 10 mM ammonium acetate and transferred to robovials (Thermo 10800107) for analysis.Samples are analyzed on the Sciex Lipidyzer Platform for targeted quantitative measurement of 1100 lipid species across 13 lipid sub-classes. Differential Mobility Device on Lipidyzer is tuned with SelexION tuning kit (Sciex 5040141). Instrument settings, tuning settings, and MRM list available upon request. Data analysis performed on Lipidyzer software. Quantitative values are normalized to milligrams of material used.

Lipid analysis

For tissue lipid measured by Folch extraction see previous description 44. Tissue and serum triglycerides and cholesterol were quantified using commercially available enzymatic kits (Wako L-type TG M, Wako cholesterol E) according to manufacturer's protocol. Hepatic cholesterol and triglyceride content were normalized to liver weight. Mice were fasted for 4 or 6 hours priorto blood and tissue collection unless otherwise noted. Oil Red O histochemistry staining were performed using frozen sections from mouse liver. Briefly, frozen sections were fixed in 10% neutral formalin for 10 min, followed by treatment in 60% isopropanol for 5 s and then staining in Oil Red O working solution (O-0625, Sigma Aldrich) for 15 min. This procedure was followed by washing with 60% isopropanol for 5 s and then water 1 min. Finally, sections were contained with Mayer's Hematoxylin for 3 min. Immunohistochemical staining of paraffin-embedded livers was performed by the UCLA Translational Pathology Core Laboratory (TPCL).

RNA-sequencing

Approximately 500µg of total RNA was isolated from frozen liver tissue of male and female mice fed chow, western, or high fat diet for 2 or 4 weeks (*n*=5 per group) using TRIzol reagent. mRNA was purified from each sample using Oligo(dT) Dynabeads (Invitrogen) and subjected to first strand cDNA synthesis and library prep using TruSeq stranded mRNA library prep kit (Illumina)and sequenced on an Illumina HiSeq 3000 with single-end 50-bp read length. Sequencing was performed by UCLA Technology Center for Genomics and Bioinformatics. The STAR ultrafast universal RNA-seq aligner v2.6d [49] was used to align the reads to a genome index that included both the genome sequence (GRCm38 mouse primary assembly) and the exon/intron structure of known mouse gene models (Gencode M18 genome annotation). Alignment files were used to generate strand-specific, gene-level count summaries with STAR's built-in gene counter. Independent filtering was applied as follows: genes with less than an

average of 1 count per sample, count outliers or low mappability (<50bp) were filtered out for downstream analysis [50,51]. Only protein-coding genes in the Gencode annotation were employed for downstream analyses. For the diet study, an additional in-house batch-correction step was implemented to correct for technical variability between libraries prepared at different times during the course of the project. Briefly, we employed one batch of samples comprising 5 independent replicates of the same condition to model and impute counts in other batches keeping the mean and dispersion distributions similar across batches. Expression estimates were computed in units of fragments per kilobase of mappable length and million counts (FPKMs). Count-based normalized and variance-stabilized data was used for all ordination, differential and clustering analyses, and all figures unless otherwise noted. Principal component analysis was performed with the function prcomp in R (https://www.R-project.org/) using standardize data as input. Samples hierarchical clustering was performed in Matlab (MATLAB, version release 2017a, The MathWorks, Inc, RRID:SCR_001622) using the pdist and seqlinkage functions. Differential expression analyses was performed with DESeq2 (Bioconductor, v3.7, RRID:SCR_015687)[51]. Several types of differential tests were performed. We used pair-wise contrasts for comparisons between levels of the same variable in a given group (e.g western diet vs. chow in males). In addition, count data were fitted to different types of multivariate models. For the diet dataset, we fitted the data to additive (~Sex+Diet) and interaction models (~Sex+Diet+Sex:Diet), and perform contrast tests with reduced models to evaluate the global effect of a variable or interaction between variables while controlling for baseline differences in the remaining variables. The same strategy was employed for the L- *Mettl14*KO dataset with Sex and *Mettl14* genotype as covariates. For each test, genes were classified as differential if adjusted p-value<0.01 (Wald test for pairwise contrasts, likelihood ratio test for multivariate models), with additional thresholds for size effect (log2 fold > 1) and baseline gene expression (FPKM>1). The pool of most variable genes in each dataset, selectedby PCA loadings or

differential expression tests, was then subjected to model-based clustering using MBCluster.Seq[52] to classify them based on their overall abundance profile across samples. Functional enrichment for genes selected in the tests and clusters above was performed with Metascape [53]. All figures were generated in Matlab (MATLAB, version release 2017a, The MathWorks, Inc, RRID:SCR_001622).

m6A-sequnecing

We followed the protocol previously outlined here [54] with minor modifications. Approximately 500µg of total RNA was extracted from frozen liver tissue (*n*=5 per group) using TRIzol reagent and purified using Oligo(dT) Dynabeads (Invitrogen). Because the large amount of input RNA required, we pooled approximately 1µg of purified mRNA was from each animal for each treatment group (*n*=5 per group). Approximately 5µg of pooled mRNA was then fragmented in fragmentation buffer (10mM Tris-HCl, pH=7.0, 10mM $ZnCl_2$) at 94°C for 2 min and 15 sec. 500ng mRNA (100ng per animal each group, pooled) was saved as input control for RNA-seq. 5µg of fragmented mRNA was incubated with 12µg of anti-m^{6}A antibody (Synaptic Systems) in 1X IP buffer (10mM Tris-HCl pH=7.4, 150mM NaCl, and 0.1% Igepal CA-630) for 2hr at 4°C on a rotating wheel. While fragmented mRNA was incubating, Protein A/G magnetic beads (Pierce) were washed twice and blocked by incubating in 1X IP buffer supplemented with 0.5mg/mL BSA for 2hr at 4°C on a rotating wheel. The m^{6}A-IP mixture was then added to the beads and incubated for an additional 2hr at 4°C on a rotating wheel. After 3 washes with IP buffer, bound mRNA was eluted by incubating with 100µL elution buffer (6.7mM N^6-Methyladenosine-5'-monophosphate salt in 1X IP buffer) for 1 hr at 4°C on a rotating wheel. Eluted mRNA fragments were collected, and the beads were incubated with an additional 100µL elution buffer for 1 hr at 4°C on a rotating wheel. The two 100µL eluates were combined and precipitated overnight at -20°C by adding one-tenth volume of 3M sodium acetate (pH=5.2) and 2.5 volumes of 100% ethanol. On average, 5µg of purified mRNA yielded tens of nanograms of

immunoprecipitated mRNA fragments. Input and immunoprecipitated RNA fragments were subjected to first-strand cDNA synthesis and library prep using TruSeq stranded mRNA library prep kit (Illumina) and sequenced on an Illumina HiSeq 3000 with single-end 50-bp read length. Raw sequence files were aligned to the GRCm38 assembly of the mouse genome using the RNA-Seq pipeline above. Peak detection was performed using MACS2 (version 2.1.1) onuniquely aligned reads using default parameters except for the --nomodel flag (--extsize=200). Both enriched and depleted peaks (using the IP libraries as reference) were identified, pooled and consolidated into a single peak library. Both enriched and depleted peaks (using the IP libraries as reference) were identified using a q-value lower than 0.05, pooled and consolidatedinto a single peak library. Downstream peak normalization and analysis was performed on the merged set of significant m^6A peaks using variance stabilized data. Peak annotation was performed with PeakAnalyzer[55], and motif enrichment with the MEME suite[56]. Peak counts per sample were summarized with STAR's built-in gene counter by re-mapping the raw sequencingfiles to the peak database. Downstream differential peak analyses and normalization followed the same rules as for the RNA-Seq data.

m^6A-IP-qPCR

The relative abundance of select lipogenic mRNAs in m^6A antibody IP samples was assessed by RT-qPCR in mouse liver. Total RNA was isolated from frozen mouse liver using TRIzol (Invitrogen). After DNase treatment, RNA was fragmented in fragmentation buffer (50mM Tris-HCl, pH 8.0, 50mM $MgCl_2$) at 94 for 3 mins. A portion of fragmented RNA was saved as input control, and the remaining fragmented RNA (approximately 10μg of total RNA) was used to perform m^6A-IP-qPCR. m^6A-immunoprecipitation was performed as described in the m^6A-seq procedure. Approximately 300μg of total RNA was used to perform m^6A-IP-qPCR. A 300ng aliquot was saved and used as input control. The remaining RNA was used m^6A-

immunoprecipitation as described in the m^{6}A-seq procedure. After overnight ethanol precipitation of immunoprecipitated RNAs, resulting pellets were suspended in 15μL RNase-free water and analyzed by qPCR. Abundance of lipogenic mRNAs in m^{6}A antibody samples was determined by normalizing to input control for each gene and expressed as the IP/input ratio or the % input.

mRNA half-life studies

Primary hepatocytes were harvested from chow-fed male WT or *Mettl14* L-KO livers as previously described [48]. After 48 h, primary hepatocytes were treated with actinomycinD (5μg/mL, Sigma) for 0h, 2h and 4h before trypsinization and collection. Total RNA was isolated by TRIzol, reverse transcribed, and analyzed by qPCR. The degradation rate of RNA was estimated as previously described [13] ($\log_2(A_t/A_0) = -kt$); where t is transcription inhibition time (h), A_t and A_0 represent mRNA quantity at time t and time 0. Two k values were calculated: time 2h versus time 0h, and time 4h versus time 0h. The final lifetime was calculated by using the average of k_{2h} and k_{4h} as previously described [13] ($t_{1/2} = 2\ln2/(k_{2h} + k_{4h})$.

Polysome profiling

Total cellular protein was isolated from approximately 300mg of frozen liver tissue in 500μL of lysis buffer (10mM Tris, pH=7.4, 150mM KCl, 5mM $MgCl_2$, 100μg/mL CHX, 0.5% Triton-X-100, 1:100 protease inhibitor (Roche), 40U/mL SUPERasin). Extracts were left on ice for 15 min with occasional pipetting and then centrifuged at 15,000xg for 15 min at 4°C. Supernatant was collected (~0.5 ml) and a small aliquot (25μL) was used to measure absorbance at 260 nm. Lysis buffer was used to equalize samples by A_{260} units. Each lysate was then split by ratio of 1:4 (Portion I/Portion II). 1 mL Trizol was added to Portion I and RNA was isolated to serve as input RNA. A 10/50% w/v sucrose gradient was prepared in lysis buffer without Triton-X. Portion II (~375μL) was then layered onto the top of the sucrose gradient by dispensing slowly

down the side of tube, making sure not to disturb gradient. Gradients were then centrifuged at 4 °C for 4 h at 27,500 r.p.m. (Beckman, rotor SW28). Aliquots of 250µL were then collected starting from the top of gradient and absorbance measured at 260nm on a nanospectrometer. Aliquots were then categorized into 3 main sub-types using the curve generated in the previous step (non-ribosome portion, 40S-80S, and polysome). Each aliquot was then combined with an equal volume of TRIzol to purify RNA and used for qPCR analysis. Abundance of lipogenic mRNAs in each fraction was determined by normalizing to input control value for each gene. We also performed the analysis by normalizing to housekeeper within the same fraction which did not alter the results.

Mitochondrial isolation and bioenergetics

Liver mitochondria isolation and respiration were measured as described previously [57]. Briefly, liver was minced, washed with PBS, and homogenized with glass Dounce homogenizer. Mitochondria were isolated by dual centrifugation (800 g and 8000 g) and respiration was obtained with a Seahorse bioanalyzer XF24 (Agilent). Fatty acid oxidation was measured in presence of 60 µM palmitoyl-carnitine, 250 µM malate and 4 µM FCCP.

ATAC-seq

ATAC-seq was optimized in liver after several modifications from original Buenstero et al. protocol [48]. A total of 100 mg of frozen liver were grinded to fine powder using cellcrusher and 1 mL of ice cold nuclei isolation buffer was added (20 mM Tris-HCl, 50 mM EDTA, 5 mM spermidine, 0.15 mM spermine, 0.1% mercaptoethanol, 40% glycerol, pH 7.5, mM EGTA, and 60 mM KCl). After 5 min of cooling on ice, cell suspension was filtered through Miracloth (Calbiochem) followed by centrifugation at 1100 × *g* for 10 min at 4 °C. Pellet was resuspended with 50 µL RSB buffer (10 mM Tris-HCl, 10 mM NaCl, 3 mM MgCl2, and pH 7.4) followed by

centrifugation at 500 × *g* for 5 min at 4 °C and resuspension in PBS. A total of 75,000 nuclei pellet were used for transposase reaction. The rest of the protocol followed that described by Buenstero. ATAC-Seq libraries were prepared using the Nextera Tn5 Transposase and DNA library preparation kit (Illumina) as described [48]. Libraries were single-end sequenced (50 bp) onan Illumina HiSeq 2000. Reads were mapped to the mouse genome (NCBI37/mm9) using Bowtie2 and were removed from the subsequent analysis if they were duplicated, mapped to mitochondrial genome, or aligned to unmapped contiguous sequences. Peak calling was performed using MACS2. The reads were converted to reads per thousand base pairs peak per million mapped reads (RPKM) by dividing by the total number of reads per sample. The average RPKM from four replicates was used to quantify the accessibility across all called peaks.

Single-molecule RNA FISH with fluorescent immunostaining:

Sequential RNA FISH with fluorescent immunostaining was performed as previously described [58]. Briefly, AML cells were transiently transfected with scrambled shRNAs or shRNAs targeting Mettl14 (mouse, TRCN0000084995). After 48 hours, the cells were washed once in PBS and then fixed in 37% formaldehyde in PBST (PBS with 0.05% Tween-20) at room temperature for 15 min. The fixing solution was removed, and ice cold methanol was immediately added to each chamber and incubated for 10 min at room temperature. Cells were rinsed once in PBS and incubated with blocking solution for 1 h at room temperature under rotation. After 1 h, the blocking solution was replaced with primary antibody (Mouse anti-DCP1a; WH0055802M6; Sigma; 1:300) and incubated for 1 h at room temperature (or overnight at 4 °C). After being washed 4 times with PBST (500 µl, 5–10 min for each wash), secondary antibody (Alexa 488 conjugate, 1:300 dilution in PBST) was added to the mixture and incubated at room temperature for 1 h. After washing 4 times with PBST (500 µl, 5–10 min for each wash), Stellaris FISH probe with Cal-Fluor610 were added and incubated 4hr at 37 °C in the dark. After washing 4 times

with PBST (500 µl, 5–10 min for each wash), anti-fade reagent (ProLong Diamond Antifade mountant with DAPI, Invitrogen) was added to mount the slides. The images were captured by a Zeiss LSM880 Confocal microscope with Airyscan and analyzed by ImageJ. Complete set of probes used provided in supplemental Table 3.

Targeted m^{6}A-modification of cellular mRNAs

TRM editor plasmid encoding nuclease-dead Cas13b and truncated Mettl3 fused to a nuclear localization sequence (pCMV-dCas13-M3nls) and gRNA expression plasmid (pU6-PspCas13b-gRNA) were obtained from Addgene (155366 and 155368; respectively). gRNAs targeting specific m^{6}A sites identified from m6A-seq on SCD1 mRNA as well as a nontargeting control gRNA were manufactured by IDT and gibson-cloned (NEB) into the gRNA expression plasmid. Targeted m^{6}A-modification of cellular mRNAs with TRM editors was then performed as previously described [26]. AML cells were plated on 6-well plates (for RNA) or 10-cm dishes (for protein and at approximately 80% confluency were transfected using Fugene HD (Promega) with 2.3µg of TRM editor plasmid and 1.1µg of gRNA expression plasmid (for 6-well) or 13.3µg of TRM editor plasmid and 5.7µg of gRNA expression plasmid (for 10-cm). At 48 hours post-transfection, protein and total RNA were harvested as described in the gene expression and immunoblot analysis section. The relative abundance of m^{6}A at different sites on *Scd1* mRNA in cells transfected with gRNAs targeting these sites or nontargeting control gRNAs was assessed by m^{6}A-IP-qPCR. To account for variability in RNA amounts, m^{6}A abundance at eachsite on *Scd1* was normalized to input or a nontargeted region on the TRM-edited transcript.

Statistical analysis

A non-paired Student's t-test was used to determine statistical significance, defined at $P < 0.05$. For multiple group experiments analysis of variance was used followed by multiple group

analysis. Unless otherwise noted, error bars represent standard error of mean. Most experiments were independently performed at least twice. Sample size is based on statistical analysis of variance and prior experience with similar in vivo studies.

Results

RNA methylation strongly enriches lipogenic transcripts and is dynamically regulated with diet

Chemical modifications on RNA are critical for the ability of cells to execute development programs and to respond to environmental challenges. To further investigate the contributions of RNA modifications in metabolic control, we performed unbiased interrogation of mRNA, lipids, and RNA methylation patterns using m^{6}A-seq [8] on livers from mice (C57BL/6) fed chow, western diet (WD) or high fat (HF) diet (Fig. 1a). As expected, HF or WD feeding led to increased weight gain, fat composition, and serum lipids at 2 weeks or 4 weeks (Supplementary Fig. 1a-i). Principal component analysis (PCA) of genome-wide RNA abundances segregates sample groups, showing that sex is the dominant factor influencing gene expression and diet composition playing a critical, but secondary role (Fig. 1b). Length of diet feeding did not a have a strong impact on gene expression. Congruently, PCA of unbiased lipidomics showed a similar trend, with sex and diet being the strongest variables influencing hepatic lipid composition (Supplementary Fig. 2a). We observed minimal shifts in most lipid species across groups (Supplementary Fig. 2b-d), but striking differences were noted for a number of neutral lipids including triglycerides (Fig. 1c). In line with PCA results, sex had a strong influence on fasting triglyceride content irrespective of diet composition with females consistently having higher levels than males (Fig. 1c).

Examination of the RNA methylation landscape in liver showed the m^{6}A modifications strongly clustered near 3' UTR regions or last exons in line with previous reports (Fig. 1d) [11,12]. In addition, we find that the RNA methylome is enriched at consensus motifs known to be bound by the METTL3/METTL14 complex (Fig. 1d) [12]. Intriguingly, PCA analysis based on global m^{6}A patterns led to stronger segregation of samples than use of gene expression alone (Fig. 1e and Supplementary Fig. 2e). We examined dynamic changes in m^{6}A profiles with dietary conditions

using several unbiased approaches, including comparing fold enrichment of m^{6}A with diet as well as examining the contributions of individual genes to variance. Regardless of approach, consistent enrichment of lipogenic genes was observed. PCA analysis showed that *Scd1* had multiple m^{6}A peaks that were in the top 20 most significant peaks contributing to variance of over 10,000 m^{6}A peaks detected (#4 and #14). Western diet (WD) feeding led to a reduction in the number of hypermethylated peaks (Fig. 1f). Comparison of chow and WD showed that differentially regulated m^{6}A peaks preferentially enriched lipogenic pathways with many lipogenic genes containing m^{6}A consensus motifs (Fig. 1g-h and Supplementary Fig. 2f-g). In agreement with these findings, m^{6}A profiles were reduced on lipogenic transcripts during WD feeding in comparison with chow (Fig. 1h). We confirmed our findings using m^{6}A-IP-RT-qPCR (Fig. 1h). A similar pattern was observed when comparing HF versus chow, where HF feeding led to a reduction in m^{6}A on mRNAs involved in fatty acid and triglyceride biosynthesis (Fig. 1h and Supplementary Fig. 2h). Taken together, our data suggest that m^{6}A peaks are strongly enriched on lipogenic mRNAs in liver and that their deposition is highly sensitive to dietary conditions.

Liver-specific deletion of the m^{6}A methylase METTL14 increases lipogenesis.

The RNA modification m^{6}A is installed by a multiprotein complex containing METTL3 and METTL14 [6,13]. Disruptions of the catalytic enzyme METTL3 or its allosteric activator METTL14 disrupts m^{6}A modifications on target RNAs [13,14]. To directly investigate the contributions of m^{6}A in hepatic lipid metabolism, we generated mice with liver-specific deletion of *Mettl14* (referred to as L-KO) by administering AAV8.tbg.cre [15,16] or AAV8.tbg.control (referred to as control or WT for paired experiments) to *Mettl14*$^{flox/flox}$ mice. We confirmed that METTL14 is readily deleted in liver but not in other tissues after 4 weeks on chow diet (Fig. 2a and Supplementary Fig. 3a). Unbiased lipidomics showed a dramatic increase in triglyceride content in *Mettl14* L-KO livers

compared with controls (Fig. 2b). The vast majority of other lipid classes were unchanged comparing WT or *Mettl14* L-KO (Fig. 2b and Supplementary Fig. 3b-c).

We did not observe changes in food intake or body weight including fat mass in the setting of loss of METTL14, suggesting that altered caloric balance is not a contributor to triglyceride accumulation (Supplementary Fig. 3d-e). Since changes in hepatic fatty acid oxidation can reciprocally affect triglyceride stores, we measured mitochondrial fatty acid metabolism in whole liver. Our results showed no change fatty acid oxidation in *Mettl14* L-KO livers compared with controls (Supplementary Fig. 3f). Measurement of genes involved in B oxidation showed no change or slight increase, suggesting that the observed increase in hepatic triglyceride stores cannot be explained by changes in beta oxidation or altered mitochondrial function (Supplementary Fig. 3g). Similarly, serum triglycerides, hepatic *Apob*, lipid export and lipolysis genes were minimally altered suggesting that impaired lipid secretion or lipid droplet breakdown is not a cause of triglyceride accumulation (Supplementary Fig. 3h-i).

Given that m^6A modifications are known to predominantly affect gene regulation by impacting mRNA stability, which can then result in corresponding changes in translation [8-10,17,18], we measured protein levels of major lipogenic genes. We observed a significant increase in protein levels of lipogenic genes in *Mettl14* L-KO compared with controls (Fig. 2c and Supplementary Fig. 3j). We confirmed this critical finding in *Mettl14* L-KO generated using crosses with Cre^{Alb} transgenics, suggesting that the effects of hepatic *Mettl14* deletion are sustained in both acute and chronic perturbations (Supplementary Fig. 3k-l). Taken together, these results suggest that enhanced fatty acid and triglyceride biosynthesis are the major contributors to triglyceride accumulation in *Mettl14* L-KO. Notably, we observe an increase in polyunsaturated fatty acids and SCD1 index or “desaturation index” in *Mettl14* L-KO in line with the observed increase in fatty acid biosynthetic enzymes (Fig. 2d) [19]. Protein levels and gene expression of other genes

involved in triglyceride metabolism such as *Apoe*, *Apob*, *Mtp* were not increased (Supplementary Fig. 3h-i). In addition, we confirmed reduced m^6A enrichment of lipogenic transcripts in *Mettl14* L-KO using m^6A-IP-RT-qPCR (Supplementary Fig. 3m-n). Collectively, our findings are consistent with idea that RNA modification act to repress lipogenic proteins in the liver and are congruous with the observed changes in m^6A landscape in response to diet.

It is well established that SREBP signaling is the dominant mechanism that influences lipogenic gene transcription and feedback regulation [1,20]. To more thoroughly investigate how m^6A modifications affect *Srebp1c* and lipogenesis, we measured mRNA levels of fatty acid biosynthetic genes. Despite the observed increase in lipogenic proteins and triglycerides levels, lipogenic mRNAs and pre-mRNA were reduced (Supplementary Fig. 4a-b). These results are in line with multiple previous studies showing that accumulation of polyunsaturated fatty acids (PUFAs) suppresses SREBP1C processing and potently feeds back on lipogenic gene transcription [1,21]. In addition, we found that loss of *Mettl14* L-KO led to a marked reduction in *Srebp1c* mRNA and nuclear SREBP1C protein (Supplementary Fig. 4a-c). Measurement of gene expression at an earlier timepoint (72 hours versus 4 weeks), where PUFA accumulation is reduced, showed an increase in lipogenic mRNAs (Supplementary Fig. 4d). Taken together, these results imply that the discordance between RNA and protein levels may be a late effect due to rampant accumulation of PUFAs. More importantly, the data suggests that a primary perturbation in fatty acid biosynthetic gene transcription is unlikely to explain the observed phenotype. We therefore reasoned that altered mRNA stability and/or protein translation may be at play, consistent with known roles of m^6A. To explore this further, we performed ribosome profiling from control or *Mettl14* L-KO livers which showed that loss of METTL14 enhances the abundance of lipogenic mRNAs but not all transcripts in the translating pool compared with controls (Fig. 2e and Supplementary Fig. 4e-f). To test whether METTL14 regulates lipogenic

genes through modulating their mRNA stability, we performed mRNA stability assays using the transcription inhibitor actinomycin D in primary hepatocytes isolated from control and *Mettl14* L-KO (Fig. 2f-g and Supplementary Fig. 4g). In the setting of transcriptional inhibition, we observed enhanced mRNA stability of lipogenic mRNAs in *Mettl14-KO* versus control hepatocytes (Fig. 2f-g and Supplementary Fig. 4g). Taken together, these results imply that increased protein levels of lipogenic genes is due to enhanced mRNA stability and downstream translation, in agreement with other evidence that internal RNA modifications impact mRNA dynamics to alter protein levels [6-8]. Complementing these results, we observed increased *Scd1* mRNA abundance in the setting of loss of *Mettl14* by fluorescence in situ hybridization (FISH) (Fig. 2h and Supplementary Fig. 4h). Interestingly, knockdown of *Mettl14* decreased localization of *Scd1* with DCP1a, a key decapping enzyme involved in mRNA degradation pathways and a marker of processing bodies (Fig. 2h). To reinforce the idea that post-transcriptional mechanisms are more critical factors in the effects of m^6A, we observed that knockdown of *Mettl14* enhances mRNA levels of lipogenic transcripts in setting of functional SREBP deficiency(*Scap*$^{-/-}$) [22,23] (Fig. 2i). Collectively, our findings suggest that increased biosynthesis of lipogenic mRNAs does not contribute to the observed increase in fatty acid and triglyceride levels in *Mettl14* L-KO and that post-transcriptional mechanisms are more critical drivers of the observed change in hepatic lipid composition.

METTL14 expression is inversely correlated with human fatty liver disease

To investigate the potential contribution of m^6A in human fatty liver disease, we examined the expression of *METTL14* in 144 liver samples from the Mexican Obesity Surgery (MOBES) cohort which includes over 500 non-related Mexican Mestizo patients (aged 18-59) [24]. Unique strengths of this study include using a gold-standard liver biopsy to confirm the presence and grading of NAFLD by Kleiner's scoring method[25]. Patients were classified in 3 groups: controls

(normal histology, n=23), simple steatosis (steatosis grade ranging from 1 to 2, n=47), and NASH (NAS> 5, n=74) based on independent scores from an experience pathologists. Clinical, metabolic and biochemical data were collected from all participants. *METTL14* expression was negatively correlated with human steatosis and NASH (Fig. 3a-b). Although triglyceride content is one of many factors that determine progression to chronic liver disease, these findings are consistent with our observed animal studies and hint at functional conservation of our proposed mechanisms.

Our results hint that modulation of m^6A levels can influence pathologic fat storage in liver, which led to tested an m^6A-based gene therapy strategy overexpressing the METTL3/METTL14/WTAP 'writer' complex in a mouse model of nonalcoholic fatty liver disease (NAFLD) (Fig. 3c). *In vivo* administration of m^6A writers using an AAV8 with a liver-specific promoter enhanced gene expression of MTA complex components (Fig 3d) and m^6A enrichment on lipogenic mRNA (Fig 3e). In addition, m^6A writers overexpression reduced lipogenic proteins (Fig. 3f) and resulted in modest decrease in Oil-red-O staining, and hepatic triglyceride content although results did not reach significance (Fig. 3g and Supplementary Fig. 4i). We observed a significant reduction in genes associated with inflammation and fibrosis (Fig.4h), key drivers of NASH progression. It should be noted that chronic overexpression of METTL14 alone was not sufficient to alter lipogenic proteins (not shown) probably since its catalytic partner METTL3 is required to efficiently enhance m^6A deposition. Recent evidence suggests that site-specific programmable RNA modification is readily achievable with minimal off target effects using a guide RNA and a dCas13-tethered writer complex [14,26,27]. To test the influence of programmable RNA methylation of lipogenesis, we introduced an RNA methylation modifying complex with individual RNA guides targeting *Scd1* m^6A sites (Fig. 3i). Targeted RNAmethylation enhanced *Scd1* m^6A levels measured by m^6A-IP-RT-qPCR and resulted in lower protein levels of SCD1 (Fig. 3j-k).

Altogether, our findings hint at strong relevance of m^{6}A in regulation of human fatty liver and that modulation of RNA methylation through multiple approaches, including an FDA approved platform, is technically feasible and efficacious in reducing important markers of chronic liver disease.

m^{6}A machinery undergoes transcriptional regulation by sex-biased transcription factors

Our global methylome analysis revealed a reduction in m^{6}A signatures with WD or HF feeding in male mice along with a potent decrease in lipogenic RNA modification (Fig. 1). In line with these results, we find that chronic WD feeding resulted in a significant increase in hepatic triglycerides in male *Mettl14* L-KO compared to controls (Fig. 4a-b), but the effect size was diminished in comparison with chow diet (Fig. 4a and Fig. 2b). Similarly, loss of *Mettl14* resulted in only a mild increase in lipogenic enzymes under WD as opposed to the dramatic effect observed under chow conditions (Fig. 4c). Consistent with chow feeding results, we did not observe major changes in other lipid species (Supplementary Fig. 5a). Interestingly, the fed state was associated with a reduction in METTL14 protein and loss of *Mettl14* resulted in no observable change in triglycerides or lipogenic proteins under refed conditions, when lipogenesis is known to be highest (Fig. 4d-e). These findings imply that m^{6}A modifications appear to be more relevant under basal fasting conditions where there is minimal demand for *de novo* fatty acid biosynthetic enzymes and that contributions of m^{6}A are diminished under western diet or refeed conditions. Thus, m^{6}A modifications act in coordinate fashion with transcriptional control mechanisms driven by SREBP1C, becoming less important when SREBP1C driven lipogenesis is at play.

To better explore the mechanism of reduction in m^{6}A with dietary challenge, we investigated the regulation of *Mettl14*. In line with the findings in Fig. 4c, we observed a decrease in the expression of the m^{6}A writer *Mettl14* with either HF or WD (Fig. 4f). We confirmed that

METTL14 protein is reduced with diet feeding (Fig. 4g). Intriguingly, the pattern of regulation in male mice was the opposite of what we observed in female mice where WD or HF feeding did not reduce *Mettl14* mRNA and protein levels and in fact trended to increase (Fig. 4f-g). We measured m^6A levels in liver using mass spectrometry which closely mirrored the *Mettl14* regulation pattern (Fig. 4h). To further explore the basis of the sex-specific differences in *Mettl14,* we used a genetic model that segregates gonadal sex from chromosomal sex effects by generating XX and XY mice with ovaries, and XX and XY mice with testes on a hyperlipidemic background [28]. We found that XX animals have higher *Mettl14* levels than XY (Fig. 5a), consistent with the observed regulation of *Mettl14* under lipid-rich conditions (Fig. 4f- g). These results imply that *Mettl14* levels are strongly influenced by chromosomal sex composition. In addition, examination of publicly available histone marks in liver under different dietary conditions led us to hypothesize that *Mettl14* may be subject to tight transcriptional regulation, since H3K27ac, a mark of transcriptional activation, was reduced with high fat feeding (Supplementary Fig. 5b). To more thoroughly explore the molecular basis of alteration inMETTL14 and m^6A with diet feeding, we performed ATAC-seq in liver under chow, WD or HF (Fig. 5b). This technique tags accessible genome sites with sequencing adaptors to provide a multidimensional view of gene regulation [29]. We did not observe dramatic changes in accessibility or chromatin rearrangements in response to diet feeding at the *Mettl14* locus (Fig. 5b). We noted, however, a single open access site was which we reasoned may be critical for docking of canonical regulatory factors or recruiting additional factors with dietary perturbations within this region. Motif interrogation of accessible regions revealed binding of a number of factors known to be involved in metabolic control including the sexually-dimorphic STAT5A and BCL6 complex[30-32]. ChIP-seq studies confirmed the binding of STAT5a and BCL6 at *METTL14* in both mice and humans (Supplementary Fig. 5c-d). Consistent with previous studies [30,31], we observed an increase in *Bcl6* levels with diet and feeding state (Fig. 5c and Supplementary Fig. 5e). Levels o

m^{6}A on lipogenic transcripts were altered comparing fast and refed state (Supplementary Fig. 5f). In addition, other MTA complex proteins were also differentially regulated with diet composition or feeding state (Supplementary Fig. 5g-h). Interestingly, binding of BCL6 has been shown to repress STAT5 targets in male but not female livers [33,34]. In addition, hepatic loss of *Bcl6* has been shown to have the opposite phenotype of the *Mettl14* L- KO where it reduces triglyceride stores [30]. These results hint that the observed reduction in

Mettl14 enzyme levels in male but not female mice with dietary challenge may be mediated by BCL6 repressive effects at the *Mettl14* promoter region. To directly test this theory, we generated *Bcl6* liver-specific knockout mice (*L-Bcl6$^{-/-}$*). When fed WD, *L-Bcl6$^{-/-}$* mice showed no differences in *Mettl14* mRNA and protein levels between the sexes in stark contrast to the differences observed in WT mice (Fig. 5d-e). Overall, these results strongly suggest the existence of sex-specific differences in RNA methylation and that m^{6}A installing machinery is subject to tight transcriptional control.

Loss of m^{6}A diminishes sex-specific differences in hepatic lipid composition and sex-biased gene expression.

Multiple studies reported on sex-specific differences in lipogenic mRNAs, their protein abundance as well as fatty liver traits across species, albeit the direction of these differences varied depending on species, genetic background, and context [5,35-37]. Female mice were noted to have higher fasting liver triglyceride than males but not necessarily all lipid species [5,36], which is in line with our findings (Fig. 6a). Transcriptome-wide gene expression analysis showed minimal differences in lipogenic mRNAs between the sexes regardless of diet (Fig. 6b). We confirmed by RT-qPCR that differences in lipogenic RNAs do not explain the observed sex-differences in fasting hepatic triglyceride content (Fig. 6c), although surprisingly females were

noted to have markedly higher protein level of fatty acid and triglyceride synthesis enzymes compared with males (Fig. 6d). *Scd1,* for example, showed higher RNA levels in males despite having higher protein levels in females (Fig. 6c-d). A similar trend was observed for other genes involved in triglyceride biosynthesis where strikingly higher protein levels were noted in females despite minimal differences in gene expression. These results imply that post-transcriptional mechanisms may be contributing to sex-specific variation in lipogenic proteins and triglyceride content. Remarkably, the protein levels of lipogenic genes were discordant with sex-differences in m^6A where males had higher m^6A levels than females on lipogenic mRNAs (Fig. 6e). These findings are in line with the above results (Fig. 2 and Fig. 3) and reinforce the idea that m^6A modifications on lipogenic transcripts may be a factor contributing to sex-differences in lipogenic protein levels and triglyceride content. To test this hypothesis more directly, we compared the effect of *Mettl14* deletion between the sexes and found that liver-specific deletion of *Mettl14* diminished the striking differences in lipogenic protein levels between male and female mice (Fig. 6f). In addition, we note that under chow conditions liver-specific deletion of *Mettl14* in male mice increased triglyceride levels, almost matching WT female mice, suggesting that m^6A activity leads to suppression of triglyceride levels and its loss leads to a more “feminized” lipid composition (Fig. 6g). Loss of m^6A control in females under chow conditions led to only a modest increase in triglyceride content (Fig. 6g). Reciprocally, loss of *Mettl14* in female mice was associated with greater differences in hepatic triglyceride compared with males under WD conditions (Fig. 6g). These results are consistent with the sex-dimorphic regulation of *Mettl14* (Fig. 5). Taken together, our findings strongly imply that sex-differences in fasting triglyceride stores are predominantly dictated by differences in RNA modifications.

To more thoroughly dissect the contributions of m^6A in hepatic sex-specific traits we performed RNA-seq on control and *Mettl14* L-KO male and female liver. PCA analysis of genome-wide

mRNA abundances showed clear segregation between groups, with m^6A unexpectedly contributing to more variability in global RNA levels than sex (Fig. 6h). Remarkably, loss of *Mettl14* abolished m^6A-based differences in variance and significantly minimized sex-driven PCA variance, suggesting a strong interaction between RNA-modifications and sex-biased gene regulation (Fig. 6h). Notably, fitting the RNA-Seq data to various multivariate additive and interaction models showed that the interaction between sex and *Mettl14* was significant for most sex-biased genes. This interaction resulted in a major reduction in sex-biased gene expression in *L-Mettl14* KO mice (Supplementary Fig. 6a-c) and hierarchical clustering revealed that L-*Mettl14* KO male mice consistently clustered with female samples (Supplementary Fig. 6c). Notably we did not observe differences in key regulators of sex-biased gene expression such as *Xist* and *Sry* (Supplementary Fig. 6d) and majority of the differentially regulated gene are not known to be direct targets of *Xist*. We used model-based clustering of all sex-biased and *Mettl14*-dependent genes to more rigorously evaluate the influence of sex on loss of m^6A. Consistent with our global analysis, a number of clusters showed dramatic abrogation of sex-biased gene expression in setting of loss of *Mettl14* (Supplementary Fig. 6e). Many of the genes in these clusters have been reported to define differences between male and female livers [38]. In addition, unsupervised interrogation of genes in these clusters revealed that multiple metabolic processes are the most enriched pathways under the interaction between m^6A and sex (Fig. 6i). In summary, our findings suggest that hepatic m^6A levels are important contributors to sex-biased gene expression and sex-specific lipid traits.

Discussion

In this work, we find that lipogenic mRNAs undergo chemical modifications that powerfully impact transcript turnover, adding another layer of dynamic gene regulation in hepatic lipid metabolism. The transcription factor SREBP1C, which is reduced during fasting and is activated by insulin and the lipid-responsive nuclear receptor Liver-X-receptor (LXR), play key roles in the induction of lipogenic genes through direct binding to lipogenic promoters [1,39]. Our model fits well with classical transcriptional mechanisms controlling lipogenesis and suggests that m^6A fine-tunes transcriptional outputs by enhancing mRNA degradation under conditions where hepatic fatty acid biogenesis may be in low demand.

A number of studies including our own have reported that female mice have higher basal fasting triglyceride levels than males but the molecular basis for these observations have been unknown [5,35-37]. We observed significant differences in the effects of m^6A between male and female mice and a strong influence of m^6A on the interaction between sex and diet composition.In male mice, loss of m^6A robustly impacted fatty acid biosynthesis and increased triglyceride stores under fasting chow conditions but its influence was diminished during western diet feeding, consistent with the observed decrease in m^6A. An opposite pattern was observed in female mice. Our findings here fill important gaps into the mysterious sex-specific differences in lipid composition but also suggest that there must be other factors beyond m^6A control that govern sex-differences in lipid composition. Although sex-differences in lipogenesis and triglyceride stores were dramatically diminished with loss of m^6A, it is important to point out that they were not entirely abrogated. Additionally, control of lipogenesis is one of many factors that influences hepatic lipid stores and it is likely that sex-differences related to other pathways suchas mitochondrial bioenergetics, nuclear receptor outputs, or lipid uptake may be important.

Finally, our study examined gene function during short-term dietary perturbations to more directly underpin mechanistic effects of m^6A. Although the influence of hepatic RNA

modifications during prolonged feeding mimicking fatty liver disease is not well-defined, we predict, based on our findings, that the female sex might confer a fitness advantage in chronic liver disease models. Despite starting with higher basal liver triglycerides, our data suggests that females are able to robustly leverage the protective effects of m^{6}A in the face of a dietary challenge whereas this unique lipid degradation pathways stalls in male mice under the same conditions. Congruent with this interpretation, we find that most human studies examining long-term dietary exposure showed higher risk of fatty liver disease in males compared with females [40]. We note, however, that disentangling the influence of covariates such as feeding state, diet composition, obesity, diabetes and alcohol consumption can be a challenge in population studies.

Contributions of m^{6}A seem to be more critical when SREBP1C driven lipogenesis is not in play, so at higher levels of gene transcription, the differential mRNA methylation is less important.

Our findings adds to other evidence suggesting that m^{6}A installation is actively regulated *in vivo* during health and disease states [13]. Notwithstanding the known limitations in detecting changes using m^{6}A -seq [41], a strength of our study is using multiple complimentary approaches including knockout models to validate our main findings. We identify the BCL6-STAT5A axis as a critical regulator of *Mettl14* and m^{6}A content in the liver but it conceivable that other m^{6}A 'writers', 'erasers' or 'readers' may undergo similar dynamic regulation by the same axis or perhaps otherrelevant transcription factors. It is also possible that *Mettl14* may undergo additional post-translational processing. Intriguingly, deletion of *Bcl6* ameliorates hepatic steatosis [30] consistent with our findings that activation of *Bcl6* inhibits m^{6}A dependent degradation of transcripts that encode regulators of lipid synthesis. It is important to point out that active regulation of m^{6}A machinery may not be the only mechanism specifying differences in lipid composition under different conditions. Indeed, our data hints that there may be ancillary variables that drive

sensitivity to m^6A effects between the sexes. For example, it is very likely that changes in stoichiometry critically dictates m^6A effects. m^6A-seq cannot reveal the precise fraction of transcript copies modified by m^6A. Differences in contributions of various m^6A readers is not examined our study. Finally, although our work focused on the role of m^6A in lipogenesis, it is important to point out that m^6A may have other contributions in metabolism. Future studies will investigate these questions more thoroughly.

Acknowledgements:

This work was supported by NIH grants DK118086, HL139549, HL149766, NS111631, CA186702, American Heart Association Transformational Project grant and Burroughs Wellcome Fund Career Award for Medical Scientists.

Author contributions:

TS conceived and supervised the study. TS and DS designed the study, guided the interpretation of the results and preparation of the manuscription. DS, XW, JK, DW performed the majority of experiments and data analysis. ZZ performed the ATAC-seq experiments and data analysis. LV and KR performed the fatty acid oxidation studies and studies determine the influence of chromosomal versus gonadal factors in Mettl14 regulation and guided the interpretation of results. KJW performed lipidomics studies and analysis. AM and SJ performed m^6A measurement via mass spectrometry. AHV analyzed the data for MBBS. JC assisted with m^6A loss of function *in vivo* studies and cellular knockdown. DC performed m^6A and RNA-seq analysis. TS wrote the manuscript with input from all authors. All authors discussed the results and approved the final version of the manuscript.

Competing Interests:

The authors declare no competing interests.

Datasets Availability:

RNA-seq, m^6A-seq, and ATAC-seq data can be accessed under NIH's SRA repository PRJNA663718 and GSE157907.

Figure Legends

Fig. 2-1 | RNA methylation strongly enriches lipogenic transcripts and is dynamically regulated with diet. a, Schematic of experimental design. **b,** Principal component analysis (PCA) of gene expression for mouse liver samples under chow (*n* = 5 females, 3 males), WD *(n* = 4 females, 5 males), and HF diet *(n* = 5 females, 4 males). The first two components (PC1, PC2) are shown along with the percent of gene expression variance explained. Clustering was obtained with data from all detected genes without additional filters. HF=High Fat, WD=Western Diet. Results are representative of 2 independent experiments. **c,** Major hepatic lipid species analyzed by lipidomics in liver harvested from male and female mice fed chow, WD, or HF (*n* = 5 per group). Results are representative of 2 independent experiments. **d,** Pie chart illustrating relative position of m^6A on immunoprecipitated transcripts and motif enrichment analysis of m^6A-containing RNAs. E-values were computed as the enrichment p-value (Fisher's Exact Test for enrichment of the motif in the positive sequences) times the number of candidate motifs tested. **e,** PCA plot for m^6A-seq in mouse liver under different conditions. Both input and m^6A immunoprecipitated samples are displayed. **f,** Scatter plot comparing m^6A enrichment between chow and WD-fed male livers. Hyper-methylated peaks with significantly higher m^6A enrichment ($\log_2 m^6A > 0$) in chow compared to WD are noted with orange dots, and hypo-methylated peaks with significantly lower m^6A enrichment ($\log_2 m^6A<0$) in chow compared to WD are noted with blue dots. Genes with no significant difference in m^6A enrichment between chow and WD are shown in orange (hypermethylated in chow compared to WD) and purple (hypomethylated in chow compared to WD). **g,** DAVID functional annotation of top 500 genes with highest m^6A enrichment in male mice fed chow versus WD diet. Gene ontology analysis was performed with –log10 (p value) plotted (x axis) as a function of classification meeting a p value of < 0.001. **h,** UCSC browser screenshots and enrichment of m^6A modification on lipogenic transcripts in male

liver as determined by m^{6}A-IP-qPCR (*n*=5 per group). The relative enrichment of m^{6}A in each sample was calculated by normalizing to tenfold input. Results are representative of 2 independent experiments. All mice were fed the indicated diet for 4 weeks and fasted for 4 hrs prior to sacrifice. Values are mean ± s.e.m. of five biological replicates (**c,h**). *P* values were calculated using one-way analysis of variance (ANOVA) followed by a multi-group comparison test (Fisher's) in **c** and **h** or unpaired two-tailed t-test in **g.** **P<0.05; ** P<0.01; *** P<0.001; **** P<0.0001.* The precise *n, P* values, and details of the statistical testing are provided in the source data file.

Fig. 2-2 | Liver-specific deletion of the m^{6}A methylase METTL14 increases de-novo lipogenesis and hepatic triglyceride content. a, Western blot from liver of WT and *Mettl14* L-KO chow-fed male mice. Equal amounts of protein were pooled from eight animals and run in triplicate. The experiment was repeated 3 times with similar results. **b,** Quantification of major lipid species detected in livers of chow-fed male WT and *Mettl14* L-KO mice by unbiased lipidomics *(n* = 8 per group). **c,** Western blot comparing lipogenic protein levels in WT versus *Mettl14* L-KO chow-fed male mice livers *(n* = 8 per group). Equal amounts of protein were pooled from eight animals per group and run in triplicate. The experiment was repeated 3 times with similar results. **d,** Unbiased lipidomic measurement from liver of PUFAs (measured by lipidomics) and *SCD1* index for WT and *Mettl14* L-KO chow-fed male mice *(n* = 8 per group). **e,** qPCR analysis of transcript levels in polysome fraction from WT and *Mettl14* L-KO chow-fed male livers *(n* = 4). The experiment was repeated 3 times with similar results. **f,** mRNA lifetime of lipogenic transcripts in primary hepatocytes harvested from livers of chow-fed WT and *Mettl14* L-KO mice *(n*=3). **g,** Quantification of transcript abundance by qPCR at 2hr and 4hr after transcription inhibition in primary hepatocytes harvested from WT and *Mettl14* L-KO chow-fed male livers *(n*=3). **h,** Single-molecule RNA FISH showing localization of *SCD1* mRNA in WT

and *Mettl14* knockdown AML cells. Images are representative of three independent experiments. **i,** qPCR analysis of lipogenic gene expression in *METTL14*-knockdown versus control in *SCAP*-KO HEK293 cells *(n*=4). The experiment was repeated 2 times with similar results. All mice were fed chow diet for the indicated time period and fasted for 4 hrs prior to sacrifice. Values are mean ± s.e.m. of 3(**f,g**), 4(**e,i**), or 8(**b,d**) independent biological replicates. *P* values were calculated using unpaired two-tailed *t*-test (**b,d,e,f,g,i**). **P<0.05; ** P<0.01; *** P<0.001; **** P<0.0001.* The precise *n, P* values, and details of the statistical testing are provided in the source data file.

Fig. 2-3 | METTL14 expression is inversely correlated with fatty liver disease and hepatic triglyceride content.

a, Pearson correlation coefficient and *P*-value comparing *METTL14* expression and NAFLD traits in humans (*n*=144) from Mexican Obesity Surgery (MOBES) Cohort. **b,** Correlation between *METTL14* expression and NAFLD from Mexican Obesity Surgery (MOBES) cohort. **c,** Schematic of gene therapy design. **d,** Gene expression in liver harvested from mice injected with AAV8.tbg.*Mettl14*, AAV8.tbg.*Mettl3 and* AAV8.tbg.*Wtap* 'writers' versus equal amount AAV8.tbg.null (*n* = 6). **e,** Quantification of m^6A abundance on lipogenic mRNAs as measured by m^6A-IP-qPCR from livers of mice injected with AAV8.tbg.*Mettl14*, AAV8.tbg.*Mettl3 and* AAV8.tbg.*Wtap* 'writers' versus equal amount AAV8.tbg.null (*n* = 4). The experiment was repeated two times with similar results. **f,** Western blot comparing lipogenic protein levels in livers of mice overexpressing m^6A writers compared to control. Equal amounts of protein were pooled from 6 animals per group and run in triplicate. The experiment was repeated three times with similar results. **g,** Oil-Red-O staining of frozen liver sections from mice in (**f**)**.** Images are representative of three independent biological replicates. Scale bar equals 100 microns (left), 50 microns (right). **h**, qPCR analysis of NASH-related fibrosis and inflammatory gene expression

from livers of mice treated with AAV overexpression of m^{6}A writers or null (*n* = 5). The experiment was repeated two times with similar results. **i,** Schematic of *in-vitro* targeted RNA modification (TRM) editor experiment. **j,** Quantification of m^{6}A abundance at *Scd1* by m^{6}A-IP-qPCR following transfection of TRM editors (*n* = 4 independent biological replicates). Values were normalized to a non-targeted m^{6}A site located in the 3' UTR of *Scd1.* **k,** Western blot comparing *SCD1* protein levels in cells transfected with gRNAs targeting A2445 or A2134 versus cells transfected with nontargeting gRNA. Samples were pooled from 3 independent biological replicates and run in duplicate. The experiment was performed three times with similar results. All mice were fed NASH diet for 8 weeks and fasted for 4 hrs prior to sacrifice. Values are mean ± s.e.m. of 4 (**e,j**), 5 (**h**), or 6 (**d**) biological replicates. *P* values were calculated using one-way analysis of variance (ANOVA) followed by multi-group comparison (Fisher's) in **b** or unpaired two-tailed *t*-test (**b**,**d,e,h,j**) **P<0.05; ** P<0.01; *** P<0.001; **** P<0.0001.* The precise *n, P* values, and details of the statistical testing are provided in the source data file.

Fig. 2-4 | m^{6}A machinery is dynamically regulated by diet and exhibits sexual dimorphism.

a, Hepatic TAG content determined by lipidomics from WT and *Mettl14* L-KO mice under WD feeding (*n* = 7 WT mice, *n = 8* KO mice). **b,** H&E staining of representative livers from male WD-fed WT or *Mettl14* L-KO mice. Images are representative of three independent biological replicates. **c,** Western blot comparing lipogenic protein levels in WT and *Mettl14* L-KO livers from chow and WD-fed male mice. Equal amounts of protein were pooled from five animals per group and run in triplicate. The experiment was repeated three times with similar results. **d,** Western blot showing *Mettl14* protein levels in liver of fasted (4 hr prior to sacrifice) versus re-fed (fasted overnight then re-fed for 4 hr) male mice. Equal amounts of protein were pooled from five animals per group and run in triplicate. The experiment was repeated three times with

similar results. **e,** Western blot showing lipogenic protein levels in livers of male WT and *Mettl14* L-KO re-fed mice (fasted overnight then re-fed for 4 hr). Equal amounts of protein were pooled from five animals per group and run in triplicate. The experiment was repeated three times with similar results. **f,** qPCR analysis of *Mettl14* expression in male and female livers under different diets (*n*=5). **g,** Western blot showing *Mettl14* protein levels in liver harvested from male and female mice fed different diets. Equal amounts of protein were pooled from five animals per group and run in triplicate. The experiment was repeated three times with similar results. **h,** LC-MS quantification of hepatic m^6A levels in male and female mice fed different diets (*n*=5). One male WD sample and one female chow sample were identified as outliers and formally excluded (Grubbs, alpha = 0.05). Mice were fed the indicated diet for 4 weeks and fasted 4-hrs prior to sacrifice except for mice in (**d,e**). Mice in (**d,e**) were fasted overnight and then re-fed for 4-hrs prior to sacrifice. Values are mean ± s.e.m. of 5 (**f,h**), 7 (**a**), or 8 (**a**) independent biological replicates. *P* values were calculated using unpaired two-tailed *t*-test (**a**) or one-way analysis of variance (ANOVA) followed by multi-group comparison (Fisher's) in **f** and **h**. **P<0.05; ** P<0.01; *** P<0.001; **** P<0.0001.* The precise *n, P* values, and details of the statistical testing are provided in the source data file.

Fig. 2-5 | m^6A machinery undergoes transcriptional regulation by sex-biased transcription factors.

a, qPCR analysis of *Mettl14* expression in livers of male and female *Apoe*$^{-/-}$ mice (*n*=3) with core four genotypes. **b,** UCSC browser screenshot of ATAC-seq data for *Mettl14* promoter region from male mouse liver under different diets and transcription factor motifs identified at this site. Screenshots are representative of five independent biological replicates. **c,** qPCR analysis of *Bcl6* expression in male mouse livers under different diets *(n*=4 WD, *n*=5 chow, HF). The experiment was repeated three times with similar results. **d,** qPCR analysis of *Mettl14*

KO mice). The experiment was repeated three times with similar results. **e,** Western blot comparing *Mettl14* protein levels in livers of male and female WD-fed *Bcl6* KO mice compared to WD-fed WT mice. Equal amounts of protein were pooled from 5 animals per group and run in triplicate. The experiment was repeated two times with similar results. All mice were fed the indicated diet for 4 weeks and were fasted for 4-hrs prior to sacrifice. Values are mean ± s.e.m. of 3 (**a**), 4 (**c**), 5 (**c, d**), or 6 (**d**) independent biological replicates. *P* values were calculated using one-way analysis of variance (ANOVA) followed with multi-group comparison (Fisher's) in **c** or two-way ANOVA followed with multi-group comparison (Tukey) in **a** and **d.** **P<0.05; ** P<0.01; *** P<0.001; **** P<0.0001.* The precise *n, P* values, and details of the statistical testingare provided in the source data file.

Fig. 2-6 | Loss of m^{6}A significantly diminishes sex-specific differences in hepatic lipid composition.

a, CIRCOS plot of hepatic lipid content under different diets *(n*=5 per group). Width of connection indicates higher TAG content. **b,** Heatmap of lipogenic gene expression in livers of male and female mice under different diets *(n*=5 per group) based on RNA-seq. **c,** qPCR analysis of lipogenic gene expression in livers of WD-fed male and female mice *(n*=4 per group). The experiment was repeated three times with similar results. **d,** Western blots comparing lipogenic protein levels in WD-fed WT male and female liver. Equal amounts of protein were pooled from eight animals per group and run in triplicate. The experiment was repeated three times with similar results. **e,** m^{6}A-IP-qPCR validation of m^{6}A enrichment on lipogenic genes in livers of male and female mice *(n*=4 per group). The experiment was repeated two times with similar results. **f,** Western blots comparing lipogenic protein levels in WD-fed *Mettl14* L-KO male and female liver. Equal amounts of protein were pooled from eight animals per group and run in triplicate. The experiment was repeated three times with similar

female livers *(n* = 6 KO chow-fed females, *n* = 7 WT and KO WD-fed males and females, *n* = 8 WT and KO chow-fed males and KO chow-fed females. **h,** Principal component analysis (PCA) of gene expression for male and female mouse livers in control and *Mettl14* L-KO samples (*n* = 7 WT males, *n* = 8 all other groups). Clustering was obtained with data from all detected genes without additional filters. **i,** Barplot of functional enrichment adjusted *P*-values (hypergeometric *P*-values after Benjamini-Hochberg correction) for genes significantly associated with sex-specific responses to *Mettl14*-L-KO (interaction genes). Ontology terms are grouped by gene member similarity. All mice were fed the indicated diet for 4 weeks and fasted for 4-hrs prior to sacrifice. Values are mean ± s.e.m. of 4 (**c,e**) or 6-8 (**g**) independent biological replicates. *P* values were calculated using unpaired two-tailed *t*-test (**c,e,g**). **P<0.05; ** P<0.01; *** P<0.001; **** P<0.0001.* The precise *n, P* values, and details of the statistical testing are provided in the source data file.

Extended Data Figure 2-1. Effect of sex and diet composition on lipid metabolism. a, Fat mass measured by MRI for male and female mice in 2-week cohort *(n*=5 per group). **b,** Body weight of male and female mice in 2-week cohort *(n*=5 per group). **c,** Quantification of serum cholesterol levels of male and female mice in 2-week cohort *(n*=5 per group). **d,** Quantification of total serum triglycerides of male and female mice in 2-week cohort *(n*=5 per group). **e,** Fat mass measured by MRI for male and female mice in 4-week cohort *(n*=5 per group). **f,** Body weight of male and female mice in 4-week cohort *(n*=5 per group). **g,** Quantification of serum cholesterol levels of male and female mice in 4-week cohort *(n*=5 per group). **h,** Quantification of total serum triglycerides of male and female mice in 4-week cohort (*n*=5 per group). **i,** Quantification of total serum NEFA from male and female mice in 4-week cohort (*n*=5 per group). Mice were fed indicated diet for 4 weeks and fasted for 4-hrs. prior to sacrifice. Values are mean ± s.e.m. of 5 independent biological replicates (**a-i**). *P* values were calculated using

*P<0.01; *** P<0.001; **** P<0.0001.* The precise *n, P* values, and details of the statistical testing are provided in the source data file.

Extended Data Figure 2-2. Lipogenic genes are highly enriched for m^{6}A under chow diet. **a,** Lipidomics PCA plot for 4-week cohort (*n*=5 per group). **b,** Quantification of major lipid species identified in lipidomics analysis of male and female mouse liver for 2-week cohort (*n*=5 per group). **c,** Lipidomics heatmap for 4-week cohort (*n*=5 per group). **d,** Quantification of major lipid species identified in lipidomics analysis of mouse liver for 4-week cohort (*n*=5 per group). **e,** Comparison of number of differentially expressed genes (>2-fold) in WD or HF versus chow-fed male livers determined by RNA-seq and m^{6}A-seq. **f,** Nucleotide sequences containing the m^{6}A motifs on lipogenic transcripts and relative position of each motif on full-length mRNA. **g,** Rank order table of genes with greatest fold-change in m^{6}A in male livers (Chow vs. Western diet). **h,** UCSC browser screenshot showing changes in m^{6}A enrichment for *DGAT2* in chow and HF diet-fed male livers. Mice were fed indicated diet for 4 weeks and fasted for 4-hrs. prior to sacrifice. Values are mean ± s.e.m. of 5 independent biological replicates (**b,d**). *P* values were calculated using one-way analysis of variance (ANOVA) with multi-group comparison (Fisher's) in **b** and **d**. **P<0.05; ** P<0.01; *** P<0.001; **** P<0.0001.* The precise *n, P* values, and details of the statistical testing are provided in the source data file.

Extended Data Figure 2-3. Loss of the m^{6}A methylase METTL14 increases lipogenesis and hepatic triglyceride accumulation. a, Western blot of *Mettl14* in other tissues from WT and *Mettl14* L-KO chow-fed males. Equal amounts of protein were pooled from four animals per group and run in triplicate. **b,** Lipidomics heatmap comparing hepatic lipidome of WT versus *Mettl14* L-KO chow-fed males. *(n*=8 per group). **c,** Quantification of various lipid species from

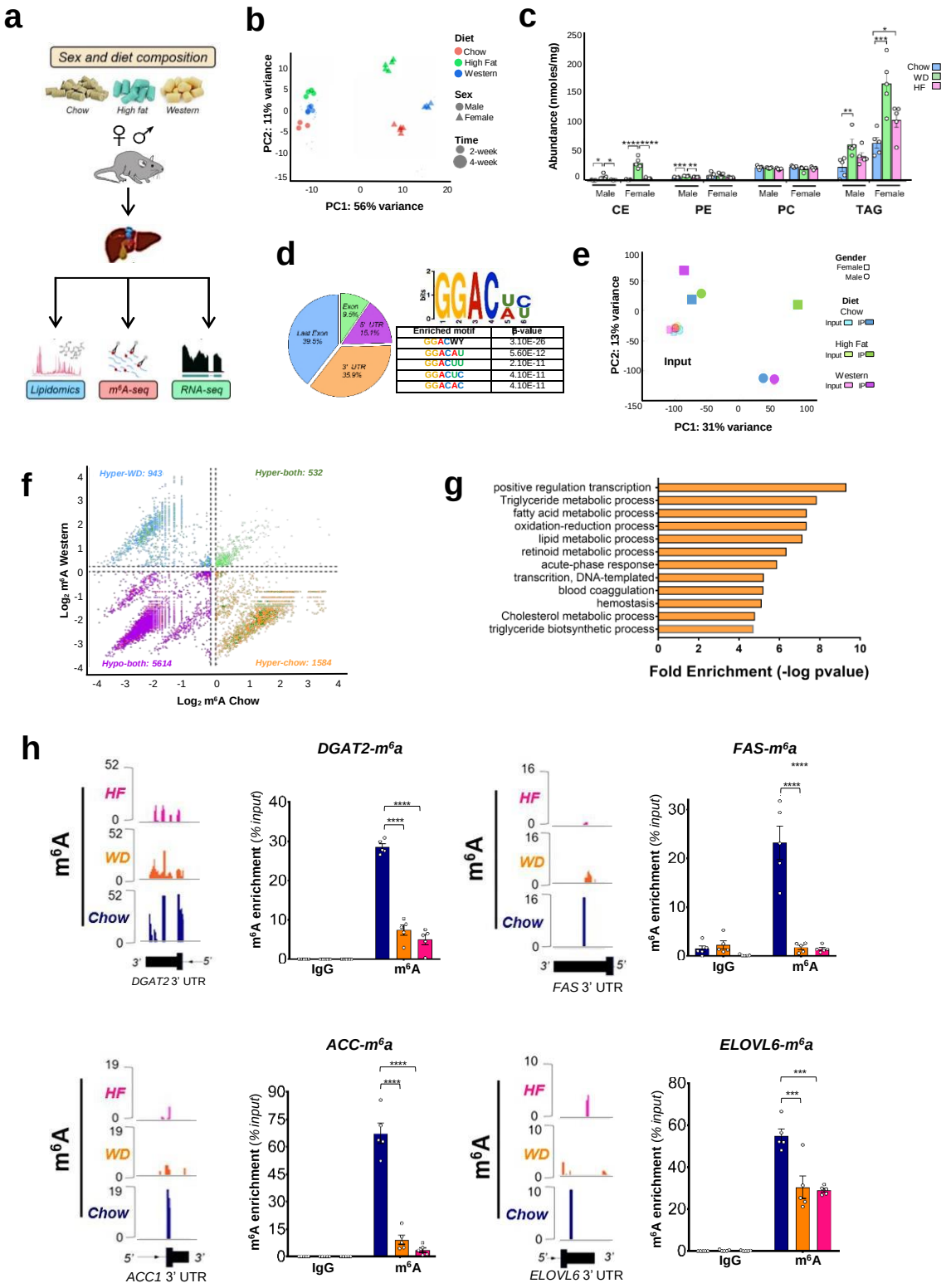
unpaired two-tailed *t*-test. Values are mean ± SEM. **d,** Quantification of total food intake for chow-fed male WT and *Mettl14* L-KO mice *(n*=8 per group). **e,** Body weight and percent fat measured by MRI for chow-fed male WT and *Mettl14* L-KO mice *(n* = 8 WT mice, *n* = 7 L-KO mice). **f,** Quantification of cellular respiratory rate in livers of chow-fed WT and *Mettl14* L-KO mice using NADH as the acceptor *(n*=5 per group). **g,** qPCR analysis of fatty acid oxidation gene expression from chow-fed WT and *Mettl14* L-KO livers *(n*=5). The experiment was repeated two times with similar results. **h,** Quantification of serum triglyceride levels from chow-fed male WT and *Mettl14* L-KO mice *(n*=8 per group). **i,** qPCR analysis of gene expression from liver of WT and *Mettl14* L-KO mice *(n*=4 per group). The experiment was repeated two times with similar results. **j,** Western blot of lipogenic protein levels in a second independent cohort of chow-fed male mice. Equal amounts of protein were pooled from eight animals per group and run in triplicate. The experiment was repeated three times with similar results. **k,** Western blot comparing levels of lipogenic proteins in the setting of chronic *Mettl14* deficiency (albumin-cre transgenics or control). Equal amounts of protein were pooled from five animals per group and run in triplicate. The experiment was repeated two times with similar results. **l,** Quantification of hepatic triglyceride concentrations from livers of albumin-cre transgenic or control mice (*n* = 4 Cre-N mice, *n* = 5 Cre-P mice). **m,** Quantification of lipogenic transcript m^6A abundance in WT and *Mettl14* L-KO chow-fed livers measured by m^6A-IP-qPCR analysis *(n* = 4). The relative enrichment of m^6A in each sample was determined by normalizing to ten-fold input. The experiment was repeated two times with similar results. **n,** Quantification of lipogenic transcript m^6A abundance in WT and *Mettl14* L-KO chow-fed livers measured by m^6A-IP-qPCR analysis *(n*=4). The relative enrichment of m^6A in each sample was determined by normalizing to *GAPDH*. The experiment was repeated two times with similar results. Mice were fed chow diet for 4 weeks and fasted for 4-hrs. prior to sacrifice. Values are mean ± s.e.m. of 4 (**l,l-n**), 5 (**f,g**), 6 (**d**), or 8 (**c,e,h**) independent biological replicates. *P* values were calculated using unpaired

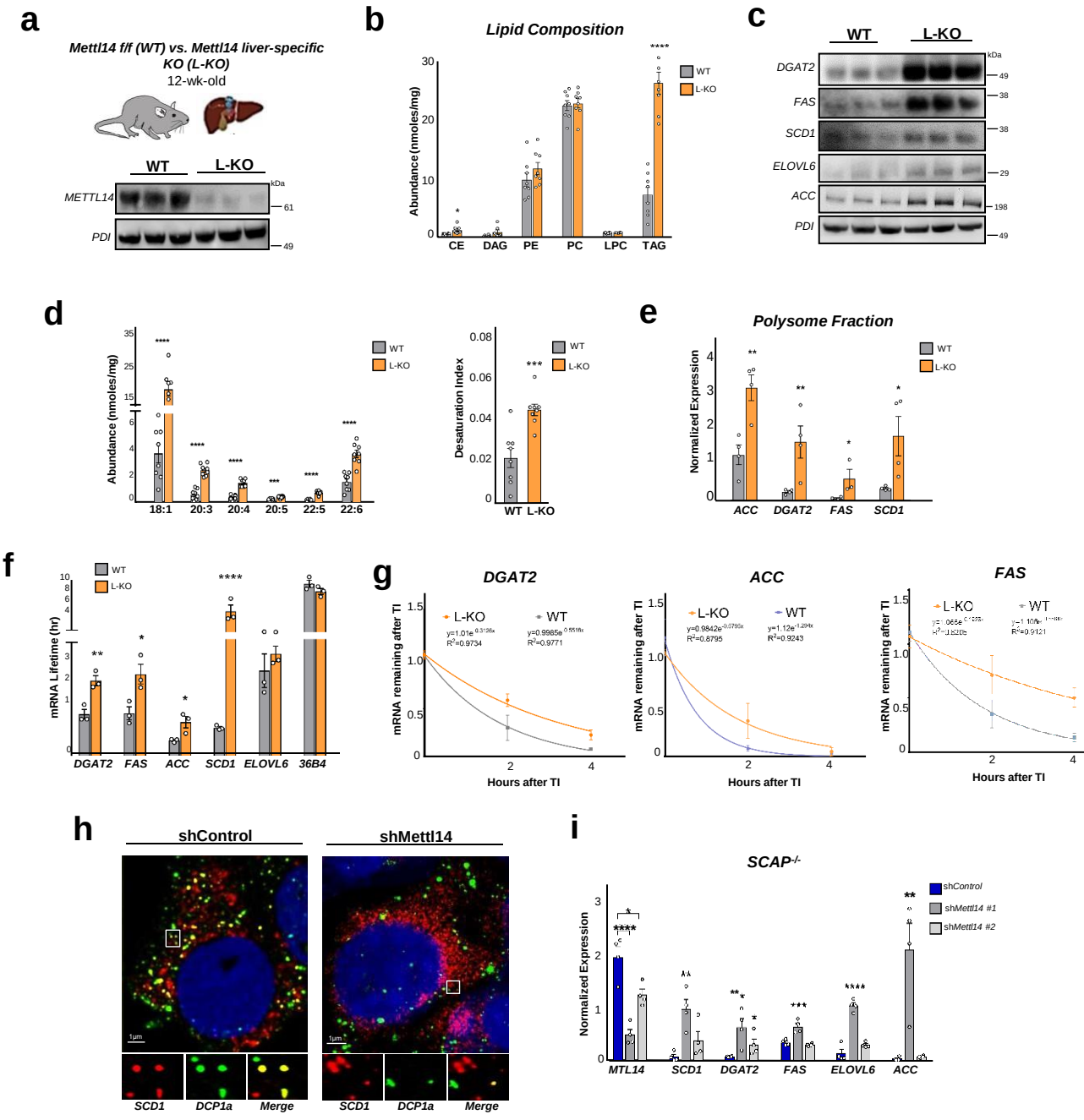
two-tailed *t*-test. (**c-i,l-n**). **P<0.05; ** P<0.01; *** P<0.001; **** P<0.0001.* The precise *n, P* values, and details of the statistical testing are provided in the source data file.

Extended Data Figure 2-4. m^{6}A regulates the stability and cytoplasmic distribution of lipogenic transcripts. a, qPCR analysis of lipogenic gene expression in livers of WT and *Mettl14* L-KO chow-fed male mice *(n*=6 per group). The experiment was repeated three times with similar results. **b,** qPCR analysis of lipogenic pre-mRNA expression using primers that amplify intronic regions from livers of WT and *Mettl14* L-KO chow-fed male mice *(n*=4 per group). The experiment was repeated three times with similar results. **c,** Western blot of SREBP-1c from nuclear fraction (mature SREBP-1 [mSREBP-1c]). mSREBP-1c levels in livers of re-fed (4 hr.) mice are shown as reference. Equal amounts of protein were pooled from five animals per group and run in triplicate. The experiment was repeated two times with similar results **d,** qPCR analysis of lipogenic transcript levels in WT or *Mettl14* L-KO livers at 72 hrs. after *Mettl14* deletion *(n*=4). **e,** Polysome profiling curve depicting three major cytoplasmic mRNA pools. **f,** qPCR analysis quantifying amount of lipogenic mRNAs in ribosome-bound fractions in livers of *Mettl14* L-KO chow-fed males compared to WT *(n*=3 per group). **g,** mRNA stability assay for *SCD1* mRNA in primary hepatocytes harvested from WT and *Mettl14* L-KO chow-fed male mice *(n* = 3 per group). **h,** qPCR analysis of *Mettl14* expression in AML cells used for single-molecule RNA FISH. *(n* = 5 per group). **i,** Quantification of hepatic triglyceride content in WD-fed male mice injected with AAV_m^{6}A writers (*Mettl14*+*Mettl3*+*WTAP*) or AAV_null (normalized to liver weight) *(n*=6 per group). Mice were fed indicated diet for 4 weeks and fasted for 4-hrs. prior to sacrifice except for mice in (**i**). Mice in (**i**) were fed NASH diet for 8-weeks and fasted 4-hrs. prior to sacrifice. Values are mean ± s.e.m. of 3 (**f,g**), 4 (**b,d**), 5 (**h**), or 6 (**a,i**) independent biological replicates. *P* values were calculated using unpaired two-tailed *t*-test. (**a,b,d,h,i**). **P<0.05; ** P<0.01; *** P<0.001; **** P<0.0001.*

Extended Data Figure 2-5. Regulation of *Mettl14* in response to dietary conditions. a, Quantification of major lipid species identified in lipidomics analysis of male WD-fed WT and *Mettl14* L-KO livers *(n* = 7 WT mice, *n* = 8 KO mice). **b,** UCSC browser screenshots showing decreased H3K27ac at *Mettl14* promoter under HF diet feeding compared to chow. Mice were fed either standard chow diet (Prolab Isopro RMH 3000, Purina) for 24 weeks or 8 weeks of standard chow diet followed by 16 weeks of HFD (Soltis *et al.* 2017. *Cell Reports*). **c,** UCSC browser screenshots showing regulation of *METTL14* promoter by *BCL6* and *STAT5a in human (*Steube *et. al* 2017. *Nat Comm ;* Gertz *et. al* 2013. *Mol. Cell; respectively*). **d,** UCSC browser screenshots showing ATAC-seq data for *Mettl14* promoter region from male mouse liver under different diets and regulation of *Mettl14* promoter by *BCL6 and STAT5a* in male and female mice fed WD (Zhang *et. al* 2011. *Mol Cel Biol*). **e,** qPCR analysis showing fast/re-feed regulation of *Bcl6* expression in livers of chow-fed males *(n*=5 per group). The experiment was repeated twice with similar results. **f,** Quantification of m^6A abundance on lipogenic transcripts infasted (4 hr) or re-fed (fasted overnight and then re-fed 4 hr) male livers as measured by m^6A- IP-qPCR (*n*=4). The experiment was repeated twice with similar results. **g,** Expression of *Mettl3,WTAP* in re-fed mice (fasted overnight and then re-fed 4 hr) compared to fasted (4 hr) mice as measured by qPCR analysis *(n*=5). **h,** qPCR analysis of *Mettl3* expression in male and female liver under various diets (*n* = 4 chow-fed females, *n* = 5 all other groups). Mice were fed the indicated diet for 4 weeks and fasted (4 hr) or re-fed (fasted overnight and then refed for 4-hrs. prior to sacrifice). Values are mean ± s.e.m. of 4 (**f**), 5 (**e,g,h**) or 7 (**a**) indpendent biological replicates. *P* values were calculated using unpaired two-tailed *t*-test (**a,e-g**) or one-way analysis of variance (ANOVA) followed by multi-group comparison (Fisher's) in **h** . **P<0.05; ** P<0.01; *** P<0.001; **** P<0.0001.* The precise *n, P* values, and details of the statistical testing are provided in the source data file.

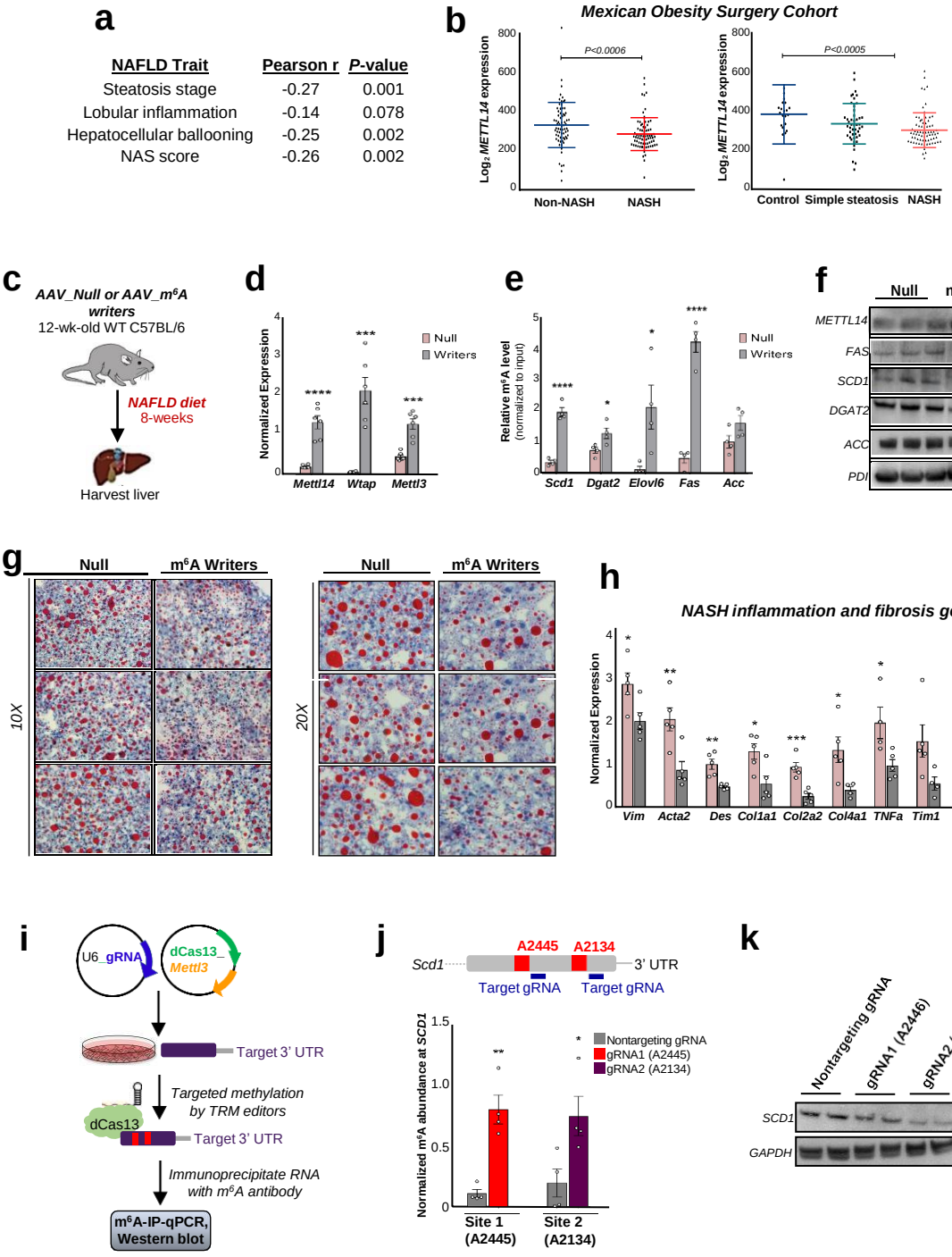
Extended Data Figure 2-6. Loss of *Mettl14* is associated with sex-biased gene expression. a, Comparison of gene expression fold changes between control and *Mettl14* L-KO mice obtained from male (x axis) or female (y axis) samples. **b,** Comparison of gene expression fold changes between male and female mice obtained from control (x axis) or *Mettl14* L-KO (y axis) samples. The linear fit of all fold changes (yellow) has a smaller slope than the Control=KO line (black), highlighting that male/female differences are higher in controls for most genes. **c,** Hierarchical clustering of samples harvested from liver based on pair-wise distances. Shown is the tree based on the Euclidean distance between all samples (*n* = 7 WT male mice, *n* = 8 mice all other groups) based on genome-wide mRNA abundance distributions**. d,** Expression of *XIST* and *SRY* in livers of WT and *Mettl14* L-KO mice as measured by qPCR (*n* = 4)**. e,** Heatmaps of selected differential genes with distinct responses to *Mettl14* L-KO. Three specific clusters of genes with sex-specific response to *Mettl14* are shown. Highlighted are the name of genes known to be involved in metabolic pathways. Gene names are ordered as in the heatmap. All mice were fed WD for 4 weeks and fasted for 4 hrs. prior to sacrifice. *P* values were calculated using unpaired two-tailed *t*-test (**d**). The precise *n, P* values, and details of the statistical testing are provided in the source data file.

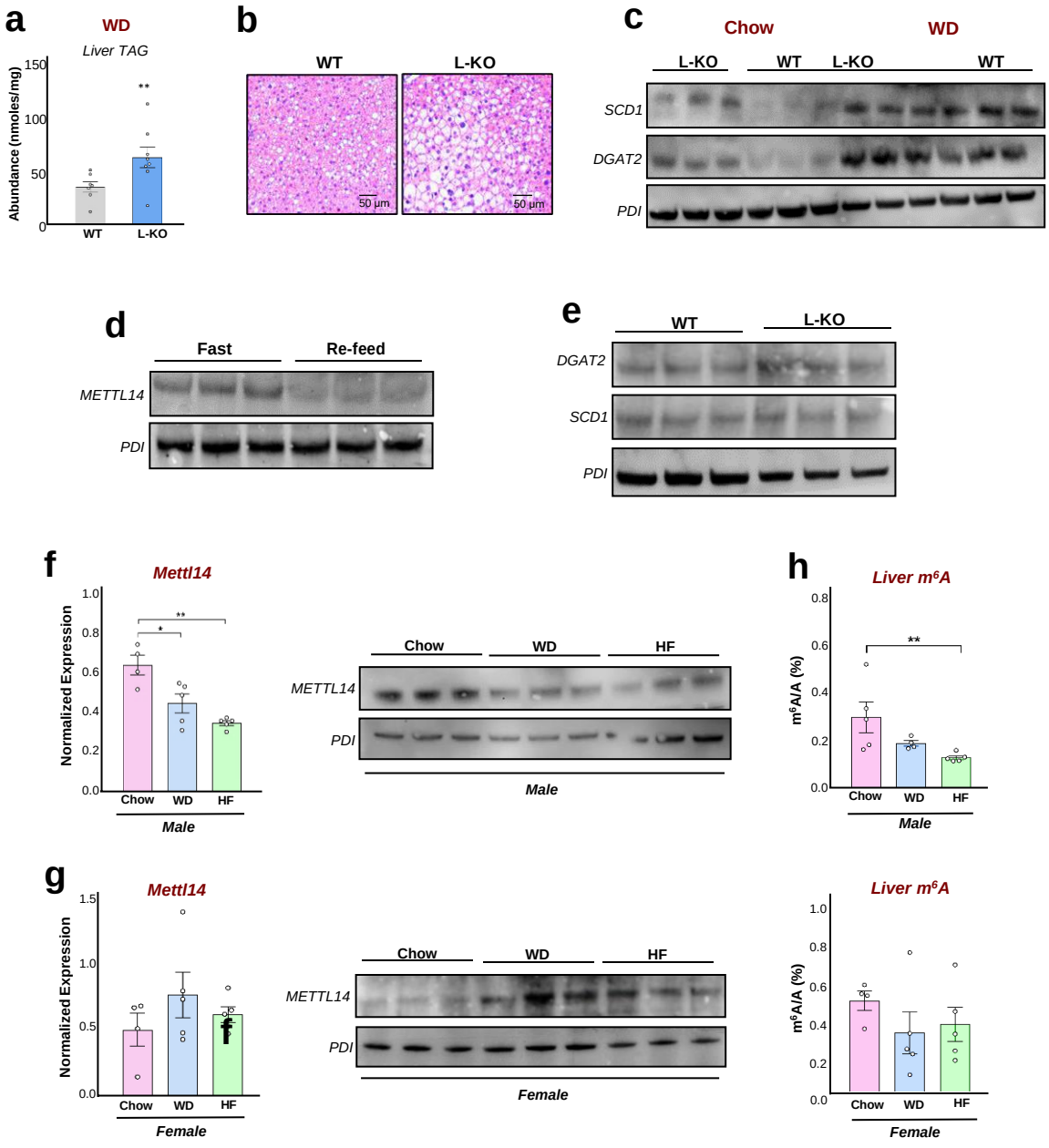

a
Sex and diet composition
Chow
High fat
Western
Lipidomics
m6A-seq
RNA-seq
b
PC2: 11% variance
PC1: 56% variance
Diet
Chow
High Fat
Western
Sex
Male
Female
Time
2-week
4-week
c
Abundance (nmoles/mg)
Chow
WD
HF
Male
Female
CE
PE
PC
TAG
d
Exon 9.5%
5' UTR 15.1%
3' UTR 35.9%
Last Exon 39.5%
Enriched motif
β-value
GGACWY 3.10E-26
GGACAU 5.60E-12
GGACUU 2.10E-11
GGACUC 4.10E-11
GGACAC 4.10E-11
e
PC2: 13% variance
PC1: 31% variance
Input
Gender
Female
Male
Diet
Chow
High Fat
Western
Input
IP
f
Hyper-WD: 943
Hyper-both: 532
Hypo-both: 5614
Hyper-chow: 1584
Log2 m6A Western
Log2 m6A Chow
g
positive regulation transcription
Triglyceride metabolic process
fatty acid metabolic process
oxidation-reduction process
lipid metabolic process
retinoid metabolic process
acute-phase response
transcrition, DNA-templated
blood coaggulation
hemostasis
Cholesterol metabolic process
triglyceride biotsynthetic process
Fold Enrichment (-log pvalue)
h
DGAT2-m6a
FAS-m6a
ACC-m6a
ELOVL6-m6a
m6A
HF
WD
Chow
m6A enrichment (% input)
IgG
m6A
DGAT2 3' UTR
FAS 3' UTR
ACC1 3' UTR
ELOVL6 3' UTR

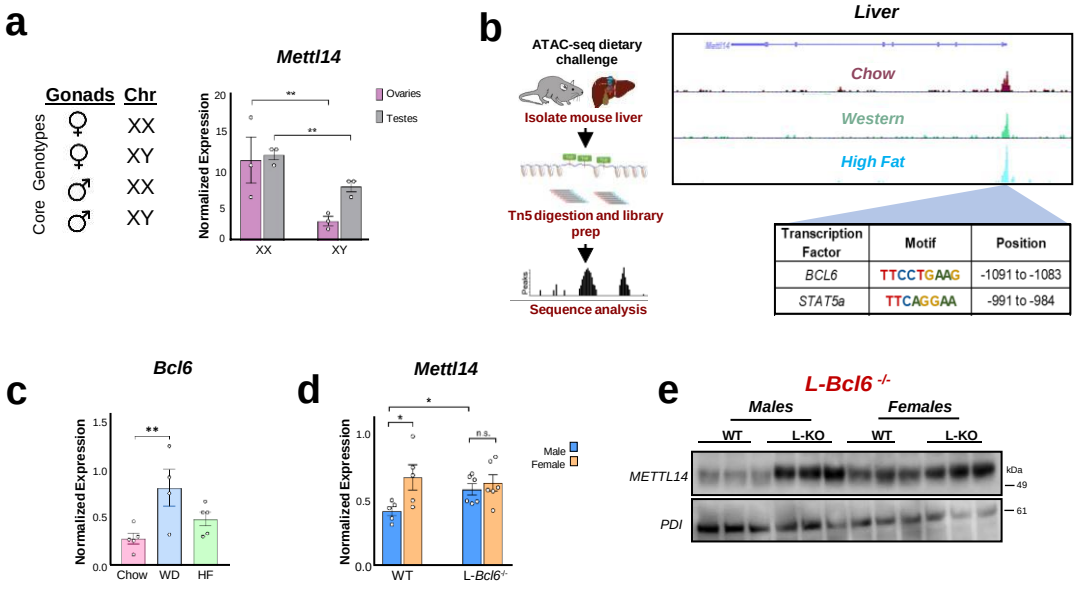

a
Mettl14 f/f (WT) vs. Mettl14 liver-specific KO (L-KO)
12-wk-old
WT
L-KO
METTL14
PDI
kDa
61
49
b
Lipid Composition
Abundance (nmoles/mg)
WT
L-KO
CE
DAG
PE
PC
LPC
TAG
c
WT
L-KO
DGAT2
FAS
SCD1
ELOVL6
ACC
PDI
kDa
d
Abundance (nmoles/mg)
18:1
20:3
20:4
20:5
22:5
22:6
Desaturation Index
e
Polysome Fraction
Normalized Expression
ACC
DGAT2
FAS
SCD1
f
mRNA Lifetime (hr)
DGAT2
FAS
ACC
SCD1
ELOVL6
36B4
g
DGAT2
ACC
FAS
mRNA remaining after TI
Hours after TI
L-KO
WT
h
shControl
shMettl14
SCD1
DCP1a
Merge
i
SCAP-/-
Normalized Expression
shControl
shMettl14 #1
shMettl14 #2
MTL14
SCD1
DGAT2
FAS
ELOVL6
ACC

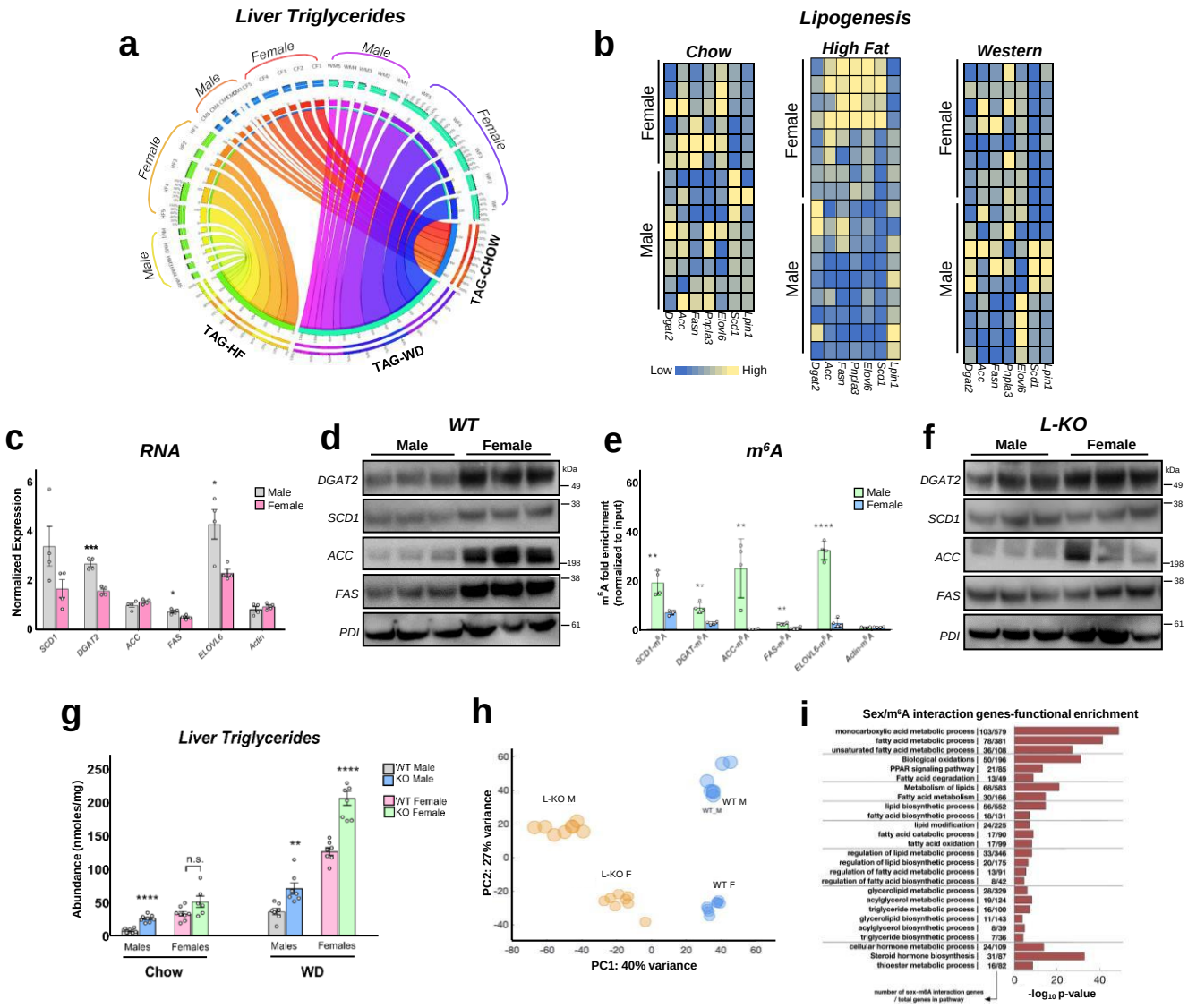
a

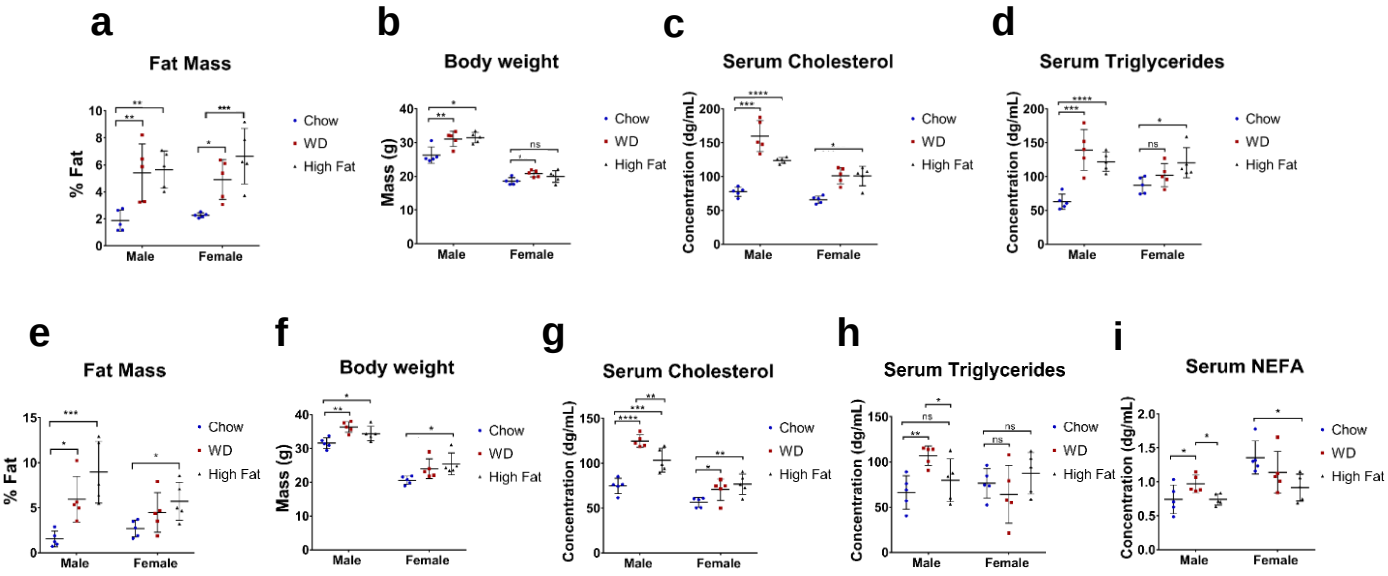
NAFLD Trait	Pearson r	*P*-value
Steatosis stage	-0.27	0.001
Lobular inflammation	-0.14	0.078
Hepatocellular ballooning	-0.25	0.002
NAS score	-0.26	0.002



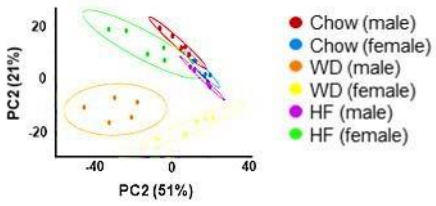

a
WD
Liver TAG
Abundance (nmoles/mg)
150
100
50
0
**
WT
L-KO
b
WT
L-KO
50 μm
50 μm
c
Chow
WD
L-KO
WT
L-KO
WT
SCD1
DGAT2
PDI
d
Fast
Re-feed
METTL14
PDI
e
WT
L-KO
DGAT2
SCD1
PDI
f
Mettl14
Normalized Expression
1.0
0.8
0.6
0.4
0.2
0.0
**
*
Chow
WD
HF
Male
Chow
WD
HF
METTL14
PDI
Male
h
Liver m6A
m6A/A (%)
0.8
0.6
0.4
0.2
0.0
**
Chow
WD
HF
Male
g
Mettl14
Normalized Expression
1.5
1.0
0.5
0.0
Chow
WD
HF
Female
Chow
WD
HF
METTL14
PDI
Female
Liver m6A
m6A/A (%)
1.0
0.8
0.6
0.4
0.2
0.0
Chow
WD
HF
Female


a
Gonads
Chr
Core Genotypes
XX
XY
XX
XY
Mettl14
Normalized Expression
Ovaries
Testes
XX
XY
b
ATAC-seq dietary challenge
Isolate mouse liver
Tn5 digestion and library prep
Sequence analysis
Liver
Chow
Western
High Fat
Transcription Factor
Motif
Position
BCL6
TTCCTGAAG
-1091 to -1083
STAT5a
TTCAGGAA
-991 to -984
c
Bcl6
Normalized Expression
Chow
WD
HF
d
Mettl14
Normalized Expression
Male
Female
WT
L-Bcl6-/-
e
L-Bcl6 -/-
Males
Females
WT
L-KO
WT
L-KO
METTL14
PDI
kDa
49
61

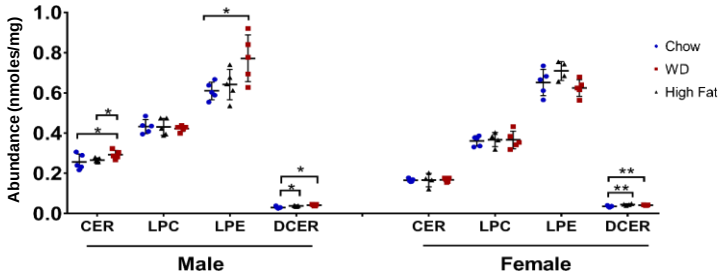

Liver Triglycerides
a
Female
Male
Male
Female
Female
Male
TAG-CHOW
TAG-HF
TAG-WD
Lipogenesis
b
Chow
High Fat
Western
Female
Male
Dgat2
Acc
Fasn
Pnpla3
Elovl6
Scd1
Lpin1
Low
High
c
RNA
Normalized Expression
Male
Female
SCD1
DGAT2
ACC
FAS
ELOVL6
Adin
d
WT
Male
Female
DGAT2
SCD1
ACC
FAS
PDI
kDa
49
38
198
38
61
e
m6A
m6A fold enrichment (normalized to input)
Male
Female
f
L-KO
Male
Female
DGAT2
SCD1
ACC
FAS
PDI
kDa
49
38
198
38
61
g
Liver Triglycerides
Abundance (nmoles/mg)
WT Male
KO Male
WT Female
KO Female
n.s.
Males
Females
Chow
WD
h
PC2: 27% variance
PC1: 40% variance
L-KO M
L-KO F
WT M
WT F
i
Sex/m6A interaction genes-functional enrichment
monocarboxylic acid metabolic process | 103/579
fatty acid metabolic process | 78/381
unsaturated fatty acid metabolic process | 36/108
Biological oxidations | 50/196
PPAR signaling pathway | 21/85
Fatty acid degradation | 13/49
Metabolism of lipids | 68/583
Fatty acid metabolism | 30/166
lipid biosynthetic process | 56/552
fatty acid biosynthetic process | 18/131
lipid modification | 24/225
fatty acid catabolic process | 17/90
fatty acid oxidation | 17/99
regulation of lipid metabolic process | 33/346
regulation of lipid biosynthetic process | 20/175
regulation of fatty acid metabolic process | 13/91
regulation of fatty acid biosynthetic process | 8/42
glycerolipid metabolic process | 28/329
acylglycerol metabolic process | 19/124
triglyceride metabolic process | 16/100
glycerolipid biosynthetic process | 11/143
acylglycerol biosynthetic process | 8/39
triglyceride biosynthetic process | 7/36
cellular hormone metabolic process | 24/109
Steroid hormone biosynthesis | 31/87
thioester metabolic process | 16/82
number of sex-m6A interaction genes / total genes in pathway
-log10 p-value


a
Fat Mass
% Fat
Male
Female
Chow
WD
High Fat
b
Body weight
Mass (g)
Male
Female
Chow
WD
High Fat
c
Serum Cholesterol
Concentration (dg/mL)
Male
Female
Chow
WD
High Fat
d
Serum Triglycerides
Concentration (dg/mL)
Male
Female
Chow
WD
High Fat
e
Fat Mass
% Fat
Male
Female
Chow
WD
High Fat
f
Body weight
Mass (g)
Male
Female
Chow
WD
High Fat
g
Serum Cholesterol
Concentration (dg/mL)
Male
Female
Chow
WD
High Fat
h
Serum Triglycerides
Concentration (dg/mL)
Male
Female
Chow
WD
High Fat
i
Serum NEFA
Concentration (dg/mL)
Male
Female
Chow
WD
High Fat

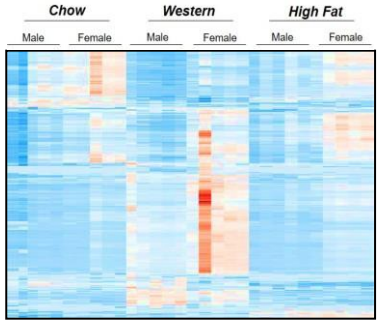
a



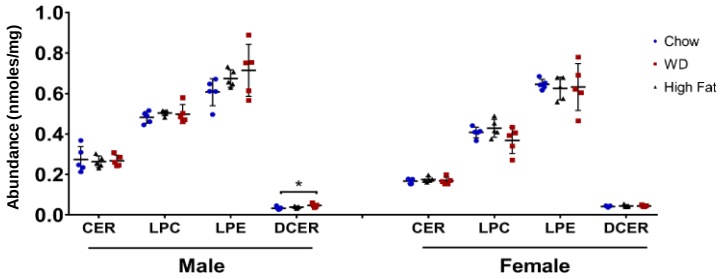
b



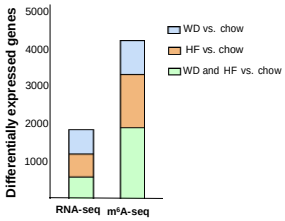
c



d



e



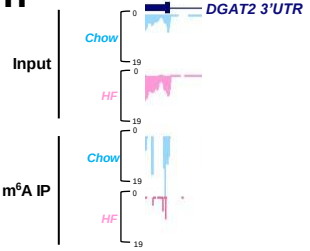
f

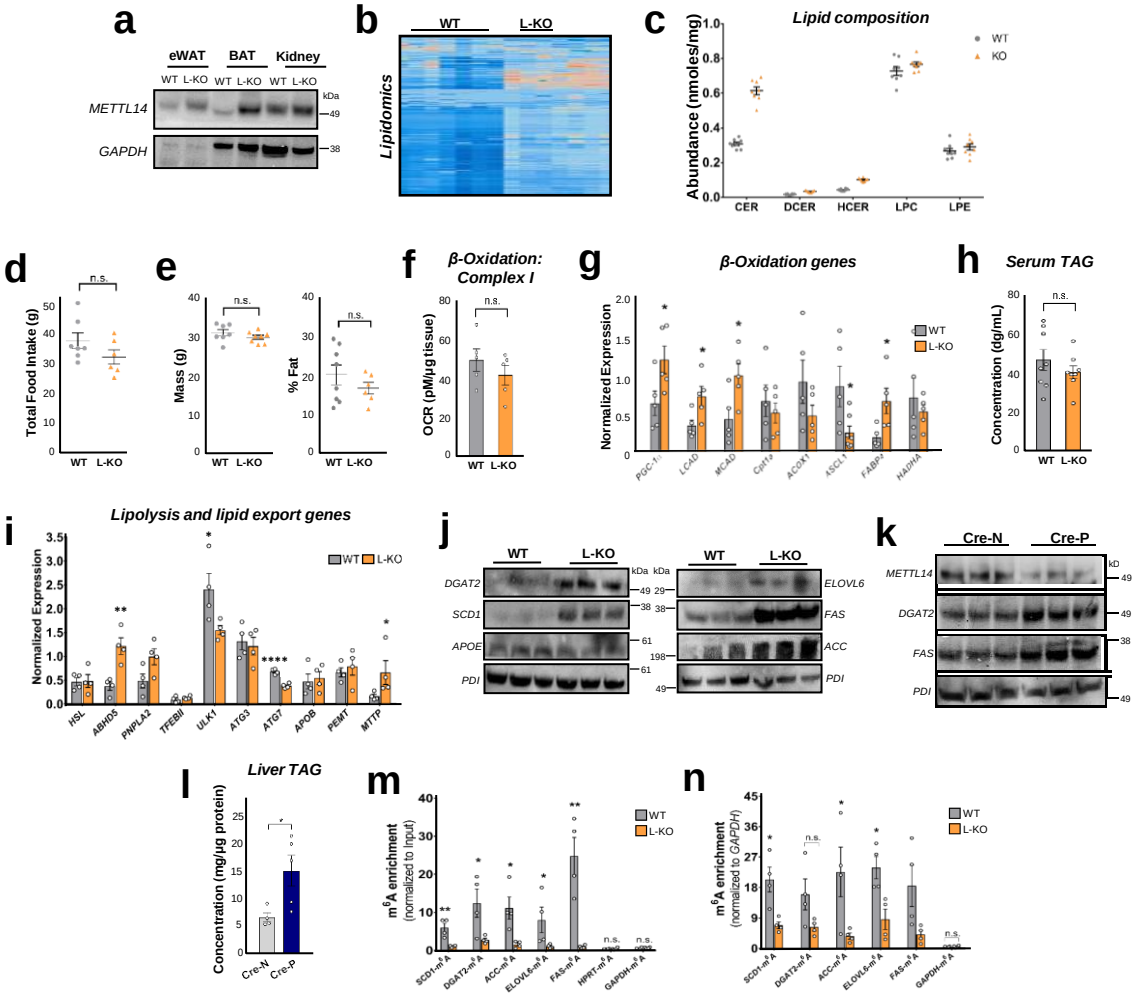
Gene	m6A motif	Position
SCD1	GGACAG	1191-1196
	GGACCC	1275-1281
	GGACAT	1604-1609
	GGACTT	1644-1649
	GGACCT	2094-2099
DGAT2	GGACTG	600-605
	GGACAC	684-689
	GGACAC	762-767
	GGACAG	884-889
ACC1	GGACCT	1364-1369
	GGACAT	1395-1400
	GGACCG	1684-1689
	GGACAC	1884-1889
FASN	GGACCT	1800-1805
ELOVL6	GGACCT	3087-3092
	GGACAA	3913-3918
	GGACGT	4071-4076
	GGACAC	4628-4633

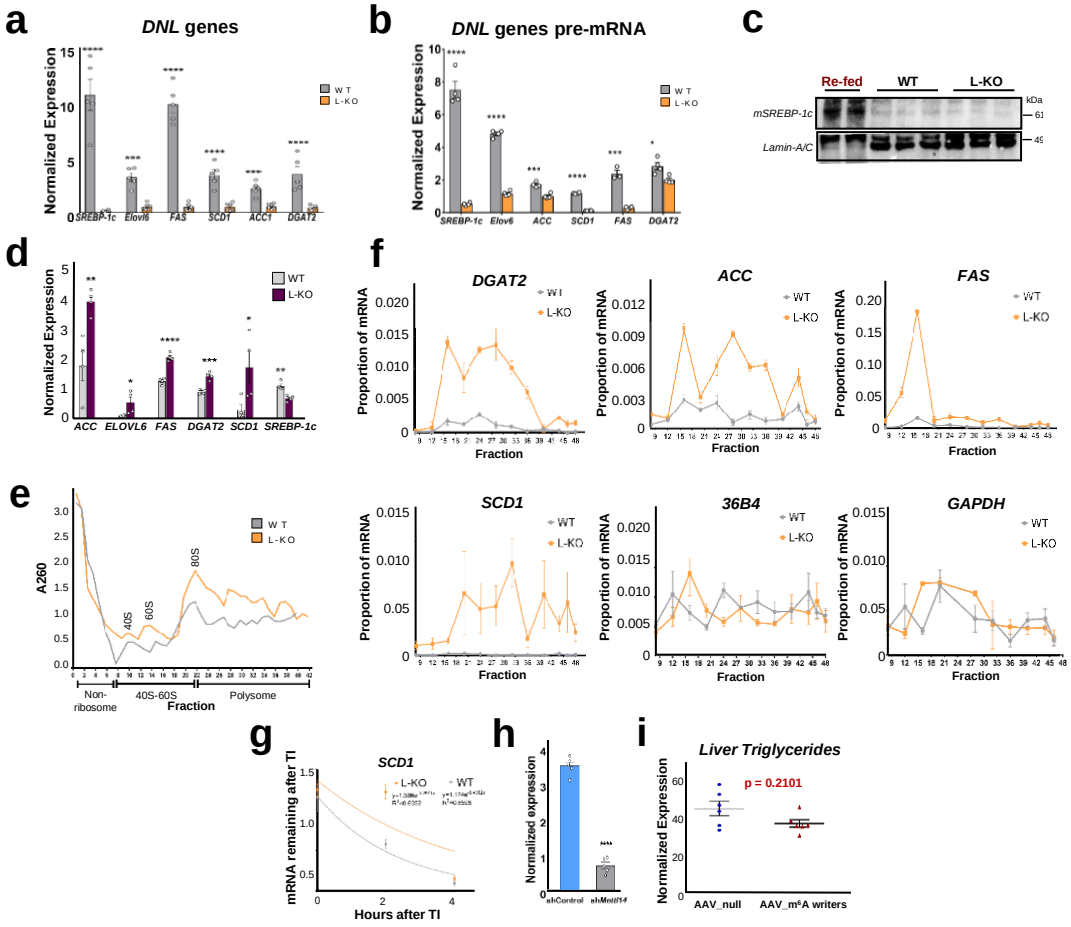
g

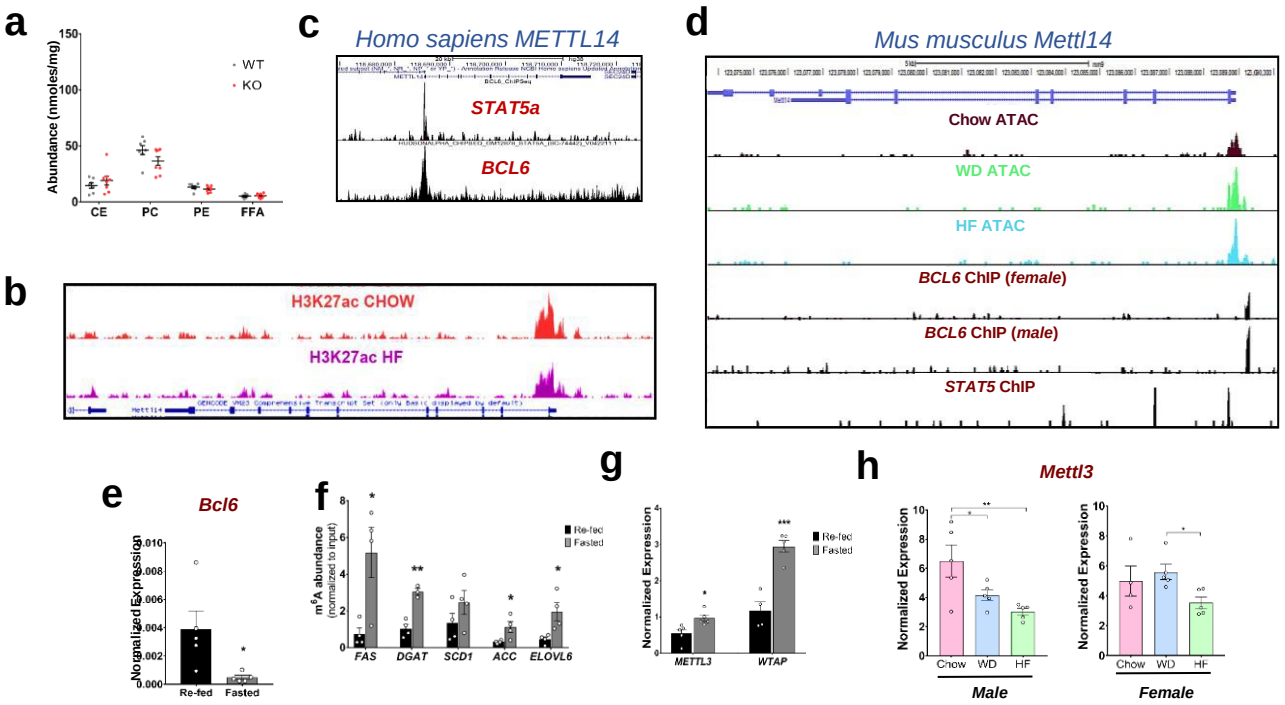
Most Fold-changed WD	
Gene	**Rank (1-10,893)**
SCD1	153
SCD1	285
SCD1	302
SCD1	316
SCD1	518
SCD1	570
DGAT2	1190

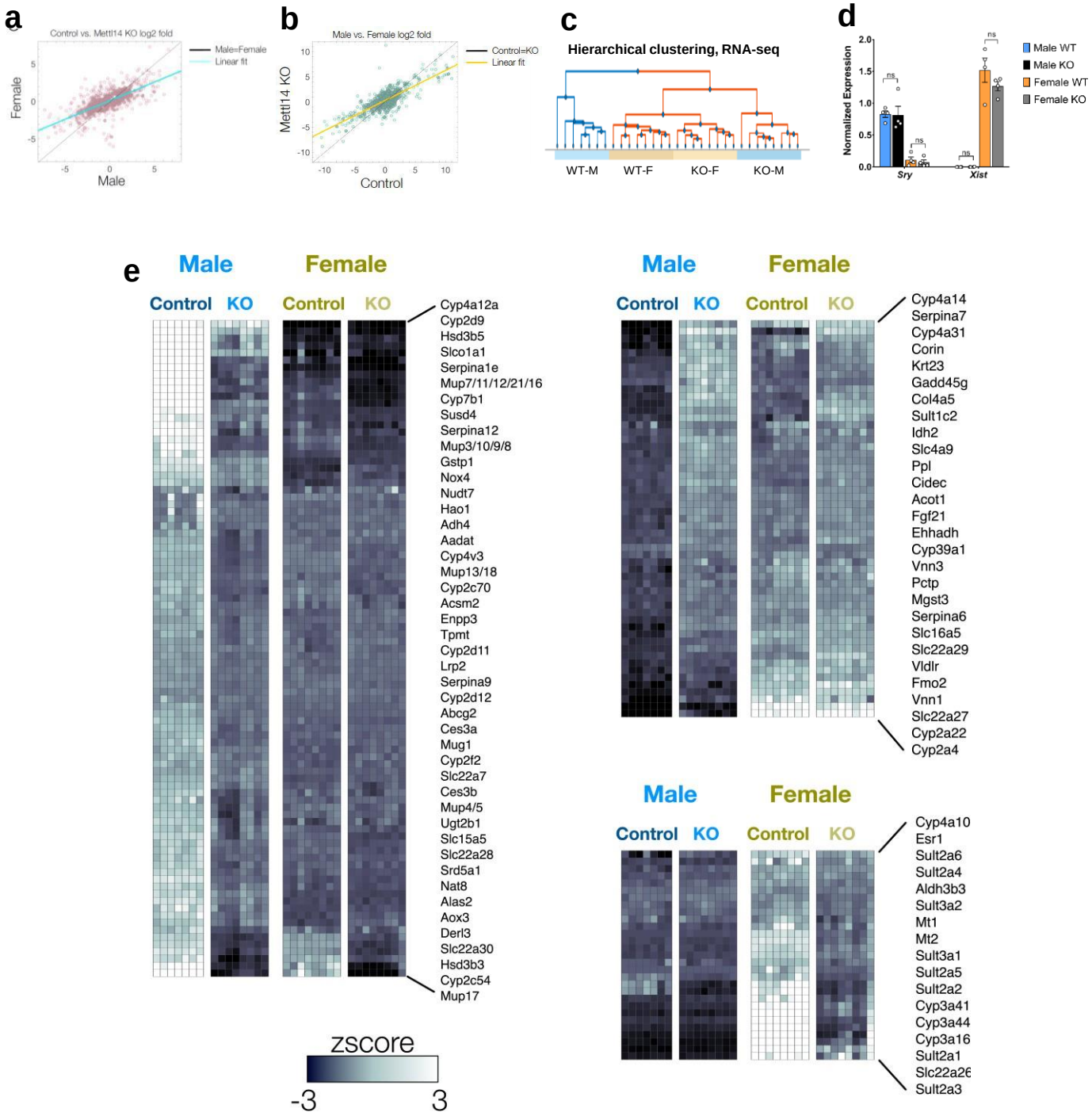
h




a
eWAT
BAT
Kidney
WT L-KO WT L-KO WT L-KO
kDa
METTL14
GAPDH
49
38
b
WT
L-KO
Lipidomics
c
Lipid composition
Abundance (nmoles/mg)
WT
KO
CER
DCER
HCER
LPC
LPE
d
Total Food Intake (g)
n.s.
WT L-KO
e
Mass (g)
% Fat
n.s.
WT L-KO
f
β-Oxidation: Complex I
OCR (pM/µg tissue)
n.s.
WT L-KO
g
β-Oxidation genes
Normalized Expression
WT
L-KO
PGC-1α
LCAD
MCAD
CPT1a
ACOX1
ASCL1
FABP2
HADHA
h
Serum TAG
Concentration (dg/mL)
n.s.
WT L-KO
i
Lipolysis and lipid export genes
Normalized Expression
WT
L-KO
HSL
ABHD5
PNPLA2
TFEB
ULK1
ATG3
ATGT
APOB
PEMT
MTTP
j
WT
L-KO
kDa
DGAT2
SCD1
APOE
PDI
ELOVL6
FAS
ACC
PDI
k
Cre-N
Cre-P
kD
METTL14
DGAT2
FAS
PDI
l
Liver TAG
Concentration (mg/µg protein)
Cre-N
Cre-P
m
m6A enrichment (normalized to Input)
WT
L-KO
SCD1-m6A
DGAT2-m6A
ACC-m6A
ELOVL6-m6A
FAS-m6A
HPRT-m6A
GAPDH-m6A
n
m6A enrichment (normalized to GAPDH)
WT
L-KO
SCD1-m6A
DGAT2-m6A
ACC-m6A
ELOVL6-m6A
FAS-m6A
GAPDH-m6A


a
DNL genes
Normalized Expression
W T
L-KO
SREBP-1c
Elovl6
FAS
SCD1
ACC1
DGAT2
b
DNL genes pre-mRNA
Normalized Expression
W T
L-KO
SREBP-1c
Elov6
ACC
SCD1
FAS
DGAT2
c
Re-fed
WT
L-KO
kDa
mSREBP-1c
Lamin-A/C
d
Normalized Expression
WT
L-KO
ACC
ELOVL6
FAS
DGAT2
SCD1
SREBP-1c
e
A260
W T
L-KO
40S
60S
80S
Non-ribosome
40S-60S
Polysome
Fraction
f
DGAT2
ACC
FAS
SCD1
36B4
GAPDH
Proportion of mRNA
WT
L-KO
Fraction
g
SCD1
mRNA remaining after TI
L-KO
WT
Hours after TI
h
Normalized expression
shControl
shMettl14
i
Liver Triglycerides
Normalized Expression
p = 0.2101
AAV_null
AAV_m6A writers


a
Abundance (nmoles/mg)
150
100
50
0
WT
KO
CE
PC
PE
FFA
c
Homo sapiens METTL14
STAT5a
BCL6
d
Mus musculus Mettl14
Chow ATAC
WD ATAC
HF ATAC
BCL6 ChIP (female)
BCL6 ChIP (male)
STAT5 ChIP
b
H3K27ac CHOW
H3K27ac HF
e
Bcl6
Normalized Expression
Re-fed
Fasted
f
m6A abundance
(normalized to input)
Re-fed
Fasted
FAS
DGAT
SCD1
ACC
ELOVL6
g
Normalized Expression
Re-fed
Fasted
METTL3
WTAP
h
Mettl3
Normalized Expression
Chow
WD
HF
Male
Female


a
Control vs. Mettl14 KO log2 fold
Male=Female
Linear fit
Female
Male
b
Male vs. Female log2 fold
Control=KO
Linear fit
Mettl14 KO
Control
c
Hierarchical clustering, RNA-seq
WT-M
WT-F
KO-F
KO-M
d
Normalized Expression
Male WT
Male KO
Female WT
Female KO
ns
Sry
Xist
e
Male
Female
Control
KO
Cyp4a12a
Cyp2d9
Hsd3b5
Slco1a1
Serpina1e
Mup7/11/12/21/16
Cyp7b1
Susd4
Serpina12
Mup3/10/9/8
Gstp1
Nox4
Nudt7
Hao1
Adh4
Aadat
Cyp4v3
Mup13/18
Cyp2c70
Acsm2
Enpp3
Tpmt
Cyp2d11
Lrp2
Serpina9
Cyp2d12
Abcg2
Ces3a
Mug1
Cyp2f2
Slc22a7
Ces3b
Mup4/5
Ugt2b1
Slc15a5
Slc22a28
Srd5a1
Nat8
Alas2
Aox3
Derl3
Slc22a30
Hsd3b3
Cyp2c54
Mup17
zscore
-3
3
Cyp4a14
Serpina7
Cyp4a31
Corin
Krt23
Gadd45g
Col4a5
Sult1c2
Idh2
Slc4a9
Ppl
Cidec
Acot1
Fgf21
Ehhadh
Cyp39a1
Vnn3
Pctp
Mgst3
Serpina6
Slc16a5
Slc22a29
Vldlr
Fmo2
Vnn1
Slc22a27
Cyp2a22
Cyp2a4
Cyp4a10
Esr1
Sult2a6
Sult2a4
Aldh3b3
Sult3a2
Mt1
Mt2
Sult3a1
Sult2a5
Sult2a2
Cyp3a41
Cyp3a44
Cyp3a16
Sult2a1
Slc22a26
Sult2a3

References

1 Horton, J. D., Goldstein, J. L. & Brown, M. S. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *The Journal of clinical investigation* **109**, 1125-1131, doi:10.1172/JCI15593 (2002).

2 Shimomura, I. *et al.* Insulin selectively increases SREBP-1c mRNA in the livers of rats with streptozotocin-induced diabetes. *Proc Natl Acad Sci U S A* **96**, 13656-13661, doi:10.1073/pnas.96.24.13656 (1999).

3 Shimano, H. *et al.* Isoform 1c of sterol regulatory element binding protein is less active than isoform 1a in livers of transgenic mice and in cultured cells. *J Clin Invest* **99**, 846-854, doi:10.1172/JCI119248 (1997).

4 Lee, K. N., Pariza, M. W. & Ntambi, J. M. Differential expression of hepatic stearoyl-CoA desaturase gene 1 in male and female mice. *Biochim Biophys Acta* **1304**, 85-88, doi:10.1016/s0005-2760(96)00145-2 (1996).

5 Horton, J. D., Shimano, H., Hamilton, R. L., Brown, M. S. & Goldstein, J. L. Disruption of LDL receptor gene in transgenic SREBP-1a mice unmasks hyperlipidemia resulting from production of lipid-rich VLDL. *J Clin Invest* **103**, 1067-1076, doi:10.1172/JCI6246 (1999).

6 Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **169**, 1187-1200, doi:10.1016/j.cell.2017.05.045 (2017).

7 Frye, M., Harada, B. T., Behm, M. & He, C. RNA modifications modulate gene expression during development. *Science* **361**, 1346-1349, doi:10.1126/science.aau1646 (2018).

8 Meyer, K. D. & Jaffrey, S. R. The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nat Rev Mol Cell Biol* **15**, 313-326, doi:10.1038/nrm3785 (2014).

9 Zaccara, S., Ries, R. J. & Jaffrey, S. R. Reading, writing and erasing mRNA methylation. *Nat Rev Mol Cell Biol* **20**, 608-624, doi:10.1038/s41580-019-0168-5 (2019).

10 Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117-120, doi:10.1038/nature12730 (2014).

11 Engel, M. *et al.* The Role of m(6)A/m-RNA Methylation in Stress Response Regulation. *Neuron* **99**, 389-403 e389, doi:10.1016/j.neuron.2018.07.009 (2018).

12 Linder, B. *et al.* Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* **12**, 767-772, doi:10.1038/nmeth.3453 (2015).

13 Weng, H. *et al.* METTL14 Inhibits Hematopoietic Stem/Progenitor Differentiation and Promotes Leukemogenesis via mRNA m(6)A Modification. *Cell Stem Cell* **22**, 191-205 e199, doi:10.1016/j.stem.2017.11.016 (2018).

14 Liu, J. *et al.* N (6)-methyladenosine of chromosome-associated regulatory RNA regulates chromatin state and transcription. *Science* **367**, 580-586, doi:10.1126/science.aay6018 (2020).

15 Sun, Z. *et al.* Hepatic Hdac3 promotes gluconeogenesis by repressing lipid synthesis and sequestration. *Nat Med* **18**, 934-942, doi:10.1038/nm.2744 (2012).

16 Lu, M. *et al.* Insulin regulates liver metabolism in vivo in the absence of hepatic Akt and Foxo1. *Nat Med* **18**, 388-395, doi:10.1038/nm.2686 (2012).

17 Ries, R. J. *et al.* m(6)A enhances the phase separation potential of mRNA. *Nature* **571**, 424-428, doi:10.1038/s41586-019-1374-1 (2019).

18 Zaccara, S. & Jaffrey, S. R. A Unified Model for the Function of YTHDF Proteins in Regulating m(6)A-Modified mRNA. *Cell* **181**, 1582-1595 e1518, doi:10.1016/j.cell.2020.05.012 (2020).

19 Attie, A. D. *et al.* Relationship between stearoyl-CoA desaturase activity and plasma triglycerides in human and mouse hypertriglyceridemia. *J Lipid Res* **43**, 1899-1907, doi:10.1194/jlr.m200189-jlr200 (2002).

20 Ou, J. *et al.* Unsaturated fatty acids inhibit transcription of the sterol regulatory element-binding protein-1c (SREBP-1c) gene by antagonizing ligand-dependent activation of the LXR. *Proc Natl Acad Sci U S A* **98**, 6027-6032, doi:10.1073/pnas.111138698 (2001).

21 Dentin, R. *et al.* Polyunsaturated fatty acids suppress glycolytic and lipogenic genes through the inhibition of ChREBP nuclear protein translocation. *J Clin Invest* **115**, 2843-2854, doi:10.1172/JCI25256 (2005).

22 Matsuda, M. *et al.* SREBP cleavage-activating protein (SCAP) is required for increased lipid synthesis in liver induced by cholesterol deprivation and insulin elevation. *Genes Dev* **15**, 1206-1216, doi:10.1101/gad.891301 (2001).

23 Shao, W., Machamer, C. E. & Espenshade, P. J. Fatostatin blocks ER exit of SCAP but inhibits cell growth in a SCAP-independent manner. *J Lipid Res* **57**, 1564-1573, doi:10.1194/jlr.M069583 (2016).

24 Xiong, X. *et al.* Landscape of Intercellular Crosstalk in Healthy and NASH Liver Revealed by Single-Cell Secretome Gene Analysis. *Mol Cell* **75**, 644-660 e645, doi:10.1016/j.molcel.2019.07.028 (2019).

25 Kleiner, D. E. *et al.* Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41**, 1313-1321, doi:10.1002/hep.20701 (2005).

26 Wilson, C., Chen, P. J., Miao, Z. & Liu, D. R. Programmable m(6)A modification of cellular RNAs with a Cas13-directed methyltransferase. *Nat Biotechnol*, doi:10.1038/s41587-020-0572-6 (2020).

27 Liu, X. M., Zhou, J., Mao, Y., Ji, Q. & Qian, S. B. Programmable RNA N(6)-methyladenosine editing by CRISPR-Cas9 conjugates. *Nat Chem Biol* **15**, 865-871, doi:10.1038/s41589-019-0327-1 (2019).

28 Link, J. C. *et al.* X chromosome dosage of histone demethylase KDM5C determines sex differences in adiposity. *J Clin Invest*, doi:10.1172/JCI140223 (2020).

29 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).

30 Sommars, M. A. *et al.* Dynamic repression by BCL6 controls the genome-wide liver response to fasting and steatosis. *Elife* **8**, doi:10.7554/eLife.43922 (2019).

31 Senagolage, M. D. *et al.* Loss of Transcriptional Repression by BCL6 Confers Insulin Sensitivity in the Setting of Obesity. *Cell Rep* **25**, 3283-3298 e3286, doi:10.1016/j.celrep.2018.11.074 (2018).

32 Kutyavin, V. I. & Chawla, A. BCL6 regulates brown adipocyte dormancy to maintain thermogenic reserve and fitness. *Proc Natl Acad Sci U S A* **116**, 17071-17080, doi:10.1073/pnas.1907308116 (2019).

33 Zhang, Y., Laz, E. V. & Waxman, D. J. Dynamic, sex-differential STAT5 and BCL6 binding to sex-biased, growth hormone-regulated genes in adult mouse liver. *Mol Cell*

34 Connerney, J., Lau-Corona, D., Rampersaud, A. & Waxman, D. J. Activation of Male Liver Chromatin Accessibility and STAT5-Dependent Gene Transcription by Plasma Growth Hormone Pulses. *Endocrinology* **158**, 1386-1405, doi:10.1210/en.2017-00060 (2017).

35 Qasem, R. J. *et al.* Decreased liver triglyceride content in adult rats exposed to protein restriction during gestation and lactation: role of hepatic triglyceride utilization. *Clin Exp Pharmacol Physiol* **42**, 380-388, doi:10.1111/1440-1681.12359 (2015).

36 Zhang, W. *et al.* FoxO1 regulates multiple metabolic pathways in the liver: effects on gluconeogenic, glycolytic, and lipogenic gene expression. *J Biol Chem* **281**, 10105-10117, doi:10.1074/jbc.M600272200 (2006).

37 Link, J. C. & Reue, K. Genetic Basis for Sex Differences in Obesity and Lipid Metabolism. *Annu Rev Nutr* **37**, 225-245, doi:10.1146/annurev-nutr-071816-064827 (2017).

38 Rinn, J. L. & Snyder, M. Sexual dimorphism in mammalian gene expression. *Trends Genet* **21**, 298-305, doi:10.1016/j.tig.2005.03.005 (2005).

39 Wang, B. & Tontonoz, P. Liver X receptors in lipid signalling and membrane homeostasis. *Nat Rev Endocrinol* **14**, 452-463, doi:10.1038/s41574-018-0037-x (2018).

40 Lonardo, A. *et al.* Sex Differences in Nonalcoholic Fatty Liver Disease: State of the Art and Identification of Research Gaps. *Hepatology* **70**, 1457-1469, doi:10.1002/hep.30626 (2019).

41 McIntyre, A. B. R. *et al.* Limits in the detection of m(6)A changes using MeRIP/m(6)A-seq. *Sci Rep* **10**, 6590, doi:10.1038/s41598-020-63355-3 (2020).

42 Wang, X. *et al.* Hepatocyte TAZ/WWTR1 Promotes Inflammation and Fibrosis in Nonalcoholic Steatohepatitis. *Cell Metab* **24**, 848-862, doi:10.1016/j.cmet.2016.09.016 (2016).

43 Rong, X. *et al.* LXRs regulate ER stress and inflammation through dynamic modulation of membrane phospholipid composition. *Cell Metab* **18**, 685-697, doi:10.1016/j.cmet.2013.10.002 (2013).

44 Sallam, T. *et al.* Feedback modulation of cholesterol metabolism by the lipid-responsive non-coding RNA LeXis. *Nature* **534**, 124-128, doi:10.1038/nature17674 (2016).

45 Tontonoz, P. *et al.* Long Noncoding RNA Facilitated Gene Therapy Reduces Atherosclerosis in a Murine Model of Familial Hypercholesterolemia. *Circulation* **136**, 776-778, doi:10.1161/CIRCULATIONAHA.117.029002 (2017).

46 Zhang, L. *et al.* Inhibition of cholesterol biosynthesis through RNF145-dependent ubiquitination of SCAP. *Elife* **6**, doi:10.7554/eLife.28766 (2017).

47 Sallam, T. *et al.* Transcriptional regulation of macrophage cholesterol efflux and atherogenesis by a long noncoding RNA. *Nat Med* **24**, 304-312, doi:10.1038/nm.4479 (2018).

48 Zhang, Z. *et al.* Collaborative interactions of heterogenous ribonucleoproteins contribute to transcriptional regulation of sterol metabolism in mice. *Nat Commun* **11**, 984, doi:10.1038/s41467-020-14711-4 (2020).

49 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

50 Casero, D. *et al.* Long non-coding RNA profiling of human lymphoid progenitor cells reveals transcriptional divergence of B cell and T cell lineages. *Nat Immunol*, doi:10.1038/ni.3299 (2015).

51 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

52 Si, Y., Liu, P., Li, P. & Brutnell, T. P. Model-based clustering for RNA-seq data. *Bioinformatics*, doi:10.1093/bioinformatics/btt632 (2013).

53 Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523, doi:10.1038/s41467-019-09234-6 (2019).

54 Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201-206, doi:10.1038/nature11112 (2012).

55 Salmon-Divon, M., Dvinge, H., Tammoja, K. & Bertone, P. PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. *BMC Bioinformatics* **11**, 415, doi:10.1186/1471-2105-11-415 (2010).

56 Bailey, T. L. & Machanick, P. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res* **40**, e128, doi:10.1093/nar/gks433 (2012).

57 Rogers, G. W. *et al.* High throughput microplate respiratory measurements using minimal quantities of isolated mitochondria. *PLoS One* **6**, e21746, doi:10.1371/journal.pone.0021746 (2011).

58 Kedersha, N. & Anderson, P. Mammalian stress granules and processing bodies. *Methods Enzymol* **431**, 61-81, doi:10.1016/S0076-6879(07)31005-7 (2007).

Chapter 3: A preclinical investigation of *MeXis* therapy in atherosclerosis

Research Letter

A preclinical investigation of *MeXis* therapy in atherosclerosis

Short title: LncRNA therapy in atherosclerosis

David A. Salisbury [1,2,3], Aneesh Kallapur[1,2,3], Josue Fraga[1,2,3], Jason Kim [1,2], Xiaohui Wu [1,2], and Tamer Sallam, M.D. Ph.D. [1,2,3*]

Affiliations:

[1] Division of Cardiology, Department of Medicine, University of California, Los Angeles, CA

[2] Molecular Biology Institute, University of California, Los Angeles, CA

[3] Molecular Biology Interdepartmental Program, University of California, Los Angeles, CA

*Corresponding author, Contact Information:
Division of Cardiology, Department of Medicine
David Geffen School of Medicine at UCLA
650 Charles E. Young South Dr. CHS A2-327
Los Angeles, CA 90095-1679
Email: tsallam@mednet.ucla.edu

Introduction

Work in the last decade established long noncoding RNAs (lncRNAs) as critical regulators of many biologic processes required for life. For example, X-inactivation and genome length regulation require the activity of lncRNAs[1]. However, the significance of lncRNAs in cardiovascular disease is still in question. Hundreds of lncRNA genes have been proposed to be critical for cardiovascular health but *in vivo* genetic perturbations have either been missing or showing subtle effects for many lncRNAs[2]. In addition, development of lncRNA-based therapies remains an aspirational goal. Our group discovered that conserved lncRNA *Mexis* acts as a key modulator of cholesterol efflux and atherosclerosis development in mice and humans[3]. *Mexis* orchestrates macrophage responses to sterol overload by boosting Abca1 expression and in part explains spatial variations in Abca1 across tissues. *Mexis* based therapies may be highly attractive since they would 1) enhance cholesterol efflux and 2) reduce inflammation. Thus, strategies to enhance *Mexis* in macrophages may in theory target multiple non-redundant pathways causality linked with risk of atherothrombotic disease.

Results

Our previous work showed that deletion of *Mexis* reduces *Abca1*, enhances inflammation and foam cell formation as well as promotes atherosclerosis development[1]. To test the therapeutic effects of lncRNAs within lesions, we use a novel genetic model that allows spatial control of *Mexis* expression from the endogenous locus. We crossed a conditional *Mexis* knockin mice with *Cre*^{LysM} to generate macrophage-specific enhanced *Mexis* expression (Figure A). Isolation of peritoneal macrophages from Cre+ mice confirmed a marked increase in *Mexis* compared with Cre- controls (Figure B). Consistent with the notion that *Mexis* acts in *trans* to boost *Abca1* levels, we observed an increase in *Abca1* mRNA and protein (Figure B). *Mexis* overexpression

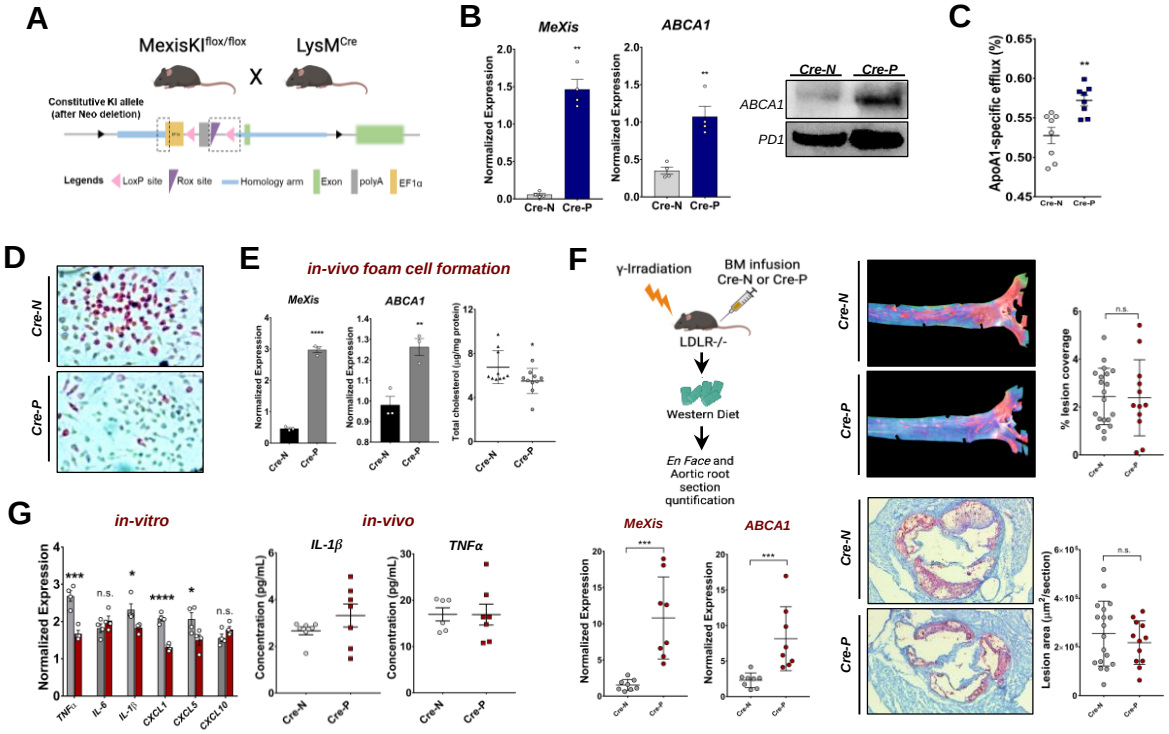
was associated with enhanced cholesterol efflux to an ApoA1 acceptor (Figure C) and reduced foam cell formation *in vitro* and *in vivo* (Figure D). To test the efficacy of Mexis-based therapy on atherosclerosis we performed a bone-marrow transplant (BMT) in *LDLR*$^{-/-}$ mice with Cre+ and Cre- marrow. *Mexis* BMT did not impact serum cholesterol, triglycerides or fasting glucose. Analysis of atherosclerosis lesions by *en face* and aortic root section did not show differences in plaque burden between groups (Figure E). Surprisingly, and in contrast to the *in vitro* results, enhanced *Mexis* expression in the atherosclerosis model led to significant elevation in systemic inflammatory markers (Figure F).

Discussion

Our results indicate that enhanced *Mexis* expression in macrophages may not be atheroprotective as hypothesized but the mechanisms at play here are highly intriguing. Despite evidence of enhanced Abca1 *in vivo* (Figure D), chronic Mexis expression was associated with unrestrained inflammation. We note that *Mexis*, like many lncRNAs, contains multiple SINE elements which are repeats thought to have arisen from integration of retroviruses into the host genome. SINE RNAs are normally expressed at low levels but enhanced SINE RNA expression triggers inflammatory signaling[4, 5]. Notably, the increase in inflammation due to chronic supraphysiologic expression of *Mexis* may have negated the potential benefit of augmenting Abca1. Our findings have important implications for proposed strategies that aim to enhance lncRNA expression for chronic disease treatment and highlights the important of careful interrogation of lncRNA effects in multiple contexts.

Figure Legends:

Figure 3-1. Sustained overexpression of *MeXis* in atherosclerosis. A. Schematic of *MeXis* knock-in strategy. **B.** expression of *MeXis* and *Abca1* in *MeXisKI*$^{f/f}$*/LysM_Cre*-positive and *MeXisKI*$^{f/f}$*/LysM_Cre*-negative donor mice as measured by qPCR and protein levels of ABCA1 in peritoneal macrophages harvested from donor *MeXisKI*$^{f/f}$*/LysM_Cre*-positive and *MeXisKI*$^{f/f}$*/LysM_Cre*-negative mice. **C.** Cholesterol efflux in the presence of ApoA1 (15µg/mL) from *MeXisKI*$^{f/f}$*/LysM*_Cre-negative and *MeXisKI*$^{f/f}$*/LysM*_Cre-positive BMDMs loaded with fluorescently-labeled cholesterol (ab196985) and treated with LXR ligand (1 µM GW3965). **D.** Oil-red-O staining of BMDMs collected from *MeXisKI*$^{f/f}$*/LysM_Cre*-positive and *MeXisKI*$^{f/f}$*/LysM_Cre*-negative mice treated with oxLDL (100ug/mL) for 72 hr. **E.** qPCR analysis of *MeXis, Abca1* expression and quantification of total cholesterol content in peritoneal macrophages collected from *MeXisKI*$^{f/f}$*/LysM_Cre*-positive and *MeXisKI*$^{f/f}$*/LysM_Cre*-negative mice injected with *AAV_PCSK9* and fed WD for 3 weeks. **F.** Schematic of BMT experiment (top left), expression of MeXis, Abca1 in aortic lesions measured by qPCR (bottom left), representative photographs and percentage of aorta surface area with atherosclerotic plaque determined through *en face* analysis from aortas of LDLR$^{-/-}$ mice irradiated with BMDMs from *MeXisKI*$^{f/f}$*LysM_Cre* mice (top right); Oil-Red-O stain of frozen sections from the aortic root and quantification of lesion area from Oil red O–stained aortic root sections of LDLR$^{-/-}$ mice irradiated with BMDMs from *MeXisKI*$^{f/f}$*LysM_Cre* mice (bottom right). **G.** qPCR analysis of inflammatory gene expression in BMDMs collected from *MeXisKI*$^{f/f}$*/LysM_Cre*-positive and *MeXisKI*$^{f/f}$*/LysM_Cre*-negative mice treated with LPS (100ng/mL) for 4 hr (left) and quantification of serum IL-1β and TNFα levels by ELISA (right).


A
MexisKI flox/flox
LysM Cre
X
Constitutive KI allele
(after Neo deletion)
Legends
LoxP site
Rox site
Homology arm
Exon
polyA
EF1α
B
MeXis
ABCA1
Normalized Expression
Cre-N
Cre-P
ABCA1
PD1
C
ApoA1-specific efflux (%)
D
Cre-N
Cre-P
E
in-vivo foam cell formation
MeXis
ABCA1
Normalized Expression
Total cholesterol (ug/mg protein)
F
γ-Irradiation
BM infusion
Cre-N or Cre-P
LDLR-/-
Western Diet
En Face and
Aortic root
section
quntification
% lesion coverage
n.s.
G
in-vitro
in-vivo
Normalized Expression
TNFα
IL-6
IL-1β
CXCL1
CXCL5
CXCL10
IL-1β
TNFα
Concentration (pg/mL)
Lesion area (μm²/section)

References:

1. Rinn JL and Chang HY. Long Noncoding RNAs: Molecular Modalities to Organismal Functions. Annual review of biochemistry. 2020;89:283-308.

2. Gao F, Cai Y, Kapranov P and Xu D. Reverse-genetics studies of lncRNAs-what we have learnt and paths forward. Genome biology. 2020;21:93.

3. Sallam T, Jones M, Thomas BJ, Wu X, Gilliland T, Qian K, Eskin A, Casero D, Zhang Z, Sandhu J, Salisbury D, Rajbhandari P, Civelek M, Hong C, Ito A, Liu X, Daniel B, Lusis AJ, Whitelegge J, Nagy L, Castrillo A, Smale S and Tontonoz P. Transcriptional regulation of macrophage cholesterol efflux and atherogenesis by a long noncoding RNA. Nat Med. 2018;24:304-312.

4. Kaneko H, Dridi S, Tarallo V, Gelfand BD, Fowler BJ, Cho WG, Kleinman ME, Ponicsan SL, Hauswirth WW, Chiodo VA, Kariko K, Yoo JW, Lee DK, Hadziahmetovic M, Song Y, Misra S, Chaudhuri G, Buaas FW, Braun RE, Hinton DR, Zhang Q, Grossniklaus HE, Provis JM, Madigan MC, Milam AH, Justice NL, Albuquerque RJ, Blandford AD, Bogdanovich S, Hirano Y, Witta J, Fuchs E, Littman DR, Ambati BK, Rudin CM, Chong MM, Provost P, Kugel JF, Goodrich JA, Dunaief JL, Baffi JZ and Ambati J. DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. Nature 2011;471:325-30.

5. Tarallo V, Hirano Y, Gelfand BD, Dridi S, Kerur N, Kim Y, Cho WG, Kaneko H, Fowler BJ, Bogdanovich S, Albuquerque RJ, Hauswirth WW, Chiodo VA, Kugel JF, Goodrich JA, Ponicsan SL, Chaudhuri G, Murphy MP, Dunaief JL, Ambati BK, Ogura Y Yoo JW, Lee DK, Provost P, Hinton DR, Nunez G, Baffi JZ, Kleinman ME and Ambati J. DICER1loss and Alu RNA induce age-related macular degeneration via the NLRP3 inflammasome and MyD88. Cell. 2012;149:847-59.

Chapter 4: A micropeptide encoded by lncRNA *AK165607* maintains homeostatic function of the lysosome and prevents accumulation of toxic protein aggregates

A micropeptide encoded by lncRNA *AK165607* maintains homeostatic function of the lysosome and prevents accumulation of toxic protein aggregates

David Salisbury[1,2,3], Jason Kim[1,3], Xiaohui Wu[1,3], and Tamer Sallam[1,3].

Affilliations:

[1]Division of Cardiology, Department of Medicine, University of California, Los Angeles, CA
[2]Molecular Biology Interdepartmental Doctoral Program, University of California, Los Angeles, CA
[3]Molecular Biology Institute, University of California, Los Angeles, CA

Abstract

Despite being annotated as noncoding, several lncRNAs have recently been shown to encode small proteins termed micropeptides. Here, we show that lncRNA *AK165607* encodes an approximately 6 kDa peptide named *ORF5. ORF5* expression is induced by inflammatory stimuli in macrophages and other immune cells. Analysis of evolutionary conservation and predicted secondary structure indicate *ORF5* forms an alpha helix and contains a conserved transmembrane domain. Mass-spectrometry identified *S100A8* as an *ORF5*-interacting protein. Loss of *ORF5* led to the intracellular accumulation of *S100A8* aggregates and triggered widespread lysosomal membrane permeabilization (LMP). Furthermore, global deletion of *ORF5* suppressed the inflammatory response in macrophages and inhibited secretion of *S100A8*. Through its conserved methionines, *ORF5* is hypothesized to provide a temporary binding groove where the hydrophobic regions of *S100A8* can transiently bind and be shielded from ROS or other forms of oxidative stress, thereby preventing premature aggregation of *S100A8*. We therefore surmise that the *ORF5* functions to maintain homeostatic function of lysosomes and prevents the accumulation of unsalvageable organelles.

Introduction

Originally termed “suicide bags” more than 50 years ago by Christian de Duve, lysosomes now play well-established roles in the degradation of both intracellular and exogenous cargo that are delivered to lysosomes during the final stage of autophagy (de Duve, 2005). In addition to this well-known catabolic function, a new function has recently been ascribed for a small subset of this ancient organelle: secretion. This particular subset of lysosomes, termed secretory lysosomes, are considered dual-function organelles because they function not only as traditional lysosomes (i.e. degradation), but also as the primary vehicles for storage and release of secretory proteins (reviewed by Blott and Griffiths, 2002). Secretory lysosomes have a central role in the immune system, as many of the effector functions of immune cells involve the release of a secretory product, such as perforin and granzymes released from cytotoxic T- cells (Isaaz, 1995). Though the exact mechanisms involved in the cellular release of misfolded proteins are not known, secretory lysosomes are proposed to be the main vehicles for this process (Borland and Vilhardt, 2017; Rabouille, 2012). Mutated aggregate-prone proteins that cause neurodegenerative diseases, such as expanded huntingtin (*HTT*) in Huntington's disease and mutated α-synuclein (*SNCA*) in Parkinson's disease, are cleared by secretory lysosomes (Jeong, 2009; Cuervo, 2004; Winslow, 2010). Accumulating evidence indicates dysfunction of the autophagy-lysosomal pathway is one of the key mechanisms underlying many common neurodegenerative diseases such as Parkinson's disease, Alzheimer's disease and Huntington's disease (Wong and Cuervo, 2010; Harris, 2012). Thus, a better understanding of the specific components involved in the lysosomal-mediated secretion of aggregation prone proteins may have important therapeutic implications for treating neurodegenerative diseases.

In macrophages, secretory lysosomes are associated with extracellular release of damage-associated molecular patterns (DAMPs), Also known as alarmins, DAMPs are cellular danger signals secreted by macrophages and other immune cells that functions as a pro-inflammatory cytokines and induce an inflammatory response in recipient cells driven by TLR4 (Vogl, 2007). An example of a pro-inflammatory DAMP that is known to be secreted by lysosomal exocytosis in macrophages is myeloid-related protein 8, also known as *MRP8* or *S100A8.* Furthermore, *S100A8* contains several hydrophobic regions that were previously shown to exhibit high propensities towards aggregate formation (Yananandra, 2009; Fritz, 2010). Binding of zinc to these regions stabilizes oligomeric forms of *S100A8* and further increases their propensity towards aggregation (Vogl, 2011). Consequently, *S100A8* has been shown to contribute to the formation of protein aggregates in neurodegenerative diseases. *S100A8* directly interacts with amyloid beta (Aβ), and zinc binding by *S100A8* has been proposed to "*seed"* the formation of Aβ plaques (Kummer, 2012; Iashchishyn, 2018)**.** Surprisingly, in addition to being the main vehicle for the secretion of *S100A8*, lysosomes have also been shown to concentrate zinc (Roh, 2012) and other metal ions within their lumen; thereby preventing harmful accumulation of these metals in the cytoplasm (Li and Kane, 2009; Polischuk, 2014). How the hydrophobic regions of *S100A8* are protected from zinc and other forms of oxidative stress while being transported to the plasma membrane in lysosomes for secretion remains unclear. Intralysosomal zinc-catalyzed oxidation of *S100A8* hydrophobic residues could trigger premature aggregation of *S100A8* within the lysosome, ultimately leading to accumulation of damaged lysosomes, impaired autophagy, and intracellular accumulation of toxic protein aggregates (Ghavami, 2010). Dysfunction of the autophagy-lysosomal pathway is thought to be one of the main mechanisms underlying common neurodegenerative diseases (Wong and Cuervo, 2010). Furthermore, given that lysosomes are also the proposed vehicles for the secretion of other aggregate-prone proteins that cause neurodegenerative diseases, such as *HTT* and *SNCA*

(Harris, 2012), a better understanding of the specific cellular factors that are involved in this process may have important therapeutic implications.

Despite being widely dismissed for years as transcriptional noise or 'junk' DNA, lncRNAs are now recognized as critical players in diverse biological processes. As a critical regulatory role for these molecules has finally begun to emerge, another novel and unexpected role for lncRNAs has also become apparent in recent years; some genes currently annotated as lncRNAs may actually encode functional micropeptides (Anderson, 2015; Bi, 2017). Micropeptides are generally classified as proteins less than 100 amino acids and vary in size from 5 to 20 kDa (Aspden, 2014). Micropeptides often exhibit a common single transmembrane domain (STMD) topology characterized by a hydrophobic alpha-helical transmembrane region separating two hydrophilic N- and C-terminal domains. Many micropeptides localize to the surface of membrane-bound organelles, such as the ER, lysosome, and mitochondria (Makarewich, 2020) where they may function as "bridge molecules" that efficiently anchor protein complexes to membranes or mediate interactions between subunits in large multisubunit protein complexes (Kim, 2000). Micropeptides can also function as chaperones that remain attached to and transiently shield the hydrophobic subunits of multimeric complexes, which can become exposed under stress, and thereby protect against protein aggregation (Muchowski, 2000)

Here we we demonstrate the macrophage-specific lncRNA *AK165607* encodes a novel micropeptide termed *ORF5.* We further demonstrate *ORF5* interacts with *S100A8* as well as additional components that are specifically associated with lysosomal-mediated secretion of aggregation-prone proteins. Loss of *ORF5* triggers accumulation of *S100A8* aggregates and

leads to widespread rupturing of lysosomes *in-vitro. In-vivo* experiments revealed loss of *ORF5* suppresses macrophage inflammatory response and impairs extracellular secretion of *S100A8.* Through its conserved methionines, we surmise *ORF5* provides a a temporary binding groove where the hydrophobic residues of *S100A8* can transiently bind and be shielded from oxidative stress, thereby preventing premature aggregation of *S100A8*. Thus, we conclude *ORF5* functions as an endogenous inhibitor of lysosome membrane permeabilization (LMP) that prevents accumulation of toxic protein aggregates, which we further surmise is essential to ensure proper secretion of inflammatory cytokines by activated macrophages.

Methods

Reagents, plasmids, and cell transfection

BB864991 was purified from LPS-treated RAW macrophages using Phusion polymerase (NEB). pSBDEST.IP (79472) and pCMV(CAT)T7-SB100 (34879) were used for overexpression of *BB864991* and originally obtained from addgene.. *EGFP-alphasynuclein-A53T (40823)* and ptf*LC3* (21074) were also obtained from addgene. Plasmids were electroporated into RAW cells using the Neon transfection system (Invitrogen). The following parameters were used for electroporation of RAW cells: pulse voltage = 1680V, pulse width = 20ms, pulse number = 2. FluoZin-3 and Lysotracker Red DND were obtained from Thermo. Acridine orange (Fisher) and Thioflavin-T (Sigma) were also used to stan cells. Sequence of gRNAs used to generate ORF5$^{-/-}$ RAW cells (Synthego) and to generate globally deficient ORF5 mice by CRISPR microinjection were: 5'-CUGGACACGUCUGCUUUGUU-3'; 5'-CUCAUGGUCUGAAUGUGCCC-3'. For immunoblot analysis, antibodies against the following proteins were used: *S100A8* (eBiosciences; 1:1,000), *actin* (Sigma; 1:10,000), *GAPDH* (Genetex, 1:500) and M2-FLAG (Sigma, 1:1000). LPS was obtained from Invitrogen. The *in vitro* translation assay was performed using the TnT Coupled Transcription/Translation System (PROMEGA).

Animals and Diets

All animals used in this study were in C57BL/6 background unless otherwise noted. *ORF5*$^{-/-}$ mice were generated by CRISPR microinjection at Embryonic Stem Cell Core (UCSD) using the following gRNAs: 5'-CUGGACACGUCUGCUUUGUU-3'; 5'-CUCAUGGUCUGAAUGUGCCC-3'. Our study used both male and female mice. Mice were housed in a temperature-controlled room under a 12-h light/12-hr dark cycle and pathogen-free conditions. Mice were fed standard chow diet (Research Diets). All mouse experiments were approved by the UCLA institutional Animal Care and Research Advisory Committee.

Cell Culture

RAW 264.7 cells were originally obtained from ATCC and cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. In some cases, cells were also treated with LPS (100ng/mL, 4hr). Bone marrow-derived macrophages (BMDMs) were isolated from femurs of male and C57BL/6 mice. Cells were treated with RBC lysis buffer for 3 min to remove red blood cells, centrifuged at 365 ´g for 5 min, and resuspended in 10% BMDM medium containing DMEM plus 2 mM L-glutamine, 00 µg/mL penicillin/streptomycin, 500 µM sodium pyruvate, and 5% v/v conditioned media containing macrophage colony
stimulating factor (M-CSF) produced by CMG cells to induce differentiation to BMDMs.. BMDMs were differentiated for 6 days prior to experiments, and medium was changed at day 4 of differentiation. All cells were incubated at 37°C in a humidified incubator containing 5% CO_2.

CRISPR/Cas9-mediated homologous recombination in iBMX cells

A single-guide RNA (sgRNA) targeting immediately upstream of the stop codon in the *BB864991* locus was cloned into the sgRNA/Cas9 expression vector px330 (addgene). The donor vector was constructed with a single FLAG epitope tag in-frame with the ORF5 coding sequence flanked by ~500 base pair homology arms specific to the *BB864991* locus. ORF5-FLAG knockin clones were generated by transient cotransfection and selected with puromycin. Detection of endogenous ORF5-FLAG peptide was performed by immunoblotting with a mouse anti-FLAG antibody (Sigma) on protein lysates immunoprecipitated with magnetic mouse anti-FLAG agarose beads (Sigma).

Gene expression and Immunoblot analysis

For gene expression analysis, total RNA was isolated using TRIzol reagent (Invitrogen) and

Green Master Mix (Diagenode) on BioRad Real-time PCR instrument. Gene expression levels were determined by using a standard curve. Each gene was normalized to the housekeeping gene 36B4, β-actin, or GAPDH. Total cellular protein was isolated from approximately 300mg of frozen liver tissue or cells of a 6-well plate using RIPA lysis buffer with protease inhibitor cocktail (Roche). Protein concentration was determined by BCA assay (Invitrogen), diluted in Nupage loading dye (Invitrogen), heated at 80 °C for 5 min, and run on 4–12% NuPAGE Bis-Tris Gel (Invitrogen) at 200V for 30min. Proteins were transferred to 0.22μM nitrocellulose membranes (Invitrogen) for 1 hr at 100V on ice and blocked with 5% milk in TBS-T to quench nonspecific protein binding. Samples were blotted with the indicated primary antibodies and detected using Alexa-488, Alexa-647, or IgG HRP-conjugated secondary antibodies. Images of uncropped blots with size markers are presented in figure S10.

RACE

The 5′ and 3′ ends of *AK165607* were defined using LPS-treated RAW macrophage RNA and the FirstChoice RLM-RACE kit (Ambion) according to the manufacturer's protocol with modifications. Briefly, for 5′ RACE, degraded mRNA 5′ ends were dephosphorylated with calf intestinal phosphotase (CIP) and full-length mRNA was uncapped with tobacco acid pyrophosphotase (TAP). Following 5′ RACE adaptor ligation, reverse transcription was performed using the SuperScript III First-Strand Synthesis system (Invitrogen) and MeXis-specific primers. For 3′ RACE, RNA was reverse transcribed using the SuperScript III First-Strand Synthesis system (Invitrogen) and adaptor-linked oligo(dT). The resulting cDNA was amplified by nested PCR across a 55–65 °C melting temperature gradient using Phusion polymerase (NEB), with the inner primers containing attB sequences. Aliquots of reactions were inspected on 1% agarose gels for product size and abundance. Products of selected PCR reactions were purified using the NucleoSpin Gel and PCR Cleanup kit (Clontech) and were

inserted into pDONR221 by Gateway cloning procedures. Cloned fragments were sequenced and then aligned to the mouse genome with the BLAST analysis tool.

Subcellular fractionation

A near confluent monolayer of CMT93 cells (3×10^8 cells) were serum starved for 2 h before homogenization to increase the lysosomal mass. The cells were rinsed two times with PBS and then collected by centrifugation. Pellets were lysed in lysis buffer (20 mM Hepes-KOH, pH 7.5, 10 mM KCl, 1.5 mM MgCl2, 1 mM EDTA, 1 mM EGTA, 250 mM sucrose, 1 mM DTT) on ice for 30 mins. Cells were centrifuged at 2600 × *g* for 10 min at 4 °C to remove large cell debris and nuclei. The supernatant was transferred to a new centrifuge tube and centrifuged at 20,000 × *g* for 20 min at 4 °C to pellet down the crude lysosomal fraction. Pellets were lysed in lysis buffer containing 1× Protease Inhibitor Cocktail and analyzed by western blotting.

Transmission Electron Microscopy

0.2×10^5 RAW cells were collected and centrifuged at 1000 *g* × 5 min. The pellet was washed in PBS and centrifuged once more at 1000 *g* for 5 min. Pellet was re-suspended in 2.5% glutaraldehyde in 0.15 M sodium cacodylate, pH 7.4 (EM-fix solution), and incubated at RT overnight. Cells were then prepared for immunostaining and transmission electron microscopy. Briefly, ultrathin sections of RAW cells were subjected to antigen retrieval with metaperiodate and then incubated overnight at 4°C with primary antibodies (anti-S100A8) diluted 1:100 in PBS. Detection was performed with rabbit secondary antibody conjugated with 5 nm colloidal gold (Electron Microscopy Sciences, Fort Washington, Pa., USA; titer 1:10–1:20). Specimens were examined in a Philips/FEI CM100 BioTWIN at 60 kV accelerating voltage. Images were recorded with a side-mounted Olympus Veleta camera.

Statistical analysis

A non-paired Student's t-test was used to determine statistical significance, defined at $P < 0.05$. For multiple group experiments analysis of variance was used followed by multiple group analysis. Unless otherwise noted, error bars represent standard error of mean. Experiments were independently performed at least twice. Sample size is based on statistical analysis of variance and prior experience with similar in vivo studies.

Results:

Production of a novel micropeptide from lncRNA *AK165607.*

lncRNA *AK165607* is located on chromosome 2 and encodes multiple alternatively spliced isoforms (Fig. 4-1A). *RACE* analysis of LPS-treated RAW cells identified the majority of these isoforms, as well as previously unannotated splice variants containing novel exons (Fig. 4-1B). One of *AK165607*-encoded transcripts, *BB864991,* is a 487 bp lncRNA. *BB864991* expressionis predominantly expressed in macrophages and is induced by various forms of inflammatory stimuli (Fig. 4-1C, D). In particular, *BB864991* expression is strongly induced by agonists of *TLR3* and *TLR4*; while *TLR7* and *TLR9* agonists repress its expression. ChIP-seq from other studies (Link, 2018) shows occupancy of NFκβ at the *BB864991* promoter increases substantially in response to LPS treatment (Fig. 4-1E). Given some lncRNAs have been found to encode functional micropeptides (Anderson, 2015; Bi, 2017), we analyzed publicly available Ribo-seq data to determine whether or not *BB864991* had been found associated with the ribosome. Surprisingly, we found evidence suggesting this lncRNA may indeed undergo translation. To determine if *BB864991* encodes a detectable micropeptide, we generated a construct that contained an in-frame FLAG tag immediately upstream of the predicted stop codon in *BB864991* and overexpressed it in HEK293 cells. Western blot analysis of the cell lysate revealed translation of an approximately 6 kDa micropeptide, which we termed *ORF5* (Fig. 4-1F). Other potential ORFs in *AK165607* that had been identified from Ribo-seq data did not show evidence of translation. Furthermore, immunofluorescence microscopy in these cells revealed the peptide localized predominantly to the cytoplasm (Fig. 4-1G). We then performed an *in-vitro* translation assay to corroborate the findings from HEK293 cells, as this method is slightly less artificial then overexpression of a FLAG-tagged construct. *In-vitro* translation revealed translation of an approximately 6 kDa peptide, consistent with what was observed in HEK293 cells (Fig. 4-1H). To more definitvely show *BB864991* does in fact encode a cryptic

micropeptide, we then generated an endogenous knock-in cell line; widely considered to be the gold standard for assaying potential production of a micropeptide from a loci of interest. After confirming successful knock-in of an in-frame 3X-FLAG tag immediately upstream of the predicted STOP codon in the endogenous *BB864991* locus, we confirmed production of a novel micropeptide from *BB864991* when an approximately 6 kDa band was detected using cell lysate harvested from the newly generated knock-in cells (Fig. 4-1I).

ORF5 is predicted to form a conserved transmembrane α-helix and interacts with _S100A8._

Structural modeling and bioinformatics analysis predicted the 37 amino acid micropeptide *ORF5* formed a transmembrane alpha helix (Fig. 4-2A). The presence of a potential transmembrane domain was confirmed using a combination of different softwares including Phobius, TMHM, and Phyre (Fig. 4-2B,C)**.** Sequence alignment of homologous ORFs identified in other species revealed the TMD-containing portion of *ORF5* was highly conserved across species (Fig. 4-2D,E). When the structure of *ORF5* was compared to structures of other proteins in the PDB, structure of *get3* bound to TMD of *pep12* was identified as top threading template (Fig. 4-2F).

To identify potential *ORF5*-interacting proteins, we performed mass-spectrometry analysis and identified *S200A8* as an interacting partner (Fig. 4-2G). Interaction between *S100A8* and *ORF5* was confirmed through co-immunoprecipitation experiments that showed *S100A8* was easily detectable by western blot using the IP eluate of cells transfected with 3X-FLAG_*ORF5* constructs (Fig. 4-2H). Helical wheel analysis, which can be used to elucidate the physical geometry of particular residues on a protein based upon predicted interactions between various regions of the given protein, revealed all of the conserved methionine residues on *ORF5* were oriented in the same direction (Fig. 4-2I). Importantly, the predicted spatial orientation of the ORF's conserved methionine resiudes wherein each residue was facing the same direction, suggests the α-helical *ORF5* peptide could potentially provide a hydrophobic groove where the

hydrophobic and highly aggregation-prone regions of *S100A8* can temporarily bind and be transiently shielded from ROS and other forms of oxidative stress, thereby preventing premature aggregation of the protein (Fig. 4-2J).

Loss of *ORF5* leads to accumulation of *S100A8* oligomers and activates autophagy.

To determine how loss of *ORF5* affected *S100A8* function, the localization of *S100A8* was determined in both WT and *ORF5*$^{-/-}$ RAW macrophages by electron microscopy. In untreated WT cells, *S100A8* was visible in distinct cytoplasmic foci as well as small patches associated with the plasma membrane (Fig. 4-3A). In untreated *ORF5*$^{-/-}$ macrophages, mild aggregation of *S100A8* was observed in the cytoplasm; although the overall morphology of these cells appeared healthy and did not differ substantially from that of WT cells. Upon LPS treatment, however, differences in cellular distribution of *S100A8* became readily apparent (Fig. 4-3B). In WT cells, LPS treatment resulted in translocation of *S100A8* to the plasma membrane where it was clearly visible in vesicles being secreted from the cell. Treatment of *ORF5*$^{-/-}$ cells with LPS resulted in the accumulation of autophagic vesicles containing *S100A8*. *ORF5* KO cells exhibited other ultrastructural features characteristic of autophagy activation, such as an increase in cytoplasmic vacuoulization, necrosis, and an increase in both the number and size of lysosomes and autolysosomal vesicles, likely due to the accumulation of toxic *S100A8* aggregates within these structures. *ORF5* KO cells also showed increased accumulation of *LC-3* (Fig. 4-3F); further corroborating the aberrant activation of autophagy observed by EM. Additionally, qPCR analysis revealed expression of autophagy genes was significantly increased in *ORF5*$^{-/-}$ RAW macrophages compared to WT (Fig. 4-3E). To confirm autophagy was actually increased in the *ORF5* KO cells, staining with LysotrackerRed (LTR) was performed. LTR exhibits enhanced fluorescence when present in highly acidic cellular compartments, such as autophagosomes and related autolysosomal vesicles where pH falls

below 4.5. Staining with LTR confirmed drastic increase in the number of autophagosomes and autolysosomal vesicles in *ORF5* KO cells, further corroborating EM, qPCR, and western blotting results (Fig. 4-3G). Importantly, the aberrant activation of autophagy could be reversed in *ORF5* KO cells by overexpression of *ORF5* (Fig. 4-3G).

Accumulation of *S100A8* aggregates has been shown to induce autophagy (Ghavani, 2010). To determine if an increase in the amount of aggregated proteins was responsible for the aberrant activation of autophagy observed in *ORF5* KO cells, cells were stained with Thioflavin-T (ThT). Thioflavin-T is a cell-permeable benzothiazole *dye* that exhibits enhanced fluorescence upon binding to aggregated proteins. Specifically, ThT recognizes the common "cross-β" architecture that is unique to aggregated proteins. ThT staining in WT and *ORF5*$^{-/-}$ macrophages revealed essentially no detectable aggregates in WT cells whereas *ORF5* KO cells exhibited a dramatic increase in intracellular protein aggregates (Fig. 4-3C). Furthermore, staining with *FluoZin3* revealed the increase in protein aggregates observed in *ORF5*$^{-/-}$ cells was accompanied by a large increase in intracellular zinc deposits (Fig. 4-3D), which could be due lysosome rupture and defective autophagy. Thus, *ORF5* protects against accumulation of *S100A8* and aberrant activation of autophagy in macrophages.

Accumulation of *S100A8* aggregates in *ORF5*$^{-/-}$ cells impairs degradative capacity of lysosomes and triggers widespread rupturing of lysosomes.

When an impairment in autophagic flux is detected, transfection of tandem RFP_GFP-tagged *LC3* can be used to visualize which step in the process of autophagy is responsible for the observed deficiency. The acidic environment of the lysosome causes the GFP signal to be quenched upon fusion of the autophagosome with the lysosome, leading to loss of the GFP

signal. When this construct was overexpressed in WT cells, enhanced RFP fluorescence was observed whereas both GFP and RFP fluorescence was observed in *ORF5* KO cells (Fig. 4-4A), indicating autophagic flux was impaired. However, marginal RFP fluorescence could still be detected in *ORF5* KO cells, indicating fusion between lysosomes and autophagosomes was still taking place; albeit to a lesser extent than in WT cells. While deficiencies in autophagosome maturation would also result in both GFP and RFP fluorescence remaining visible in the *ORF5* KO cells, these cells showed increased accumulation of *LC-3* and exhibited a drastic increase in the number of autophagosomes and autolysosomal vesicles following exposure to LPS, suggesting autophagosome maturation was not compromised in these cells. Based on this, we hypothesized a reduction in the number of intact lysosomes could be responsible for the failure to quench GFP fluorescence as was observed in *ORF5* KO cells. Western blot analysis and qPCR analysis of gene expression revealed markers of lysosomal stress were elevated at both the RNA and protein levels in *ORF5* KO cells (Fig. 4-4B,C). Furthermore, western blotting of lysosome-enriched fraction revealed a marked accumulation of *S100A8* oligomers in the lysosomes of *ORF5*$^{-/-}$ cells (Fig. 4-4D). Additionally, TEM analysis revealed a massive accumulation of swollen and ruptured lysosomes in *ORF5*$^{-/-}$ cells treated with LPS (Fig. 4-4E), strongly suggesting accumulation of *S100A8* aggregates could trigger lysosome rupture. *ORF5* KO cells also showed signs of lysophagy (Fig. 4-4E), evident by the appearance of the hallmark double-bilayer autophagosomal membrane which begins to form around ruptured lysosomes amd mediates their clearance by autophagy. In contrast, LPS treatment in WT cells resulted in the formation of numerous endolysosomes and endolysosomal vesicles (Fig. 4-4E), which indicates successful clearance and secretion of *S100A8*. Thus, these results strongly suggest that accumulation of *S100A8* aggregates in the lysosomes of *ORF5* KO cells leads to widespread lysosome membrane permeabilization (LMP) and the accumulation of unsalvageable lysosomes. In support of this hypothesis, a substantial reduction in the number of

intact lysosomes was observed in ORF5-/- macrophages stained with acridine orange (Fig. 4-4F). To determine if lysosomal-mediated secretion of other aggregation-prone proteins was also compromised in *ORF5* KO cells, both WT and KO cells were transfected with a plasmid for overexpressing the *SNCA* (A93T) mutant that forms aggregates in PD. While only minor accumulation of the *SNCA*(A93T) mutant was observed in WT cells, aggregates of *SNCA*(A93T) accumulated rapidly throughout the cytoplasm in *ORF5* KO cells (Fig. 4-4G), suggesting secretion of other proteins transported by secretory lysosomes is also likely to be impaired in *ORF5* KO macrophages. Collectively, these results imply hydrophobic domain shielding of *S100A8* hydrophobic regions by *ORF5* within lysosomes is essential for maintaininghomeostatic function of the lysosome and ensuring proper removal of protein aggregates (Fig. 4-4H).

Inflammatory response and secretion of *S100A8* are impaired in *ORF5*$^{-/-}$ mice.

To determine consequence of *ORF5* deficiency *in-vivo*, mice globally deficient in *ORF5* were generated through the deletion of a 550 bp region of *BB864991* encoding *ORF5* by CRISPR microinjection (Fig. 4-5A). qPCR analysis confirmed successful deletion of *ORF5* in these animals. Given the well-established role of *S100A8* as a pro-inflammatory DAMP that amplifies the *TLR4*-driven inflammatory response, we compared the expression of inflammatory cytokine genes in BMDMs harvested from WT and *ORF5*$^{-/-}$ mice. Following LPS treatment, *ORF5*$^{-/-}$ BMDMs exhibited a suppressed inflammatory response (Fig. 4-5B). To determine if the suppressed inflammatory response of *ORF5*$^{-/-}$ cells was specific to *TLR4*-driven inflammation ora common feature in response other types of inflammatory stimuli as well, WT and KO BMDMs were treated with various TLR agonists. qPCR analysis of inflammatory cytokine production revealed the suppressed inflammatory response observed in ORF5-/- BMDMs treated with LPS was also evident following exposure to inflammatory stimuli that activate TLR7 and TLR9

signalling (Fig. 4-5C). Interestingly, however, expression of inflammatory cytokines was actually higher in *ORF5*$^{-/-}$ BMDMs compared to WT when the inflammatory response was triggered using compounds that specifically activate *TLR1/2* signaling pathways (Fig. 4-5C). Thus, the inflammatory response of *ORF5*$^{-/-}$ macrophages is suppressed in *TLR4*, *TLR7*, and *TLR9*-driven inflammation; whereas *TLR1/2*-driven inflammation is associated with an enhanced inflammatory response in *ORF5*$^{-/-}$ cells. Given the function of *S100A8* as a pro-inflammatory paracrine factor that augments the inflammatory signaling cascade by binding to *TLR4* on the surface of neighboring cells, we reasoned one explanation for the suppressed inflammatory response in *ORF5*$^{-/-}$ macrophages could be that secretion of *S100A8* had become impaired in these cells. *LTR* staining revealed an increase in the number of autophagosomes and autophagic vesicles in *ORF5*$^{-/-}$ BMDMs compared to WT (Fig. 4-5D), further implying *S100A8* may have prematurely aggregated in *ORF5*$^{-/-}$ BMDMs, triggering the aberrant activation of autophagy. Quantification of amount of secreted *S100A8* in supernatants from WT and *ORF5*$^{-/-}$ BMDMs treated with LPS revealed secretion of *S100A8* was severely impaired in *ORF5*$^{-/-}$ mice (Fig. 4-5E). Importantly, addition of recombinant *S100A8* to media of *ORF5*$^{-/-}$ cells restored the inflammatory response (Fig. 4-5F). These results strongly suggest *ORF5* plays an essential rolein maintaining homeostatic function of lysosomes *in-vivo,* which also appears to be essential forensuring proper induction of the inflammatory response in activated macrophages.

Discussion

Structural modeling and bioinformatic analysis indicate *ORF5* forms a conserved transmembrane alpha-helix. Although sequence conservation amongst micropeptides is low, a significant portion of micropeptides exhibit a common single transmembrane (STMD) topology,

and the C-terminal domains. Thus, the predicted structure and topology of *ORF5* is consistent with that of other lncRNA-encoded micropeptides (Anderson, 2015; Bi, 2017, Jackson, 2018). Based on this topology, many STMD-containing micropeptides have been proposed to function as chaperones that tether individual subunits of larger, multimeric protein complexes to the proper cellular membranes (Karaoglu, 1997). Several lines of evidence suggests *ORF5* may function in a similar capacity and mediate the attachment of *S100A8* to the lysosomal and/or plasma membrane. *S100A8* is predominantly localized in the cytoplasm but rapidly translocates to the plasma membrane for secretion upon rise in intracellular calcium levels (Roth, 1993). However, the mechanism by which *S100A8* reaches the plasma membrane and remains anchored is unclear, as *S100A8* lacks a TMD or conventional sorting sequence (Nacken, 2000). Thus, we propose one function of *ORF5* may be to chaperone *S100A8* to the proper lysosomal vesicle for secretion, analogous to what has been described for other STMD-containing micropeptides

The list of biological functions for STMD-containing micropeptides continues to grow. In addition to their roles in anchoring membrane protein complexes to the proper cellular membrane (Karaoglu, 1997) and acting as "bridge" molecules that mediate assembly of larger protein complexes (Kim, 2000), STMD-containing micropeptides have also been shown to transiently shield hydrophobic surfaces and effectively prevent uncontrolled or premature aggregation of hydrophobic subunits or assembly intermediates (Muchowski, 2000). One group of STMD-containing micropeptides for which this particular function has been well studied are the small heat shock proteins (HSPs). Small HSPs are a set of evolutionarily conserved proteins that help counteract heat and oxidative stress (Sundby, 2005). HSPs contain a set of conserved methionines located on one side of an amphipathic helix; thus forming a binding groove lined with methionines (Muchowski, 2000). Previous work has

demonstrated this binding groove provides a hydrophobic surface onto which hydrophobic regions of partially denatured substrate proteins can be bound and thereby rescued from aggregation. Structural modeling suggests *ORF5* may also serve a function analogous to that of HSPs: hydrophobic domain shielding. *ORF5* likewise contains a set of conserved methionines. Interestingly, the methionine residues in *ORF5* are all located on the same side of the alpha helix; conveniently placing these residues in the ideal orientation to transiently shield hydrophobic regions of *S100A8* which have been associated with an increased propensity towards aggregate formation (Vogl, 2011). Given that *S100A8* is transported to the plasma membrane for secretion in lysosomes, which also accumulate high concentrations of zinc and other metal ions within their lumen, shielding of *S100A8* hydrophobic resisdues by *ORF5* would be critical in this context for preventing premature aggregation of the protein while its being transported in lysosomes to the plasma membrane for secretion. Interestingly, structural modeling also suggested *ORF5* may bind divalent metal ions, such as zinc. Furthermore, the location of all the methionine residues on the same side of the alpha-helix further suggests *ORF5* could also prevent premature aggregation of *S100A8* by direct chelation of the bound zinc ion, although additional biochemical studies are needed to confirm this hypothesis.

Earlier studies that examined the localization of *S100A8* found it primarily localized to cytoplasmic foci and plasma membrane patches in untreated macrophages (Chakraborty, 2015). However, when cells were stimulated with *TNFα*, *S100A8* began to accumulate in lysosomal vesicles. Subsequently, accumulation of *S100A8* in the lysosomal compartment led to increased secretion of the protein. In WT cells, transmission electron microscopy (TEM) revealed *S100A8* was primarily localized to distinct foci in the cytoplasm as well as small patches at the plasma membrane, which is consistent with earlier studies. Upon LPS stimulation of WT cells, *S100A8* then translocated to the plasma membrane and was readily detected in

secretory vesicles, which is also consistent with previous results. In ORF5 KO cells, LPS treatment resulted in aberrant activation of autophagy and accumulation of autophagosomes and autolysosomes containing aggregated *S100A8*. Multiple lines of evidence suggests the appearance of these structures in *ORF5*$^{-/-}$ cells was due to excessive oxidation and aggregation of *S100A8* hydrophobic residues, which is predicted to occur in the absence of *ORF5. S100A8* has previously been shown to induce apoptosis and autophagy (Ghavami, 2010). Treatment of cells with *S100A8/A9* induced *LC3-β* cleavage, increased *Atg12-Atg5* formation, and increased *Beclin-1* expression. *S100A8* has also been shown to promote autophagy in leukemia cells (Yang, 2014). Thus, the increase in LTR staining, enhanced expression of autophagy genes, and accumulation of *LC3* observed in *ORF5* KO cells are all consistent with previous results examining the effect of excessive *S100A8*. The increase in the number of autophagosomes and autolysosomal vesicles containing undigested cellular material observed in *ORF5* KO cells also implies lysosome function is impaired in these cells. Lysosomal stress genes were increased in *ORF5* KO cells, as were the levels of several proteins often used as makers of lysosomal stress. Furthermore, swollen/enlarged lysosomes were abundant in *ORF5* KO cells after LPS treatment, and these cells demonstrated clear evidence of widespread lysophagy and LMP by TEM analysis. Lastly, transfection of cells with *LC3*-RFP-GFP, which is commonly used to determine the specific molecular mechanism contributing to impaired autophagic flux, revealed a marked reduction in the number of intact lysosomes in *ORF5*$^{-/-}$ cells but not in WT cells. The reduction in number of intact lysosomes observed in *ORF5*$^{-/-}$ cells was further corroborated by acridine orange (AO) staining, a common technique used to visualize the number of intact versus ruptured lysosomes. Previous work suggests aggregation of *S100A8* in the lysosome islikely the main explanation for this phenomena, as earlier work demonstrated *S100A8*-induced cell death could be reversed when cells were pretreated with the lysosomal hydrogen pump inhibitor Baf-A1 (Ghavami, 2010), thereby confirming a role for the lysosomal pathway in *S100A8*-induced

autophagy and cell death. Cathepsin inhibitors that block LMP also reversed the phenotype; further corroborating defects in the lysosomal pathway and lysosome rupture as key contributors to *S100A8*-induced autophagy and cell death. Furthermore, *S100A8* was previously shown to be secreted by alternative pathway of secretion involving secretory lysosomes (Chakraborty, 2015). Other studies have confirmed lysosomal localization of *S100A8*and revealed that accumulation of *S100A8* and *S100A9* in the lysosomal compartment was alsoassociated with induction of their release from cells. Furthermore, an inhibitor of lysosomal activity could effectively modulate the extracellular release of *S100A8* and *S100A9,* clearly demonstrating that *S100A8* accumulates in the lysosomal compartment prior to its extracellular secretion. Thus, the most likely explanation for structural features observed by TEM in *ORF5* KO cells is the accumulation of oxidized *S100A8* in lysosomal vesicles leading to widespread lysosome rupture and aberrant activation of autophagy. In the absence of *ORF5*, we propose intralysosomal zinc-catalyzed oxidation reactions of *S100A8* results in excessive oxidation and premature aggregation of the protein within the lysosome, ultimately culminating with the appearance of swollen lysosomes and widespread LMP. Furthermore, the reduction in number of intact lysosomes observed in *ORF5* KO cells may also explain the subsequent accumulation of autophagosomes containing undigested material that was observed in these cells, as increased LMP would also reduce the number of intact lysosomes available for fusion with autophagosomes thereby leading to the accumulation of autophagosomes that contain undigested cellular material.

Secretory lysosomes are also proposed to be main vehicles for release of aggregation-prone proteins, especially those associated with neurodegenerative diseases in humans. For example, α-synuclein is released in vesicular structures upon various cellular stresses independent of classical exocytosis pathways (Jang, 2010; Lee, 2014). Similarly, mutant *Htt* is released by a mechanism that involves fusion of the lysosomes with the plasma membrane (Trajkovic, 2017).

Interestingly, several components specifically associated with the release of aggregation-prone proteins were identified as *ORF5*-interacting partners by mass-spectrometry. For example, *ORF5* was shown to interact with *SNCA*, the aggregation of which represents the key and defining feature of Parkinson's disease. We therefore wondered if the impairment in lysosomal-mediated secretion observed in *ORF5* KO cells was specific for the secretion of *S100A8* or if defects in secretion of other aggregation prone proteins secreted by this same pathway were also evident in *ORF5* KO cells. When mutant *SNCA* was overexpressed in WT and *ORF5* KO macrophages, intracellular accumulation and reduced secretion of the aggregation-prone *SNCA*(A93T) mutant was observed for *ORF5* KO cells; whereas accumulation of *SNCA*(A93T) in WT cells was found to be negligible. Thus, we believe *ORF5* may have a broader function wherein hydrophobic domain shielding by the micropeptide represents an essential and previously unrecognized mechanism for preventing premature aggregation and defects in secretion of other proteins that are intrinsically aggregation-prone like *S100A8*. Furthermore, it will be interesting to see if *ORF5*$^{-/-}$ mice display altered susceptibility with regard to developing neurodegenerative diseases.

Lysosomal and autophagy dysfunction occurs both in lysosomal storage diseases (LSDs) as well in many common neurodegenerative diseases, which are ultimately characterized by defective cellular clearance and subsequent accumulation of toxic material. Here, we demonstrate production of an STMD-containing micropeptide from lncRNA *AK165607*. By transiently shielding hydrophobic domains of *S100A8* and preventing premature aggregation, *ORF5* functions to maintain homeostatic function of lysosomes thereby ensuring lysosomal secretion of *S100A8,* and potentially other aggregation prone proteins, remains intact. Pharmacological manipulation of *ORF5* activity as a means of enhancing lysosomal function may therefore represent a promising therapeutic avenue for the treatment of protein aggregation disorders.

Figure Legends

Figure 4-1: Production of a novel micropeptide from lncRNA *AK165607.* A. Schematic of *AK165607* locus that encodes *ORF5.* B. Current annotations of *AK165607*-derived transcripts (top) and *RACE* annotations (bottom); transcripts matching current annotations (dark red), novel transcripts (blue), novel exons (yellow). C. Expression of *BB864991* in various tissues. D. Expression of *BB864991* in RAW cells treated with different TLR agonists. E. ChIP-seq showing increased NFκβ binding to *BB864991* promoter following LPS treatment. F. Western blot showing expression of a 6 kDa peptide in cells transfected with *ORF5*-3X_FLAG constructs. G. Confocal microscopy showing localization of 3X-FLAG tagged *ORF5* in HeLa cells. H. *in-vitro* translation assay showing translation of *ORF5* from *BB864991*. I. Schematic of FLAG knock-in strategy and validation of successful FLAG insertion into endogenous *BB864991* locus by WB iniBMX cells.

Figure 4-2: *ORF5* forms an evolutionarily conserved transmembrane α-helix and interacts with *S100A8.* A. Predicted secondary structure and proposed model of *ORF5*; c=coil, h=helix. B. Predicted secondary structure of *ORF5* using *Phyre.* C. Alignment and *in-silico* translation of homologous ORFs identified in other species and secondary structure prediction (*Phobius*). D. Alignment and *in-silico* translation of homologous ORFs identified in other species showing conservation and location of conserved TMD. E. Alignment and *in-silico* translation of homologous ORFs identified in other species showing evolutionary conservation of *ORF5*. Black boxes indicate identical or conserved residues. F. Top-threading template in PDB and schematic depicting shielding of *get3* hydrophobic residues by TMD of *pep12.* G. Detection of 3X-FLAG_*ORF5* in cell lysate and IP eluate from cells transfected with pDEST_*ORF5* or pDEST_empty that were used for mass-spectrometry analysis. H. Validation of *S100A8:ORF5*

interaction by co-immunoprecipitation showing detection of *S100A8* in IP eluate of cells transfected with pDEST_*ORF5* but not pDEST_empty. I. Helical wheel prediction of spatial orientation and geometry of conserved methionine residues in *ORF5.* J. Model depicting proposed role of *ORF5* transiently shielding hydrophobic domains of *S100A8.*

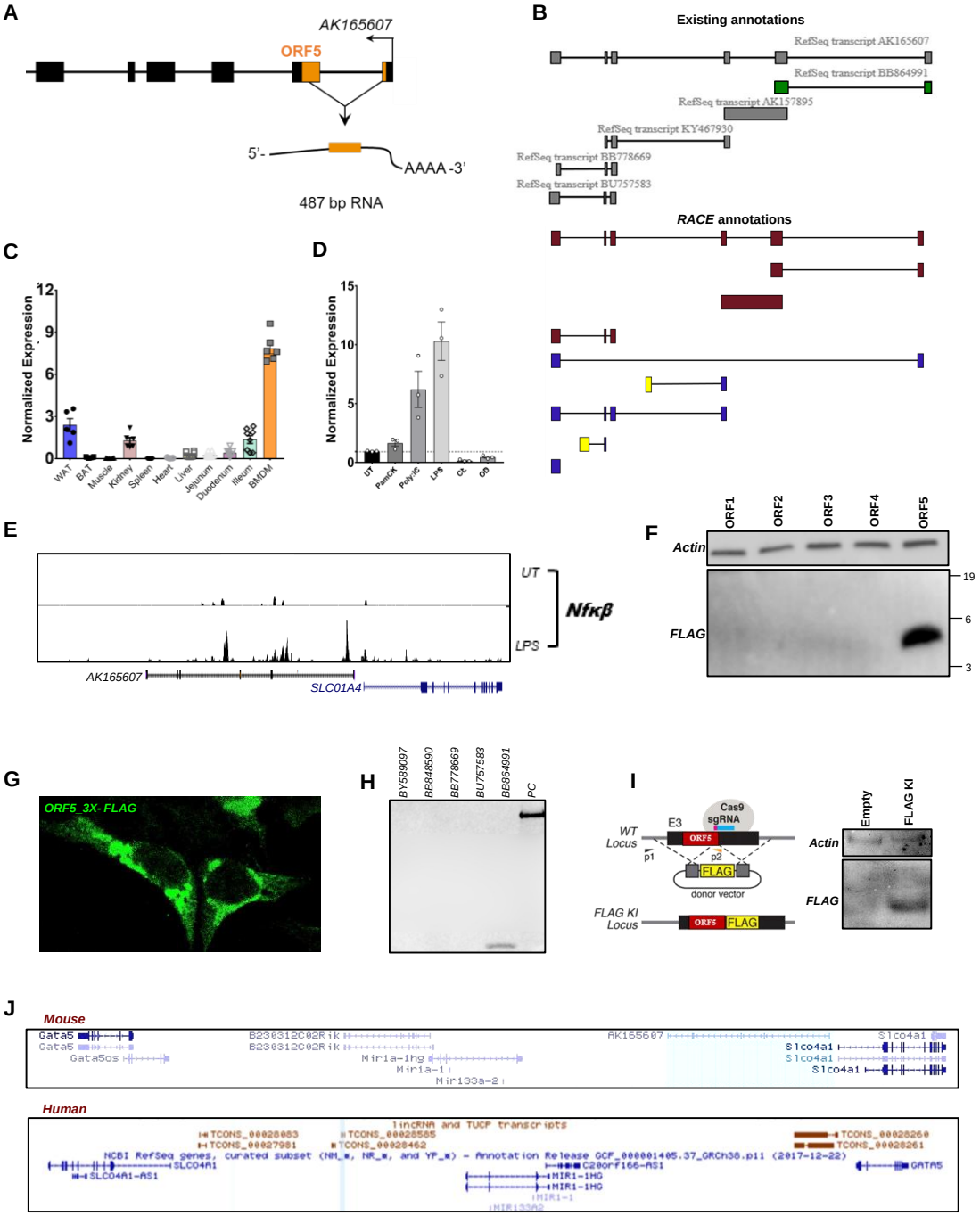
Figure 4-3: Loss of *ORF5* leads to accumulation of *S100A8* oligomers and activates autophagy. A. EM images of WT and *ORF5*$^{-/-}$ RAW macrophages, untreated. B. EM images ofWT (top) and *ORF5*$^{-/-}$ (bottom)RAW macrophages treated with LPS (100ng/mL) for 4hr. c.Staining of intracellular protein aggregates using *Thioflavin-T* dye in WT and *ORF5*$^{-/-}$ RAW macrophages treated with LPS (100ng/mL) for 4hr. D. Staining of intracellular zinc deposits using *Fluoro-Zin3* dye in WT and *ORF5*$^{-/-}$ RAW macrophages treated with LPS (100ng/mL) for 4hr. E. qPCR analysis of autophagy genes in WT and *ORF5*$^{-/-}$ RAW macrophages treated with LPS (100ng/mL) for 4hr. F. Western blot of *LC3* from lysates of WT and *ORF5*$^{-/-}$ RAW macrophages treated with LPS (100ng/mL) for 4hr. G. Visualization of # of acidic vesicles (autophagosomes, autolysosomal vesicles) in WT and *ORF5*$^{-/-}$ RAW macrophages treated withLPS (100ng/mL) for 4hr and stained with *Lysotracker Red DND-99.*

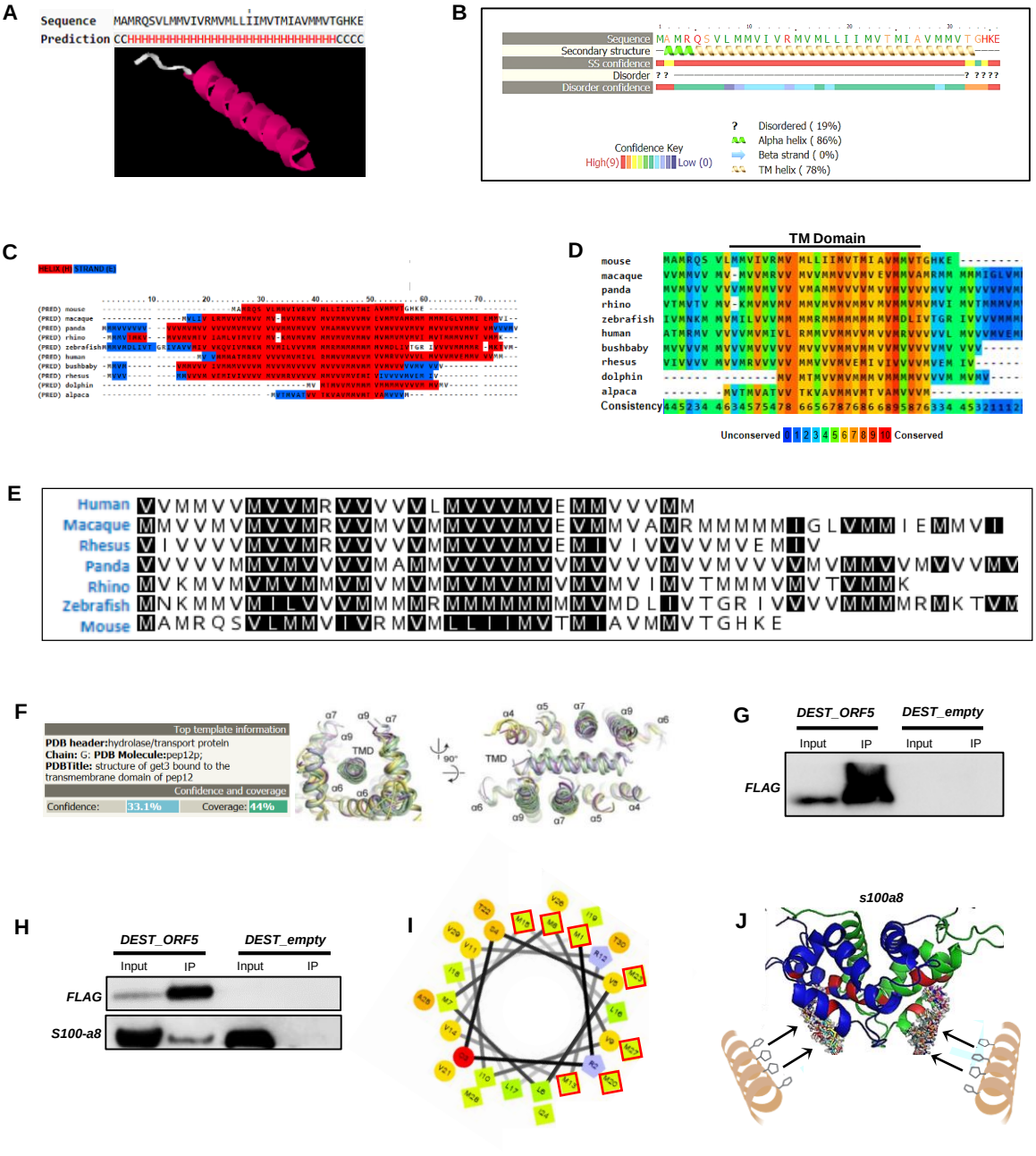
Figure 4-4: Accumulation of *S100A8* aggregates in *ORF5*$^{-/-}$ cells impairs degradative capacity of lysosomes and triggers widespread lysosome rupture. A. WT and *ORF5*$^{-/-}$ RAW cells transfected with *LC3*-GFP-RFP and treated with LPS (100ng/mL) for 4 hrs. B. Western blot of lysosomal stress markers using cell lysates from WT and *ORF5*$^{-/-}$ RAW cells treated with LPS (100ng/mL) for 4hr. C. qPCR analysis of lysosomal stress genes expression in WT and *ORF5*$^{-/-}$ RAW cells treated with LPS (100ng/mL) for 4hr. D. Western blot of various *S100A8* isoforms using lysosomal fractions from WT and *ORF5*$^{-/-}$ RAW cells treated with LPS (100ng/mL) for 4hr. E. EM images showing accumulation of *S100A8* in lysosomes and increase in the # of ruptured lysosomes in *ORF5*$^{-/-}$ RAW macrophages treated with LPS (100ng/mL) for

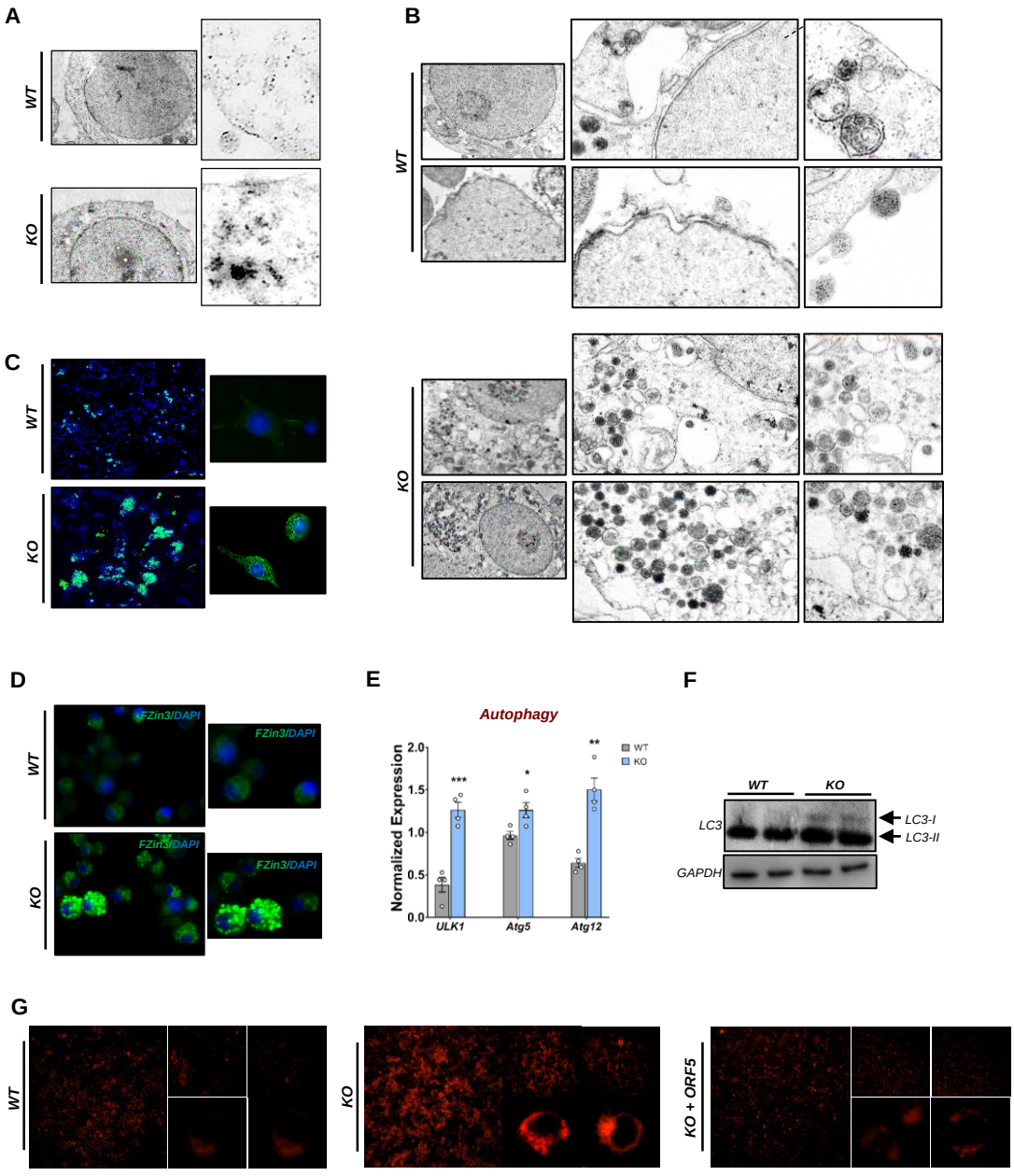
4hr. F. Visualization of # of intact lysosomes in WT and *ORF5*$^{-/-}$ RAW macrophages treated with LPS (100ng/mL) for 4hr and stained with acridine orange. G. Accumulation of mutant *SNCA*(A93T) in WT and *ORF5*$^{-/-}$ RAW macrophages transfected with *SNCA*(A93T) and treated with LPS (100ng/mL) for 4hr. H. Model depicting proposed role of *ORF5* in maintaining homeostatic function of lysosomes.

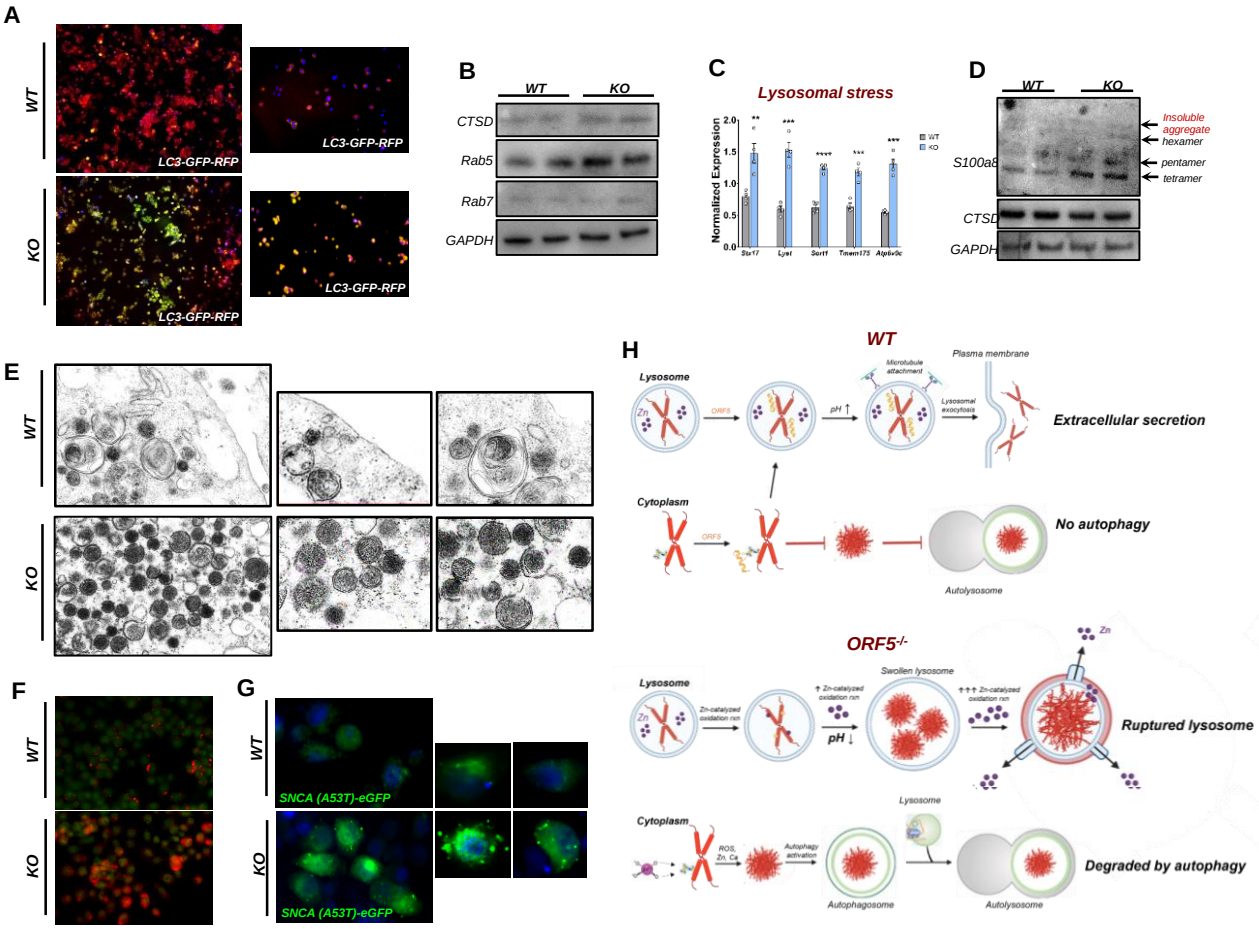
Figure 4-5: Inflammatory response and secretion of *S100A8* are impaired in *ORF5*$^{-/-}$ mice.

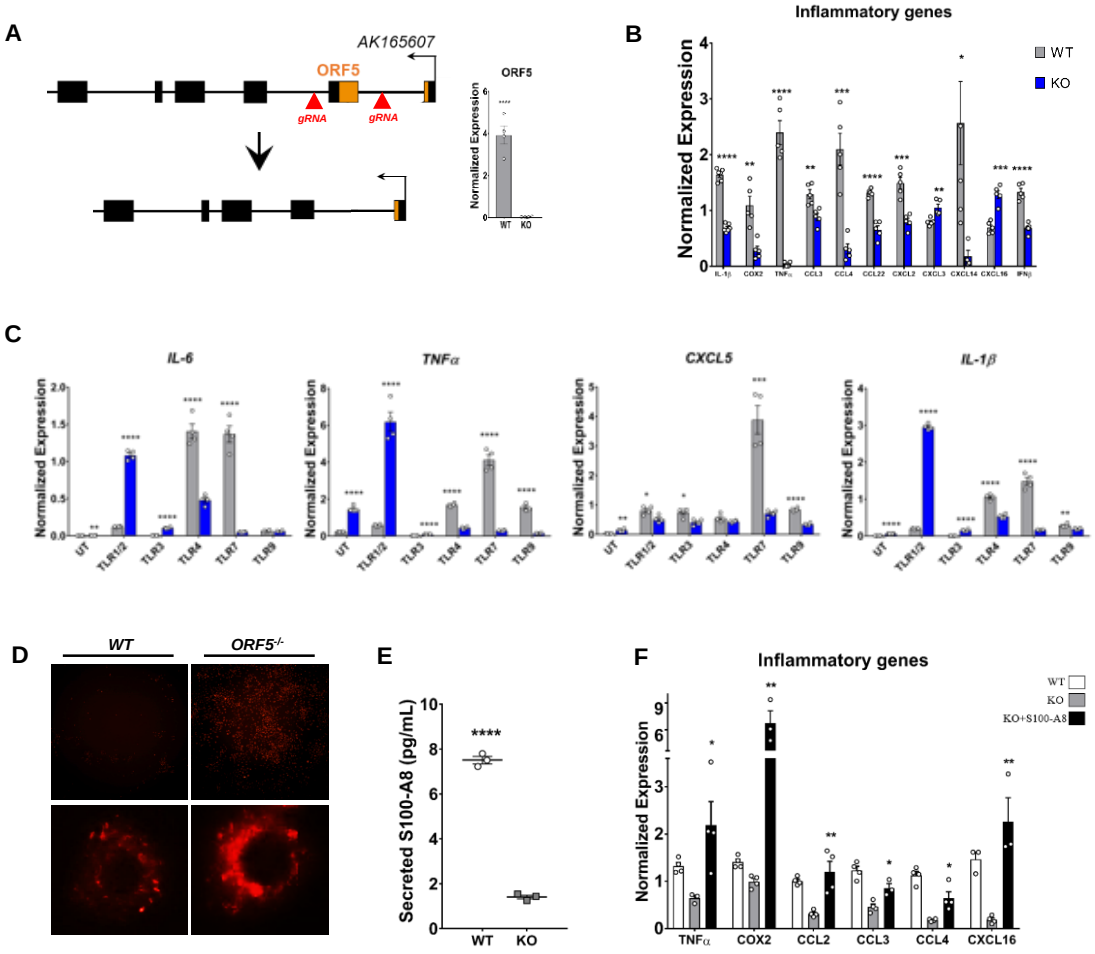
A. Schematic depicting generation of *ORF5*$^{-/-}$ mice and qPCR analysis showing deletion of *ORF5*. B. qPCR analysis of inflammatory gene expression in WT and *ORF5*$^{-/-}$ macrophages treated with LPS (100ng/mL) for 4hr. C. qPCR analysis of inflammatory gene expression in WT and *ORF5*$^{-/-}$ macrophages treated with various TLR agonists for 4hr. D. Visualization of the # of acidic vesicles in WT and *ORF5*$^{-/-}$ macrophages treated with LPS (100ng/mL) for 4hr and stained with *Lysotracker Red DND-99*. E. Quantification of secreted *S100A8* levels by ELISA using supernatant from WT and *ORF5*$^{-/-}$ macrophages treated with LPS (100ng/mL) for 4hr. F. qPCR analysis of inflammatory gene expression in WT and *ORF5*$^{-/-}$ macrophages treated with LPS (100ng/mL) for 4hr with or without addition of recombinant *S100A8.*


A
AK165607
ORF5
5'-
AAAA-3'
487 bp RNA
B
Existing annotations
RefSeq transcript AK165607
RefSeq transcript BB864991
RefSeq transcript AK157895
RefSeq transcript KY467930
RefSeq transcript BB778669
RefSeq transcript BU757583
RACE annotations
C
Normalized Expression
WAT
BAT
Muscle
Kidney
Spleen
Heart
Liver
Jejunum
Duodenum
Ileum
BMDM
D
Normalized Expression
UT
PamCK
PolyIC
LPS
CL
OD
E
UT
LPS
Nfκβ
AK165607
SLC01A4
F
ORF1
ORF2
ORF3
ORF4
ORF5
Actin
FLAG
19
6
3
G
ORF5_3X-FLAG
H
BY589097
BB848590
BB778669
BU757583
BB864991
PC
I
WT Locus
E3
Cas9 sgRNA
ORF5
p1
p2
FLAG
donor vector
FLAG KI Locus
ORF5
FLAG
Empty
FLAG KI
Actin
FLAG
J
Mouse
Gata5
Gata5
Gata5os
B230312C02Rik
B230312C02Rik
Mir1a-1hg
Mir1a-1
Mir133a-2
AK165607
Slco4a1
Slco4a1
Slco4a1
Human
lincRNA and TUCP transcripts
TCONS_00028603
TCONS_00027981
TCONS_00028565
TCONS_00028462
TCONS_00028260
TCONS_00028261
NCBI RefSeq genes, curated subset (NM_*, NR_*, and YP_*) - Annotation Release GCF_000001405.37_GRCh38.p11 (2017-12-22)
SLC04A1
SLC04A1-AS1
C20orf166-AS1
GATA5
MIR1-1HG
MIR1-1HG
MIR1-1
MIR133A2


A
Sequence MAMRQSVLMMVIVRMVMLLIIMVTMIAVMMVTGHKE
Prediction CCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCC
B
Sequence
Secondary structure
SS confidence
Disorder
Disorder confidence
Confidence Key
High(9) Low (0)
? Disordered (19%)
Alpha helix (86%)
Beta strand (0%)
TM helix (78%)
C
HELIX (H) STRAND (E)
D
TM Domain
mouse
macaque
panda
rhino
zebrafish
human
bushbaby
rhesus
dolphin
alpaca
Consistency
Unconserved 0 1 2 3 4 5 6 7 8 9 10 Conserved
E
Human
Macaque
Rhesus
Panda
Rhino
Zebrafish
Mouse
F
Top template information
PDB header:hydrolase/transport protein
Chain: G: PDB Molecule:pep12p;
PDBTitle: structure of get3 bound to the transmembrane domain of pep12
Confidence and coverage
Confidence: 33.1% Coverage: 44%
TMD
90°
G
DEST_ORF5
DEST_empty
Input IP Input IP
FLAG
H
DEST_ORF5
DEST_empty
Input IP Input IP
FLAG
S100-a8
I
J
s100a8


A
WT
KO
B
WT
KO
C
WT
KO
D
FZin3/DAPI
WT
KO
E
Autophagy
Normalized Expression
WT
KO
ULK1
Atg5
Atg12
F
WT
KO
LC3
LC3-I
LC3-II
GAPDH
G
WT
KO
KO + ORF5


A
WT
KO
LC3-GFP-RFP
LC3-GFP-RFP
LC3-GFP-RFP
LC3-GFP-RFP
B
WT
KO
CTSD
Rab5
Rab7
GAPDH
C
Lysosomal stress
Normalized Expression
WT
KO
Stx17
Lyst
Scrt1
Tmem175
Atp6v0c
D
WT
KO
S100a8
CTSD
GAPDH
Insoluble aggregate
hexamer
pentamer
tetramer
E
WT
KO
F
WT
KO
G
WT
KO
SNCA (A53T)-eGFP
SNCA (A53T)-eGFP
H
WT
Lysosome
Zn
ORF5
pH ↑
Microtubule attachment
Plasma membrane
Lysosomal exocytosis
Extracellular secretion
Cytoplasm
ORF5
No autophagy
Autolysosome
ORF5-/-
Lysosome
Zn
Zn-catalyzed oxidation rxn
↑ Zn-catalyzed oxidation rxn
pH ↓
Swollen lysosome
↑↑↑ Zn-catalyzed oxidation rxn
Zn
Ruptured lysosome
Cytoplasm
ROS, Zn, Ca
Autophagy activation
Lysosome
Autophagosome
Autolysosome
Degraded by autophagy


A
AK165607
ORF5
gRNA
gRNA
ORF5
Normalized Expression
WT
KO
B
Inflammatory genes
Normalized Expression
WT
KO
IL-1β
COX2
TNFα
CCL3
CCL4
CCL22
CXCL2
CXCL3
CXCL14
CXCL16
IFNβ
C
IL-6
TNFα
CXCL5
IL-1β
Normalized Expression
UT
TLR1/2
TLR3
TLR4
TLR7
TLR9
D
WT
ORF5-/-
E
Secreted S100-A8 (pg/mL)
WT
KO
F
Inflammatory genes
Normalized Expression
WT
KO
KO+S100-A8
TNFα
COX2
CCL2
CCL3
CCL4
CXCL16

References:

Anderson, Douglas M., et al. "A micropeptide encoded by a putative long noncoding RNA regulatesmuscle `performance." *Cell* 160.4 (2015): 595-606.

Aspden, Julie L., et al. "Extensive translation of small open reading frames revealed by Poly-Ribo-Seq." *Elife* 3 (2014): e03528.

Bi, Pengpeng, et al. "Control of muscle formation by the fusogenic micropeptide myomixer." *Science* 356.6335 (2017): 323-327.

Blott, E. J., & Griffiths, G. M. (2002). Secretory lysosomes. *Nature reviews Molecular cell biology*, *3*(2), 122-131.

Borland, H., & Vilhardt, F. (2017). Prelysosomal compartments in the unconventional secretion of amyloidogenic seeds. *International journal of molecular sciences*, *18*(1), 227.
de Duve, C. (2005). The lysosometurns fifty. *Nat. Cell Biol.* 7, 847–849.

Chakraborty, P., Bjork, P., Källberg, E., Olsson, A., Riva, M., Mörgelin, M., ... & Leanderson, T. (2015). Vesicular location and transport of S100A8 and S100A9 proteins in monocytoid cells. *PloS one*, *10*(12), e0145217.

Fritz, G., Botelho, H. M., Morozova-Roche, L. A., & Gomes, C. M. (2010). Natural and amyloid self-assembly of S100 proteins: structural basis of functional diversity. *The FEBS journal*, *277*(22), 4578-4590.

Ghavami, S., Eshragi, M., Ande, S. R., Chazin, W. J., Klonisch, T., Halayko, A. J., ... & Los, M. (2010). S100A8/A9 induces autophagy and apoptosis via ROS-mediated cross-talk between mitochondria and lysosomes that involves BNIP3. *Cell research*, *20*(3), 314-331.

Harris, H., & Rubinsztein, D. C. (2012). Control of autophagy as a therapy for neurodegenerative disease. *Nature reviews neurology*, *8*(2), 108-117.

Iashchishyn, I. A., Sulskis, D., Nguyen Ngoc, M., Smirnovas, V., and Morozova-Roche, L. A. (2017). Finke-Watzky Two-Step Nucleation-Autocatalysis Model of S100A9 Amyloid Formation: Protein Misfolding as “Nucleation” Event. *ACS Chem Neurosci* 8, 2152–2158.

Isaaz, S., Baetz, K., Olsen, K., Podack, E., & Griffiths, G. M. (1995). Serial killing by cytotoxic T lymphocytes: T cell receptor triggers degranulation, re-filling of the lytic granules and secretion of lytic proteins via a non-granule pathway. *European journal of immunology*, *25*(4), 1071-1079.

Jackson, Ruaidhri, et al. "The translation of non-canonical open reading frames controls mucosal immunity." *Nature* 564.7736 (2018): 434-438.

Jang, A., Lee, H. J., Suk, J. E., Jung, J. W., Kim, K. P., & Lee, S. J. (2010). Non-classical exocytosis of α-synuclein is sensitive to folding states and promoted under stress conditions. *Journal of neurochemistry*, *113*(5), 1263-1274.

Jeong, H., Then, F., Melia Jr, T. J., Mazzulli, J. R., Cui, L., Savas, J. N., ... & Krainc, D. (2009). Acetylation targets mutant huntingtin to autophagosomes for degradation. *Cell*, *137*(1), 60-72.

Karaoglu, D., Kelleher, D. J., & Gilmore, R. (1997). The highly conserved Stt3 protein is a subunit of the yeast oligosaccharyltransferase and forms a subcomplex with Ost3p and Ost4p. *Journal of Biological Chemistry*, *272*(51), 32513-32520.

Kim, H., Park, H., Montalvo, L., & Lennarz, W. J. (2000). Studies on the role of the hydrophobic domain of Ost4p in interactions with other subunits of yeast oligosaccharyl transferase. *Proceedings of the National Academy of Sciences*, *97*(4), 1516-1520.

Kummer, M. P., Vogl, T., Axt, D., Griep, A., Vieira-Saecker, A., Jessen, F., ... & Heneka, M. T. (2012). Mrp14 deficiency ameliorates amyloid β burden by increasing microglial phagocytosis and modulation of amyloid precursor protein processing. *Journal of Neuroscience*, *32*(49), 17824-17829.

Lee, H. J., Savelieff, M. G., Kang, J., Brophy, M. B., Nakashige, T. G., Lee, S. J. C., et al. (2018). Calprotectin influences the aggregation of metal-free and metal-bound amyloid-beta by direct interaction. *Metallomics* 10, 1116–1127.

Li, S. C., & Kane, P. M. (2009). The yeast lysosome-like vacuole: endpoint and crossroads. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1793*(4), 650-663.

Link, V. M., Duttke, S. H., Chun, H. B., Holtman, I. R., Westin, E., Hoeksema, M. A., ... & Glass, C. K. (2018). Analysis of genetically diverse macrophages reveals local and domain-wide mechanisms that control transcription factor binding and function. *Cell*, *173*(7), 1796-1809.

Makarewich, Catherine A. "The hidden world of membrane microproteins." *Experimental Cell Research* 388.2 (2020): 111853.

Muchowski, P. J., Schaffar, G., Sittler, A., Wanker, E. E., Hayer-Hartl, M. K., & Hartl, F. U. (2000). Hsp70 and hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. *Proceedings of the National Academy of Sciences*, *97*(14), 7841-7846.

Nacken, W., Sopalla, C., Pröpper, C., Sorg, C., & Kerkhoff, C. (2000). Biochemical characterization of the murine S100A9 (MRP14) protein suggests that it is functionally equivalent to its human counterpart despite its low degree of sequence homology. *European journal of biochemistry*, *267*(2), 560-565.

Polishchuk, E. V., Concilli, M., Iacobacci, S., Chesi, G., Pastore, N., Piccolo, P., ... & Polishchuk, R. S. (2014). Wilson disease protein ATP7B utilizes lysosomal exocytosis to maintain copper homeostasis. *Developmental cell*, *29*(6), 686-700.

Rabouille, C., Malhotra, V., & Nickel, W. (2012). Diversity in unconventional protein secretion. *Journal of cell science*, *125*(22), 5251-5255.

Roth, J., Burwinkel, F., van den Bos, C., Goebeler, M., Vollmer, E., & Sorg, C. (1993). MRP8 and MRP14, S-100-like proteins associated with myeloid differentiation, are translocated to plasma membrane and intermediate filaments in a calcium-dependent manner.

Sundby, C., Härndahl, U., Gustavsson, N., Åhrman, E., & Murphy, D. J. (2005). Conserved methionines in chloroplasts. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1703*(2), 191-202.

Trajkovic, K., Jeong, H., & Krainc, D. (2017). Mutant huntingtin is secreted via a late endosomal/lysosomal unconventional secretory pathway. *Journal of Neuroscience*, *37*(37), 9000-9012.

Vogl, T., Tenbrock, K., Ludwig, S., Leukert, N., Ehrhardt, C., Van Zoelen, M. A., ... & Roth, J. (2007). Mrp8 and Mrp14 are endogenous activators of Toll-like receptor 4, promoting lethal, endotoxin-induced shock. *Nature medicine*, *13*(9), 1042-1049.

Vogl, T., Leukert, N., Barczyk, K., Strupat, K., & Roth, J. (2006). Biophysical characterization of S100A8 and S100A9 in the absence and presence of bivalent cations. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1763*(11), 1298-1306.

Von Bergen, M., Friedhoff, P., Biernat, J., Heberle, J., Mandelkow, E. M., & Mandelkow, E. (2000). Assembly of τ protein into Alzheimer paired helical filaments depends on a local sequence motif (306VQIVYK311) forming β structure. *Proceedings of the National Academy of Sciences*, *97*(10), 5129-5134.

Winslow, A. R., Chen, C. W., Corrochano, S., Acevedo-Arozena, A., Gordon, D. E., Peden, A. A., ... & Rubinsztein, D. C. (2010). α-Synuclein impairs macroautophagy: implications for Parkinson's disease. *Journal of Cell Biology*, *190*(6), 1023-1037.

Wong, E., & Cuervo, A. M. (2010). Autophagy gone awry in neurodegenerative diseases. *Nature neuroscience*, *13*(7), 805-811.

Yanamandra, K., Alexeyev, O., Zamotin, V., Srivastava, V., Shchukarev, A., Brorsson, A. C., ... &

Morozova-Roche, L. A. (2009). Amyloid formation by the pro-inflammatory S100A8/A9 proteins in the ageing prostate. *PloS one*, *4*(5), e5562.

Yang, M., Zeng, P., Kang, R., Yu, Y., Yang, L., Tang, D., & Cao, L. (2014). S100A8 contributes to drug resistance by promoting autophagy in leukemia cells. *PloS one*, *9*(5), e97242.

Chapter 5: Conclusions and Future Direction

In addition to their “well-recognized” role as “junk” DNA, the list of biological processes in which regulatory RNAs have been shown to execute essential functions has grown to cover nearly all fundamental processes that take place within complex eukaryotes. Unfortunately, the prevalence of both Fatty Liver Disease and atherosclerotic CVD have also increased by a similar magnitude over the past decades. Furthermore, at least with regard to Fatty Liver Disease, there are there are currently no approved pharmacological therapies, emphasizing the growing need for the development of effective therapeutics (Musso, 2012). Given the concentrations of blood cholesterol prevalent in most contemporary nations far surpasses the biological needs of the organism and that this trend which will almost certainly persist and become further exacerbated with time, the need for developing more costly and effective forms of therapy in this arena will likely continue to grow as well (Schreiner, 2016; Hopstock, 2017). RNA-regulation has recently emerged as a key regulatory mechanism that is essential for the maintenance of lipid homeostasis in higher eukaryotes (Salisbury, 2021). Thus, a greater understanding of the diverse mechanisms by which regulatory RNAs function to maintain tight control of lipid metabolism by fine-tuning cellular pathways governing metabolic homeostasis in response to changing environmental cues will likely have important therapeutic implications.

In chapter 2, we demonstrate that a small, previously undetectable RNA modification can potently impact the development of fatty liver disease and protects against hepatic lipid overload (Salisbury, 2021). m^6A, which is enriched on lipogenic transcripts in the livers of fasted mice, effectively triages these transcripts for degradation in cytoplasmic P-bodies, which we show to be critical for proper suppression of *de-novo lipogenesis (DNL)* during fasting and thus prevention of hepatic triglyceride (TG) accumulation. Interestingly, failure to suppress lipogenesis during fasting is thought to be a key feature in many NAFLD patients (Diraison, 2003; Lambert, 2014); highlighting the need to fully understand the molecular mechanisms that

regulate this process. However, at least in chow-fed animals, m^6A-mediated suppression of *DNL* during fasting seems to be less relevant in females where TG accumulation was also observed following *Mettl14* deletion but occurred to a much lesser extent than was observed in male livers. Although expression of certain lipogenic transcripts themselves may differ at baseline between males and females, as does the expression of m^6A writers and erasers, the precise mechanisms underlying differential enrichment of the m^6A modification at the same sites and on the same transcripts between males and females clearly warrant further investigation. One hypothesis is that specific RNA-binding proteins, which may bind 3' UTR regions thereby preventing or promoting deposition of m^6A, may be differentially expressed between the sexes and thus contribute to these differences in m^6A enrichment. Data from the Mexican Obesity Surgery (MOBES) cohort suggests that m^6A levels are correlated with development of fatty liver disease in humans, as we demonstrated expression of *Mettl14* was inversely associated with the progression from a healthy liver to steatosis and ultimately to NASH. It will therefore be interesting to see whether m^6A levels correlate with other complex metabolic or lipid traits and potentially contributes to the development of additional metabolic-related disorders characterized by abnormal accumulation of lipid. Hybrid Mouse Diversity Panel (HMDP) analysis has proven to be a powerful tool for understanding the effect of minor genetic variations and environmental cues in the pathophysiology of complex diseases (Parker, 2019). By correlating minor genetic variations inherently present in a particular mouse strain with various metabolic parameters like blood glucose or serum cholesterol levels, HMDP analysis can help successfully associate these minor genetic alterations with a specific metabolic phenotype or disease that would otherwise be difficult to untangle. Thus, by measuring m^6A levels in these different mouse strains and determining if they correlate with specific metabolic traits or disease susceptibility, the repertoire of processes in which regulatory RNA modifications play an essential role is sure to expand.

In chapter 3, we present results of a pre-clinical model of sustained *MeXis* overexpression in the setting of atherosclerosis. Our previous work showed that deletion of *MeXis* reduces *Abca1*, enhances inflammation and foam cell formation, as well as promoting atherosclerosis development (Sallam, 2018). Thus, the regulatory RNA *MeXis* may represent an ideal target for therapeutics aimed at lowering cholesterol content of macrophages. To test the therapeutic effects of this lncRNA within lesions, we use a novel genetic mouse model that allows spatial control of *MeXis* expression from the endogenous locus. Although this model was sufficient to increase both mRNA and protein levels of *Abca1* in *MeXis*-overexpressing mice, the increase was not sufficient to protect against development of atherosclerosis; as *LDLR*$^{-/-}$ mice that were irradiated and transfused with bone marrow from either WT or *MeXis*-overexpressing mice did not display any significant differences in the onset and development of atherosclerosis. Unexpectedly, 8 mice that had all been transplanted with bone marrow from *MeXis*-overexpressing animals died suddenly around the 16th week of the study; reducing the statistical power of our study and potentially obscuring any subtle differences in atherosclerotic development that may have been present. Reasons why bone marrow specifically from *MeXis*-overexpressing mice would be toxic to a small subset of animals while bone marrow from WT animals would not have this effect are far from clear. However, *MeXis*, like many lncRNAs, contains multiple small-interspersed nuclear elements (SINE) elements that have arisen throughout course of evolution via integration of retroviral genomes into the host genome (Faulkner, 2009). Typically, SINE-containing RNAs are normally expressed at low levels and have thus not been directly linked to cellular toxicity. However, several studies have shown sustained expression of these SINE-containing lncRNAs can trigger an inflammatory response (Kaneko, 2011; Tarallo, 2012). Interestingly, expression of inflammatory cytokines was actually slightly increased in *MeXis* overexpressing mice compared to controls. Furthermore, we used *MeXis*-overexpressing mice that were homozygous for *Cre* recombinase, meaning these mice

were not just expressing supraphysiological levels of *MeXis* but were also expressing supraphysiological levels of its associated SINE elements. Thus, the increase in inflammation, most likely due to chronic supraphysiological expression of SINE elements located on the *MeXis* transcript, is one explanation why sustained overexpression of this lncRNA did not protect against atherosclerosis despite augmenting *Abca1* protein and RNA levels. Furthermore, our findings have unexpected but fundamentally important implications for future studies in which SINE-containing lncRNAs are expressed at supraphysiological levels for a sustained period of time.

In chapter 4, we report the discovery of a novel lncRNA-encoded micropeptide that functions to maintain homeostatic function of lysosomes; thereby preventing toxic accumulation of insoluble protein aggregates within cells. Lysosomes are proposed to be the main vehicle for the release of aggregation prone proteins and dysfunction of the autophagy-lysosome pathway is widely recognized as a key mechanism most neurodegenerative diseases characterized by the accumulation of protein aggregates (Wong and Cuervo, 2010; Harris, 2012). Whether *ORF5* or other yet to be discovered endogenous inhibitors of lysosome rupture can be therapeutically targeted and used to develop novel forms for treating neurodegenerative disease in humans is an exciting area that requires further investigations. Additionally, whether the proposed mechanism of hydrophobic domain shielding of *S100A8* is specific to only *S100A8* or representative of a more general but unrecognized mechanism essential for ensuring proper secretion of other intrinsically aggregation-prone proteins remains to be determined. *S100A8* was shown to be secreted by alternative pathway of secretion involving secretory lysosomes that requires intact microtubule network (Chakraborty, 2015). Interestingly, mass-spectrometry analysis revealed *ORF5* specifically interacts with other components associated with lysosome-mediated release of other aggregation prone proteins. For example, *ORF5* was shown to

interact with *SNCA*, the aggregation of which represents the hallmark feature of Parkinson's disease. Other components specifically associated with lysosomal-mediated secretion of aggregation-prone proteins that were shown to ineract with *ORF5* includes *Serpin1B, TUBAC, TGM3, Dennd5a, EML4,* and *AP5M1.* Furthermore, the loss-of-function phenotype associated with some of the proteins, such as *AP5M1*, is identical to the phenotype observed in *ORF5*$^{-/-}$ cells: aberrant activation of autophagy, accumulation of swollen lysosomes containing undigested cellular material, and necrosis (Khundadze, 2019). Thus, hydrophobic domain of additional aggregation-prone proteins secreted via lysosomal exocytosis by small peptides like *ORF5* could therefore represent a novel and previously unrecognized mechanism for preventing premature aggregation of proteins that would otherwise occur within lysosomes and could directly promote the development of neurodegenerative disease.

In conclusion, work presented here ascribes novel and essential roles for multiple regulatory RNAs in diverse aspects of metabolic homeostasis. Although each regulatory RNA described here utilizes a different mechanism to fine-tune metabolic pathways and specify higher order assemblies and complex interactions within the vast cellular proteome, all of these examples clearly illustrate that a molecule widely considered to represent nothing more than junk for more than half a century can actually execute critical functions that are essential for proper regulation of fundamental biological processes.

References:

Chakraborty, P., Bjork, P., Källberg, E., Olsson, A., Riva, M., Mörgelin, M., ... & Leanderson, T. (2015). Vesicular location and transport of S100A8 and S100A9 proteins in monocytoid cells. *PloS one*, *10*(12), e0145217.

Diraison, F., Moulin, P. H., & Beylot, M. (2003). Contribution of hepatic de novo lipogenesis and reesterification of plasma non esterified fatty acids to plasma triglyceride synthesis during non-alcoholic fatty liver disease. *Diabetes & metabolism*, *29*(5), 478-485.

Faulkner, G. J., Kimura, Y., Daub, C. O., Wani, S., Plessy, C., Irvine, K. M., ... & Carninci, P. (2009). The regulated retrotransposon transcriptome of mammalian cells. *Nature genetics*, *41*(5), 563-571.

Harris, H., & Rubinsztein, D. C. (2012). Control of autophagy as a therapy for neurodegenerative disease. *Nature reviews neurology*, *8*(2), 108-117.

Hopstock, L. A., Bønaa, K. H., Eggen, A. E., Grimsgaard, S., Jacobsen, B. K., Løchen, M. L., ... & Wilsgaard, T. (2017). Longitudinal and secular trends in total cholesterol levels and impact of lipid-lowering drug use among Norwegian women and men born in 1905–1977 in the population-based Tromsø Study 1979–2016. *BMJ open*, *7*(8), e015001.

Kaneko, H., Dridi, S., Tarallo, V., Gelfand, B. D., Fowler, B. J., Cho, W. G., ... & Ambati, J. (2011). DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. *Nature*, *471*(7338), 325-330.

Khundadze, M., Ribaudo, F., Hussain, A., Rosentreter, J., Nietzsche, S., Thelen, M., ... & Hübner, C. A. (2019). A mouse model for SPG48 reveals a block of autophagic flux upon disruption of adaptor protein complex five. *Neurobiology of disease*, *127*, 419-431.

Lambert, J. E., Ramos–Roman, M. A., Browning, J. D., & Parks, E. J. (2014). Increased de novo lipogenesis is a distinct characteristic of individuals with nonalcoholic fatty liver disease. *Gastroenterology*, *146*(3), 726-735.

Musso, G., Cassader, M., Rosina, F., & Gambino, R. (2012). Impact of current treatments on liver disease, glucose metabolism and cardiovascular risk in non-alcoholic fatty liver disease (NAFLD): a systematic review and meta-analysis of randomised trials. *Diabetologia*, *55*(4), 885-904.

Parker, B. L., Calkin, A. C., Seldin, M. M., Keating, M. F., Tarling, E. J., Yang, P., ... & Drew, B. G. (2019). An integrative systems genetic analysis of mammalian lipid metabolism. *Nature*, *567*(7747), 187-193.

Sallam, T., Jones, M., Thomas, B. J., Wu, X., Gilliland, T., Qian, K., ... & Tontonoz, P. (2018). Transcriptional regulation of macrophage cholesterol efflux and atherogenesis by a long noncoding RNA. *Nature medicine*, *24*(3), 304-312.

Salisbury, D. A., Casero, D., Zhang, Z., Wang, D., Kim, J., Wu, X., ... & Sallam, T. (2021). Transcriptional regulation of N6-methyladenosine orchestrates sex-dimorphic metabolic traits. *Nature Metabolism*, *3*(7), 940-953.

Schreiner, P. J., Jacobs Jr, D. R., Wong, N. D., & Kiefe, C. I. (2016). Twenty-five year secular trends in lipids and modifiable risk factors in a population-based biracial cohort: the Coronary Artery Risk Development in Young Adults (CARDIA) study, 1985–2011. *Journal of the American Heart Association*, *5*(7), e003384.

Tarallo, V., Hirano, Y., Gelfand, B. D., Dridi, S., Kerur, N., Kim, Y., ... & Ambati, J. (2012). DICER1 loss and Alu RNA induce age-related macular degeneration via the NLRP3 inflammasome and MyD88. *Cell*, *149*(4), 847-859.

Wong, E., & Cuervo, A. M. (2010). Autophagy gone awry in neurodegenerative diseases. *Nature neuroscience*, *13*(7), 805-811.